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

by

Christian M. Dahl
Department of Economics
Purdue University

and

Yu Qin
Department of Statistics
Purdue University

Purpose

- Introduce and motivate the parametric random field regression model in econometrics

- Characterize the large sample properties of the estimated parameters in the random field regression model

Problem of Interest

Consider the sequence of observables $\{y_t, \boldsymbol{x}_{t\cdot}\}_{t=1}^{T}$, with y_t generated according to

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

where $\psi : \mathbb{R}^k \rightarrow \mathbb{R}$ and e_t is an unobservable i.i.d. random variable.

- $\psi(\cdot)$ typically unknown
- $\boldsymbol{x}_{t\cdot} \in \mathbb{R}^k$ typically unknown
- Stochastic properties of e_t typically unknown

Objectives:

- Obtain consistent estimators of the unknown function $\psi(\cdot)$
- Develop hypothesis testing procedures regarding the shape of $\psi(\cdot)$
- Find hypothesis testing procedures in relation to the elements of $\boldsymbol{x}_t$.
- Derive testing procedures for the distributional assumptions on e_t?

Flexible nonparametric methods offer a consistent estimate of $\psi(\cdot)$ within a broad class of nonlinear relations:

- Kernel methods (Nadaraya/Watson 1964; Härdle, 1990; Fan, et. al., 1996), Series expansion (Gallant and Nychka, 1987), Wavelets (Donoho et. al., 1995), Nearest neighbor (Yakowitz, 1987; Mizrach, 1992), Smoothing splines (Reinsch, 1967; Eubank, 1988; Wahba, 1990),

- Hamilton (2001): Above methods sacrifice many of the benefits of parametric methods:
  - The methods are not readily adapted for the hypothesis testing and model simplification that are quite necessary in order to make sense of a multivariate nonlinear relation

The random field regression model

- Let $y_t \in \mathbb{R}$, $\boldsymbol{x}_{t\cdot} \in \mathbb{R}^k$ for $t = 1, ..., T$ and consider the model

$$y_t = \underbrace{\boldsymbol{x}_{t\cdot}\boldsymbol{\beta}}_{\text{obs. linear part}} + \underbrace{\lambda m(\boldsymbol{g} \odot \boldsymbol{x}_{t\cdot})}_{\text{unobs. nonlin. part}} + \underbrace{\widetilde{\epsilon}_t}_{\text{N}(0,\sigma^2)}$$

- Parameters of interest:
  - $\boldsymbol{\beta} \in \mathbb{R}^k$ and $\sigma^2 \in \mathbb{R}_+$
  - $\lambda \in \mathbb{R}_+$: Weight of the nonlinear component
  - $\boldsymbol{g} \in \mathbb{R}^k_+$: Curvature of the nonlinear component

- A second order stationary and isotropic Gaussian random field

$$m(\boldsymbol{z}) = m(\underbrace{\boldsymbol{z}}_{\substack{\text{location:} \\ \boldsymbol{z} \in \mathbb{R}^k}}, \underbrace{\omega}_{\substack{\text{elementary outcome:} \\ \omega \in \Omega}})$$

with

$$\begin{aligned} m(\boldsymbol{z}) &\sim \mathsf{N}(0,1), \\ \mathsf{E}(m(\boldsymbol{z})m(\boldsymbol{w})) &= \boldsymbol{H}(h), \end{aligned}$$

and $h \equiv \frac{1}{2}[(\boldsymbol{z}-\boldsymbol{w})'(\boldsymbol{z}-\boldsymbol{w})]^{\frac{1}{2}}$

- The spherical covariance function (typical element for $t, s = 1, ..., T$)

$$
\begin{aligned}
\boldsymbol{H}(h_{ts}) &= \begin{cases} G_{k-1}(h_{ts}, 1)/G_{k-1}(0, 1) & \text{if } h_{ts} \le 1 \\ 0 & \text{if } h_{ts} > 1 \end{cases} \\
G_k(h_{ts}, r) &= \int_{h_{ts}}^{r} (r^2 - w^2)^{k/2} dw \\
h_{ts} &= \frac{1}{2}[(\boldsymbol{g} \odot \boldsymbol{x}_{t\cdot} - \boldsymbol{g} \odot \boldsymbol{x}_{s\cdot})'(\boldsymbol{g} \odot \boldsymbol{x}_{t\cdot} - \boldsymbol{g} \odot \boldsymbol{x}_{s\cdot})]^{\frac{1}{2}}
\end{aligned}
$$

- Examples:

$$
\begin{aligned}
\boldsymbol{H}(h_{ts}) &= \begin{cases} 1 - h_{ts} & if\ h_{ts} \le 1 \\ 0 & if\ h_{ts} > 1 \end{cases} \quad \text{for } k = 1. \\
\boldsymbol{H}(h_{ts}) &= \begin{cases} 1 - \frac{2}{\pi}[h_{ts}(1 - h_{ts}^2)^{\frac{1}{2}} + \sin^{-1}(h_{ts})] & if\ h_{ts} \le 1 \\ 0 & if\ h_{ts} > 1 \end{cases} \quad \text{for } k = 2.
\end{aligned}
$$

Spline Smoothers: A Gaussian Random Field Model

- Consider the random field regression model

$$\begin{aligned} y_t &= f(x_t) + \widetilde{\epsilon}_t, \quad t = 1, 2, .., T. \\ f(x_t) &= x_t \beta_1 + \lambda m(x_t), \quad x_t \in [0, 1] \\ m(x_t) &= \int_0^1 (x_t - h)_+ dW(h) \\ \mathsf{E}\left(m(x_t) m(x_s)\right) &= \int_0^1 (x_t - h)_+ (x_s - h)_+ dh \end{aligned}$$

- $W(h)$ is a Wiener process with unit dispersion

- $m(x_t) \sim \mathsf{N}\left(0, \frac{1}{3} |x_t|^3\right)$

Wahba (1978): Let $\widehat{f}_S(x_t)$ be the minimum variance, unbiased, linear estimator of $f(x_t)$ given the observations $\{y_t, x_t\}_{t=1}^{T}$. Then $\widehat{f}_S(x_t)$ will equal the *cubic spline smoother*, and be the minimizer of

$$\frac{1}{T}\sum_{t=1}^{T}(y_t - f(x_t))^2 + S\int_0^1 \mathrm{D}^2 f(u)du$$

with

$$S = \sigma^2(T\lambda^2)^{-1}$$

and where $\mathrm{D}^2 f(u)$ denotes the second order derivative of $f(u)$ with respect to u.

Key Question: What random field is appropriate to employ, or what is the logical way to generalize the univariate Brownian motion to k dimensions?

Wecker and Ansley (1983): View the latent stochastic process as a part of the actual data-generating process, with the properties of the latent process regarded as population parameters to be estimated by maximum likelihood.

Hamilton (2001): View the latent stochastic process as an *approximation* of the actual data generating process, with the properties of the latent process regarded as population parameters to be estimated by maximum likelihood (or Bayesian methods).

Examples: Accuracy of the likelihood approach

- AR:

$$y_t = 0.6 y_{t-1} + \epsilon_t$$

- TAR

$$y_t = 0.9 y_{t-1} 1\left(|y_{t-1}| \leq 1\right) - 0.3 y_{t-1} 1\left(|y_{t-1}| > 1\right) + \epsilon_t$$

- LSTAR

$$\begin{aligned} y_t &= 1.8 y_{t-1} - 1.06 y_{t-2} \\ &\quad + \left(0.02 - 0.9 y_{t-1} + 0.795 y_{t-2}\right) F(y_{t-1}) + \epsilon_t \\ F(y_{t-1}) &= \left[1 + \exp\left(-100\left(y_{t-1} - 0.02\right)\right)\right]^{-1} \end{aligned}$$

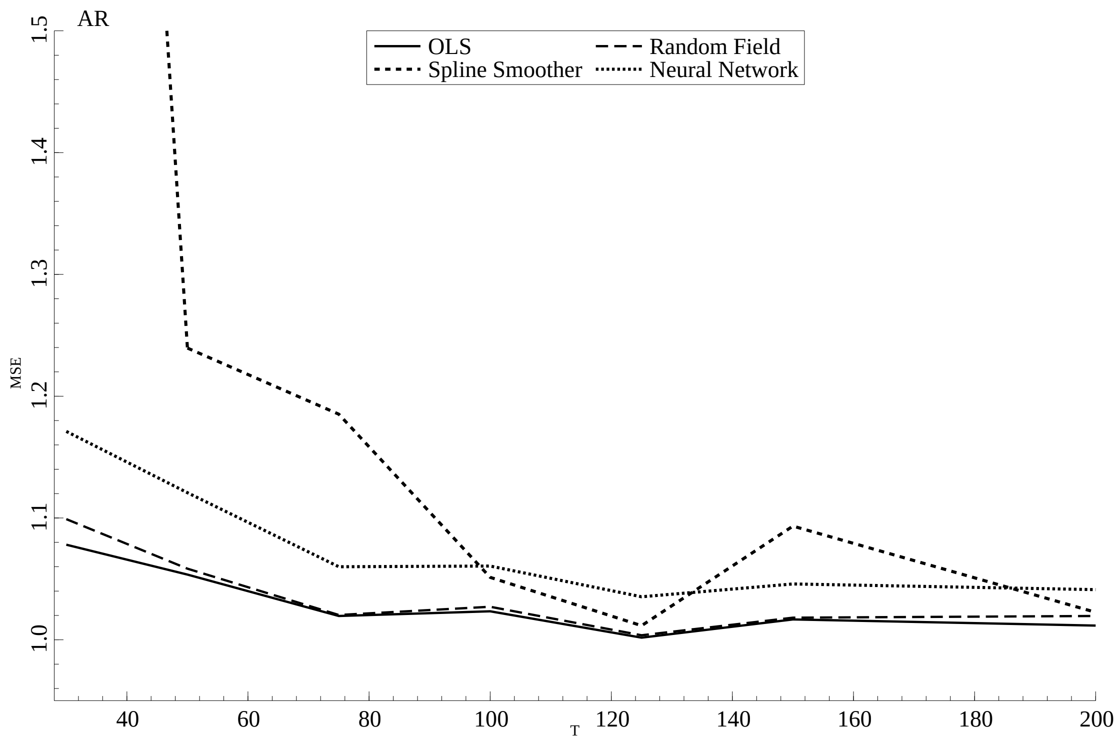

AR
OLS
Random Field
Spline Smoother
Neural Network
MSE
1.0
1.1
1.2
1.3
1.4
1.5
T
40
60
80
100
120
140
160
180
200

Figure 1: AR

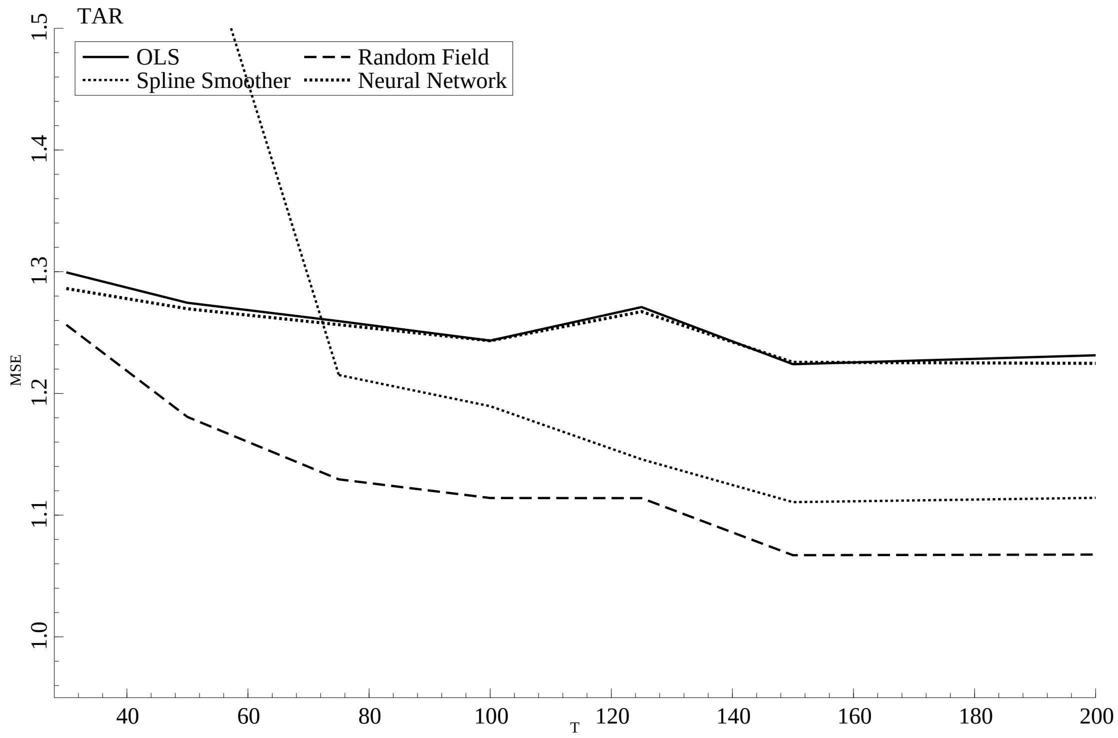

TAR
OLS
Random Field
Spline Smoother
Neural Network
MSE
T
1.0
1.1
1.2
1.3
1.4
1.5
40
60
80
100
120
140
160
180
200

Figure 2: TAR

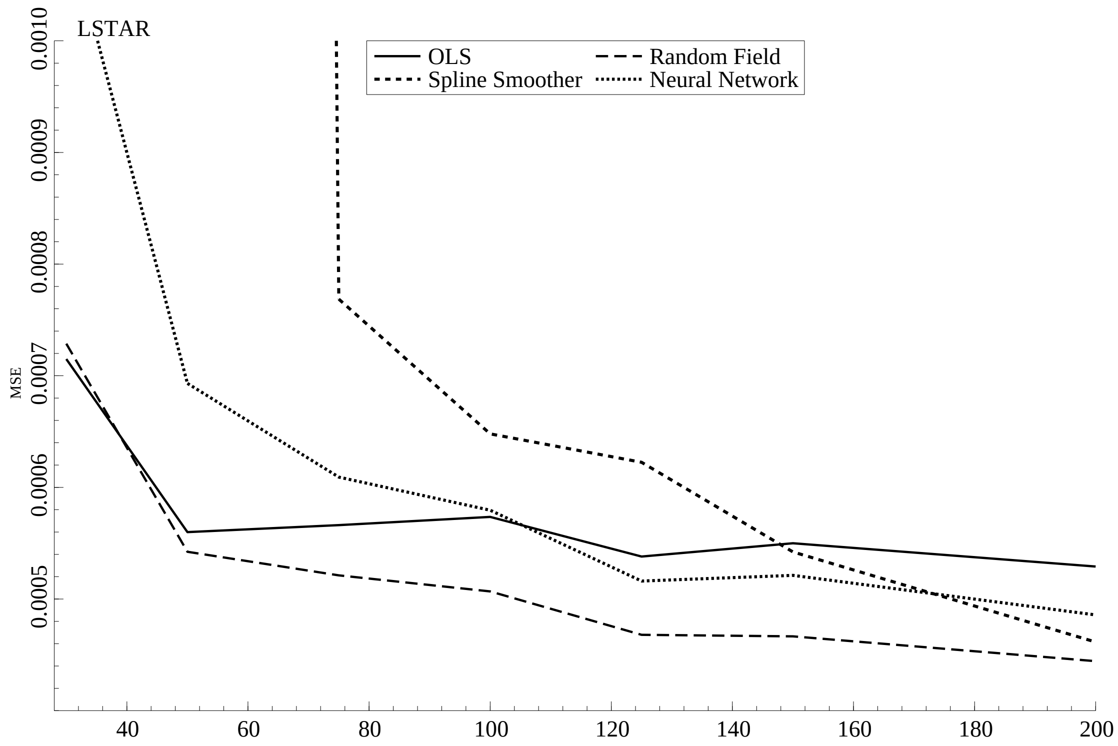


Figure 3: LSTAR

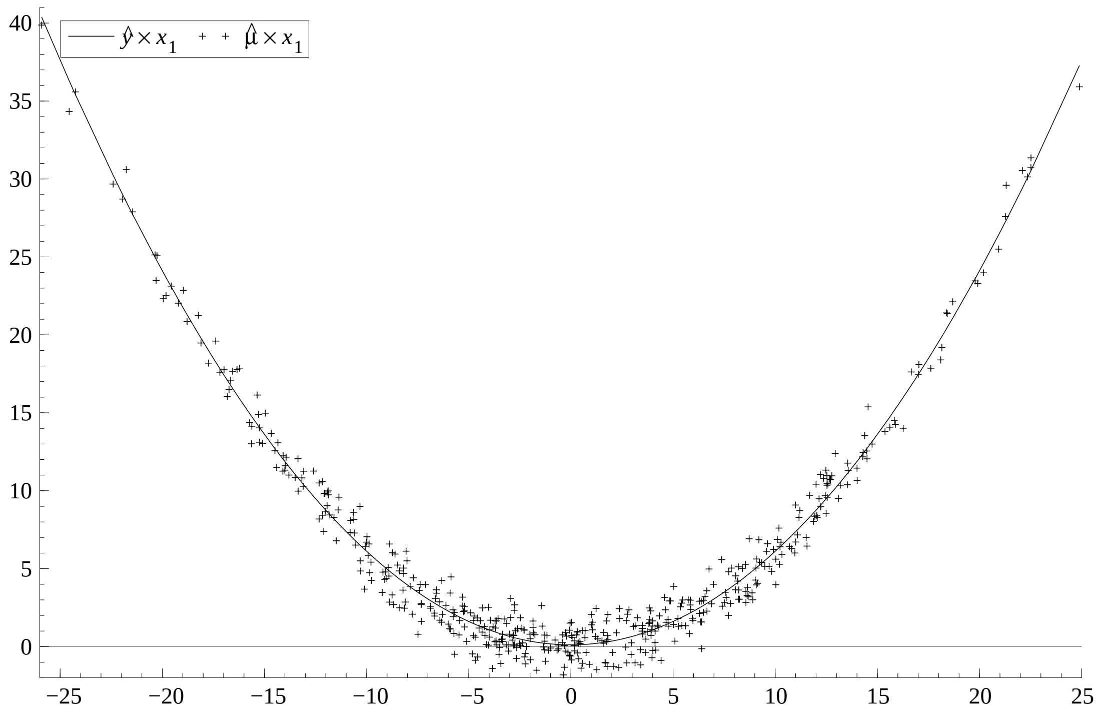


Figure 4: $y_t = 0.06x_{1t}^2 + 0.4\bar{x}_2 + \varepsilon_t$

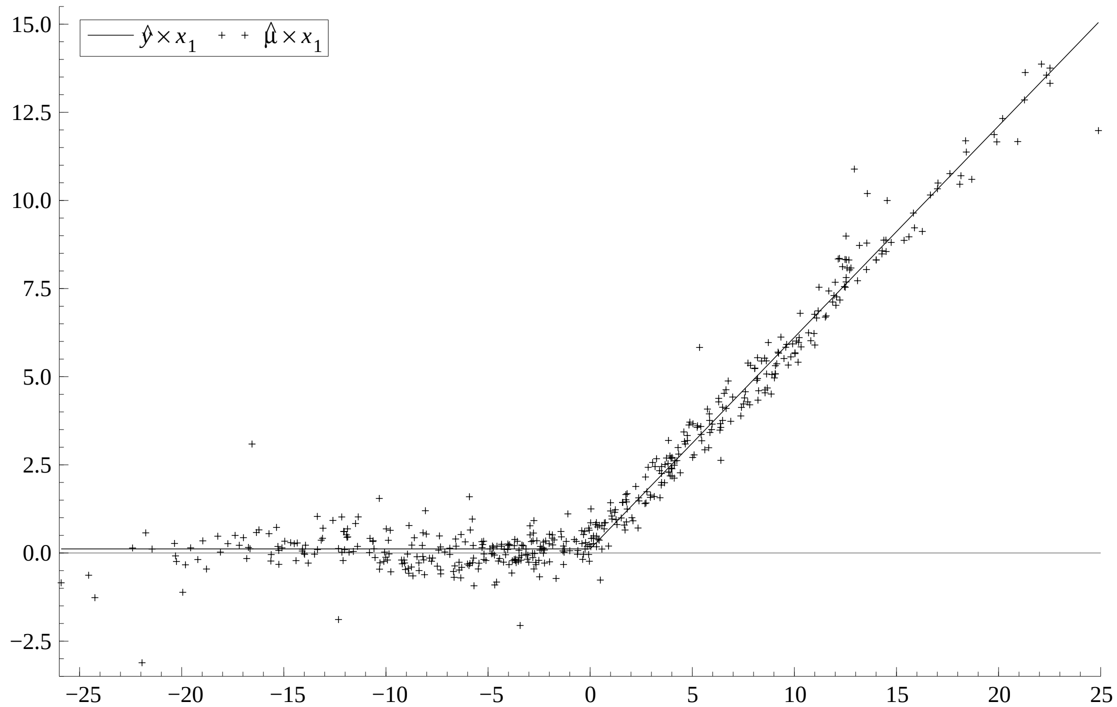


Figure 5: $y_t = 0.6x_{1t}1_{x_{1t}>0} + 0.4\bar{x}_2 + \varepsilon_t$

Estimation

- The GLS representation of the random field writes $((\lambda_1, \sigma_1) \equiv (\lambda^2, \sigma^2))$

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &\sim \mathsf{N}(\mathbf{0}, \lambda_1 \boldsymbol{H} + \sigma_1 \boldsymbol{I}) \end{aligned}$$

  with joint log likelihood function

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

Mardia and Marshall (1984): Provides conditions under which the maximum likelihood estimates of $\boldsymbol{\theta} = (\lambda_1, \sigma_1)'$ and $\boldsymbol{\beta}$ will be weakly consistent and asymptotically normal distributed.

- Essential requirements
  - The random field regression model is the actual data generating process
  - $\boldsymbol{x}_{t\cdot}$ is sampled on a grid with fixed spacing
- Assumptions seem too restrictive for most economic applications!

Hamilton (2001): Provides condition under which the maximum likelihood approach yields a consistent estimator of the conditional mean function ($\widehat{\mu}_t$), when the actual data generating mechanism is $y_t = \psi(\boldsymbol{x}_{t\cdot}) + e_t$ given by

$$\widehat{\mu}_t = x_{t\cdot}\widehat{\boldsymbol{\beta}}_T + \widehat{\lambda}_1 \widehat{\boldsymbol{H}} \left(\widehat{\lambda}_1 \widehat{\boldsymbol{H}} + \widehat{\sigma}_1 \boldsymbol{I}_T \right)^{-1} \left(y_t - x_{t\cdot}\widehat{\boldsymbol{\beta}}_T \right)$$

- Essential requirements
  - Smoothness of $\psi(\cdot)$
  - Denseness of $\boldsymbol{x}_{t\cdot}$

- Recall the objectives:
  - Obtain consistent estimators of the unknown function $\psi(\cdot)$?
  - Develop hypothesis testing procedures regarding the shape of $\psi(\cdot)$?
  - Find hypothesis testing procedures in relation to the elements of $\boldsymbol{x}_{t\cdot}$?

Requirement: Characterize the asymptotic distribution of the estimated parameters.

Proof of Consistency

Assumption 1 (parameter space): The parameter vector $\boldsymbol{g} = (g_1, g_2, ..., g_k)'$ consists of predetermined constants. In particular, $g_i = \frac{1}{2\sqrt{ks_i^2}}$, for $i = 1, \ldots, k$, 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 ith explanatory variable.

Assumption 2 (compactness, parameter space): 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 $\overline{\boldsymbol{\theta}} = (\overline{\lambda}, \overline{\sigma})$, such that $\lambda_1 \in [\underline{\lambda}, \overline{\lambda}]$ and $\sigma_1 \in [\underline{\sigma}, \overline{\sigma}]$.

Definition 1 (dgp, denseness): The deterministic sequence $\{\boldsymbol{x}_{t\cdot}\}$, with $\boldsymbol{x}_{t\cdot} \in A$ 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 $f : A \to \mathbb{R}$ such that $f(\boldsymbol{x}) > 0$ for all $\boldsymbol{x}$ and such that for any $\epsilon > 0$, and any continuous $\phi : A \times A \times \Gamma_0 \to \mathbb{R}$ there exists an N such that for all $T \geq N$,

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

- Denseness versus i.i.d?

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

$$l(\boldsymbol{x}_{t\cdot}; \lambda_1) = \int_A \phi(\boldsymbol{x}_{t\cdot}, \boldsymbol{x}; \lambda_1) f(\boldsymbol{x}) d\boldsymbol{x}$$

Assumption 3 (dgp, representability, distribution): Let $\boldsymbol{x}_{t\cdot}$ be dense according to Definition 1 and let y_t be generated according to $y_t = \psi(\boldsymbol{x}_{t\cdot}) + e_t$, for $t = 1, 2, ..., T$, where $\psi : \mathbb{R}^k \to \mathbb{R}$ is given as $\psi(\boldsymbol{x}_{t\cdot}) = \boldsymbol{x}_{t\cdot}\boldsymbol{\beta} + l(\boldsymbol{x}_{t\cdot}; \lambda_1)$ and $l(\boldsymbol{x}_{t\cdot}; \lambda_1)$ 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.

The link between the DGP and the approximating process?

- Suggest a function $\phi(\cdot, \cdot; \lambda_1)$ such that $l(\cdot)$ is representable with respect to $\phi(\cdot)$ and compute (based on the sample)

$$l_T(\boldsymbol{x}_{t\cdot}; \lambda_1) = \frac{1}{T} \sum_{s=1}^{T} \phi(\boldsymbol{x}_{t\cdot}, \boldsymbol{x}_{s\cdot}; \lambda_1),$$

- Provided that $\boldsymbol{x}_{t\cdot}$ is dense, $\lim_{T\to\infty} l_T(\boldsymbol{x}_{t\cdot}; \lambda_1) \xrightarrow{p} l(\boldsymbol{x}_{t\cdot}; \lambda_1)$ uniformly on $A \times \Gamma_0$ for $\forall t$

- Hamilton's key result. If $l(\boldsymbol{x}_{t\cdot}; \lambda_1)$ is a Taylor series or Fourier sine series it is representable in terms of elements of the spherical covariance function.

Assumption 4 (dgp, second order moments): Let $\boldsymbol{\Psi}(\boldsymbol{X}) = (\psi(\boldsymbol{x}_{1\cdot}), ..., \psi(\boldsymbol{x}_{T\cdot}))'$ where $\psi(\boldsymbol{x}_{t\cdot})$ is defined as in Assumption 3. Assume: *i*. $\lim_{T\to\infty} \frac{1}{T}\boldsymbol{X}'\boldsymbol{X}$ converges to a finite nonsingular matrix. *ii*. $\lim_{T\to\infty} \frac{1}{T}\boldsymbol{\Psi}(\boldsymbol{X})'\boldsymbol{\Psi}(\boldsymbol{X})$ converges to a finite scalar uniformly in $\boldsymbol{\beta}$. *iii*. $\lim_{T\to\infty} \frac{1}{T}\boldsymbol{X}'\boldsymbol{\Psi}(\boldsymbol{X})$ converges to a finite $(k,1)$ vector uniformly in $\boldsymbol{\beta}$.

Consistency of the misspecified OLS estimator: White (1980) shows that under A1-A4 the OLS estimator $\widehat{\boldsymbol{\beta}}$ is consistent with respect to a fixed population magnitude $\boldsymbol{\beta}^* \in B \subseteq \mathbb{R}^k$ given as

$$\boldsymbol{\beta}^* = \left(\lim_{T\to\infty} \frac{1}{T}\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \lim_{T\to\infty} \frac{1}{T}\boldsymbol{X}'\boldsymbol{\Psi}(\boldsymbol{X}).$$

Theorem 1 (consistency): Let Θ and B be sets in finite-dimensional real spaces and let $\{Q_T(\boldsymbol{\theta}, \boldsymbol{\beta})\}_T$ be the sequence of objective functions. Then $\widehat{\boldsymbol{\theta}} = \arg\max\, Q_T(\boldsymbol{\theta}, \boldsymbol{\beta}^*)$ satisfies $\widehat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}^*$ if the following conditions hold: *i.* Θ and B are compact sets. *ii.* $\widehat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}^* \in \text{int}(B)$. *iii.* $Q_T(\boldsymbol{\theta}, \boldsymbol{\beta})$ is a continuous measurable function for $\forall T$.

iv. $\exists Q^*(\boldsymbol{\theta}, \boldsymbol{\beta})$ such that $Q_T(\boldsymbol{\theta}, \boldsymbol{\beta}) \xrightarrow{p} Q^*(\boldsymbol{\theta}, \boldsymbol{\beta})$ as $T \to \infty$ uniformly on $\Theta \times B$.

v. There exists a unique maximizer $\boldsymbol{\theta}^* \in \text{int}(\Theta)$ of $Q^*(\boldsymbol{\theta}, \boldsymbol{\beta}^*)$.

Proof of condition iv.

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

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

The "new" objective function: Ignoring the constant term and after dividing by T, the log likelihood function can be represented as a sum of double arrays, i.e., as

$$\begin{aligned} Q_T(\boldsymbol{\theta}, \boldsymbol{\beta}) &= -\frac{1}{2T}\sum_{t=1}^{T} q_{tT}(\boldsymbol{\theta}, \boldsymbol{\beta}) \\ q_{tT}(\boldsymbol{\theta}, \boldsymbol{\beta}) &= \log(\lambda_1 h_{tT} + \sigma_1) + w_{tT}^2 (\lambda_1 h_{tT} + \sigma_1)^{-1}, \end{aligned}$$

- w_{tT} is a "rotation" of residuals ($\boldsymbol{w} = \boldsymbol{P}\boldsymbol{v}$) such that $w_{tT} \sim \mathrm{IN}(b_{tT}, \sigma_e^2)$ where $\boldsymbol{b} = \boldsymbol{P}\,(\boldsymbol{\Psi}(\boldsymbol{X}) - \boldsymbol{X}\boldsymbol{\beta})$

- Challenge: Many properties of h_{tT} and w_{tT} unknown $\rightarrow$ Difficult (impossible!) to find suitable LLN (CLT).

Theorem 4: Let Assumptions 1 - 4 hold and define $Q_T^*(\boldsymbol{\theta}, \boldsymbol{\beta}) = \mathsf{E}\left(Q_T\left(\boldsymbol{\theta}, \boldsymbol{\beta}\right)\right)$. Then

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

where

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

To complete the proof of condition iv. we need to show that $Q_T^*(\boldsymbol{\theta}, \boldsymbol{\beta})$ has a limit uniformly on $\Theta \times B$ as $T \to \infty$.

Proposition 1 Consider $\Psi(\boldsymbol{X}) = (\psi(\boldsymbol{x}_1), ..., \psi(\boldsymbol{x}_T))'$ and write

$$\begin{aligned} \boldsymbol{S} &= \Psi(\boldsymbol{X}) - \boldsymbol{\mu}(\boldsymbol{X}) \\ &= \Psi(\boldsymbol{X}) - \left(\boldsymbol{X}\boldsymbol{\beta} + \lambda_1 \boldsymbol{H} \left(\lambda_1 \boldsymbol{H} + \sigma_1 \boldsymbol{I}_T\right)^{-1} (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\right) \end{aligned}$$

Then

$$\frac{1}{T}\mathsf{E}(\boldsymbol{S}'\boldsymbol{S}) = \frac{1}{T}\sum_{t=1}^{T} \frac{\sigma_1^2 b_t^2}{(\lambda_1 h_t + \sigma_1)^2} + \frac{1}{T}\sum_{t=1}^{T} \frac{\sigma_e^2 \lambda_1^2 h_t^2}{(\lambda_1 h_t + \sigma_1)^2}$$

From Hamilton (2001) $\lim_{T\to\infty} \frac{1}{T}\mathsf{E}(\boldsymbol{S}'\boldsymbol{S}) = 0$ (uniformly in $\boldsymbol{\theta}, \boldsymbol{\beta}$ and $\boldsymbol{X}$), hence

$$\begin{aligned} \lim_{T\to\infty} \sup_{\Theta\times B} \frac{1}{T}\sum_{t=1}^{T} \frac{b_t^2}{(\lambda_1 h_t + \sigma_1)^2} &\to 0, \\ \lim_{T\to\infty} \sup_{\Theta\times B} \frac{1}{T}\sum_{t=1}^{T} \frac{h_t^2}{(\lambda_1 h_t + \sigma_1)^2} &\to 0. \end{aligned}$$

Proposition 2: Define $R_T \equiv \frac{1}{T}\sum_{t=1}^{T} \log(\lambda_1 h_t + \sigma_1)$, $U_T \equiv \frac{1}{T}\sum_{t=1}^{T} \frac{b_t^2}{\lambda_1 h_t + \sigma_1}$, $R_T^{i.j} = \frac{\partial^{i+j} R_T}{(\partial\lambda_1)^i(\partial\sigma_1)^j}$, and $U_T^{i.j} = \frac{\partial^{i+j} U_T}{(\partial\lambda_1)^i(\partial\sigma_1)^j}$. Furthermore, denote $R^{i.j} = \lim_{T\to\infty} \sup_{\Theta} R_T^{i.j}$ and $U^{i.j} = \lim_{T\to\infty} \sup_{\Theta\times B} U_T^{i.j}$. 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

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

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 $\mathcal{F}_T^{I.J}$, defined as

$$\mathcal{F}_T^{I.J} = \left\{ \{R_T\}_T , \left\{ R_T^{i.j} \right\}_T , \{U_T\}_T , \left\{ U_T^{i.j} \right\}_T ; \ i = 0, 1, ...I, j = 0, 1, ..., J \right\}.$$

Given the assumptions of Proposition 1, each element of $\mathcal{F}_T^{4.4}$ is composed of a sequence of uniformly bounded functions on $\Theta \times B$.

Lemma 7: Let $\mathcal{F}_T^{I.J}$ be defined as in Lemma 6. Given the assumptions of Proposition 1, each element of $\mathcal{F}_T^{3.3}$ is composed of a sequence of equicontinuous functions on $\Theta \times B$.

Rudin (1976) Theorem 7.16 : Uniform convergence of a sequence of functions implies the convergence of the sequence of the integrated functions. Convergence of the sequence of differentiated functions is not guaranteed.

Theorem 5: Let $\{f_n(u)\}_n$ and $\{\mathsf{D} f_n(u)\}_n$ be two equicontinuous sequences of functions on a compact set. Then, if $f_n(u)$ converges pointwisely, both $f_n(u)$ and $\mathsf{D} f_n(u)$ converges uniformly. In particular, if $\lim_{n\to\infty} f_n(u) = f_0(u)$, then

$$\lim_{n\to\infty} \mathsf{D} f_n(u) = \mathsf{D} f_0(u)$$

uniformly.

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

$$\left\{\{R_T\}_T, \left\{R_T^{i.j}\right\}_T; i = 0, 1, ...4, j = 0, 1, ..., 4\right\},$$

and

$$\left\{\{U_T\}_T, \left\{U_T^{i.j}\right\}_T; i = 0, 1, ...4, j = 0, 1, ..., 4\right\},$$

converges uniformly on $\Theta \times B$, then each element of $F_T^{3.3}$ converges uniformly on $\Theta \times B$.

Remark:

- To show convergence of $Q_T^*(\boldsymbol{\theta}, \boldsymbol{\beta})$, we need only the convergence of the 3 element from the $F_T^{3,3}$ family.

- Other elements will be frequently used in the proofs of the convergence of the gradient and Hessian of $Q_T^*(\boldsymbol{\theta}, \boldsymbol{\beta})$

- No closed form solutions of the limit of $Q_T^*(\boldsymbol{\theta}, \boldsymbol{\beta})$ and its derivatives are provided

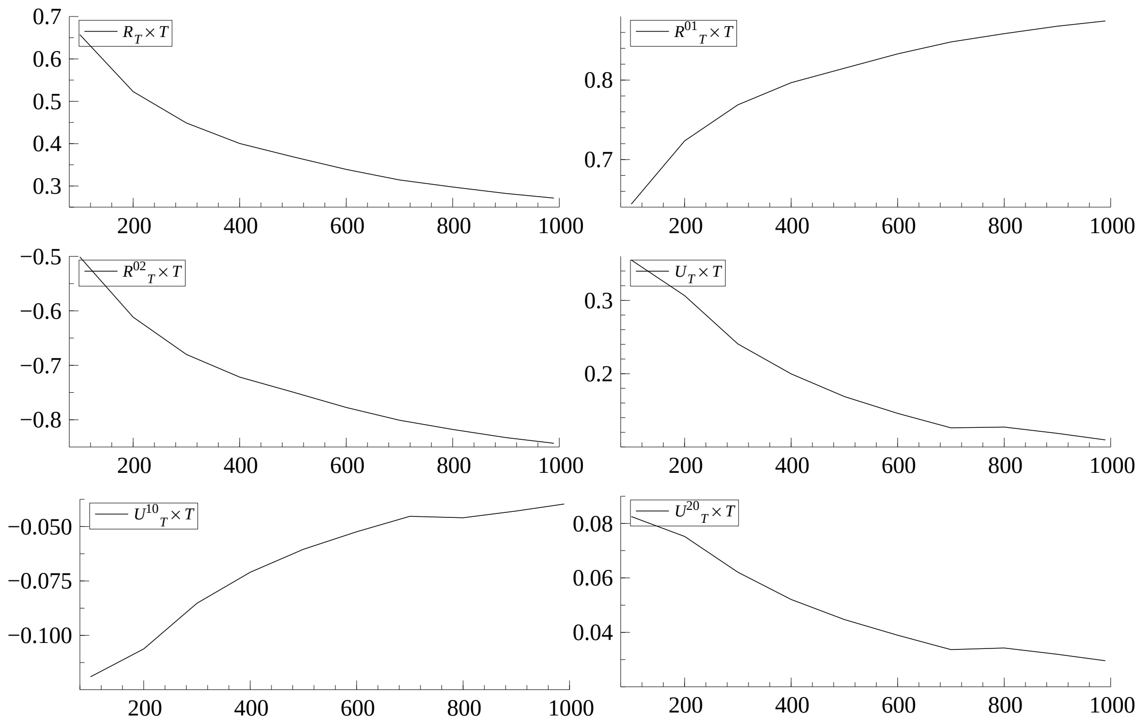


Figure 6: $y_t = 0.6x_{1t}1_{x_{1t}>0} + 0.4x_{2t} + \varepsilon_t$

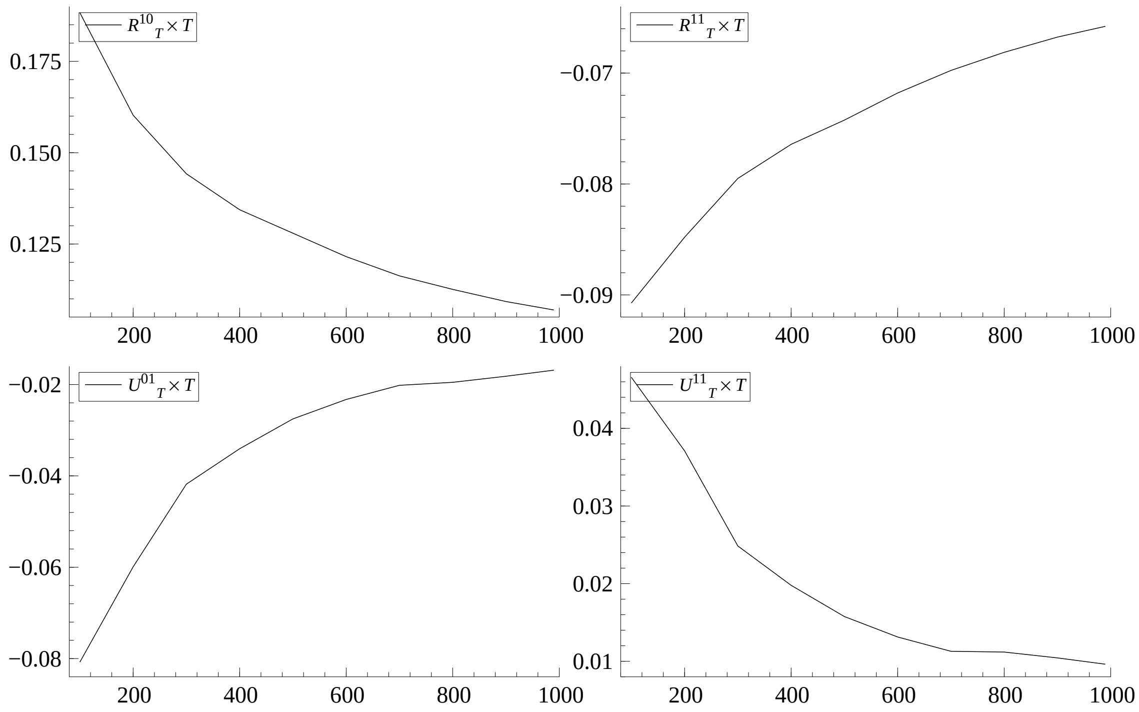


Figure 7: $y_t = 0.6x_{1t}1_{x_{1t}>0} + 0.4x_{2t} + \varepsilon_t$

Proof of condition iv. (Identification)

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

$$\lim_{T\to\infty} \boldsymbol{\mathcal{H}}(\boldsymbol{\theta}, \boldsymbol{\beta}^*) = \begin{pmatrix} \lim_{T\to\infty} -\frac{1}{2} U_T^{2.0} & 0 \\ 0 & \lim_{T\to\infty} -\frac{1}{2}\left(R_T^{0.2} + \sigma_e^2 R_T^{0.3}\right) \end{pmatrix}$$

- It is not possible to check the condition of Theorem 7 prior to estimation.

- After estimation verify the condition $\widehat{\sigma}_e^2 > \frac{1}{2}\widehat{\sigma}_1$ where $\widehat{\sigma}_e^2 = \frac{1}{T}\sum_{t=1}^{T} \widehat{\mu}_t^2 / \left(\widehat{\sigma}_1 \widehat{R}_T^{0.2}\right)$

Asymptotic Normality

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

$$\sqrt{T}\left(\widehat{\boldsymbol{\zeta}}_T - \boldsymbol{\zeta}^*\right) \xrightarrow{d} \mathsf{N}(\mathbf{0}, \boldsymbol{M}^*), \tag{1}$$

In particular,

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

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

- It can be expected that $\widehat{\boldsymbol{M}}_{11} = \boldsymbol{M}_{11}(\widehat{\boldsymbol{\zeta}}_T, \widehat{\sigma}_e^2)$ will give an over confident estimate in small samples.

- Use the sample counterparts $\mathsf{D}^2_{\boldsymbol{\theta\theta}} Q_T(\widehat{\boldsymbol{\zeta}})$, $\mathsf{D}^2_{\boldsymbol{\theta\theta}} Q_T(\widehat{\boldsymbol{\zeta}})$, $\mathsf{D}^2_{\boldsymbol{\theta\beta}} Q_T(\widehat{\boldsymbol{\zeta}})$ and the sample variance of $\boldsymbol{g}_t(\boldsymbol{\theta}, \boldsymbol{\beta})$ instead of $\frac{1}{T}\boldsymbol{\Sigma}^*$.

Simulation Experiments

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	$y_t = 3\sin(x_{1t} + x_{2t}) + e_t$

Figure 8: Asymptotic and simulated densities.

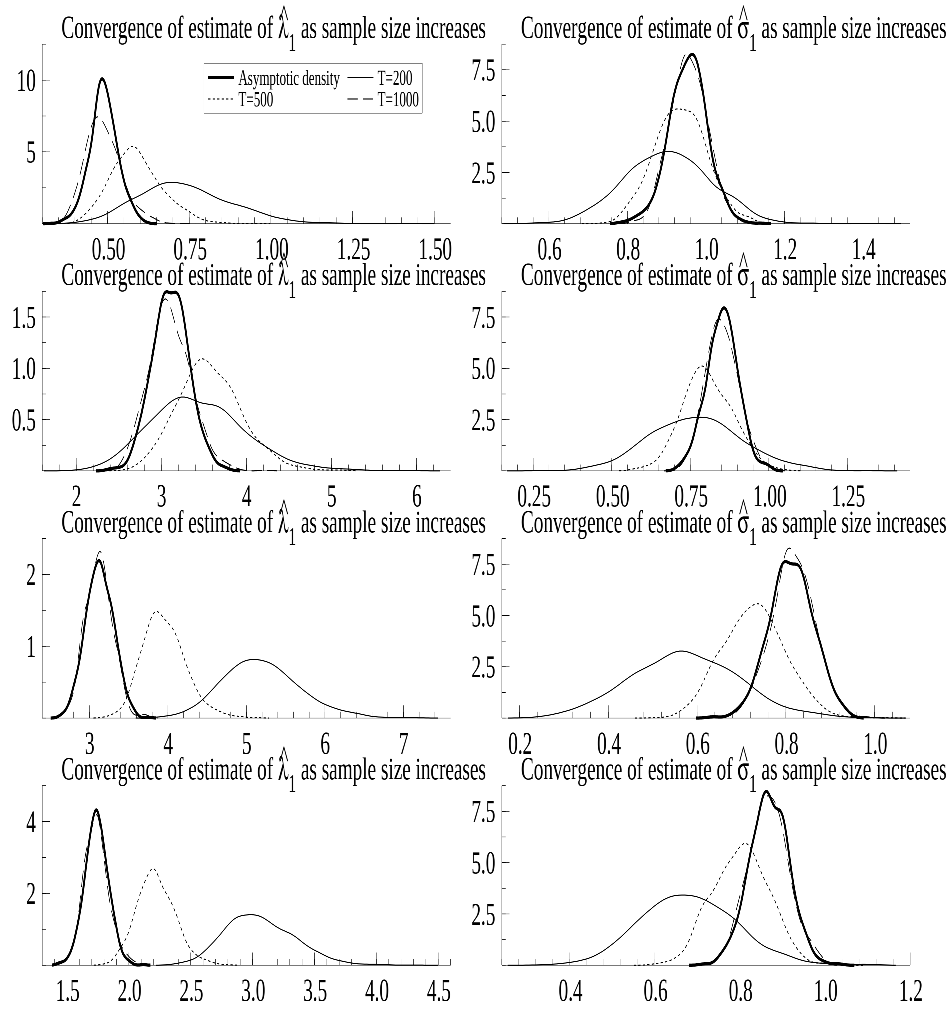


Table 2: Alternative estimates of the standard errors for $T = 200$.

Index	"$True$"	$\widehat{A\text{var}}$	Avar	$Bootstrap$	$Hessian$
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

Constructing confidence bands about the true conditional mean function:

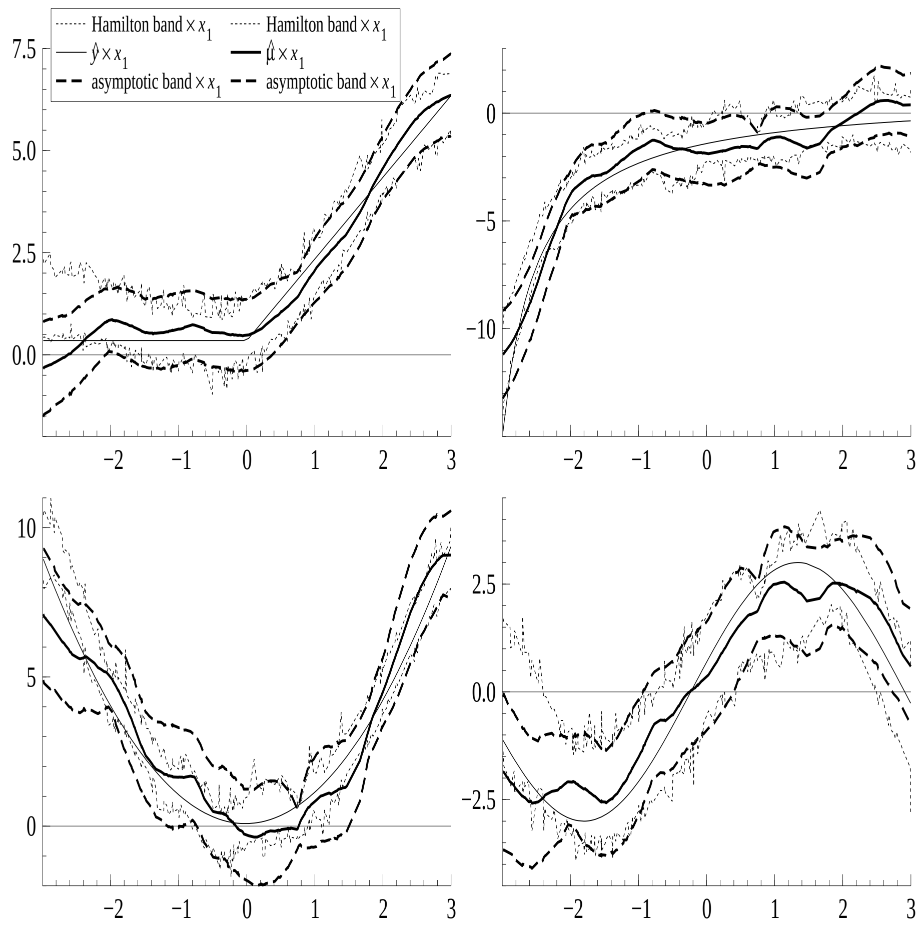
1. Fix the parameter vector $\boldsymbol{\beta}$ at $\widehat{\boldsymbol{\beta}}_T$, and sample $\boldsymbol{\theta}^j$ from N $\left(\hat{\boldsymbol{\theta}}_T, \frac{1}{T}\widehat{\boldsymbol{M}_{11}}(\widehat{\boldsymbol{\zeta}}_T, \hat{\sigma}_e^2)\right)$.

2. For each $\boldsymbol{\theta}^j$, generate the predictor $\widehat{\xi_s^j}(\boldsymbol{x}_{s\cdot}^*)$

3. Sort the values $\widehat{\xi_s^j}(\boldsymbol{x}_{s\cdot}^*)$ in an ascending order as $\left(\widehat{\xi_s^{j}}^{1}(\boldsymbol{x}_{s\cdot}^*), \ldots, \widehat{\xi_s^{j}}^{N}(\boldsymbol{x}_{s\cdot}^*)\right)$

4. The confidence band is then given as $\left(\widehat{\xi_s^{j}}^{\alpha/2N}(\boldsymbol{x}_{s\cdot}^*), \widehat{\xi_s^{j}}^{(1-\alpha/2)N)}(\boldsymbol{x}_{s\cdot}^*)\right)$ for $s = 1, 2, ..., S$.

Figure 9: 90% confidence band when $x_2 = \bar{x}_2$



Summary

- Large sample behavior of the maximum likelihood estimates of the unknown parameters in a non-locally misspecified random field regression model is characterized.

- Based on some "new" 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.

- Confidence bands constructed using the asymptotic distribution of the estimated parameters has good coverage of actual conditional mean function.

- 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.