

# NONPARAMETRIC ESTIMATION OF MARGINAL EFFECTS IN REGRESSION-SPLINE RANDOM EFFECTS MODELS

SHUJIE MA, JEFFREY S. RACINE, AND AMAN ULLAH

ABSTRACT. We consider a B-spline regression approach towards nonparametric modelling of a random effects (error component) model. We focus our attention on the estimation of marginal effects (derivatives) and their asymptotic properties. Theoretical underpinnings are provided, finite-sample performance is evaluated via Monte Carlo simulation, and an application that examines the contribution of different types of public infrastructure on private production is investigated using panel data comprising the 48 contiguous states in the US over the period 1970-1986.

JEL Classification: C14, C23.

## 1. INTRODUCTION

This paper is concerned with nonparametric estimation of marginal effects in so-called ‘random effects’ panel data settings. There exists an extensive literature on nonparametric estimation of conditional mean functions using regression spline methods, and a principal focus of this literature is on estimation of spline coefficients in particular. However, applied econometricians are primarily interested in ‘marginal effects’, or equivalently, derivatives of the conditional mean function, and this has received nowhere near the attention that estimation of the conditional mean function itself has garnered. Second, in empirical work econometricians often find the covariance matrix of errors in their regression spline models to be a non-scalar identity matrix whereas the existing literature on this tends to assume a scalar covariance matrix. Third, nonparametric regression splines have often been considered with only a single-predictor instead of multiple-predictors. In view of all these, an objective of this paper is to develop the estimation of marginal effects for the nonparametric spline regression model under a non-scalar covariance error matrix and with multiple-predictors. While this is done in the context of panel random effect models, the results we develop are general and

---

*Date:* September 20, 2019.

*Key words and phrases.* GLS, Consistent, B-spline.

Ullah gratefully acknowledges research support from the academic senate, UCR. Racine would like to gratefully acknowledge support from the Social Sciences and Humanities Research Council of Canada (SSHRC:www.sshrc.ca), and the Shared Hierarchical Academic Research Computing Network (SHARCNET:www.sharcnet.ca). Ma’s research was partially supported by NSF grants DMS 1306972 and 1712558. We would like to thank John Kealey for his helpful comments and feedback, and Wei Lin for discussions on the subject matter of this paper. We are also grateful to the editor and two referees for their helpful comments.

applicable to a class of econometric models including cross sectional model with heteroscedasticity, regression models with serial correlation, and seemingly unrelated regression models, among others.

The parametric random effects model has been extensively studied from a theoretical perspective, and is widely used by practitioners; see Baltagi (2013) and Hsiao (2003) by way of illustration. It is well known that the generalized least squares (GLS) estimator of a random effects model is asymptotically efficient. Practitioners often find themselves in need of nonparametric methods for estimating the regression function flexibly; see Wang (2003), Wang, Carroll & Lin (2005), Henderson, Carroll & Li (2008) and Henderson et al. (2008) for nonparametric kernel regression methods with panel data or longitudinal data. Moreover, Lin, Wang, Welsh & Carroll (2004) compared smoothing splines with kernels in univariate nonparametric regression.

It is worth noting that regression splines as another popular nonparametric smoothing tool are also well-suited to the estimation of random effects models because they involve simple least squares procedures (i.e., maximum likelihood presuming normality), are simple to interpret, and are widely embraced by practitioners. For excellent reviews of regression spline methods see Stone (1985, 1994), Wang & Yang (2007, 2009), Schumaker (2007), Prautzsch, Boehm & Paluszny (2002), Ma & Yang (2011), and Haupt, Kagerer & Steiner (2014). In econometrics, there exists an extensive related literature on semiparametric and nonparametric series ('sieve') estimation; see Newey (1991, 1994, 1999), Andrews (1991), Shen & Wong (1994), Chen (2007), Horowitz (2009), and Phillips & Liao (2014) by way of illustration.

Regression splines are often restricted to settings which involve univariate unknown smooth functions, see Huang, Zhang & Zhou, Wang, Song & Wang and Ma, Huang & Tsai for their applications in analysis of longitudinal data. However, in many econometrics applications, multiple predictors are present. One such case is the multivariate panel data model with random effects, which is the focus of this paper. To this end, we propose to estimate the multivariate unknown regression function via tensor-product B-splines. We also provide a simple expression for the spline estimator of its partial derivatives, which are often required to investigate by economic applications. The results presented here are also applicable to regression splines with a general class of error covariance matrices. The theoretical properties of the proposed estimators are provided. Finite-sample performance is assessed via simulation, while an empirical application of the proposed

estimator is presented for a panel data random effects econometric model with four explanatory variables.

Related work has considered the efficient nonparametric estimation of conditional mean functions with panel data and within-group correlations via kernel smoothing. These include Wang (2003) who considered nonparametric kernel-based marginal estimation for longitudinal/clustered data, Wang et al. (2005) who studied marginal generalized semiparametric partially linear models for clustered data, Lin et al. (2004) who showed that a smoothing spline estimator is asymptotically equivalent to a recently proposed seemingly unrelated kernel estimator of the univariate unknown function in Wang (2003) for any working covariance matrix, and Henderson et al. (2008) who proposed an iterative nonparametric kernel estimator for estimating nonparametric panel data models with fixed effects, by way of illustration.

The rest of the paper proceeds as follows. Section 2 introduces the proposed procedure and outlines theoretical underpinnings; proofs are relegated to Appendix A. Section 3 assesses the finite-sample performance of the proposed approach via Monte Carlo simulation and compares it with popular parametric approaches. Section 4 considers an illustrative application, while Section 5 presents some concluding remarks.

## 2. METHODOLOGY

In this section, we provide the theoretical foundations of the proposed approach. Four main theorems are developed. Theorem 2.1 presents the asymptotic variance of the spline regression function along with the order of its bias term, which is asymptotically negligible. We then establish the asymptotic normal distribution for the spline regression function in Theorem 2.2. Theorem 2.3 presents the asymptotic variance for the estimator of the derivative of the regression spline along with the order of its bias, while Theorem 2.4 establishes the spline derivative estimator's asymptotic normality. In what follows, we assume that  $T$  is fixed, but that  $n$  goes to infinity.

We consider a nonparametric one-way error component model written as

$$(1) \quad Y_{it} = m(\mathbf{X}_{it}) + \varepsilon_{it},$$

where  $i = 1, \dots, n$ ,  $t = 1, \dots, T$ ,  $Y_{it}$  is the endogenous variable,  $\mathbf{X}_{it} = (X_{it1}, \dots, X_{itd})^\top$  is a vector of  $d$  exogenous variables, and  $m(\cdot)$  is an unknown smooth function. Assume for  $1 \leq s \leq d$ , each  $X_{its}$  is distributed on a compact interval  $[a_s, b_s]$ , and without loss of generality, we take all intervals  $[a_s, b_s] = [0, 1]$ . Moreover,  $\mathbf{X}_i = (\mathbf{X}_{i1}, \dots, \mathbf{X}_{iT})^\top$ , for  $i = 1, \dots, n$ , are independent and have the same distribution. We allow  $\varepsilon_{it}$  to follow the random effects specification

$$(2) \quad \varepsilon_{it} = u_i + v_{it},$$

where  $u_i$  is i.i.d.  $(0, \sigma_u^2)$ ,  $v_{it}$  is i.i.d.  $(0, \sigma_v^2)$ ,  $u_i$  and  $v_{it}$  are uncorrelated for all  $i$  and  $t$ , and  $\sigma_u^2$  and  $\sigma_v^2$  satisfy  $0 < \sigma_u^2 < \infty$  and  $0 < \sigma_v^2 < \infty$ . Let  $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \dots, \varepsilon_{iT})^\top$  be a  $T \times 1$  vector, and  $\boldsymbol{\varepsilon}_i$ , for  $i = 1, \dots, n$ , are independent. Then  $\mathbf{V} \equiv E(\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^\top)$  takes the form

$$(3) \quad \mathbf{V} = \sigma_v^2 \mathbf{I}_T + \sigma_u^2 \mathbf{1}_T \mathbf{1}_T^\top,$$

where  $\mathbf{I}_T$  is an identity matrix of dimension  $T$  and  $\mathbf{1}_T$  is a  $T \times 1$  column vector of ones. The covariance matrix for  $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1^\top, \dots, \boldsymbol{\varepsilon}_n^\top)^\top$  is

$$(4) \quad \boldsymbol{\Omega} = E(\boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^\top) = \mathbf{I}_N \otimes \mathbf{V}, \boldsymbol{\Omega}^{-1} = \mathbf{I}_N \otimes \mathbf{V}^{-1},$$

where  $\mathbf{A} \otimes \mathbf{B}$  is the Kronecker product of two matrices  $\mathbf{A}$  and  $\mathbf{B}$ . By simple linear algebra,  $\mathbf{V}^{-1} = (V_{it'})_{t,t'=1}^\top = V_1 \mathbf{I}_T + V_2 \mathbf{1}_T \mathbf{1}_T^\top$  with  $V_1 = \sigma_v^{-2}$  and  $V_2 = -(\sigma_v^2 + \sigma_u^2 T)^{-1} \sigma_u^2 \sigma_v^{-2}$ .

We use regression B-splines to estimate the mean function  $m(\cdot)$  and its first derivative. Let  $N = N_n$  be the number of interior knots and let  $q$  be the spline order. Divide  $[0, 1]$  into  $(N + 1)$  subintervals  $I_j = [\chi_j, \chi_{j+1})$ ,  $j = 0, \dots, N - 1$ ,  $I_N = [\chi_N, 1]$ , where  $\{\chi_j\}_{j=1}^N$  is a sequence of interior knots, given as

$$\chi_{-(q-1)} = \dots = \chi_0 = 0 < \chi_1 < \dots < \chi_N < 1 = \chi_{N+1} = \dots = \chi_{N+q}.$$

Equally-spaced knots are used in order to simplify the proof, but our asymptotic results can be extended to cover alternative regular knot sequences. Define the  $q$ -th order B-spline basis as  $B_{s,q} = \{B_j(x_s) : 1 - q \leq j \leq N\}^\top$  (de Boor (2001, p. 89)). Let  $G_{s,q} = G_{s,q}^{(q-2)}$  be the space spanned by  $B_{s,q}$ , and let  $G_q$  be the tensor product of  $G_{1,q}, \dots, G_{d,q}$ , which is the space of functions spanned

by

$$\begin{aligned} \mathcal{B}_q(\mathbf{x}) &= B_{1,q} \otimes \cdots \otimes B_{d,q} \\ &= \left[ \left\{ \prod_{s=1}^d B_{j_s,q}(x_s) : 1-q \leq j_s \leq N, 1 \leq s \leq d \right\}^\top \right]_{\mathbf{K}_n \times 1} \\ &= \left[ \{\mathcal{B}_{j_1, \dots, j_d, q}(\mathbf{x}) : 1-q \leq j_s \leq N, 1 \leq s \leq d\}^\top \right]_{\mathbf{K}_n \times 1}, \end{aligned}$$

where  $\mathbf{x} = (x_1, \dots, x_d)^\top$  and  $\mathbf{K}_n = (N+q)^d$ . Let  $\mathbf{B}_q = [\{\mathcal{B}_q(\mathbf{X}_{11}), \dots, \mathcal{B}_q(\mathbf{X}_{nT})\}^\top]_{nT \times \mathbf{K}_n}$ , where  $\mathbf{X}_{it} = (X_{it1}, \dots, X_{itd})^\top$ . Then  $m(\mathbf{x})$  can be approximated by  $\mathcal{B}_q(\mathbf{x})^\top \boldsymbol{\beta}$ , where  $\boldsymbol{\beta}$  is a  $\mathbf{K}_n \times 1$  vector. Letting  $\mathbf{Y} = \left[ \left\{ (Y_{it})_{1 \leq t \leq T, 1 \leq i \leq n} \right\}^\top \right]_{nT \times 1}$ , we estimate  $\boldsymbol{\beta}$  by minimizing the weighted least squares criterion,

$$\{\mathbf{Y} - \mathbf{B}_q \boldsymbol{\beta}\}^\top \boldsymbol{\Omega}^{-1} \{\mathbf{Y} - \mathbf{B}_q \boldsymbol{\beta}\}.$$

Then the estimator  $\hat{\boldsymbol{\beta}}$  of  $\boldsymbol{\beta}$  solves the estimating equations  $\mathbf{B}_q^\top \boldsymbol{\Omega}^{-1} \{\mathbf{Y} - \mathbf{B}_q \boldsymbol{\beta}\} = 0$ , which gives the GLS estimator

$$(5) \quad \hat{\boldsymbol{\beta}} = \left( \mathbf{B}_q^\top \boldsymbol{\Omega}^{-1} \mathbf{B}_q \right)^{-1} \mathbf{B}_q^\top \boldsymbol{\Omega}^{-1} \mathbf{Y}.$$

The estimator of  $m(\mathbf{x})$  is then given by  $\hat{m}(\mathbf{x}) = \mathcal{B}_q(\mathbf{x})^\top \hat{\boldsymbol{\beta}}$ . In de Boor (2001, p. 116), it is shown that the first derivative of a spline function can be expressed in terms of a spline of one order lower. For any function  $s(\mathbf{x}) \in G_q$  that can be expressed by  $s(\mathbf{x}) = \sum_{j_1, \dots, j_d} a_{j_1, \dots, j_d} B_{j_1, q}(x_1) \cdots B_{j_d, q}(x_d)$ , the first derivative of  $s(\mathbf{x})$  with respect to  $x_s$  is

$$\frac{\partial s}{\partial x_s}(\mathbf{x}) = \sum_{j_s=2-q}^N \sum_{1-q \leq j_{s'} \leq N, 1 \leq s' \neq s \leq d} a_{j_1, \dots, j_d}^{(1s)} B_{j_s, q-1}(x_s) \prod_{s' \neq s} B_{j_{s'}, q}(x_{s'}),$$

in which  $a_{j_1, \dots, j_d}^{(1s)} = (q-1)(a_{j_1, \dots, j_s, \dots, j_d} - a_{j_1, \dots, j_s-1, \dots, j_d}) / (\chi_{j_s+q-1} - \chi_{j_s})$ , for  $2-q \leq j_s \leq N$  and  $1 \leq s' \neq s \leq d$ ,  $1-q \leq j_{s'} \leq N$ . Let  $\mathbf{L}_n = (N+q)^{d-1}(N+q-1)$ , and

$$\mathcal{B}_{s, q-1}(\mathbf{x}) = \left[ \{B_{j_1, q}(x_1) \cdots B_{j_s, q-1}(x_s) \cdots B_{j_d, q}(x_d)\}^\top_{1-q \leq j_{s'} \leq N, s' \neq s, 2-q \leq j_s \leq N} \right]_{\mathbf{L}_n \times 1}.$$

For  $1 \leq s \leq d$ ,  $\frac{\partial m}{\partial x_s}(\mathbf{x})$ , which is the first derivative of  $m(\mathbf{x})$  with respect to  $x_s$ , is estimated as

$$\widehat{\frac{\partial m}{\partial x_s}}(\mathbf{x}) = \mathcal{B}_{s, q-1}(\mathbf{x})^\top D_s \left( \mathbf{B}_q^\top \boldsymbol{\Omega}^{-1} \mathbf{B}_q \right)^{-1} \mathbf{B}_q^\top \boldsymbol{\Omega}^{-1} \mathbf{Y},$$

in which  $D_s = \left\{ \mathbf{I}_{(N+q)^{s-1}} \otimes M_1 \otimes \mathbf{I}_{(N+q)^{d-s}} \right\}_{\mathbf{L}_n \times \mathbf{K}_n}$ , and

$$M_1 = (q-1) \begin{pmatrix} \frac{-1}{\chi_1 - \chi_{2-q}} & \frac{1}{\chi_1 - \chi_{2-q}} & 0 & \cdots & 0 \\ 0 & \frac{-1}{\chi_2 - \chi_{3-q}} & \frac{1}{\chi_2 - \chi_{3-q}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{-1}{\chi_{N+q-1} - \chi_N} & \frac{1}{\chi_{N+q-1} - \chi_N} \end{pmatrix}_{(N+q-1) \times (N+q)}.$$

Let  $\nabla m(\mathbf{x})$  be the gradient vector of  $m(\mathbf{x})$ . The estimator of  $\nabla m(\mathbf{x})$  is

$$\begin{aligned} \widehat{\nabla m}(\mathbf{x}) &= \left\{ \frac{\widehat{\partial m}}{\partial x_1}(\mathbf{x}), \dots, \frac{\widehat{\partial m}}{\partial x_d}(\mathbf{x}) \right\}^\top \\ (6) \quad &= \mathcal{B}_{q-1}^*(\mathbf{x})^\top \left( \mathbf{B}_q^\top \boldsymbol{\Omega}^{-1} \mathbf{B}_q \right)^{-1} \mathbf{B}_q^\top \boldsymbol{\Omega}^{-1} \mathbf{Y}, \end{aligned}$$

in which  $\mathcal{B}_{q-1}^*(\mathbf{x}) = \left[ \left\{ D_{1,1}^\top \mathcal{B}_{q-1,1}(\mathbf{x}), \dots, D_{1,d}^\top \mathcal{B}_{q-1,d}(\mathbf{x}) \right\} \right]_{\mathbf{K}_n \times d}$ . For any  $\mu \in (0, 1]$ , we denote by  $C^{0,\mu} [0, 1]^d$  the space of order  $\mu$  Hölder continuous functions on  $[0, 1]^d$ , i.e.,

$$C^{0,\mu} [0, 1]^d = \left\{ \phi : \|\phi\|_{0,\mu} = \sup_{\mathbf{x} \neq \mathbf{x}', \mathbf{x}, \mathbf{x}' \in [0,1]^d} \frac{|\phi(\mathbf{x}) - \phi(\mathbf{x}')|}{\|\mathbf{x} - \mathbf{x}'\|_2^\mu} < +\infty \right\}$$

in which  $\|\mathbf{x}\|_2 = \left( \sum_{s=1}^d x_s^2 \right)^{1/2}$  is the Euclidean norm of  $\mathbf{x}$ , and  $\|\phi\|_{0,\mu}$  is the  $C^{0,\mu}$ -norm of  $\phi$ .

Given a  $d$ -tuple  $\alpha = (\alpha_1, \dots, \alpha_d)$  of non-negative integers, let  $[\alpha] = \alpha_1 + \dots + \alpha_d$  and let  $D^\alpha$  denote the differential operator defined by  $D^\alpha = \frac{\partial^{[\alpha]}}{\partial x_1^{\alpha_1} \dots \partial x_d^{\alpha_d}}$ . The assumptions needed for the asymptotic results are listed below:

- (A1) The regression function  $D^\alpha m \in C^{0,1} [0, 1]^d$  for all  $\alpha$  with  $[\alpha] = p-1$  and for a given integer  $p \geq d/2$ , the spline order satisfies  $q \geq p$ .
- (A2) The marginal density  $f(\mathbf{x})$  of  $\mathbf{X}$  satisfies  $f(\mathbf{x}) \in C [0, 1]^d$  and  $f(\mathbf{x}) \in [c_f, C_f]$  for constants  $0 < c_f \leq C_f < \infty$ .
- (A3) There exists a positive value  $\eta > 0$  and a finite positive  $M_\eta$  such that  $E(\varepsilon_{it})^{2(2+\eta)} < M_\eta$  for all  $i$  and  $t$ .
- (A4) As  $n \rightarrow \infty$ , the number of interior knots  $N$  satisfies  $N^{-1} = o\{n^{-1/(2p+d)}\}$  and  $N^d = o\{n^{1/2} (\log n)^{-1}\}$ .

Define

$$(7) \quad \boldsymbol{\Sigma} = E \left( \mathbf{B}_q^\top \boldsymbol{\Omega}^{-1} \mathbf{B}_q \right), \sigma_n^2(\mathbf{x}) = \mathbf{B}_q(\mathbf{x})^\top \boldsymbol{\Sigma}^{-1} \mathbf{B}_q(\mathbf{x}),$$

$$(8) \quad \Phi_n(\mathbf{x}) = \left\{ \mathbf{B}_{q-1}^*(\mathbf{x})^\top \boldsymbol{\Sigma}^{-1} \mathbf{B}_{q-1}^*(\mathbf{x}) \right\}_{d \times d}.$$

Assumption (A1) is a smoothness condition placed on the regression function, Assumption (A2) is a condition for the design density function, and Assumption (A3) is a moment condition on the error terms. These assumptions are commonly invoked in the nonparametric smoothing literature. The number of interior knots  $N$  increases with the sample size  $n$  and Assumption (A4) presents the order requirement for  $N$ . Theorem 2.1 provides the asymptotic variance of the spline estimator of the regression function, and the order of its bias term. Based on Assumption (A4), the asymptotic bias is negligible. Hence, as a further step, we establish in Theorem 2.2 the asymptotic normal distribution of the regression function estimator.

**Theorem 2.1.** *Under Assumptions (A1)-(A4), as  $n \rightarrow \infty$ ,*

- (i)  $\text{Var} \{ \widehat{m}(\mathbf{x}) \} = \sigma_n^2(\mathbf{x}) + o(n^{-1}N^d)$  for  $\sigma_n^2(\mathbf{x})$  given in (7), and there exist constants  $0 < c_\sigma < C_\sigma < \infty$  such that

$$c_\sigma n^{-1}N^d \leq \inf_{\mathbf{x} \in [0,1]^d} \sigma_n^2(\mathbf{x}) \leq \sup_{\mathbf{x} \in [0,1]^d} \sigma_n^2(\mathbf{x}) \leq C_\sigma n^{-1}N^d,$$

- (ii)  $\sup_{\mathbf{x} \in [0,1]^d} |E \{ \widehat{m}(\mathbf{x}) \} - m(\mathbf{x})| = O(N^{-p})$ .

**Theorem 2.2.** *Under Assumptions (A1)-(A4), as  $n \rightarrow \infty$ ,*

$$\sigma_n^{-1}(\mathbf{x}) \{ \widehat{m}(\mathbf{x}) - m(\mathbf{x}) \} \longrightarrow \mathbf{N}(0, 1).$$

Similarly, Theorem 2.3 gives the asymptotic variance of the spline estimator of the derivative of  $m(\mathbf{x})$ ,  $\widehat{\nabla m}(\mathbf{x})$ , and the order of its bias. Theorem 2.4 further establishes the asymptotic normality of  $\widehat{\nabla m}(\mathbf{x})$ .

**Theorem 2.3.** *Under Assumptions (A1)-(A4), as  $n \rightarrow \infty$ ,*

- (i)  $\text{Var} \left\{ \widehat{\nabla m}(\mathbf{x}) \right\} = \Phi_n(\mathbf{x}) + o(n^{-1}N^{2+d})$  for  $\Phi_n(\mathbf{x})$  given in (8), and there exist constants  $0 < c_\Phi < C_\Phi < \infty$  such that  $c_\Phi \mathbf{I}_d \leq n(N+1)^{-2-d} \Phi_n(\mathbf{x}) \leq C_\Phi \mathbf{I}_d$ , and

$$(ii) \sup_{\mathbf{x} \in [0,1]^d} \left\| E \left\{ \widehat{\nabla m}(\mathbf{x}) \right\} - \nabla m(\mathbf{x}) \right\|_2 = O \{ N^{-(p-1)} \}.$$

**Theorem 2.4.** *Under Assumptions (A1)-(A4), as  $n \rightarrow \infty$ ,*

$$\Phi_n^{-1/2}(\mathbf{x}) \left\{ \widehat{\nabla m}(\mathbf{x}) - \nabla m(\mathbf{x}) \right\} \longrightarrow \mathbf{N}(\mathbf{0}_d, \mathbf{I}_d),$$

in which  $\mathbf{0}_d$  is a  $d \times 1$  vector of 0's.

Proofs of these theorems are presented in Appendix A. It is worth mentioning that although we assume that the true variance-covariance matrix of the error terms has the exchangeable structure given in (3), our GLS estimation method indeed can be applied to settings with any given variance-covariance matrix satisfying a mild condition.

**2.1. Efficiency of the Proposed Approach: GLS versus OLS.** The proposed GLS estimator has smaller asymptotic variance than the ordinary least squares (OLS) estimator. To see this, using the Cholesky decomposition, we have  $\boldsymbol{\Omega} = \mathbf{C}\mathbf{C}^\top$ . Then the GLS estimator is the same as the OLS estimator by using  $\mathbf{Y}^* = \mathbf{C}^{-1}\mathbf{Y}$ ,  $\mathbf{B}_q^* = \mathbf{C}^{-1}\mathbf{B}_q$  and  $\boldsymbol{\varepsilon}^* = \mathbf{C}^{-1}\boldsymbol{\varepsilon}$ , where  $\text{Var}(\boldsymbol{\varepsilon}^*) = \mathbf{I}$ , which is

$$\widehat{\boldsymbol{\beta}} = (\mathbf{B}_q^{*\top} \mathbf{B}_q^*)^{-1} \mathbf{B}_q^{*\top} \mathbf{Y}^*.$$

The OLS estimator is

$$\widetilde{\boldsymbol{\beta}} = (\mathbf{B}_q^{*\top} \boldsymbol{\Omega} \mathbf{B}_q^*)^{-1} \mathbf{B}_q^{*\top} \boldsymbol{\Omega} \mathbf{Y}^* = \{ (\mathbf{B}_q^{*\top} \mathbf{B}_q^*)^{-1} \mathbf{B}_q^{*\top} + \mathbf{D} \} \mathbf{Y}^*,$$

where

$$\mathbf{D} = (\mathbf{B}_q^{*\top} \boldsymbol{\Omega} \mathbf{B}_q^*)^{-1} \mathbf{B}_q^{*\top} \boldsymbol{\Omega} - (\mathbf{B}_q^{*\top} \mathbf{B}_q^*)^{-1} \mathbf{B}_q^{*\top}.$$

Then  $\mathbf{D}\mathbf{B}_q^* = \mathbf{0}$ . Moreover,  $\text{Var}(\widehat{\boldsymbol{\beta}}|\mathbf{X}) = (\mathbf{B}_q^{*\top} \mathbf{B}_q^*)^{-1}$ , and

$$\begin{aligned} \text{Var}(\widetilde{\boldsymbol{\beta}}|\mathbf{X}) &= (\mathbf{B}_q^{*\top} \mathbf{B}_q^*)^{-1} + (\mathbf{B}_q^{*\top} \mathbf{B}_q^*)^{-1} (\mathbf{D}\mathbf{B}_q^*)^\top + (\mathbf{D}\mathbf{B}_q^*) (\mathbf{B}_q^{*\top} \mathbf{B}_q^*)^{-1} + \mathbf{D}\mathbf{D}^\top \\ &= (\mathbf{B}_q^{*\top} \mathbf{B}_q^*)^{-1} + \mathbf{D}\mathbf{D}^\top. \end{aligned}$$

Therefore,

$$\mathcal{B}_q(\mathbf{x})^\top \text{Var}(\widetilde{\boldsymbol{\beta}}|\mathbf{X}) \mathcal{B}_q(\mathbf{x}) - \mathcal{B}_q(\mathbf{x})^\top \text{Var}(\widehat{\boldsymbol{\beta}}|\mathbf{X}) \mathcal{B}_q(\mathbf{x}) = \mathcal{B}_q(\mathbf{x})^\top \mathbf{D}\mathbf{D}^\top \mathcal{B}_q(\mathbf{x}) \geq 0.$$



**2.2. GLS estimator with unknown covariance.** In practice,  $\widehat{\boldsymbol{\beta}}$  is infeasible as it depends on  $\mathbf{V}$  given in (3) which involves unknown parameters  $\sigma_v^2$  and  $\sigma_u^2$ . We estimate  $\sigma_v^2$  and  $\sigma_u^2$  by the following way. Let  $\widetilde{\varepsilon}_{it} = Y_{it} - \widetilde{m}(\mathbf{X}_{it})$ , where  $\widetilde{m}(\mathbf{X}_{it}) = \mathcal{B}_q(\mathbf{X}_{it})^\top \widetilde{\boldsymbol{\beta}}$  and  $\widetilde{\boldsymbol{\beta}}$  is the OLS estimator:  $\widetilde{\boldsymbol{\beta}} = (\mathbf{B}_q^\top \mathbf{B}_q)^{-1} \mathbf{B}_q^\top \mathbf{Y}$ . Thus,  $\sigma_u^2$  is estimated by

$$\widehat{\sigma}_u^2 = \{T(T-1)\}^{-1} n^{-1} \sum_{t \neq t'} \sum_{i=1}^n \widetilde{\varepsilon}_{it} \widetilde{\varepsilon}_{it'},$$

and  $\sigma_v^2$  is estimated by

$$\widehat{\sigma}_v^2 = (nT)^{-1} \sum_{t=1}^T \sum_{i=1}^n \widetilde{\varepsilon}_{it}^2 - \widehat{\sigma}_u^2.$$

Based on the estimated  $\widehat{\sigma}_u^2$  and  $\widehat{\sigma}_v^2$ , we obtain the feasible GLS estimator

$$\widehat{\boldsymbol{\beta}}^F = \left( \mathbf{B}_q^\top \widehat{\boldsymbol{\Omega}}^{-1} \mathbf{B}_q \right)^{-1} \mathbf{B}_q^\top \widehat{\boldsymbol{\Omega}}^{-1} \mathbf{Y},$$

where  $\widehat{\boldsymbol{\Omega}} = \mathbf{I}_N \otimes \widehat{\mathbf{V}}$  and  $\widehat{\mathbf{V}} = \widehat{\sigma}_v^2 \mathbf{I}_T + \widehat{\sigma}_u^2 \mathbf{1}_T \mathbf{1}_T^\top$ . The estimator of  $m(\mathbf{x})$  is then given by  $\widehat{m}^F(\mathbf{x}) = \mathcal{B}_q(\mathbf{x})^\top \widehat{\boldsymbol{\beta}}^F$ , and the estimator of  $\nabla m(\mathbf{x})$  is

$$\widehat{\nabla m}^F(\mathbf{x}) = \mathcal{B}_{q-1}^*(\mathbf{x})^\top \left( \mathbf{B}_q^\top \widehat{\boldsymbol{\Omega}}^{-1} \mathbf{B}_q \right)^{-1} \mathbf{B}_q^\top \widehat{\boldsymbol{\Omega}}^{-1} \mathbf{Y}.$$

**Theorem 2.5.** *Under the same assumptions as given in Theorem 2.1, as  $n \rightarrow \infty$ ,*

- 1 (i)  $\sup_{\mathbf{x} \in [0,1]^d} \text{Var} \{ \widehat{m}^F(\mathbf{x}) \} = O(n^{-1} N^d)$  and  $\sup_{\mathbf{x} \in [0,1]^d} |E \{ \widehat{m}^F(\mathbf{x}) \} - m(\mathbf{x})| = O(N^{-p})$ ;  
(ii)  $\sigma_n^{-1}(\mathbf{x}) \{ \widehat{m}^F(\mathbf{x}) - m(\mathbf{x}) \} \rightarrow \mathbf{N}(0, 1)$ , where  $\sigma_n^2(\mathbf{x})$  is given in (7).
- 2 (i)  $\sup_{\mathbf{x} \in [0,1]^d} \left\| \text{Var} \{ \widehat{\nabla m}^F(\mathbf{x}) \} \right\|_2 = O(n^{-1} N^{2+d})$  and  $\sup_{\mathbf{x} \in [0,1]^d} \left\| E \{ \widehat{\nabla m}^F(\mathbf{x}) \} - \nabla m(\mathbf{x}) \right\|_2 = O\{N^{-(p-1)}\}$ ;  
(ii)  $\Phi_n^{-1/2}(\mathbf{x}) \{ \widehat{\nabla m}^F(\mathbf{x}) - \nabla m(\mathbf{x}) \} \rightarrow \mathbf{N}(\mathbf{0}_d, \mathbf{I}_d)$ , where  $\Phi_n(\mathbf{x})$  is given in (8).

Let  $\widehat{\boldsymbol{\Sigma}}_n = \mathbf{B}_q^\top \widehat{\boldsymbol{\Omega}}^{-1} \mathbf{B}_q$ ,  $\widehat{\sigma}_n^2(\mathbf{x}) = \mathcal{B}_q(\mathbf{x})^\top \widehat{\boldsymbol{\Sigma}}_n^{-1} \mathcal{B}_q(\mathbf{x})$  and  $\widehat{\Phi}_n(\mathbf{x}) = \left\{ \mathcal{B}_{q-1}^*(\mathbf{x})^\top \widehat{\boldsymbol{\Sigma}}_n^{-1} \mathcal{B}_{q-1}^*(\mathbf{x}) \right\}_{d \times d}$ .

Then we have the following results.

**Theorem 2.6.** *Under the same assumptions as given in Theorem 2.1, as  $n \rightarrow \infty$ ,* (i)  $\sigma_n^{-1}(\mathbf{x}) \{ \widehat{m}^F(\mathbf{x}) - m(\mathbf{x}) \} \rightarrow \mathbf{N}(0, 1)$ ; (ii)  $\widehat{\Phi}_n^{-1/2}(\mathbf{x}) \{ \widehat{\nabla m}^F(\mathbf{x}) - \nabla m(\mathbf{x}) \} \rightarrow \mathbf{N}(\mathbf{0}_d, \mathbf{I}_d)$ .

Based on the results given in Theorem 2.6, we can construct pointwise confidence intervals for  $m(\mathbf{x})$  and  $\nabla m(\mathbf{x})$ .

### 3. FINITE-SAMPLE PERFORMANCE

In this section, we undertake Monte Carlo simulations designed to assess the finite-sample performance of the proposed approach. We consider both nonlinear and linear data generating processes (DGPs), and include parametric linear pooled and random effects models by way of comparison. When the simulated DGP is linear, the linear parametric model that uses this information will naturally be more efficient than the nonparametric B-spline approach, so we can quantify the efficiency loss arising from ignorance of the underlying DGP. When the simulated DGP is nonlinear, the misspecified linear model will be inconsistent while the proposed approach remains consistent, so we can quantify the impact of incorrect parametric specification and highlight potential benefits associated with adopting the proposed approach.

**3.1. Nonlinear DGP.** We simulate data according to

$$\begin{aligned} Y_{it} &= m(\mathbf{X}_{it}) + \varepsilon_{it} \\ &= 1 + 2 \cos(2\pi \mathbf{X}_{it}) + u_i + v_{it}, \end{aligned}$$

where  $\mathbf{X} \sim U[0, 1]$ ,  $u_i \sim N(0, \sigma_u^2)$  and  $v_{it} \sim N(0, \sigma_v^2)$ . We let  $\sigma_u = (0, 1)$ ,  $\sigma_v = 1$ ,  $n = (10, 20, 30)$  and  $T = (10, 20, 30)$ , so the panel size  $nT$  ranges from 100 to 900.

For each of the  $M = 10,000$  Monte Carlo replications, we estimate a parametric linear pooled model ('Linear Pooled'), a parametric linear random effects model ('Linear RE'), a nonparametric B-spline pooled model ('B-spline Pooled'), the proposed nonparametric B-spline random effects model ('B-spline RE'). For the B-spline model, the spline degree and number of interior knots are selected via cross-validation. For each replication, we compute the root mean square error (RMSE) for  $\widehat{m}(\mathbf{x})$  as

$$\text{RMSE} = \sqrt{(nT)^{-1} \sum_{i=1}^n \sum_{t=1}^T (\widehat{m}(\mathbf{X}_{it}) - m(\mathbf{X}_{it}))^2}.$$

and RMSE for  $\widehat{\nabla m}(\mathbf{x})$  as

$$\text{RMSE} = \sqrt{(nT)^{-1} \sum_{i=1}^n \sum_{t=1}^T (\widehat{\nabla m}(\mathbf{X}_{it}) - \nabla m(\mathbf{X}_{it}))^2}.$$

We summarize the nonlinear DGP Monte Carlo results in tables 1 and 2. Table 1 reports median RMSE from the  $M$  Monte Carlo replications for each model. Table 2 reports median RMSE over all  $M$  Monte Carlo replications relative to that for the proposed random effects B-spline approach. In the latter table, numbers greater than one indicate that the method specified in each column heading is less efficient than the proposed method. When  $\sigma_u^2 = 0$ , there are no random effects present, so we can gauge any loss arising from the presumption of random effects when in fact there are none (note that the random effects models are estimated as *if* random effects were present). Given that the underlying DGP used here is nonlinear, we can assess how the proposed method performs as  $n$  and  $T$  increase, thereby contrasting the consistency of the proposed method with the inconsistency of its misspecified parametric counterpart.

TABLE 1. Nonlinear DGP RMSE (median taken over all  $M$  Monte Carlo replications). Columns 1-6 present RMSE for  $\widehat{m}(\mathbf{x})$ , Columns 7-12 RMSE for  $\widehat{\nabla m}(\mathbf{x})$ .

$n$	$T$	B-spline Pooled	B-spline RE	Linear Pooled	Linear RE	$n$	$T$	B-spline Pooled	B-spline RE	Linear Pooled	Linear RE
$\sigma_u = 0$						$\sigma_u = 0$					
10	10	0.270	0.271	1.406	1.406	10	10	4.372	4.375	8.894	8.893
10	20	0.193	0.194	1.410	1.410	10	20	2.868	2.878	8.890	8.890
10	30	0.156	0.157	1.411	1.411	10	30	2.223	2.229	8.889	8.889
20	10	0.193	0.193	1.410	1.410	20	10	2.868	2.886	8.890	8.889
20	20	0.133	0.133	1.413	1.413	20	20	1.958	1.963	8.889	8.889
20	30	0.110	0.110	1.413	1.413	20	30	1.661	1.660	8.888	8.888
30	10	0.156	0.157	1.411	1.411	30	10	2.223	2.218	8.889	8.889
30	20	0.110	0.110	1.413	1.413	30	20	1.661	1.666	8.888	8.888
30	30	0.090	0.091	1.413	1.413	30	30	1.416	1.418	8.887	8.887
$\sigma_u = 1$						$\sigma_u = 1$					
10	10	0.451	0.392	1.436	1.433	10	10	5.039	4.649	8.894	8.890
10	20	0.362	0.311	1.437	1.436	10	20	4.211	3.847	8.892	8.891
10	30	0.324	0.282	1.437	1.436	10	30	3.437	2.663	8.890	8.888
20	10	0.320	0.274	1.426	1.425	20	10	4.258	3.946	8.892	8.890
20	20	0.262	0.226	1.428	1.427	20	20	2.861	2.225	8.890	8.889
20	30	0.231	0.201	1.427	1.426	20	30	2.179	1.829	8.888	8.887
30	10	0.263	0.225	1.423	1.422	30	10	3.661	2.802	8.889	8.887
30	20	0.215	0.184	1.423	1.423	30	20	2.199	1.847	8.889	8.888
30	30	0.187	0.162	1.423	1.422	30	30	1.834	1.546	8.886	8.885

TABLE 2. Nonlinear DGP Relative Median RMSE (relative to the proposed B-spline random effects estimator). Numbers greater than one indicate that the estimator listed in the column heading is less efficient than the proposed estimator. Columns 1-5 present RMSE for  $\widehat{m}(\mathbf{x})$ , Columns 6-10 RMSE for  $\widehat{\sqrt{V}}m(\mathbf{x})$ .

$n$	$T$	B-spline Pooled	Linear Pooled	Linear RE	$n$	$T$	B-spline Pooled	Linear Pooled	Linear RE
$\sigma_u = 0$					$\sigma_u = 0$				
10	10	0.996	5.192	5.194	10	10	0.999	2.033	2.033
10	20	0.998	7.282	7.283	10	20	0.997	3.089	3.089
10	30	0.998	9.002	9.002	10	30	0.997	3.988	3.988
20	10	0.999	7.291	7.291	20	10	0.994	3.081	3.080
20	20	1.000	10.584	10.584	20	20	0.998	4.529	4.529
20	30	0.999	12.856	12.856	20	30	1.000	5.354	5.354
30	10	0.999	9.010	9.011	30	10	1.002	4.007	4.007
30	20	0.999	12.853	12.853	30	20	0.997	5.336	5.336
30	30	0.999	15.607	15.607	30	30	0.999	6.267	6.267
$\sigma_u = 1$					$\sigma_u = 1$				
10	10	1.149	3.660	3.653	10	10	1.084	1.913	1.912
10	20	1.162	4.616	4.612	10	20	1.095	2.312	2.311
10	30	1.147	5.093	5.090	10	30	1.291	3.339	3.338
20	10	1.168	5.200	5.195	20	10	1.079	2.254	2.253
20	20	1.161	6.318	6.316	20	20	1.286	3.995	3.994
20	30	1.149	7.083	7.081	20	30	1.192	4.860	4.859
30	10	1.170	6.328	6.326	30	10	1.306	3.172	3.172
30	20	1.170	7.753	7.750	30	20	1.191	4.813	4.813
30	30	1.150	8.763	8.761	30	30	1.186	5.748	5.748

3.2. **Linear DGP.** We simulate data according to

$$\begin{aligned} Y_{it} &= m(\mathbf{X}_{it}) + \varepsilon_{it} \\ &= 1 + 2\mathbf{X}_{it} + u_i + v_{it}, \end{aligned}$$

where  $\mathbf{X} \sim U[-1, 1]$ ,  $u_i \sim N(0, \sigma_u^2)$  and  $v_{it} \sim N(0, \sigma_v^2)$ . We let  $n = (10, 20, 30)$ ,  $T = (10, 20, 30)$ ,  $\sigma_u = (0, 1)$ , and  $\sigma_v = 1$ .

We summarize the linear DGP Monte Carlo results in tables 3 and 4. Given that the underlying DGP for this simulation is linear, we can gauge the loss arising from not knowing the true functional form of the DGP.

3.3. **Discussion.** Perhaps surprisingly, there is virtually no loss in efficiency for either the parametric or the proposed nonparametric approach when one falsely presumes that random effects

TABLE 3. Linear DGP RMSE (median taken over all  $M$  Monte Carlo replications).  
Columns 1-6 present RMSE for  $\widehat{m}(\mathbf{x})$ , Columns 7-12 RMSE for  $\widehat{\nabla m}(\mathbf{x})$ .

$n$	$T$	B-spline Pooled	B-spline RE	Linear Pooled	Linear RE	$n$	$T$	B-spline Pooled	B-spline RE	Linear Pooled	Linear RE
$\sigma_u = 0$						$\sigma_u = 0$					
10	10	0.176	0.177	0.118	0.118	10	10	0.226	0.228	0.117	0.119
10	20	0.122	0.122	0.083	0.084	10	20	0.157	0.157	0.082	0.084
10	30	0.098	0.098	0.068	0.069	10	30	0.120	0.121	0.066	0.067
20	10	0.122	0.122	0.083	0.084	20	10	0.157	0.158	0.082	0.083
20	20	0.085	0.085	0.059	0.059	20	20	0.106	0.106	0.057	0.057
20	30	0.069	0.069	0.049	0.049	20	30	0.088	0.088	0.047	0.047
30	10	0.098	0.098	0.068	0.069	30	10	0.120	0.120	0.066	0.066
30	20	0.069	0.069	0.049	0.049	30	20	0.088	0.088	0.047	0.047
30	30	0.057	0.057	0.039	0.039	30	30	0.071	0.071	0.038	0.038
$\sigma_u = 1$						$\sigma_u = 1$					
10	10	0.369	0.315	0.268	0.251	10	10	0.336	0.244	0.160	0.125
10	20	0.310	0.266	0.242	0.229	10	20	0.212	0.157	0.112	0.084
10	30	0.288	0.252	0.233	0.227	10	30	0.170	0.124	0.092	0.068
20	10	0.259	0.219	0.193	0.176	20	10	0.224	0.161	0.111	0.086
20	20	0.226	0.193	0.178	0.168	20	20	0.149	0.106	0.082	0.060
20	30	0.203	0.179	0.168	0.162	20	30	0.121	0.085	0.066	0.048
30	10	0.209	0.176	0.155	0.144	30	10	0.181	0.132	0.093	0.070
30	20	0.183	0.157	0.144	0.136	30	20	0.120	0.084	0.066	0.049
30	30	0.165	0.143	0.134	0.129	30	30	0.099	0.071	0.054	0.039

are present; see e.g. the column with heading ‘B-spline Pooled’,  $\sigma_u = 0$ , in tables 2 and 4. When random effects are present, however, all random effects approaches are more efficient than their pooled counterparts.

When the DGP is linear (tables 3 and 4), the loss in efficiency arising from not knowing the true DGP diminishes with the panel size ( $nT$ ). All approaches are consistent in this case (the parametric model is correctly specified). But when the DGP is nonlinear (tables 1 and 2), the proposed approach is consistent while the parametric approach is inconsistent (RMSE does not fall as either  $n$  or  $T$  increases).

#### 4. APPLICATION TO PUBLIC CAPITAL PRODUCTIVITY

We now consider a popular panel dataset that covers all 48 states in the continental U.S. over the period 1970-1986 and revisit a previously-examined relationship between public capital and private sector output. The questions of whether public sector capital is productive and whether there is

TABLE 4. Linear DGP Relative Median RMSE (relative to the proposed B-spline random effects estimator). Numbers greater than one indicate that the estimator listed in the column heading is less efficient than the proposed estimator. Columns 1-5 present RMSE for  $\widehat{m}(\mathbf{x})$ , Columns 6-10 RMSE for  $\widehat{\nabla m}(\mathbf{x})$ .

$n$	$T$	B-spline Pooled	Linear Pooled	Linear RE	$n$	$T$	B-spline Pooled	Linear Pooled	Linear RE
$\sigma_u = 0$					$\sigma_u = 0$				
10	10	0.995	0.664	0.669	10	10	0.990	0.513	0.520
10	20	0.998	0.683	0.685	10	20	0.998	0.524	0.534
10	30	1.003	0.697	0.701	10	30	0.992	0.542	0.548
20	10	0.998	0.683	0.686	20	10	0.994	0.522	0.526
20	20	1.000	0.691	0.693	20	20	0.998	0.540	0.540
20	30	0.999	0.701	0.701	20	30	1.000	0.538	0.539
30	10	1.000	0.695	0.697	30	10	1.000	0.546	0.549
30	20	0.998	0.700	0.701	30	20	0.999	0.537	0.541
30	30	1.000	0.689	0.690	30	30	0.998	0.535	0.532
$\sigma_u = 1$					$\sigma_u = 1$				
10	10	1.171	0.850	0.796	10	10	1.375	0.656	0.511
10	20	1.168	0.909	0.862	10	20	1.354	0.712	0.535
10	30	1.141	0.924	0.899	10	30	1.377	0.743	0.550
20	10	1.180	0.882	0.805	20	10	1.395	0.691	0.533
20	20	1.170	0.921	0.872	20	20	1.397	0.770	0.563
20	30	1.138	0.939	0.905	20	30	1.431	0.781	0.568
30	10	1.189	0.882	0.820	30	10	1.378	0.710	0.535
30	20	1.168	0.920	0.871	30	20	1.420	0.785	0.578
30	30	1.152	0.933	0.898	30	30	1.392	0.757	0.554

a role for the public sector in encouraging private economic performance have been the subject of much debate and have received extensive attention from economists. Some claim that public capital has played a significant role in boosting private sector output, while others hold the opposite view that public capital has been detrimental to private sector productivity. Notwithstanding these contradictory conclusions, much of this work has been conducted within a narrow parametric framework that assumes a Cobb-Douglas specification of the underlying production function. A crucial shortcoming of the Cobb-Douglas functional form is its reliance on the dubious assumption of constant elasticities across all states and years. We follow Baltagi & Pinnoi (1995) in what follows.

The astute reader may note that, theoretically, we treat the case of increasing  $n$  and fixed  $T$ , while this is the case where  $n$ , the number of states, is fixed. However, in any analysis, for a given

sample the number of cross-sectional units is naturally pre-determined, so we beg the reader’s forgiveness and trust they will indulge us on this issue.

We estimate the B-spline random effects analogue of the model that appears in Baltagi & Pinnoi (1995),

$$(9) \quad \log(\text{gsp})_{it} = m(\log(\text{pcap})_{it}, \log(\text{pc})_{it}, \log(\text{emp})_{it}, \text{unemp}_{it}) + \varepsilon_{it},$$

where  $\varepsilon_{it} = u_i + v_{it}$  follows a one-way error components (random effects) specification,  $E(v_{it}|x_i, u_i) = 0$  (here the  $x_i$  are the state-level predictors listed in (9)),  $E(u_i|x_i) = E(u_i) = 0$ , and  $m(\cdot)$  is an unknown function to be estimated along with its partial derivatives (i.e., ‘elasticities’ for predictors appearing in log form). The variables in the model are ‘gsp’ (gross state product), ‘pcap’ (private capital stock), ‘pc’ (public capital stock), ‘emp’ (employment) and ‘unemp’ (state unemployment rate). The B-spline degree and knots are selected via the cross-validation procedure that is outlined in Ma, Racine & Yang (2015). By spline theory, any continuous function satisfying a mild condition can be well approximated by a combination of a sufficient number of spline basis functions. However, using more spline functions may over fit the data while choosing less spline functions may under fit the data. Thus, selecting the right number of spline functions by a data-driven method is necessary. We apply the cross-validation method which shows a good numerical performance, but deriving the corresponding theory such as selection consistency is very challenging as the variance and bias of the spline estimate involve the number of basis functions in a complicated fashion. Note that Ma & Racine (2013) used the cross-validation to select variables, but here we include all the covariates in our analysis without variable selection. We use a tensor B-spline basis and least squares cross-validation (we also present as results for Hurvich, Simonoff & Tsai’s (1998)  $AIC_c$  approach). Mean and median elasticity values are reported for the B-spline models, and for the sake of comparison, we also present results from the pooled linear and linear random effects models that assume constant elasticities across all states and years.

Figure B.1 in Appendix B presents mean elasticities by year along with 95% asymptotic confidence intervals when the least-squares cross-validation method is used to select the B-spline degree and number of knots. Figure B.3 presents mean elasticities by year along with 95% asymptotic confidence intervals when the least-squares cross-validation procedure imposes the restriction that

all variables must be included (minimum degree  $\geq 1$ ). Figures B.5 and B.7 presents the same mean elasticities as figures B.1 and B.3 but in this instance the  $AIC_c$  approach is used instead of least-squares cross-validation. Note that all figures are relegated to Appendix B.

We also present boxplots in figures B.2, B.4, B.6 and B.8 which are comparable to figures B.1, B.3, B.5 and B.7. These figures present robust pictures of the heterogeneity of elasticity measures among states. Table 5 presents a summary of elasticity estimates (constant over year and state) from the linear OLS and linear random effects models, and also presents the mean and median elasticity estimates over state and year from the B-spline model. Table 6 presents the same results but forcing all variables to be included (minimum spline degree allowed is 1). Both tables 5 and 6 use least squares cross-validation to determine the spline order and degree. Tables 7 and 8 present similar results, but using  $AIC_c$  to determine the spline order and degree.

**4.1. Discussion.** The search for an optimal spline degree and number of interior knots is computationally challenging. Search is conducted over integer space and the objective function is non-differentiable and non-convex. We have four predictors, so if we conduct a search allowing the spline degree to range from 0-15, and the number of interior knots to range from 0-15, then there are over 4 billion possibilities ( $16^8 = 4,294,967,296$ ), rendering an exhaustive search infeasible (we allow the degree and the number of interior knots to differ for each predictor). We therefore make use of the NOMAD solver ('non smooth optimization via mesh adaptive direct search', Le Digabel (2011)), which allows one to conduct a search over high-dimensional constrained integer spaces. We re-run the NOMAD solver starting from a large number of initial values, which produced the results that are reported below.

Some general patterns emerge, regardless of whether one forces all predictors to be included in the model, or whether one uses least-squares as opposed to  $AIC_c$ -based cross-validation methods.

- (1) There is substantial heterogeneity across states in the year-by-year elasticity estimates.
- (2) The mean and median elasticity with respect to public capital ( $\log(\text{pc})$ ) is uniformly lower than in the linear random effects model that assumes constant elasticities across states and year.



- (3) The mean and median elasticity with respect to private capital ( $\log(\text{pcap})$ ) is, on balance, higher than in the linear random effects model that assumes constant elasticities across states and year.
- (4) The mean and median elasticity with respect to employment ( $\log(\text{emp})$ ) is, on balance, higher than in the linear random effects model that assumes constant elasticities across states and year.
- (5) Some states display negative elasticities with respect to both public and private capital. These are consistent with the empirical (i.e., raw or non-model based estimates given by, e.g.,  $(\log(\text{gsp}_{it}) - \log(\text{gsp}_{it-1})) / (\log(\text{pc}_{it}) - \log(\text{pc}_{it-1}))$ ), but not reported here).
- (6) The robust summaries (medians, boxplots) of the nonparametric elasticity estimates by year are very similar, regardless of whether degree  $\geq 0$  or  $\geq 1$ , or whether least-squares versus  $\text{AIC}_c$  cross-validation is used.

TABLE 5. Comparison of elasticity estimates when least-squares cross-validation is used. Models are pooled linear OLS ('Linear OLS'), linear random effects model ('Linear RE'), and B-spline random effects model ('B-spline RE', mean and median values).

	Linear OLS	Linear RE	B-spline Mean RE	B-spline Median RE
$\log(\text{pcap})$	0.1550	0.0044	0.0230	0.0118
$\log(\text{pc})$	0.3092	0.3106	0.2243	0.2651
$\log(\text{emp})$	0.5939	0.7297	0.8438	0.8348
unemp	-0.0067	-0.0062	NA	NA
CV-score	0.002576108			
CV-degree	1 3 1 0			
CV-segments	1 1 14 1			

## 5. CONCLUDING REMARKS

Applied econometricians typically focus on the estimation of 'marginal effects' or derivatives of the conditional mean function. Though the parametric random effects model has been extensively studied, many practical problems require nonparametric estimates. It is well known that the generalized least squares (GLS) estimator of a random effects model is asymptotically efficient. In this paper, we propose an approach based on regression splines that allows us to directly implement the GLS estimator of the model's parameters. In addition, the estimators are computationally

TABLE 6. Comparison of elasticity estimates when least-squares cross-validation is used. Models are pooled linear OLS ('Linear OLS'), linear random effects model ('Linear RE'), and B-spline random effects model ('B-spline RE', mean and median values). Model is constrained to include all variables (minimum degree is 1).

	Linear OLS	Linear RE	B-spline Mean RE	B-spline Median RE
log(pcap)	0.1550	0.0044	0.0117	0.0439
log(pc)	0.3092	0.3106	0.2836	0.2720
log(emp)	0.5939	0.7297	0.7706	0.7623
unemp	-0.0067	-0.0062	-0.0068	-0.0051
CV-score	0.002812414			
CV-degree	1 2 1 1			
CV-segments	1 1 11 1			

TABLE 7. Comparison of elasticity estimates when  $AIC_c$  cross-validation is used. Models are pooled linear OLS ('Linear OLS'), linear random effects model ('Linear RE'), and B-spline random effects model ('B-spline RE', mean and median values).

	Linear OLS	Linear RE	B-spline Mean RE	B-spline Median RE
log(pcap)	0.1550	0.0044	-0.0980	0.0350
log(pc)	0.3092	0.3106	0.2335	0.2396
log(emp)	0.5939	0.7297	0.8061	0.8120
unemp	-0.0067	-0.0062	NA	NA
$AIC_c$ -score	-4.919433			
$AIC_c$ -degree	2 2 2 0			
$AIC_c$ -segments	1 1 12 1			

TABLE 8. Comparison of elasticity estimates when  $AIC_c$  cross-validation is used. Models are pooled linear OLS ('Linear OLS'), linear random effects model ('Linear RE'), and B-spline random effects model ('B-spline RE', mean and median values). Model is constrained to include all variables (minimum degree is 1).

	Linear OLS	Linear RE	B-spline Mean RE	B-spline Median RE
log(pcap)	0.1550	0.0044	0.0570	0.0784
log(pc)	0.3092	0.3106	0.2916	0.2627
log(emp)	0.5939	0.7297	0.6931	0.7397
unemp	-0.0067	-0.0062	-0.0078	-0.0057
$AIC_c$ -score	-4.820016			
$AIC_c$ -degree	2 2 1 1			
$AIC_c$ -segments	1 1 8 1			

attractive and simple both to implement and to interpret. We establish theoretical properties of the proposed estimators, assess their finite-sample performance via simulation, and illustrate their

application via a widely-studied dataset. The results presented here are also applicable to regression splines with a general class of error covariance matrices.

## REFERENCES

- Andrews, D. W. K. (1991), ‘Asymptotic normality of series estimators for nonparametric and semiparametric regression models’, *Econometrica* **59**, 307–345.
- Baltagi, B. H. (2013), *Econometric Analysis of Panel Data*, fifth edn, Wiley.
- Baltagi, B. H. & Pinnoi, N. (1995), ‘Public capital stock and state productivity growth: Further evidence from an error components model’, *Empirical Economics* **20**(2), 351–359.
- Chen, X. (2007), Large sample sieve estimation of semi-nonparametric models, in ‘Handbook of Econometrics’, Vol. 6, pp. 5549–5632.
- de Boor, C. (2001), *A Practical Guide to Splines*, Springer.
- Demko, S. (1986), ‘Spectral bounds for  $|a^{-1}|_{\infty}$ ’, *Journal of Approximation Theory* **48**, 207–212.
- Haupt, H., Kagerer, K. & Steiner, W. J. (2014), ‘Smooth quantile-based modeling of brand sales, price and promotional effects from retail scanner panels’, *Journal of Applied Econometrics* **29**(6), 1007–1028.
- Henderson, D., Carroll, R. . J. & Li, Q. (2008), ‘Nonparametric estimation and testing of fixed effects panel data models’, *Journal of Econometrics* **144**(1), 257–275.
- Horowitz, J. L. (2009), *Semiparametric and Nonparametric Methods in Econometrics*, Springer-Verlag.
- Hsiao, C. (2003), *Analysis of Panel Data*, second edn, Econometric Society Monographs.
- Huang, J. Z., Zhang, L. & Zhou, L. (2007), ‘Efficient estimation in marginal partially linear models for longitudinal/clustered data using splines’, *Scandinavian Journal of Statistics* **34**(1), 451–477.
- Hurvich, C. M., Simonoff, J. S. & Tsai, C. L. (1998), ‘Smoothing parameter selection in nonparametric regression using an improved akaike information criterion’, *Journal of the Royal Statistical Society B* **60**, 271–293.
- Le Digabel, S. (2011), ‘Algorithm 909: NOMAD: Nonlinear optimization with the MADS algorithm’, *ACM Transactions on Mathematical Software* **37**(4), 44:1–44:15.
- Lin, X., Wang, N., Welsh, A. & Carroll, R. . J. (2004), ‘Equivalent kernels of smoothing splines in nonparametric regression for clustered data’, *Biometrika* **91**(1), 177–193.
- Ma, S., Huang, Z. & Tsai, C. (2016), ‘Parameter estimation for a generalized semiparametric model with repeated measurements’, *Annals of the Institute of Statistical Mathematics* **68**(1), 725–764.
- Ma, S. & Racine, J. S. (2013), ‘Additive regression splines with irrelevant categorical and continuous regressors’, *Statistica Sinica* **23**, 515–541.
- Ma, S., Racine, J. S. & Yang, L. (2015), ‘Spline regression in the presence of categorical predictors’, *Journal of Applied Econometrics* **30**, 703–717.

- Ma, S. & Yang, L. (2011), 'A jump-detecting procedure based on spline estimation', *Journal of Nonparametric Statistics* **23**(1), 67–81.
- Newey, W. K. (1991), 'Uniform convergence in probability and stochastic equicontinuity', *Econometrica* pp. 1161–1167.
- Newey, W. K. (1994), 'The asymptotic variance of semiparametric estimators', *Econometrica* **62**, 1349–1382.
- Newey, W. K. (1999), 'Consistency of two-step sample selection estimators despite misspecification of distribution', *Economics Letters* **63**(2), 129–132.
- Phillips, P. & Liao, Z. (2014), Series estimation of stochastic processes: Recent developments and econometric applications, in 'Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics', Oxford University Press, pp. 377–420.
- Prautzsch, H., Boehm, W. & Paluszny, M. (2002), *Bezier and B-Spline Techniques*, Springer.
- Schumaker, L. (2007), *Spline Functions: Basic Theory*, third edn, Cambridge Mathematical Library.
- Shen, X. & Wong, W. H. (1994), 'Convergence rate of sieve estimates', *The Annals of Statistics* **22**(2), 580–615.
- Stone, C. J. (1985), 'Additive regression and other nonparametric models', *Annals of Statistics* **13**, 689–705.
- Stone, C. J. (1994), 'The use of polynomial splines and their tensor products in multivariate function estimation', *Annals of Statistics* **22**, 118–184.
- Wang, L. & Yang, L. (2007), 'Spline-backfitted kernel smoothing of nonlinear additive autoregression model', *Annals of Statistics* **35**, 2474–2503.
- Wang, L. & Yang, L. (2009), 'Spline estimation of single index model', *Statistica Sinica* **19**, 765–783.
- Wang, N. (2003), 'Marginal nonparametric kernel regression accounting for within-subject correlation', *Biometrika* **90**(1), 43–52.
- Wang, N., Carroll, R. . J. & Lin, X. (2005), 'Efficient semiparametric marginal estimation for longitudinal/clustered data', *Journal of the American Statistical Association* **100**(469), 147–157.
- Wang, S., Song, Q. & Wang, L. (2013), 'Simultaneous variable selection and estimation in semiparametric modeling of longitudinal/clustered data', *Bernoulli* **19**(1), 252–274.

## APPENDIX A. PROOFS OF MAIN PROPOSITIONS

For any vector  $\boldsymbol{\zeta} = (\zeta_1, \dots, \zeta_s) \in \mathbb{R}^s$ , denote the norm by  $\|\boldsymbol{\zeta}\|_r = (|\zeta_1|^r + \dots + |\zeta_s|^r)^{1/r}$ ,  $1 \leq r < +\infty$ ,  $\|\boldsymbol{\zeta}\|_\infty = \max(|\zeta_1|, \dots, |\zeta_s|)$ . For any matrix  $\mathbf{A} = (A_{ij})_{i=1, j=1}^{s, s'}$ , denote  $\|\mathbf{A}\|_\infty = \max_i \sum_{j=1}^{s'} |A_{ij}|$  and  $\|\mathbf{A}\| = \sup\{\|\mathbf{Ax}\|/\|\mathbf{x}\|_2: \mathbf{x} \in \mathbb{R}^{s'} \text{ with } \|\mathbf{x}\|_2 \neq 0\}$ . For any functions  $\phi, \varphi$ , define the empirical inner product and norm as  $\langle \phi, \varphi \rangle_{n, tt'} = n^{-1} \sum_{i=1}^n \phi(\mathbf{X}_{it}) \varphi(\mathbf{X}_{it'}) V_{tt'}$ ,  $\|\phi\|_{n, tt'}^2 = \langle \phi, \phi \rangle_{n, tt'}$ . If the functions  $\phi, \varphi$  are  $L_2$ -integrable, we define the theoretical inner product and the corresponding norm as  $\langle \phi, \varphi \rangle_{tt'} = E(\langle \phi, \varphi \rangle_{n, tt'})$ ,  $\|\phi\|_{tt'}^2 = E(\|\phi\|_{n, tt'}^2)$ . We denote by the same letters  $c, C$ , any positive constants without distinction. For positive numbers  $a_n$  and  $b_n$ ,  $n \geq 1$ , let  $a_n \asymp b_n$  denote that  $\lim_{n \rightarrow \infty} a_n/b_n = c$ , where  $c$  is some nonzero constant. Let  $h = 1/(N+1)$  be the distance between neighboring knots. Denote by  $[a]$  the largest integer not greater than the real number  $a$ . For any  $x_s \in [0, 1]$ , its location and relative position indices  $j(x_s), \delta(x_s)$  are defined as

$$(A.1) \quad j(x_s) = j_n(x_s) = \min\{[x_s/h], N\}, \delta_s(x) = \{x_s - \chi_{j(x_s)}\}/h.$$

It is clear that  $\chi_{j_n(x_s)} \leq x_s < \chi_{j_n(x_s)+1}$ ,  $0 \leq \delta(x_s) < 1, \forall x_s \in [0, 1)$ , and  $\delta(1) = 1$ .

$\widehat{\boldsymbol{\beta}}$  in (5) can be decomposed into  $\widehat{\boldsymbol{\beta}}_m$  and  $\widehat{\boldsymbol{\beta}}_\varepsilon$ , such that  $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}_m + \widehat{\boldsymbol{\beta}}_\varepsilon$ , where

$$(A.2) \quad \widehat{\boldsymbol{\beta}}_m = \boldsymbol{\Sigma}_n^{-1} \mathbf{B}_q^\top \boldsymbol{\Omega}^{-1} \mathbf{m}, \widehat{\boldsymbol{\beta}}_\varepsilon = \boldsymbol{\Sigma}_n^{-1} \mathbf{B}_q^\top \boldsymbol{\Omega}^{-1} (\mathbf{u} + \mathbf{v}),$$

for  $\boldsymbol{\Sigma}_n = \mathbf{B}_q^\top \boldsymbol{\Omega}^{-1} \mathbf{B}_q$ , in which  $\mathbf{m} = \{m(\mathbf{X}_{11}), \dots, m(\mathbf{X}_{nT})\}^\top$ ,  $\mathbf{u} = \{u_1, \dots, u_n\}^\top \otimes \mathbf{1}_T$ , and  $\mathbf{v} = \{v_{11}, \dots, v_{nT}\}^\top$ . Then  $\widehat{m}(\mathbf{x}) = \widehat{m}_m(\mathbf{x}) + \widehat{m}_\varepsilon(\mathbf{x})$ , in which

$$(A.3) \quad \widehat{m}_m(\mathbf{x}) = \mathcal{B}_q(\mathbf{x})^\top \widehat{\boldsymbol{\beta}}_m, \widehat{m}_\varepsilon(\mathbf{x}) = \mathcal{B}_q(\mathbf{x})^\top \widehat{\boldsymbol{\beta}}_\varepsilon.$$

Define

$$\widetilde{\boldsymbol{\beta}}_\varepsilon = \boldsymbol{\Sigma}_n^{-1} \mathbf{B}_q^\top \boldsymbol{\Omega}^{-1} (\mathbf{u} + \mathbf{v}) = \boldsymbol{\Sigma}_n^{-1} \mathbf{B}_q^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\varepsilon}, \widetilde{m}_\varepsilon(\mathbf{x}) = \mathcal{B}_q(\mathbf{x})^\top \widetilde{\boldsymbol{\beta}}_\varepsilon.$$

We first present a Bernstein's inequality in Lemma A.1, which will be used throughout the proofs.

**Lemma A.1.** (*Bosq 1998, Theorem 1.2*). *Let  $\xi_i, 1 \leq i \leq n$ , be independent random variables with  $E(\xi_i) = 0$  and  $E(\xi_i^2) = \sigma_i^2$ , and let  $S_n = \sum_{i=1}^n \xi_i$ . Suppose  $E|\xi_i|^k \leq c^{k-2} k! E\xi_i^2$  for some  $c > 0$ , for*

$i = 1, \dots, n$ ,  $k = 3, 4, \dots$  (Cramer's conditions), then

$$P(|S_n| \geq t) \leq 2 \exp\left\{-\frac{t^2}{4 \sum_{i=1}^n \sigma_i^2 + 2ct}\right\}.$$

**Lemma A.2.** *Under Assumptions (A2) and (A4), as  $n \rightarrow \infty$ ,*

$$\begin{aligned} & \max_{1 \leq t \leq T} \max_{j_1, \dots, j_d, j'_1, \dots, j'_d} \left| \left\langle \mathcal{B}_{j_1, \dots, j_d, q}, \mathcal{B}_{j'_1, \dots, j'_d, q} \right\rangle_{n, tt} - \left\langle \mathcal{B}_{j_1, \dots, j_d, q}, \mathcal{B}_{j'_1, \dots, j'_d, q} \right\rangle_{tt} \right| \\ &= O_{a.s.} \left\{ \left( n^{-1} h^d \log n \right)^{1/2} \right\}, \\ & \max_{1 \leq t \neq t' \leq T} \max_{j_1, \dots, j_d, j'_1, \dots, j'_d} \left| \left\langle \mathcal{B}_{j_1, \dots, j_d, q}, \mathcal{B}_{j'_1, \dots, j'_d, q} \right\rangle_{n, tt'} - \left\langle \mathcal{B}_{j_1, \dots, j_d, q}, \mathcal{B}_{j'_1, \dots, j'_d, q} \right\rangle_{tt'} \right| \\ &= O_{a.s.} \left\{ \left( n^{-1} h^{2d} \log n \right)^{1/2} \right\}, \end{aligned}$$

*Proof of Lemma A.2.* Let

$$\zeta_{j_1, \dots, j_d, j'_1, \dots, j'_d, itt'} = \mathcal{B}_{j_1, \dots, j_d, q}(\mathbf{X}_{it}) \mathcal{B}_{j'_1, \dots, j'_d, q}(\mathbf{X}_{it'}) - E \left\{ \mathcal{B}_{j_1, \dots, j_d, q}(\mathbf{X}_{it}) \mathcal{B}_{j'_1, \dots, j'_d, q}(\mathbf{X}_{it'}) \right\}.$$

For  $t = t'$ , when  $|j_s - j'_s| > q - 1$  for some  $1 \leq s \leq d$ ,  $\zeta_{j_1, \dots, j_d, j'_1, \dots, j'_d, itt} = 0$ , when  $|j_s - j'_s| \leq q - 1$  for all  $1 \leq s \leq d$ , by the properties of the B-spline basis, there exist constants  $0 < c_{B,k} < C_{B,k} < \infty$  and  $0 < c'_B < C'_B < \infty$ , such that  $c_{B,k} h^d \leq E \left| \mathcal{B}_{j_1, \dots, j_d, q}(\mathbf{X}_{it}) \mathcal{B}_{j'_1, \dots, j'_d, q}(\mathbf{X}_{it}) \right|^k \leq C_{B,k} h^d$  and  $c'_B h^{dk} \leq \left| E \left\{ \mathcal{B}_{j_1, \dots, j_d, q}(\mathbf{X}_{it}) \mathcal{B}_{j'_1, \dots, j'_d, q}(\mathbf{X}_{it}) \right\} \right|^k \leq C'_B h^{dk}$ , thus

$$E \zeta_{j_1, \dots, j_d, j'_1, \dots, j'_d, itt}^2 \geq c_{B,k} h^d - C'_B h^{dk} \geq c_{\zeta^2} h^d,$$

for some constant  $0 < c_{\zeta^2} < \infty$ , and  $E \zeta_{j_1, \dots, j_d, j'_1, \dots, j'_d, itt}^2 \leq C_{B,k} h^d$ .

$$\begin{aligned} E \left| \zeta_{j_1, \dots, j_d, j'_1, \dots, j'_d, itt} \right|^k &\leq 2^{k-1} \left[ E \left| \mathcal{B}_{j_1, \dots, j_d, q}(\mathbf{X}_{it}) \mathcal{B}_{j'_1, \dots, j'_d, q}(\mathbf{X}_{it}) \right|^k \right. \\ &\quad \left. + \left| E \left\{ \mathcal{B}_{j_1, \dots, j_d, q}(\mathbf{X}_{it}) \mathcal{B}_{j'_1, \dots, j'_d, q}(\mathbf{X}_{it}) \right\} \right|^k \right] \\ &\leq 2^{k-1} \left( C_{B,k} h^d + C'_B h^{dk} \right) \leq c_{\zeta^k} h^d, \end{aligned}$$

for some constant  $0 < c_{\zeta^k} < \infty$ . For  $t \neq t'$ , there exist constants  $0 < c'_{B,k} < C'_{B,k} < \infty$  and  $0 < c'_B < C'_B < \infty$ , such that  $c'_{B,k} h^{2d} \leq E \left| \mathcal{B}_{j_1, \dots, j_d, q}(\mathbf{X}_{it}) \mathcal{B}_{j'_1, \dots, j'_d, q}(\mathbf{X}_{it'}) \right|^k \leq C'_{B,k} h^{2d}$  and

$c_B'' h^{2dk} \leq \left| E\{\mathcal{B}_{j_1, \dots, j_d, q}(\mathbf{X}_{it}) \mathcal{B}_{j'_1, \dots, j'_d, q}(\mathbf{X}_{it'})\} \right|^k \leq C_B'' h^{2dk}$ . Following the same reasoning as above, one can prove that there exist constants  $0 < c'_{\zeta_2}$  and  $c'_{\zeta_k} < \infty$  such that  $E\zeta_{j_1, \dots, j_d, j'_1, \dots, j'_d, itt'}^2 \geq c'_{\zeta_2} h^{2d}$ , and  $E\left| \zeta_{j_1, \dots, j_d, j'_1, \dots, j'_d, itt'} \right|^k \leq c'_{\zeta_k} h^{2d}$ . Thus, there exists a constant  $c = \max\left(c_{\zeta^k} c_{\zeta_2}^{-1}, c'_{\zeta_k} c_{\zeta_2}^{-1}\right)$ , such that  $E\left| \zeta_{j_1, \dots, j_d, j'_1, \dots, j'_d, itt'} \right|^k \leq ck! E\zeta_{j_1, \dots, j_d, j'_1, \dots, j'_d, itt'}^2 < \infty$ , for  $k \geq 3$ . Then by Bernstein's inequality given in Lemma A.1,

$$\begin{aligned}
 P\left(n^{-1} \left| \sum_{i=1}^n \zeta_{j_1, \dots, j_d, j'_1, \dots, j'_d, itt'} \right| \geq \left\{ c' n^{-1} h^d \log n \right\}^{1/2}\right) &\leq 2 \exp \left\{ -\frac{c' n h^d \log n}{4C_{B,2} n h^d + 2c \{c' n h^d \log n\}^{1/2}} \right\} \\
 &= 2n^{-c'(4C_{B,2})^{-1}} \\
 &\leq 2n^{-4}, \text{ for any } c' \geq 16C_{B,2},
 \end{aligned}$$

which implies

$$\sum_{n=1}^{\infty} P \left[ \max_{1 \leq t \leq T} \max_{j_1, \dots, j_d, j'_1, \dots, j'_d} \left| n^{-1} \sum_{i=1}^n \zeta_{j_1, \dots, j_d, j'_1, \dots, j'_d, itt'} \right| \geq \left\{ c' n^{-1} h^d \log n \right\}^{1/2} \right] \leq 2T \sum_{n=1}^{\infty} \mathbf{K}_n^2 n^{-4} < \infty,$$

where the last inequality holds because  $T$  is fixed and  $\mathbf{K}_n = o(n)$ . Thus, the Borel-Cantelli Lemma implies that

$$(A.4) \quad \max_{1 \leq t \leq T} \max_{j_1, \dots, j_d, j'_1, \dots, j'_d} \left| n^{-1} \sum_{i=1}^n \zeta_{j_1, \dots, j_d, j'_1, \dots, j'_d, itt'} \right| = O_{a.s.} \{ (n^{-1} h^d \log n)^{1/2} \}.$$

Following the same reasoning, one can prove that for

$$\max_{1 \leq t \neq t' \leq T} \max_{j_1, \dots, j_d, j'_1, \dots, j'_d} \left| n^{-1} \sum_{i=1}^n \zeta_{j_1, \dots, j_d, j'_1, \dots, j'_d, itt'} \right| = O_{a.s.} \{ (n^{-1} h^{2d} \log n)^{1/2} \}.$$

□

**Lemma A.3.** *Under Assumptions (A2) and (A4), for  $\sigma_n^2(\mathbf{x})$  defined in (7),  $\sigma_n^2(\mathbf{x}) = \text{Var}\{\tilde{m}_\varepsilon(\mathbf{x})\}$ , and there exist constants  $0 < c_\sigma < C_\sigma < \infty$  such that as  $n \rightarrow \infty$ ,*

$$c_\sigma n^{-1} h^{-d} \leq \inf_{\mathbf{x} \in [0,1]^d} \sigma_n^2(\mathbf{x}) \leq \sup_{\mathbf{x} \in [0,1]^d} \sigma_n^2(\mathbf{x}) \leq C_\sigma n^{-1} h^{-d}.$$

*Proof of Lemma A.3.* For any vector  $\mathbf{a}_n = \{a_{j_1, \dots, j_d} : 1 - q \leq j_s \leq N, 1 \leq s \leq d\}^\top \in \mathbb{R}^{\mathbf{K}_n}$  and  $\|\mathbf{a}_n\|_2 \neq 0$ ,

$$\begin{aligned} n^{-1} \mathbf{a}_n^\top \Sigma \mathbf{a}_n &= \sum_{t, t'=1}^T \sum_{j_1, \dots, j_d, j'_1, \dots, j'_d} a_{j_1, \dots, j_d} a_{j'_1, \dots, j'_d} E \left\{ \mathcal{B}_{j_1, \dots, j_d, q}(\mathbf{X}_{it}) \mathcal{B}_{j'_1, \dots, j'_d, q}(\mathbf{X}_{it'}) V_{tt'} \right\} \\ &= \sum_{t=1}^T \sum_{j_1, \dots, j_d, |j_s - j'_s| \leq q-1} a_{j_1, \dots, j_d} a_{j'_1, \dots, j'_d} E \left\{ \mathcal{B}_{j_1, \dots, j_d, q}(\mathbf{X}_{it}) \mathcal{B}_{j'_1, \dots, j'_d, q}(\mathbf{X}_{it}) \right\} (V_1 + V_2) \\ &\quad + \sum_{t \neq t'} \sum_{j_1, \dots, j_d, j'_1, \dots, j'_d} a_{j_1, \dots, j_d} a_{j'_1, \dots, j'_d} E \left\{ \mathcal{B}_{j_1, \dots, j_d, q}(\mathbf{X}_{it}) \mathcal{B}_{j'_1, \dots, j'_d, q}(\mathbf{X}_{it'}) \right\} V_2. \end{aligned}$$

Moreover,  $E \left\{ \mathcal{B}_{j_1, \dots, j_d, q}(\mathbf{X}_{it}) \mathcal{B}_{j'_1, \dots, j'_d, q}(\mathbf{X}_{it}) \right\} \asymp h^d$  and  $E \left\{ \mathcal{B}_{j_1, \dots, j_d, q}(\mathbf{X}_{it}) \mathcal{B}_{j'_1, \dots, j'_d, q}(\mathbf{X}_{it'}) \right\} \asymp h^{2d}$  for  $t \neq t'$ . Thus,

$$\begin{aligned} n^{-1} \mathbf{a}_n^\top \Sigma \mathbf{a}_n &\asymp Th^d \sum_{j_1, \dots, j_d, |j_s - j'_s| \leq q-1} a_{j_1, \dots, j_d} a_{j'_1, \dots, j'_d} (V_1 + V_2) \\ &\quad + T(T-1) h^{2d} \sum_{j_1, \dots, j_d, j'_1, \dots, j'_d} a_{j_1, \dots, j_d} a_{j'_1, \dots, j'_d} V_2 = \Delta_n. \end{aligned}$$

Then  $\Delta_n \leq Th^d (V_1 + V_2) (2q-1)^d \mathbf{a}_n^\top \mathbf{a}_n + T(T-1) h^{2d} V_2 \mathbf{K}_n \mathbf{a}_n^\top \mathbf{a}_n \asymp h^d \mathbf{a}_n^\top \mathbf{a}_n$ , and  $\Delta_n \geq Th^d (V_1 + V_2) (2q-1)^d \mathbf{a}_n^\top \mathbf{a}_n \asymp h^d \mathbf{a}_n^\top \mathbf{a}_n$ . Therefore, for large enough  $n$ , there exist constants  $0 < c_B < C_B < \infty$  such that

$$(A.5) \quad c_B h^d \mathbf{a}_n^\top \mathbf{a}_n \leq n^{-1} \mathbf{a}_n^\top \Sigma \mathbf{a}_n \leq C_B h^d \mathbf{a}_n^\top \mathbf{a}_n.$$

This leads to

$$(A.6) \quad C_B^{-1} h^{-d} \mathbf{a}_n^\top \mathbf{a}_n \leq n \mathbf{a}_n^\top \Sigma^{-1} \mathbf{a}_n \leq c_B^{-1} h^{-d} \mathbf{a}_n^\top \mathbf{a}_n,$$

for any vector  $\mathbf{a}_n \in \mathbb{R}^{\mathbf{K}_n}$  and  $\mathbf{a}_n \neq \mathbf{0}_{\mathbf{K}_n}$ . As a result, we have for any  $\mathbf{x} \in [0, 1]^d$ ,

$$C_B^{-1} h^{-d} \mathcal{B}_q(\mathbf{x})^\top \mathcal{B}_q(\mathbf{x}) \leq n \mathcal{B}_q(\mathbf{x})^\top \Sigma^{-1} \mathcal{B}_q(\mathbf{x}) \leq c_B^{-1} h^{-d} \mathcal{B}_q(\mathbf{x})^\top \mathcal{B}_q(\mathbf{x}),$$



and by B-spline properties, for large enough  $n$ , there exist constants  $0 < c_\sigma < C_\sigma < \infty$  such that

$$\begin{aligned} \inf_{\mathbf{x} \in [0,1]^d} \mathcal{B}_q(\mathbf{x})^\top \Sigma^{-1} \mathcal{B}_q(\mathbf{x}) &\geq c_B^{-1} n^{-1} h^{-d} & \inf_{\mathbf{x} \in [0,1]^d} \mathcal{B}_q(\mathbf{x})^\top \mathcal{B}_q(\mathbf{x}) &\geq c_\sigma n^{-1} h^{-d} \\ \sup_{\mathbf{x} \in [0,1]^d} \mathcal{B}_q(\mathbf{x})^\top \Sigma^{-1} \mathcal{B}_q(\mathbf{x}) &\leq c_B^{-1} n^{-1} h^{-d} & \sup_{\mathbf{x} \in [0,1]^d} \mathcal{B}_q(\mathbf{x})^\top \mathcal{B}_q(\mathbf{x}) &\leq C_\sigma n^{-1} h^{-d}. \end{aligned}$$

□

**Lemma A.4.** *Under Assumptions (A2)-(A4), as  $n \rightarrow \infty$ ,  $\sigma_n^{-1}(\mathbf{x}) \tilde{m}_\varepsilon(\mathbf{x}) \rightarrow N(0,1)$ , and  $\sup_{\mathbf{x} \in [0,1]^d} |\sigma_n^{-1}(\mathbf{x}) \{\hat{m}_\varepsilon(\mathbf{x}) - \tilde{m}_\varepsilon(\mathbf{x})\}| = O_{a.s.}(n^{-1/2} h^{-d} \log n) = o_{a.s.}(1)$ .*

*Proof of Lemma A.4.*

$$\begin{aligned} &\sigma_n^{-1}(\mathbf{x}) \tilde{m}_\varepsilon(\mathbf{x}) \\ &= \sigma_n^{-1}(\mathbf{x}) \mathcal{B}_q(\mathbf{x})^\top \tilde{\boldsymbol{\beta}}_\varepsilon \\ &= \sigma_n^{-1}(\mathbf{x}) \sum_{i=1}^n V_1 \mathcal{B}_q(\mathbf{x})^\top \Sigma^{-1} \sum_{t=1}^T \mathcal{B}_q(\mathbf{X}_{it}) (u_i + v_{it}) \\ &\quad + \sigma_n^{-1}(\mathbf{x}) \sum_{i=1}^n V_2 \mathcal{B}_q(\mathbf{x})^\top \Sigma^{-1} \sum_{t=1}^T \sum_{t'=1}^T \mathcal{B}_q(\mathbf{X}_{it}) (v_{it'} + u_i) \\ &= \sum_{i=1}^n \sigma_n^{-1}(\mathbf{x}) \mathcal{B}_q(\mathbf{x})^\top \Sigma^{-1} \sum_{t=1}^T \mathcal{B}_q(\mathbf{X}_{it}) \left[ \left\{ V_1 (u_i + v_{it}) + V_2 \sum_{t'=1}^T (v_{it'} + u_i) \right\} \right] \\ &= \sum_{i=1}^n \epsilon_i, \end{aligned}$$

and  $\{\epsilon_i\}_{i=1}^n$  are i.i.d. random variables with  $E(\epsilon_i) = 0$ . Moreover,  $E\{\sigma_n^{-1}(\mathbf{x}) \tilde{m}_\varepsilon(\mathbf{x})\}^2 = 1$ . By a central limit theorem we have that as  $n \rightarrow \infty$ ,  $\sigma_n^{-1}(\mathbf{x}) \tilde{m}_\varepsilon(\mathbf{x}) \rightarrow N(0,1)$ . Denote  $\varsigma_{it} = T^{-1} \sum_{t'=1}^T \nu_{it,it'} \varepsilon_{it'}$ ,  $\vartheta_{i,j_1,\dots,j_d} = T^{-1} \sum_t \mathcal{B}_{j_1,\dots,j_d}(\mathbf{X}_{it}) \varsigma_{it}$  and  $\vartheta_i = \{\vartheta_{i,j_1,\dots,j_d}\}^\top$ . Then  $\vartheta_i$  for  $i = 1, \dots, n$  are independent. Since

$$\begin{aligned} n^{-1} T^{-2} \mathbf{B}_q^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\varepsilon} &= n^{-1} T^{-2} \sum_{i=1}^n \sum_{t,t'=1}^T \mathcal{B}_q(\mathbf{X}_{it}) \nu_{it,it'} \varepsilon_{it'} \\ &= n^{-1} T^{-2} \sum_{i=1}^n \sum_t \mathcal{B}_q(\mathbf{X}_{it}) \sum_{t'=1}^T \nu_{it,it'} \varepsilon_{it'} \\ \text{(A.7)} \quad &= n^{-1} \sum_{i=1}^n T^{-1} \sum_t \mathcal{B}_q(\mathbf{X}_{it}) \varsigma_{it} = n^{-1} \sum_{i=1}^n \vartheta_i, \end{aligned}$$

where  $\nu_{it,it'}$  is the  $(it, it')$ <sup>th</sup> component in  $\boldsymbol{\Omega}^{-1}$  given in (4). Note that  $\nu_{it,it'} = 0$  for  $i \neq i'$ . Moreover,  $E(\varsigma_{it}) = 0$ ,  $E(\vartheta_{i,j_1,\dots,j_d}) = 0$  and

$$\begin{aligned} E(\vartheta_{i,j_1,\dots,j_d})^2 &= E(T^{-2} \sum_{t,t'=1}^T \mathcal{B}_{j_1,\dots,j_d}(\mathbf{X}_{it}) \nu_{it,it'} \varepsilon_{it'})^2 \\ &\leq [T^{-2} \sum_{t,t'=1}^T [E\{\mathcal{B}_{j_1,\dots,j_d}(\mathbf{X}_{it}) \nu_{it,it'} \varepsilon_{it'}\}^2]^{1/2}]^2 \leq ch^d \end{aligned}$$

for some constant  $0 < c < \infty$ . Let  $D_n = n^\alpha h^{d/2}$  with  $\alpha$  satisfying  $\alpha < 1/2$ ,  $1/2 + (2 + \eta)^{-1} < 2\alpha$ , and  $1/2\eta + 1 < (3 + 2\eta)\alpha$ , which are satisfied by any  $\eta > 0$  given in Assumption (A3). By Assumption (A4),  $D_n \rightarrow \infty$  as  $n \rightarrow \infty$ . Write  $\varsigma_{it} = \varsigma_{it,1}^{D_n} + \varsigma_{it,2}^{D_n}$ , where  $\varsigma_{it,1}^{D_n} = \varsigma_{it}\{|\varsigma_{it}| > D_n\}$  and  $\varsigma_{it,2}^{D_n} = \varsigma_{it}\{|\varsigma_{it}| \leq D_n\}$ . Then

$$(A.8) \quad \vartheta_{i,j_1,\dots,j_d} = \vartheta_{i,j_1,\dots,j_d,1} + \vartheta_{i,j_1,\dots,j_d,2} + \vartheta_{i,j_1,\dots,j_d,3},$$

where  $\vartheta_{i,j_1,\dots,j_d,1} = T^{-1} \sum_t^T \mathcal{B}_{j_1,\dots,j_d}(\mathbf{X}_{it}) \varsigma_{it,1}^{D_n}$ ,

$$\vartheta_{i,j_1,\dots,j_d,2} = T^{-1} \sum_t^T \mathcal{B}_{j_1,\dots,j_d}(\mathbf{X}_{it}) \varsigma_{it,2}^{D_n} - E(T^{-1} \sum_t^T \mathcal{B}_{j_1,\dots,j_d}(\mathbf{X}_{it}) \varsigma_{it,2}^{D_n}),$$

and  $\vartheta_{i,j_1,\dots,j_d,3} = E(T^{-1} \sum_t^T \mathcal{B}_{j_1,\dots,j_d}(\mathbf{X}_{it}) \varsigma_{it,2}^{D_n})$ . Since  $\sup_{i,t,t'} |\nu_{it,it'}| \leq (V_1 + V_2) < \infty$ , this together with Assumption (A3) implies that

$$(E|\varsigma_{it}|^{2(2+\eta)})^{1/2(2+\eta)} \leq T^{-1} \sum_{t'=1}^T (E|\nu_{it,it'} \varepsilon_{it'}|^{2(2+\eta)})^{1/2(2+\eta)} \leq (V_1 + V_2) M_\eta^{1/2(2+\eta)}.$$

Thus  $E|\varsigma_{it}|^{2(2+\eta)} \leq (V_1 + V_2)^{2(2+\eta)} M_\eta$ . Let  $A_n = \{\max_{j_1,\dots,j_d} |n^{-1} \sum_{i=1}^n \vartheta_{i,j_1,\dots,j_d,1}| > n^{-m}\}$  for any  $m > 0$ . Since  $|\varsigma_{it}| \leq D_n$  implies  $\varsigma_{it,1}^{D_n} = 0$  and thus implies  $A_n^C$ , then

$$\begin{aligned} \sum_{n=1}^{\infty} P(A_n) &\leq \sum_{n=1}^{\infty} P(|\varsigma_{it}| > D_n) \leq \sum_{n=1}^{\infty} \frac{E|\varsigma_{it}|^{2(2+\eta)}}{D_n^{2(2+\eta)}} \\ &\leq (V_1 + V_2)^{2(2+\eta)} M_\eta \sum_{n=1}^{\infty} n^{-2\alpha(2+\eta)} h^{-d(2+\eta)} \\ &\leq (V_1 + V_2)^{2(2+\eta)} M_\eta \sum_{n=1}^{\infty} (n^{-1/2} h^{-d})^{2+\eta} n^{-1-\delta'} < \infty, \end{aligned}$$

by the constraint  $1/2 + (2 + \eta)^{-1} < 2\alpha$  and Assumption (A4), for some finite constant  $\delta' > 0$ . By Borel-Cantelli Lemma, we have that for any  $m > 0$ ,

$$(A.9) \quad \max_{j_1, \dots, j_d} |n^{-1} \sum_{i=1}^n \vartheta_{i, j_1, \dots, j_d, 1}| = O_{a.s.}(n^{-m}).$$

Moreover,

$$\begin{aligned} \vartheta_{i, j_1, \dots, j_d, 3} &= E(\vartheta_{i, j_1, \dots, j_d}) - E\{T^{-1} \sum_t^T \mathcal{B}_{j_1, \dots, j_d}(\mathbf{X}_{it}) \varsigma_{it, 1}^{D_n}\} \\ &= -E\{T^{-1} \sum_t^T \mathcal{B}_{j_1, \dots, j_d}(\mathbf{X}_{it}) \varsigma_{it, 1}^{D_n}\}. \end{aligned}$$

By B-spline properties, we have  $\max_{j_1, \dots, j_d} E|\mathcal{B}_{j_1, \dots, j_d}(\mathbf{X}_{it})| \leq c'h^d$  for some constant  $0 < c' < \infty$ .

Since

$$\begin{aligned} E|\varsigma_{it, 1}^{D_n}| &= |E[\varsigma_{it} \{|\varsigma_{it}| > D_n\}]| \leq (E|\varsigma_{it}|^{2(2+\eta)})^{1/2(2+\eta)} \{P(|\varsigma_{it}| > D_n)\}^{(3+2\eta)/2(2+\eta)} \\ &\leq (E|\varsigma_{it}|^{2(2+\eta)})^{1/2(2+\eta)} (E|\varsigma_{it}|^{2(2+\eta)} D_n^{-2(2+\eta)})^{(3+2\eta)/2(2+\eta)} \\ &\leq E|\varsigma_{it}|^{2(2+\eta)} D_n^{-(3+2\eta)} \leq (V_1 + V_2)^{2(2+\eta)} M_\eta n^{-\alpha(3+2\eta)} h^{-(3+2\eta)d/2}. \end{aligned}$$

Then

$$\begin{aligned} \max_{j_1, \dots, j_d} \left| E\{T^{-1} \sum_t^T \mathcal{B}_{j_1, \dots, j_d}(\mathbf{X}_{it}) \varsigma_{it, 1}^{D_n}\} \right| &\leq T^{-1} \sum_t^T \max_{j_1, \dots, j_d} E|\mathcal{B}_{j_1, \dots, j_d}(\mathbf{X}_{it}) \varsigma_{it, 1}^{D_n}| \\ &\leq c'(V_1 + V_2)^{2(2+\eta)} n^{-\alpha(3+2\eta)} h^{-(1/2+\eta)d} \\ &= o\left\{\left(n^{-1} h^d \log n\right)^{1/2}\right\}. \end{aligned}$$

by the constraint  $1/2\eta + 1 < (3 + 2\eta)\alpha$  and Assumption (A4). Therefore,

$$(A.10) \quad \begin{aligned} &\max_{j_1, \dots, j_d} |n^{-1} \sum_{i=1}^n \vartheta_{i, j_1, \dots, j_d, 3}| \\ &\leq n^{-1} \sum_{i=1}^n \max_{j_1, \dots, j_d} \left| E\{T^{-1} \sum_t^T \mathcal{B}_{j_1, \dots, j_d}(\mathbf{X}_{it}) \varsigma_{it, 1}^{D_n}\} \right| = o\left\{\left(n^{-1} h^d \log n\right)^{1/2}\right\}. \end{aligned}$$

Moreover, there exists a constant  $0 < c'' < \infty$  such that  $E|\mathcal{B}_{j_1, \dots, j_d}(\mathbf{X}_{it})|^2 \leq c''h^d$ , and thus

$$\begin{aligned} E(\vartheta_{i, j_1, \dots, j_d}^2) &\leq c''h^d E(\varsigma_{it})^2 \leq c''h^d (V_1 + V_2)^2 E|\varepsilon_{it'}|^2 \\ &\leq c''(V_1 + V_2)^2 M_\eta^{1/(2+\eta)} h^d = c'''h^d, \end{aligned}$$

where  $c''' = c''(V_1 + V_2)^2 M_\eta^{1/(2+\eta)}$ . Also,

$$\begin{aligned} E(\vartheta_{i,j_1,\dots,j_d,1}^2) &\leq c'' h^d E(\varsigma_{it})^2 \leq c'' h^d (E|\varsigma_{it}|^{2(2+\eta)})^{2/2(2+\eta)} P(|\varsigma_{it}| > D_n)^{(1+\eta)/(2+\eta)} \\ &\leq c'' h^d (E|\varsigma_{it}|^{2(2+\eta)})^{2/2(2+\eta)} (E|\varsigma_{it}|^{2(2+\eta)} / D_n^{2(2+\eta)})^{(1+\eta)/(2+\eta)} \\ &= E|\varsigma_{it}|^{2(2+\eta)} D_n^{-2(1+\eta)} \leq (V_1 + V_2)^{2(2+\eta)} M_\eta (n^\alpha h^{d/2})^{-2(1+\eta)} = o(h^d), \end{aligned}$$

and

$$E(\vartheta_{i,j_1,\dots,j_d,3}^2) \leq \left| E\left\{ T^{-1} \sum_t^T \mathcal{B}_q(\mathbf{X}_{it}) \varsigma_{it,1}^{D_n} \right\} \right|^2 = o(n^{-1} h^d \log n) = o(h^d).$$

Thus, for sufficiently large  $n$ ,

$$\begin{aligned} E(\vartheta_{i,j_1,\dots,j_d,2}^2) &\leq 3\{E(\vartheta_{i,j_1,\dots,j_d}^2) + E(\vartheta_{i,j_1,\dots,j_d,1}^2) + E(\vartheta_{i,j_1,\dots,j_d,3}^2)\} \\ &\leq 4c''' h^d. \end{aligned}$$

Moreover, by the definition of  $\vartheta_{i,j_1,\dots,j_d,2}$ , we have  $|\vartheta_{i,j_1,\dots,j_d,2}| \leq 2D_n |T^{-1} \sum_t^T \mathcal{B}_{j_1,\dots,j_d}(\mathbf{X}_{it})| \leq 2D_n$ , and thus  $E|\vartheta_{i,j_1,\dots,j_d,2}|^k \leq (2D_n)^{k-2} k! E|\vartheta_{i,j_1,\dots,j_d,2}|^2$ . Therefore, by Bernstein's inequality given in Lemma A.1, we have for a given finite constant  $C > 0$ ,

$$P\left(|\sum_{i=1}^n \vartheta_{i,j_1,\dots,j_d,2}| \geq C \left(nh^d \log n\right)^{1/2}\right) \leq 2 \exp\left\{-\frac{C^2 nh^d \log n}{16c''' nh^d + 4CD_n (nh^d \log n)^{1/2}}\right\}.$$

Since  $D_n (nh^d \log n)^{1/2} = o(nh^d)$ , then for sufficiently large  $n$ , we have

$$P\left(|\sum_{i=1}^n \vartheta_{i,j_1,\dots,j_d,2}| \geq C \left(nh^d \log n\right)^{1/2}\right) \leq 2 \exp\left\{-\frac{C^2 nh^d \log n}{17c''' nh^d}\right\} = 2n^{-C^2/(17c''')} < 2n^{-4},$$

for  $C > 2\sqrt{17c'''}.$  Then

$$\begin{aligned} \sum_{n=1}^{\infty} P\left(\max_{j_1,\dots,j_d} |n^{-1} \sum_{i=1}^n \vartheta_{i,j_1,\dots,j_d,2}| \geq C \left(n^{-1} h^d \log n\right)^{1/2}\right) \\ \leq 2 \sum_{n=1}^{\infty} \mathbf{K}_n n^{-4} < \infty. \end{aligned}$$

Thus, the Borel-Cantelli Lemma implies that

$$(A.11) \quad \max_{j_1,\dots,j_d} |n^{-1} \sum_{i=1}^n \vartheta_{i,j_1,\dots,j_d,2}| = O_{a.s.}\{(n^{-1} h^d \log n)^{1/2}\}.$$

Therefore, by (A.7), (A.9), (A.10) and (A.11), we have  $\|n^{-1}\mathbf{B}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon}\|_\infty = O_{a.s.}\left\{T^2(n^{-1}h^d \log n)^{1/2}\right\}$ .

Since  $T$  is fixed, then

$$(A.12) \quad \begin{aligned} & \left\|n^{-1}\mathbf{B}^\top\boldsymbol{\Omega}^{-1}(\mathbf{u} + \mathbf{v})\right\|_\infty = \left\|n^{-1}\mathbf{B}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon}\right\|_\infty \\ & = O_{a.s.}\left\{T^2(n^{-1}h^d \log n)^{1/2}\right\} = O_{a.s.}\left\{(n^{-1}h^d \log n)^{1/2}\right\}. \end{aligned}$$

For any vector  $\mathbf{a}_n = \{a_{j_1, \dots, j_d} : 1 - q \leq j_s \leq N, 1 \leq s \leq d\}^\top \in \mathbb{R}^{\mathbf{K}_n}$  with  $\|\mathbf{a}_n\|_2 = 1$ , we have

$$\begin{aligned} & |\mathbf{a}_n^\top n^{-1}(\boldsymbol{\Sigma}_n - \boldsymbol{\Sigma})\mathbf{a}_n| \\ & = \left| \sum_{j_1, \dots, j_d, j'_1, \dots, j'_d} a_{j_1, \dots, j_d} a_{j'_1, \dots, j'_d} \sum_{t, t'=1}^T \left( \left\langle \mathcal{B}_{j_1, \dots, j_d, q}, \mathcal{B}_{j'_1, \dots, j'_d, q} \right\rangle_{n, tt'} - \left\langle \mathcal{B}_{j_1, \dots, j_d, q}, \mathcal{B}_{j'_1, \dots, j'_d, q} \right\rangle_{tt'} \right) V_{tt'} \right| \\ & \leq \left| \sum_{j_1, \dots, j_d, |j_s - j'_s| \leq q-1} a_{j_1, \dots, j_d} a_{j'_1, \dots, j'_d} \sum_{t=1}^T \left( \left\langle \mathcal{B}_{j_1, \dots, j_d, q}, \mathcal{B}_{j'_1, \dots, j'_d, q} \right\rangle_{n, tt} - \left\langle \mathcal{B}_{j_1, \dots, j_d, q}, \mathcal{B}_{j'_1, \dots, j'_d, q} \right\rangle_{tt} \right) \right| \\ & \quad \times (V_1 + V_2) \\ & \quad + \left| \sum_{j_1, \dots, j_d, j'_1, \dots, j'_d} a_{j_1, \dots, j_d} a_{j'_1, \dots, j'_d} \sum_{t \neq t'} \left( \left\langle \mathcal{B}_{j_1, \dots, j_d, q}, \mathcal{B}_{j'_1, \dots, j'_d, q} \right\rangle_{n, tt'} - \left\langle \mathcal{B}_{j_1, \dots, j_d, q}, \mathcal{B}_{j'_1, \dots, j'_d, q} \right\rangle_{tt'} \right) \right| V_2 \\ & \leq T(2q-1)^d \mathbf{a}_n^\top \mathbf{a}_n \max_{1 \leq t \leq T} \max_{j_1, \dots, j_d, j'_1, \dots, j'_d} \left| \left( \left\langle \mathcal{B}_{j_1, \dots, j_d, q}, \mathcal{B}_{j'_1, \dots, j'_d, q} \right\rangle_{n, tt} - \left\langle \mathcal{B}_{j_1, \dots, j_d, q}, \mathcal{B}_{j'_1, \dots, j'_d, q} \right\rangle_{tt} \right) \right| \\ & \quad \times (V_1 + V_2) \\ & \quad + T^2 \mathbf{K}_n \mathbf{a}_n^\top \mathbf{a}_n \max_{t \neq t'} \max_{j_1, \dots, j_d, j'_1, \dots, j'_d} \left| \left\langle \mathcal{B}_{j_1, \dots, j_d, q}, \mathcal{B}_{j'_1, \dots, j'_d, q} \right\rangle_{n, tt'} - \left\langle \mathcal{B}_{j_1, \dots, j_d, q}, \mathcal{B}_{j'_1, \dots, j'_d, q} \right\rangle_{tt'} \right| V_2. \end{aligned}$$

The above result together with Lemma A.2 and Assumption (A4) imply that

$$(A.13) \quad \begin{aligned} |\mathbf{a}_n^\top n^{-1}(\boldsymbol{\Sigma}_n - \boldsymbol{\Sigma})\mathbf{a}_n| & = O_{a.s.}\left\{(n^{-1}h^d \log n)^{1/2}\right\} + O(h^{-d})O_{a.s.}\left\{(n^{-1}h^{2d} \log n)^{1/2}\right\} \\ & = O_{a.s.}\left\{(n^{-1} \log n)^{1/2}\right\} = o_{a.s.}(h^d). \end{aligned}$$

By the above result and (A.5), we have with probability 1, as  $n \rightarrow \infty$ ,  $c_B h^d \mathbf{a}_n^\top \mathbf{a}_n \leq n^{-1} \mathbf{a}_n^\top \boldsymbol{\Sigma}_n \mathbf{a}_n \leq C_B h^d \mathbf{a}_n^\top \mathbf{a}_n$ , and thus for any vector  $\mathbf{a}_n \in \mathbb{R}^{\mathbf{K}_n}$  and  $\mathbf{a}_n \neq \mathbf{0}$ ,

$$(A.14) \quad C_B^{-1} h^{-d} \mathbf{a}_n^\top \mathbf{a}_n \leq \mathbf{a}_n^\top (n^{-1} \boldsymbol{\Sigma}_n)^{-1} \mathbf{a}_n \leq c_B^{-1} h^{-d} \mathbf{a}_n^\top \mathbf{a}_n.$$

Therefore, by the result in Demko (1986) together with (A.6) and (A.14), we have

$$(A.15) \quad \|(n^{-1}\boldsymbol{\Sigma})^{-1}\|_{\infty} = O(h^{-d}), \|(n^{-1}\boldsymbol{\Sigma}_n)^{-1}\|_{\infty} = O_{a.s.}(h^{-d}).$$

Following the same reasoning as the proof for (A.13), we have  $\|n^{-1}(\boldsymbol{\Sigma}_n - \boldsymbol{\Sigma})\|_{\infty} = O_{a.s.}\left\{(n^{-1}\log n)^{1/2}\right\}$ . Therefore,

$$\begin{aligned} & \sup_{\mathbf{x} \in [0,1]^d} |\widehat{m}_{\varepsilon}(\mathbf{x}) - \widetilde{m}_{\varepsilon}(\mathbf{x})| \\ &= \sup_{\mathbf{x} \in [0,1]^d} \left| \mathcal{B}_q(\mathbf{x})^{\top} (\boldsymbol{\Sigma}_n^{-1} - \boldsymbol{\Sigma}^{-1}) \mathbf{B}_q^{\top} \boldsymbol{\Omega}^{-1} (\mathbf{u} + \mathbf{v}) \right| \\ &\leq \sup_{\mathbf{x} \in [0,1]^d} \|\mathcal{B}_q(\mathbf{x})\|_1 \|n^{-1}(\boldsymbol{\Sigma}_n - \boldsymbol{\Sigma})\|_{\infty} \|(n^{-1}\boldsymbol{\Sigma}_n)^{-1}\|_{\infty} \|(n^{-1}\boldsymbol{\Sigma})^{-1}\|_{\infty} \left\| n^{-1} \mathbf{B}_q^{\top} \boldsymbol{\Omega}^{-1} (\mathbf{u} + \mathbf{v}) \right\|_{\infty} \\ &= O_{a.s.}\left\{(n^{-1}\log n)^{1/2}\right\} O_{a.s.}(h^{-2d}) O_{a.s.}\left\{(n^{-1}h^d \log n)^{1/2}\right\} \\ &= O_{a.s.}\left(n^{-1}h^{-3d/2} \log n\right). \end{aligned}$$

By the above result and Lemma A.3, one has  $\sup_{\mathbf{x} \in [0,1]^d} |\sigma_n^{-1}(\mathbf{x}) \{\widehat{m}_{\varepsilon}(\mathbf{x}) - \widetilde{m}_{\varepsilon}(\mathbf{x})\}| = O_{a.s.}(n^{-1/2}h^{-d} \log n) = o_{a.s.}(1)$  by Assumption (A4). Therefore, by Slutsky's theorem, one has as  $n \rightarrow \infty$ ,  $\sigma_n^{-1}(\mathbf{x}) \widehat{m}_{\varepsilon}(\mathbf{x}) \rightarrow N(0, 1)$ .  $\square$

**Lemma A.5.** *Under Assumptions (A1), (A2) and (A4), as  $n \rightarrow \infty$ ,  $\sup_{\mathbf{x} \in [0,1]^d} |\widehat{m}_m(\mathbf{x}) - m(\mathbf{x})| = O_{a.s.}(h^p)$ .*

*Proof of Lemma A.5.* By Theorem 12.8 and (13.69) of de Boor (2001), there exists  $\boldsymbol{\beta} \in \mathbb{R}^{K_n}$  such that  $\sup_{\mathbf{x} \in [0,1]^d} \left| \mathcal{B}_q(\mathbf{x})^{\top} \boldsymbol{\beta} - m(\mathbf{x}) \right| = O(h^p)$ . Moreover, by B-spline properties, we have  $\sup_{\mathbf{x} \in [0,1]^d} \|\mathcal{B}_q(\mathbf{x})\|_1 = O(1)$  and  $\max_{1 \leq i \leq n, 1 \leq t \leq T} \|E\{\mathcal{B}_q(\mathbf{X}_{it})\}\|_{\infty} = O(h^d)$ . Following the same reasoning as the proof for Lemma A.2, we have

$$\max_{1 \leq t \leq T} \left\| n^{-1} \sum_{i=1}^n [\mathcal{B}_q(\mathbf{X}_{it}) - E\{\mathcal{B}_q(\mathbf{X}_{it})\}] \right\|_{\infty} = O_{a.s.}\left\{(n^{-1}h^d \log n)^{1/2}\right\}.$$

Thus

$$\begin{aligned}
& \left\| n^{-1} \sum_{i=1}^n \sum_{t=1}^T \mathcal{B}_q(\mathbf{X}_{it}) \right\| \leq T \max_{1 \leq t \leq T} \left\| n^{-1} \sum_{i=1}^n \mathcal{B}_q(\mathbf{X}_{it}) \right\|_{\infty} \\
& \leq T \max_{1 \leq t \leq T} \left\| n^{-1} \sum_{i=1}^n [\mathcal{B}_q(\mathbf{X}_{it}) - E\{\mathcal{B}_q(\mathbf{X}_{it})\}] \right\|_{\infty} + T \max_{1 \leq i \leq n, 1 \leq t \leq T} \|E\{\mathcal{B}_q(\mathbf{X}_{it})\}\|_{\infty} \\
& = O_{a.s.} \left\{ T \left( n^{-1} h^d \log n \right)^{1/2} \right\} + O(Th^d) = O(Th^d) = O(h^d),
\end{aligned}$$

by Assumption (A4) and the condition that  $T$  is fixed. Then,  $\sup_{\mathbf{x} \in [0,1]^d} |\widehat{m}_m(\mathbf{x}) - m(\mathbf{x})|$  equals to

$$\begin{aligned}
& \sup_{\mathbf{x} \in [0,1]^d} \left| \mathcal{B}_q(\mathbf{x})^{\top} \Sigma_n^{-1} \left( \mathbf{B}_q^{\top} \Omega^{-1} \mathbf{m} \right) - m(\mathbf{x}) \right| \\
& \leq \sup_{\mathbf{x} \in [0,1]^d} \left| \mathcal{B}_q(\mathbf{x})^{\top} \Sigma_n^{-1} \left( \mathbf{B}_q^{\top} \Omega^{-1} \mathbf{m} \right) - \mathcal{B}_q(\mathbf{x})^{\top} \Sigma_n^{-1} \left( \mathbf{B}_q^{\top} \Omega^{-1} \mathbf{B}_q \boldsymbol{\beta} \right) \right| \\
& \quad + \sup_{\mathbf{x} \in [0,1]^d} \left| \mathcal{B}_q(\mathbf{x})^{\top} \Sigma_n^{-1} \left( \mathbf{B}_q^{\top} \Omega^{-1} \mathbf{B}_q \boldsymbol{\beta} \right) - m(\mathbf{x}) \right| \\
\text{(A.16)} \quad & = \sup_{\mathbf{x} \in [0,1]^d} \left| \mathcal{B}_q(\mathbf{x})^{\top} \Sigma_n^{-1} \left\{ \mathbf{B}_q^{\top} \Omega^{-1} \left( \mathbf{m} - \mathbf{B}_q^{\top} \boldsymbol{\beta} \right) \right\} \right| + \sup_{\mathbf{x} \in [0,1]^d} \left| \mathcal{B}_q(\mathbf{x})^{\top} \boldsymbol{\beta} - m(\mathbf{x}) \right|.
\end{aligned}$$

Thus, by the above results together with (A.15), we have that with probability 1, there exist constants  $C_1, C_2 \in (0, \infty)$  such that

$$\begin{aligned}
& \sup_{\mathbf{x} \in [0,1]^d} |\widehat{m}_m(\mathbf{x}) - m(\mathbf{x})| \\
& \leq C_1 \sup_{\mathbf{x} \in [0,1]^d} \|\mathcal{B}_q(\mathbf{x})\|_1 \| (n^{-1} \Sigma_n)^{-1} \|_{\infty} \left\| n^{-1} \sum_{i=1}^n \sum_{t=1}^T \mathcal{B}_q(\mathbf{X}_{it}) \right\|_{\infty} h^p + C_2 h^p \\
& = O(1) O(h^{-d}) O(h^d) O(h^p) + O(h^p) = O(h^p).
\end{aligned}$$

□

*Proof of Theorems 2.1 and 2.2.* Theorems 2.1 and 2.2 follow from Lemmas A.3, A.4, and A.5. □

*Proof of Theorems 2.3 and 2.4.* The proofs of Theorems 2.3 and 2.4 follow reasoning that is similar to the proofs of Theorems 2.1 and 2.2, and are omitted here. □

*Proof of Theorem 2.5.* We will show the results in given in 1. Proving the results in 2 will follow the same strategy and thus is omitted. By the definition of  $\sigma_u^2$ , we have  $\sigma_u^2 = \{T(T-1)\}^{-1} n^{-1} \sum_{t \neq t'} \sum_{i=1}^n E(\varepsilon_{it} \varepsilon_{it'})$ . Define

$$\widetilde{\sigma}_u^2 = \{T(T-1)\}^{-1} n^{-1} \sum_{t \neq t'} \sum_{i=1}^n \varepsilon_{it} \varepsilon_{it'}.$$

Then  $\tilde{\sigma}_u^2 - \sigma_u^2$  can be written as

$$\tilde{\sigma}_u^2 - \sigma_u^2 = n^{-1} \sum_{i=1}^n \{T(T-1)\}^{-1} \sum_{t \neq t'} \{\varepsilon_{it}\varepsilon_{it'} - E(\varepsilon_{it}\varepsilon_{it'})\}.$$

Let  $\xi_i = \{T(T-1)\}^{-1} \sum_{t \neq t'} \{\varepsilon_{it}\varepsilon_{it'} - E(\varepsilon_{it}\varepsilon_{it'})\}$ . Then

$$(A.17) \quad \tilde{\sigma}_u^2 - \sigma_u^2 = n^{-1} \sum_{i=1}^n \xi_i.$$

It is clear that  $E(\xi_i) = 0$ . Moreover,  $\xi_i$  is a function of  $\varepsilon_i$ , and  $\varepsilon_i$  for  $1 \leq i \leq n$  are independent, so that  $\xi_i$  for  $1 \leq i \leq n$  are independent. Let  $D_n = n^\alpha$  with  $\alpha < 1/2$ ,  $\alpha(2+\eta) > 1$  and  $\alpha(1+\eta) > 1/2$ , which are satisfied by any  $\eta > 0$  given in Assumption (A3). Write

$$(A.18) \quad \xi_i = \xi_{i,1}^{D_n} + \xi_{i,2}^{D_n} + \xi_{i,3}^{D_n},$$

where  $\xi_{i,1}^{D_n} = \xi_i \{|\xi_i| > D_n\}$ ,  $\xi_{i,2}^{D_n} = \xi_i \{|\xi_i| \leq D_n\} - \xi_{i,3}^{D_n}$ ,  $\xi_{i,3}^{D_n} = E[\xi_i \{|\xi_i| \leq D_n\}]$ . Since

$$(A.19) \quad \begin{aligned} (E|\xi_i|^{2+\eta})^{1/(2+\eta)} &= [E\{\{T(T-1)\}^{-1} \sum_{t \neq t'} \varepsilon_{it}\varepsilon_{it'} - \sigma_u^2\}^{2+\eta}]^{1/(2+\eta)} \\ &\leq \{T(T-1)\}^{-1} (E|\sum_{t \neq t'} \varepsilon_{it}\varepsilon_{it'}|^{2+\eta})^{1/(2+\eta)} + \sigma_u^2 \\ &\leq \{T(T-1)\}^{-1} \sum_{t \neq t'} (E|\varepsilon_{it}\varepsilon_{it'}|^{2+\eta})^{1/(2+\eta)} + \sigma_u^2 \\ &\leq \{T(T-1)\}^{-1} \sum_{t \neq t'} (E|\varepsilon_{it}|^{2(2+\eta)})^{1/2} (E|\varepsilon_{it'}|^{2(2+\eta)})^{1/2} + \sigma_u^2 \\ &\leq \{T(T-1)\}^{-1} \{T(T-1)\} M_\eta^{1/2} M_\eta^{1/2} + \sigma_u^2, \end{aligned}$$

where the last inequality follows from Assumption (A3), then

$$E|\xi_i|^{2+\eta} \leq (M_\eta + \sigma_u^2)^{(2+\eta)}.$$

let  $A_n = \{n^{-1} \sum_{i=1}^n \xi_{i,1}^{D_n} > n^{-m}\}$  for any given  $m > 0$ . Since  $|\xi_i| \leq D_n$  implies that  $\xi_{i,1}^{D_n} = 0$  and thus  $A_n^C$ , then

$$\begin{aligned} \sum_{n=1}^{\infty} P(A_n) &\leq \sum_{n=1}^{\infty} P(|\xi_i| > D_n) \\ &\leq \sum_{n=1}^{\infty} \frac{E|\xi_i|^{2+\eta}}{D_n^{2+\eta}} \leq (M_\eta + \sigma_u^2)^{(2+\eta)} \sum_{n=1}^{\infty} n^{-\alpha(2+\eta)} < \infty. \end{aligned}$$



Therefore, by Borel-Cantelli Lemma, we have that for any  $m > 0$ ,

$$(A.20) \quad |n^{-1} \sum_{i=1}^n \xi_{i,1}^{D_n}| = O_{a.s.}(n^{-m}).$$

Moreover,  $\xi_{i,3}^{D_n} = E(\xi_i) - E[\xi_i \{|\xi_i| > D_n\}] = -E[\xi_i \{|\xi_i| > D_n\}]$ . Since

$$(A.21) \quad \begin{aligned} |E[\xi_i \{|\xi_i| > D_n\}]| &\leq \{E|\xi_i|^{(2+\eta)}\}^{1/(2+\eta)} \{P(|\xi_i| > D_n)\}^{(1+\eta)/(2+\eta)} \\ &\leq \{E|\xi_i|^{(2+\eta)}\}^{1/(2+\eta)} (E|\xi_i|^{(2+\eta)} D_n^{-(2+\eta)})^{(1+\eta)/(2+\eta)} \\ &\leq E|\xi_i|^{(2+\eta)} D_n^{-(1+\eta)} \leq (M_\eta + \sigma_u^2)^{(2+\eta)} n^{-\alpha(1+\eta)}, \end{aligned}$$

then

$$(A.22) \quad \begin{aligned} |n^{-1} \sum_{i=1}^n \xi_{i,3}^{D_n}| &\leq n^{-1} \sum_{i=1}^n |\xi_{i,3}^{D_n}| = n^{-1} \sum_{i=1}^n |E[\xi_i \{|\xi_i| > D_n\}]| \\ &\leq (M_\eta + \sigma_u^2)^{(2+\eta)} n^{-\alpha(1+\eta)} = o(n^{-1/2}). \end{aligned}$$

Following the same reasoning as the proof for (A.19), we have

$$\begin{aligned} (E\xi_i^2)^{1/2} &\leq \{T(T-1)\}^{-1} \sum_{t \neq t'} (E|\varepsilon_{it}\varepsilon_{it'}|^2)^{1/2} + \sigma_u^2 \\ &\leq \{T(T-1)\}^{-1} \sum_{t \neq t'} [\{E(\varepsilon_{it}^2)^{2+\eta}\}^{1/(2+\eta)} \{E(\varepsilon_{it'}^2)^{2+\eta}\}^{1/(2+\eta)}]^{1/2} + \sigma_u^2 \\ &\leq M_\eta^{1/2(2+\eta)} M_\eta^{1/2(2+\eta)} + \sigma_u^2 = M_\eta^{1/(2+\eta)} + \sigma_u^2, \end{aligned}$$

so that  $E\xi_i^2 \leq (M_\eta^{1/(2+\eta)} + \sigma_u^2)^2 < \infty$ . Also,

$$\begin{aligned} E(\xi_{i,1}^{D_n})^2 &\leq (E|\xi_i|^{2+\eta})^{2/(2+\eta)} P(|\xi_i| > D_n)^{\eta/(2+\eta)} \\ &\leq (E|\xi_i|^{2+\eta})^{2/(2+\eta)} (E|\xi_i|^{2+\eta} / D_n^{2+\eta})^{\eta/(2+\eta)} \\ &= E|\xi_i|^{2+\eta} D_n^{-\eta} \leq (M_\eta + \sigma_u^2)^{(2+\eta)} n^{-\alpha\eta}, \end{aligned}$$

and by (A.21),

$$E(\xi_{i,3}^{D_n})^2 \leq |E[\xi_i \{|\xi_i| > D_n\}]|^2 \leq (M_\eta + \sigma_u^2)^{2(2+\eta)} n^{-2\alpha(1+\eta)}.$$

Hence, for sufficiently large  $n$ ,

$$E(\xi_{i,2}^{D_n})^2 \leq 3\{E\xi_i^2 + E(\xi_{i,1}^{D_n})^2 + E(\xi_{i,3}^{D_n})^2\} \leq 4(M_\eta^{1/(2+\eta)} + \sigma_u^2)^2.$$

Moreover, by definition of  $\xi_{i,2}^{D_n}$ , we have  $E|\xi_{i,2}^{D_n}|^k \leq (2D_n)^{k-2} E(\xi_{i,2}^{D_n})^2 \leq (2D_n)^{k-2} k! E(\xi_{i,2}^{D_n})^2$ . Thus, by Bernstein's inequality given in Lemma A.1, we have for a given finite constant  $C > 0$ ,

$$P\left(\left|\sum_{i=1}^n \xi_{i,2}^{D_n}\right| \geq C(n \log n)^{1/2}\right) \leq 2 \exp\left\{-\frac{C^2 n \log n}{16n(M_\eta^{1/(2+\eta)} + \sigma_u^2)^2 + 4CD_n(n \log n)^{1/2}}\right\}.$$

Since  $D_n(n \log n)^{1/2} = n^{\alpha+1/2}\{\log(n)\}^{1/2} = o(n)$ , then for sufficiently large  $n$ , we have

$$\begin{aligned} & P\left(\left|\sum_{i=1}^n \xi_{i,2}^{D_n}\right| \geq C(n \log n)^{1/2}\right) \\ & \leq 2 \exp\left\{-\frac{C^2 n \log n}{17n(M_\eta^{1/(2+\eta)} + \sigma_u^2)^2}\right\} = 2n^{-C^2/\{17(M_\eta^{1/(2+\eta)} + \sigma_u^2)^2\}}. \end{aligned}$$

Choose a  $C$  satisfying  $C > 17^{1/2}(M_\eta^{1/(2+\eta)} + \sigma_u^2)$ . Then

$$\sum_{n=1}^{\infty} P\left(\left|\sum_{i=1}^n \xi_{i,2}^{D_n}\right| \geq C(n \log n)^{1/2}\right) < \infty,$$

so that

$$(A.23) \quad n^{-1} \left|\sum_{i=1}^n \xi_{i,2}^{D_n}\right| = O_{a.s.}\{n^{-1/2}(\log n)^{1/2}\}.$$

Therefore, the results (A.17), (A.18), (A.20), (A.22) and (A.23) immediately imply that

$$|\tilde{\sigma}_u^2 - \sigma_u^2| = O_{a.s.}\{n^{-1/2}(\log n)^{1/2}\}.$$

Moreover,

$$\begin{aligned} |\tilde{\sigma}_u^2 - \hat{\sigma}_u^2| & \leq \{T(T-1)\}^{-1} n^{-1} \sum_{i=1}^n \sum_{t=1}^T |2m(\mathbf{X}_{it}) \{\tilde{m}(\mathbf{X}_{it}) - m(\mathbf{X}_{it})\}| \\ & \quad + \{T(T-1)\}^{-1} n^{-1} \sum_{i=1}^n \sum_{t=1}^T \{\tilde{m}(\mathbf{X}_{it}) - m(\mathbf{X}_{it})\}^2. \end{aligned}$$

It is clear that the results in Lemmas A.3, A.4 and A.5 still hold by letting  $\mathbf{\Omega} = \mathbf{I}$ , and we have  $\sup_{\mathbf{x} \in [0,1]^d} |\tilde{m}(\mathbf{x}) - m(\mathbf{x})| = O_{a.s.}(n^{-1/2}N^{d/2}(\log n)^{1/2} + N^{-p})$ . Thus  $|\tilde{\sigma}_u^2 - \hat{\sigma}_u^2| = O_{a.s.}(n^{-1/2}N^{d/2}(\log n)^{1/2} + N^{-p})$ , and

$$(A.24) \quad |\hat{\sigma}_u^2 - \sigma_u^2| \leq |\hat{\sigma}_u^2 - \tilde{\sigma}_u^2| + |\tilde{\sigma}_u^2 - \sigma_u^2| = O_{a.s.}\left(n^{-1/2}N^{d/2}(\log n)^{1/2} + N^{-p}\right).$$

Similarly, we have

$$(A.25) \quad |\hat{\sigma}_v^2 - \sigma_v^2| = O_{a.s.} \left( n^{-1/2} N^{d/2} (\log n)^{1/2} + N^{-p} \right).$$

By the same strategy as given in A.3, we decompose as  $\hat{m}^F(\mathbf{x}) = \hat{m}_m^F(\mathbf{x}) + \hat{m}_\varepsilon^F(\mathbf{x})$ , where

$$\hat{m}_m^F(\mathbf{x}) = \mathcal{B}_q(\mathbf{x})^\top \hat{\beta}_m^F, \hat{m}_\varepsilon^F(\mathbf{x}) = \mathcal{B}_q(\mathbf{x})^\top \hat{\beta}_\varepsilon^F,$$

and

$$\hat{\beta}_m^F = \hat{\Sigma}_n^{-1} \mathbf{B}_q^\top \hat{\Omega}^{-1} \mathbf{m}, \hat{\beta}_\varepsilon^F = \hat{\Sigma}_n^{-1} \mathbf{B}_q^\top \hat{\Omega}^{-1} \varepsilon,$$

in which  $\hat{\Sigma}_n = \mathbf{B}_q^\top \hat{\Omega}^{-1} \mathbf{B}_q$ . By the definitions of  $\hat{m}_\varepsilon^F(\mathbf{x})$  and  $\hat{m}_\varepsilon(\mathbf{x})$ , we have

$$\hat{m}_\varepsilon^F(\mathbf{x}) - \hat{m}_\varepsilon(\mathbf{x}) = \omega_1(\mathbf{x}) + \omega_2(\mathbf{x}),$$

where

$$\omega_1(\mathbf{x}) = \mathcal{B}_q(\mathbf{x})^\top \Sigma_n^{-1} \mathbf{B}_q^\top (\hat{\Omega}^{-1} - \Omega^{-1}) \varepsilon,$$

$$\omega_2(\mathbf{x}) = \mathcal{B}_q(\mathbf{x})^\top (\hat{\Sigma}_n^{-1} - \Sigma_n^{-1}) \mathbf{B}_q^\top \hat{\Omega}^{-1} \varepsilon.$$

Denote  $\Delta = \{\Delta_{it, it'}\} = \hat{\Omega}^{-1} - \Omega^{-1}$ . Then

$$\begin{aligned} n^{-1} \mathbf{B}_q^\top (\hat{\Omega}^{-1} - \Omega^{-1}) \varepsilon &= n^{-1} \sum_{i, t, t'} \mathcal{B}_q(\mathbf{X}_{it'}) \Delta_{it, it'} \varepsilon_{it} \\ &= (\hat{\sigma}_u^2 + \hat{\sigma}_v^2 - \sigma_u^2 - \sigma_v^2) n^{-1} \sum_i \sum_t \mathcal{B}_q(\mathbf{X}_{it}) \varepsilon_{it} + (\hat{\sigma}_u^2 - \sigma_u^2) n^{-1} \sum_i \sum_{t \neq t'} \mathcal{B}_q(\mathbf{X}_{it'}) \varepsilon_{it}. \end{aligned}$$

By following the same arguments as the proof for (A.12) with  $\Omega^{-1}$  replaced by  $\mathbf{I}_{NT}$  or  $\mathbf{I}_N \otimes \mathbf{1}_T \mathbf{1}_T^\top$ ,

we have

$$\begin{aligned} \left\| n^{-1} \sum_i \sum_t \mathcal{B}_q(\mathbf{X}_{it}) \varepsilon_{it} \right\|_\infty &= O_{a.s.} \left\{ \left( n^{-1} h^d \log n \right)^{1/2} \right\}, \\ \left\| n^{-1} \sum_i \sum_{t \neq t'} \mathcal{B}_q(\mathbf{X}_{it'}) \varepsilon_{it} \right\|_\infty &= O_{a.s.} \left\{ \left( n^{-1} h^d \log n \right)^{1/2} \right\}. \end{aligned}$$

Thus, by the above results together with (A.24) and (A.25), we have

$$\begin{aligned} \|n^{-1}\mathbf{B}_q^\top(\widehat{\boldsymbol{\Omega}}^{-1} - \boldsymbol{\Omega}^{-1})\boldsymbol{\varepsilon}\|_\infty &= O\left(n^{-1/2}N^{d/2}(\log n)^{1/2} + N^{-p}\right) O_{a.s.}\left\{\left(n^{-1}h^d \log n\right)^{1/2}\right\} \\ &= O_{a.s.}(n^{-1} \log n). \end{aligned}$$

The above result and (A.15) lead to

$$\begin{aligned} \sup_{\mathbf{x} \in [0,1]^d} |\omega_1(\mathbf{x})| &\leq \sup_{\mathbf{x} \in [0,1]^d} \|\mathcal{B}_q(\mathbf{x})\|_1 \|n\boldsymbol{\Sigma}_n^{-1}\|_\infty \|n^{-1}\mathbf{B}_q^\top(\widehat{\boldsymbol{\Omega}}^{-1} - \boldsymbol{\Omega}^{-1})\boldsymbol{\varepsilon}\|_\infty \\ \text{(A.26)} \qquad \qquad \qquad &= O_{a.s.}(n^{-1}h^{-d} \log n) = o_{a.s.}(n^{-1/2}h^{-d/2}), \end{aligned}$$

and

$$\begin{aligned} \|n^{-1}\mathbf{B}_q^\top \widehat{\boldsymbol{\Omega}}^{-1} \boldsymbol{\varepsilon}\|_\infty &\leq \|n^{-1}\mathbf{B}_q^\top(\widehat{\boldsymbol{\Omega}}^{-1} - \boldsymbol{\Omega}^{-1})\boldsymbol{\varepsilon}\|_\infty + \|n^{-1}\mathbf{B}_q^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\varepsilon}\|_\infty \\ &= O_{a.s.}(n^{-1} \log n) + O_{a.s.}\left\{\left(n^{-1}h^d \log n\right)^{1/2}\right\} \\ &= O_{a.s.}\left\{\left(n^{-1}h^d \log n\right)^{1/2}\right\}. \end{aligned}$$

Moreover,

$$\begin{aligned} &\|n^{-1}(\widehat{\boldsymbol{\Sigma}}_n - \boldsymbol{\Sigma}_n)\|_\infty \\ &\leq (|\widehat{\sigma}_u^2 - \sigma_u^2| + |\widehat{\sigma}_v^2 - \sigma_v^2|) \{T(2q-1)^d \max_{1 \leq t \leq T} \max_{j_1, \dots, j_d, j'_1, \dots, j'_d} |n^{-1} \sum_{i=1}^n \mathcal{B}_{j_1, \dots, j_d, q}(\mathbf{X}_{it}) \mathcal{B}_{j'_1, \dots, j'_d, q}(\mathbf{X}_{it})| \\ &\quad + T^2 \mathbf{K}_n \max_{t \neq t'} \max_{j_1, \dots, j_d, j'_1, \dots, j'_d} |n^{-1} \sum_{i=1}^n \mathcal{B}_{j_1, \dots, j_d, q}(\mathbf{X}_{it}) \mathcal{B}_{j'_1, \dots, j'_d, q}(\mathbf{X}_{it'})|\}. \end{aligned}$$

Since  $E\{\mathcal{B}_{j_1, \dots, j_d, q}(\mathbf{X}_{it}) \mathcal{B}_{j'_1, \dots, j'_d, q}(\mathbf{X}_{it})\} = O(h^d)$  and  $E\{\mathcal{B}_{j_1, \dots, j_d, q}(\mathbf{X}_{it}) \mathcal{B}_{j'_1, \dots, j'_d, q}(\mathbf{X}_{it'})\} = O(h^{2d})$  for  $t \neq t'$ , this result together with Lemma A.2 implies that there exists a constant  $C \in (0, \infty)$  such that with probability 1,

$$\begin{aligned} \|n^{-1}(\widehat{\boldsymbol{\Sigma}}_n - \boldsymbol{\Sigma}_n)\|_\infty &\leq C(|\widehat{\sigma}_u^2 - \sigma_u^2| + |\widehat{\sigma}_v^2 - \sigma_v^2|) \{T(2q-1)^d h^d + T^2 \mathbf{K}_n h^{2d}\} \\ &= O\left(n^{-1/2}N^{d/2}(\log n)^{1/2} + N^{-p}\right) O\left(N^{-d}\right). \end{aligned}$$

Hence,

$$\begin{aligned}
& \sup_{\mathbf{x} \in [0,1]^d} |\omega_2(\mathbf{x})| \\
& \leq \sup_{\mathbf{x} \in [0,1]^d} \|\mathcal{B}_q(\mathbf{x})\|_1 \|(n^{-1}\widehat{\Sigma}_n)^{-1}\|_\infty \|(n^{-1}\Sigma_n)^{-1}\|_\infty \|n^{-1}(\widehat{\Sigma}_n - \Sigma_n)\|_\infty \|n^{-1}\mathbf{B}_q^\top \widehat{\Omega}^{-1} \boldsymbol{\varepsilon}\|_\infty \\
& = O_{a.s.}(N^d) O_{a.s.}(N^d) O_{a.s.}(n^{-1/2}N^{d/2}(\log n)^{1/2} + N^{-p}) O_{a.s.}(N^{-d}) O_{a.s.}\left\{\left(n^{-1}h^d \log n\right)^{1/2}\right\} \\
& \text{(A.27)} \\
& = O_{a.s.}(n^{-1/2}N^{d/2}) O_{a.s.}\left\{n^{-1/2}N^{d/2} \log(n)\right\} = o_{a.s.}(n^{-1/2}N^{d/2}).
\end{aligned}$$

Therefore, the above results imply that

$$\begin{aligned}
\sup_{\mathbf{x} \in [0,1]^d} |\widehat{m}_\varepsilon^F(\mathbf{x}) - \widehat{m}_\varepsilon(\mathbf{x})| & \leq \sup_{\mathbf{x} \in [0,1]^d} |\omega_1(\mathbf{x})| + \sup_{\mathbf{x} \in [0,1]^d} |\omega_2(\mathbf{x})| \\
& = o_{a.s.}(n^{-1/2}N^{d/2}).
\end{aligned}$$

Next, by the definitions of  $\widehat{m}_m^F(\mathbf{x})$  and  $\widehat{m}_m(\mathbf{x})$  and the same decomposition as given in (A.16), we have

$$\begin{aligned}
& \sup_{\mathbf{x} \in [0,1]^d} |\{\widehat{m}_m^F(\mathbf{x}) - m(\mathbf{x})\} - \{\widehat{m}_m(\mathbf{x}) - m(\mathbf{x})\}| \\
& \leq \sup_{\mathbf{x} \in [0,1]^d} |\mathcal{B}_q(\mathbf{x})^\top \widehat{\Sigma}_n^{-1} \{\mathbf{B}_q^\top \widehat{\Omega}^{-1} (\mathbf{m} - \mathbf{B}_q^\top \boldsymbol{\beta})\} - \mathcal{B}_q(\mathbf{x})^\top \Sigma_n^{-1} \{\mathbf{B}_q^\top \Omega^{-1} (\mathbf{m} - \mathbf{B}_q^\top \boldsymbol{\beta})\}| \\
& \quad + 2 \sup_{\mathbf{x} \in [0,1]^d} |\mathcal{B}_q(\mathbf{x})^\top \boldsymbol{\beta} - m(\mathbf{x})|.
\end{aligned}$$

Following the same reasoning as the proofs for (A.26) and (A.27), we have

$$\begin{aligned}
& \sup_{\mathbf{x} \in [0,1]^d} |\mathcal{B}_q(\mathbf{x})^\top \Sigma_n^{-1} \mathbf{B}_q^\top (\widehat{\Omega}^{-1} - \Omega^{-1}) (\mathbf{m} - \mathbf{B}_q^\top \boldsymbol{\beta})| \\
& = O_{a.s.}(h^{-d}) O_{a.s.}(h^d) O_{a.s.}(n^{-1/2}N^{d/2}(\log n)^{1/2} + N^{-p}) O_{a.s.}(N^{-p}) = o_{a.s.}(N^{-p}), \\
& \sup_{\mathbf{x} \in [0,1]^d} |\mathcal{B}_q(\mathbf{x})^\top (\widehat{\Sigma}_n^{-1} - \Sigma_n^{-1}) \mathbf{B}_q^\top \widehat{\Omega}^{-1} (\mathbf{m} - \mathbf{B}_q^\top \boldsymbol{\beta})| \\
& = O_{a.s.}(N^d) O_{a.s.}(n^{-1/2}N^{d/2}(\log n)^{1/2} + N^{-p}) O_{a.s.}(N^{-d}) O_{a.s.}(N^d) O_{a.s.}(N^{-d}) O_{a.s.}(N^{-p}) \\
& = o_{a.s.}(N^{-p}).
\end{aligned}$$

Thus, the above results imply that

$$\begin{aligned}
& \sup_{\mathbf{x} \in [0,1]^d} |\widehat{m}_m^F(\mathbf{x}) - \widehat{m}_m(\mathbf{x})| \\
&= \sup_{\mathbf{x} \in [0,1]^d} |\{\widehat{m}_m^F(\mathbf{x}) - m(\mathbf{x})\} - \{\widehat{m}_m(\mathbf{x}) - m(\mathbf{x})\}| \\
&= o_{a.s.}(N^{-p}) + O_{a.s.}(N^{-p}) = o_{a.s.}\left(n^{-1/2}N^{d/2}\right).
\end{aligned}$$

Therefore,

$$\begin{aligned}
& \sup_{\mathbf{x} \in [0,1]^d} |\widehat{m}^F(\mathbf{x}) - \widehat{m}(\mathbf{x})| \\
&\leq \sup_{\mathbf{x} \in [0,1]^d} |\widehat{m}_\varepsilon^F(\mathbf{x}) - \widehat{m}_\varepsilon(\mathbf{x})| + \sup_{\mathbf{x} \in [0,1]^d} |\widehat{m}_m^F(\mathbf{x}) - \widehat{m}_m(\mathbf{x})| = o_{a.s.}\left(n^{-1/2}N^{d/2}\right).
\end{aligned}$$

Hence, the result 1 of Theorem 2.5 follows directly from the above result and Theorem (2.1).  $\square$

*Proof of Theorem 2.6.* By the result given in Theorem 2.5, there exists a constant  $C \in (0, \infty)$  such that

$$\begin{aligned}
& \sup_{\mathbf{x} \in [0,1]^d} E|\widehat{\sigma}_n^{-1}(\mathbf{x})\{\widehat{m}(\mathbf{x}) - m(\mathbf{x})\} - \sigma_n^{-1}(\mathbf{x})\{\widehat{m}(\mathbf{x}) - m(\mathbf{x})\}| \\
&\leq \sup_{\mathbf{x} \in [0,1]^d} \{E|\widehat{\sigma}_n^{-1}(\mathbf{x}) - \sigma_n^{-1}(\mathbf{x})|^2\}^{1/2} \{E|\widehat{m}(\mathbf{x}) - m(\mathbf{x})|^2\}^{1/2} \\
&\leq C^* n^{-1/2} N^{d/2} \sup_{\mathbf{x} \in [0,1]^d} \{E|\widehat{\sigma}_n^{-1}(\mathbf{x}) - \sigma_n^{-1}(\mathbf{x})|^2\}^{1/2}.
\end{aligned}$$

Moreover, by the result in Lemma A.3, we have

$$\begin{aligned}
\sup_{\mathbf{x} \in [0,1]^d} |\widehat{\sigma}_n(\mathbf{x}) - \sigma_n(\mathbf{x})| &\leq \sup_{\mathbf{x} \in [0,1]^d} \sigma_n^{-1}(\mathbf{x}) |\widehat{\sigma}_n^2(\mathbf{x}) - \sigma_n^2(\mathbf{x})| \\
&\leq c_\sigma^{-1/2} n^{1/2} N^{d/2} \sup_{\mathbf{x} \in [0,1]^d} |\widehat{\sigma}_n^2(\mathbf{x}) - \sigma_n^2(\mathbf{x})|,
\end{aligned}$$

and by (A.24) and (A.25), we have  $\sup_{\mathbf{x} \in [0,1]^d} |\widehat{\sigma}_n^2(\mathbf{x}) - \sigma_n^2(\mathbf{x})| = o_{a.s.}(1)$ . The above results imply that

$$\sup_{\mathbf{x} \in [0,1]^d} E|\widehat{\sigma}_n^{-1}(\mathbf{x})\{\widehat{m}(\mathbf{x}) - m(\mathbf{x})\} - \sigma_n^{-1}(\mathbf{x})\{\widehat{m}(\mathbf{x}) - m(\mathbf{x})\}| = o(1).$$

Then (i) in Theorem 2.6 follows from Slutsky's theorem. The proof of (ii) follows the same procedure and thus is omitted.  $\square$

APPENDIX B. FIGURES AND TABLES

**B.1. Least Squares Cross Validation (degree  $\geq 0$ ).** Figures B.1 and B.2 present results for the case where least-squares cross-validation is used and where the degree of the polynomial can be zero, thereby allowing for the removal of variables from the resulting estimate. Figure B.1 presents the mean elasticities along with asymptotic 95% confidence intervals, while Figure B.2 presents boxplot summaries which highlight the median and interquartile ranges.

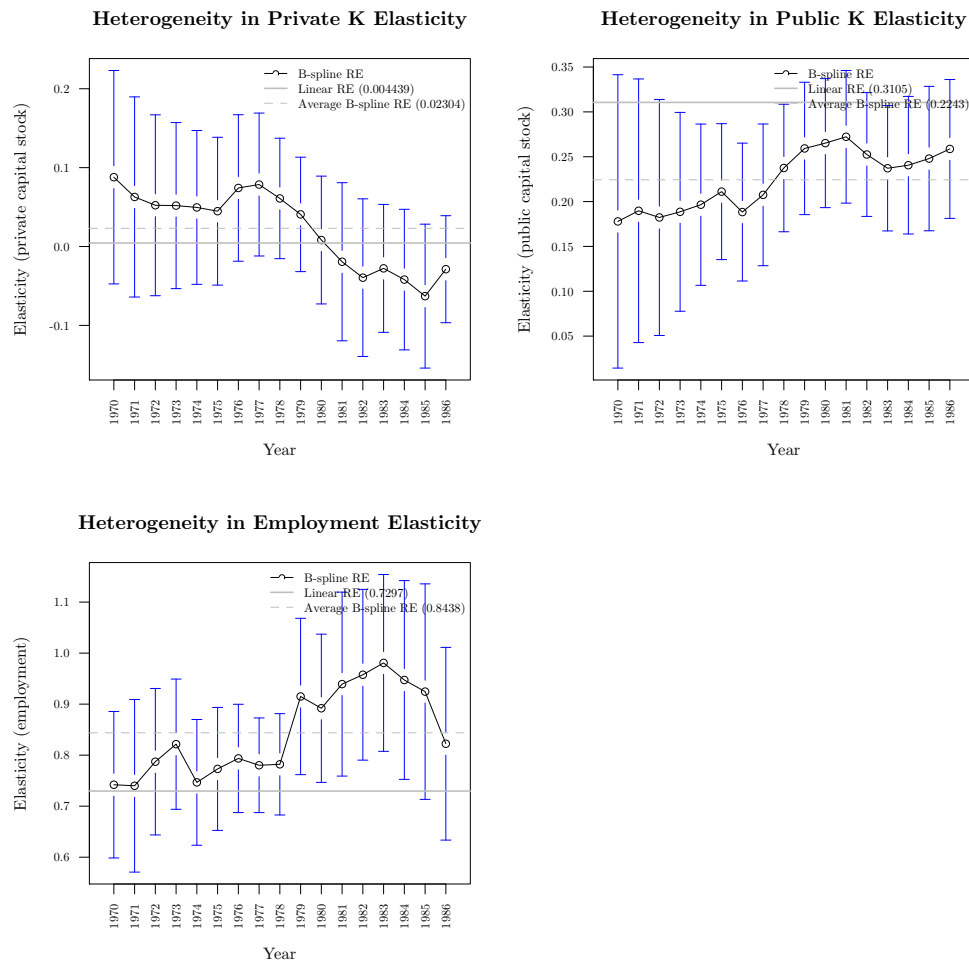


FIGURE B.1. Heterogeneity in B-spline RE elasticity estimates by year (solid line with dots is the average taken over state elasticities by year, vertical bars show 95% confidence intervals). Horizontal lines represent Linear RE estimates and average of mean B-spline RE estimates taken over both year and state.

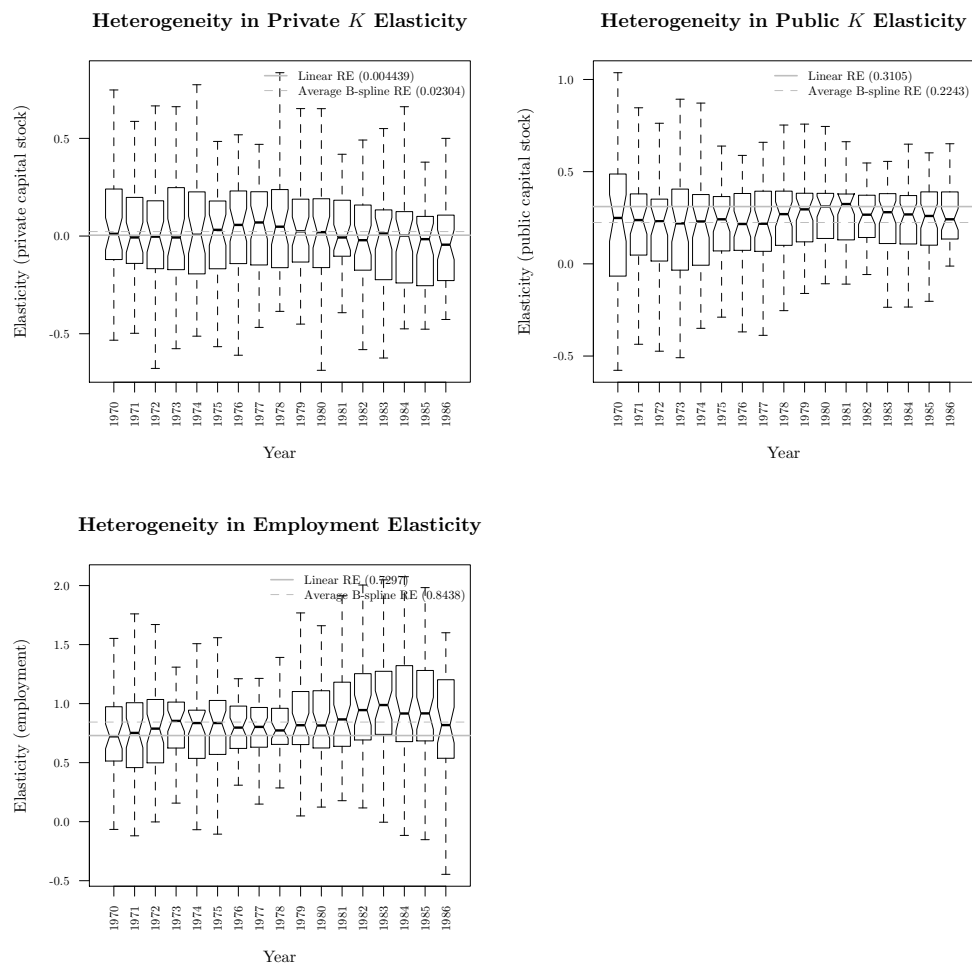


FIGURE B.2. Heterogeneity in B-spline RE elasticity estimates by year (boxplots). Horizontal lines represent Linear RE estimates and average of mean B-spline RE estimates taken over both year and state.



**B.2. Least Squares Cross Validation, All Variables Forced to Be Relevant (degree  $\geq 1$ ).**

Figures B.3 and B.4 present results for the case where least-squares cross-validation is used and where the degree of the polynomial is  $\geq 1$ , thereby forcing all variables to be included in the resulting estimate. Figure B.3 presents the mean elasticities along with asymptotic 95% confidence intervals, while Figure B.4 presents boxplot summaries which highlight the median and interquartile ranges.

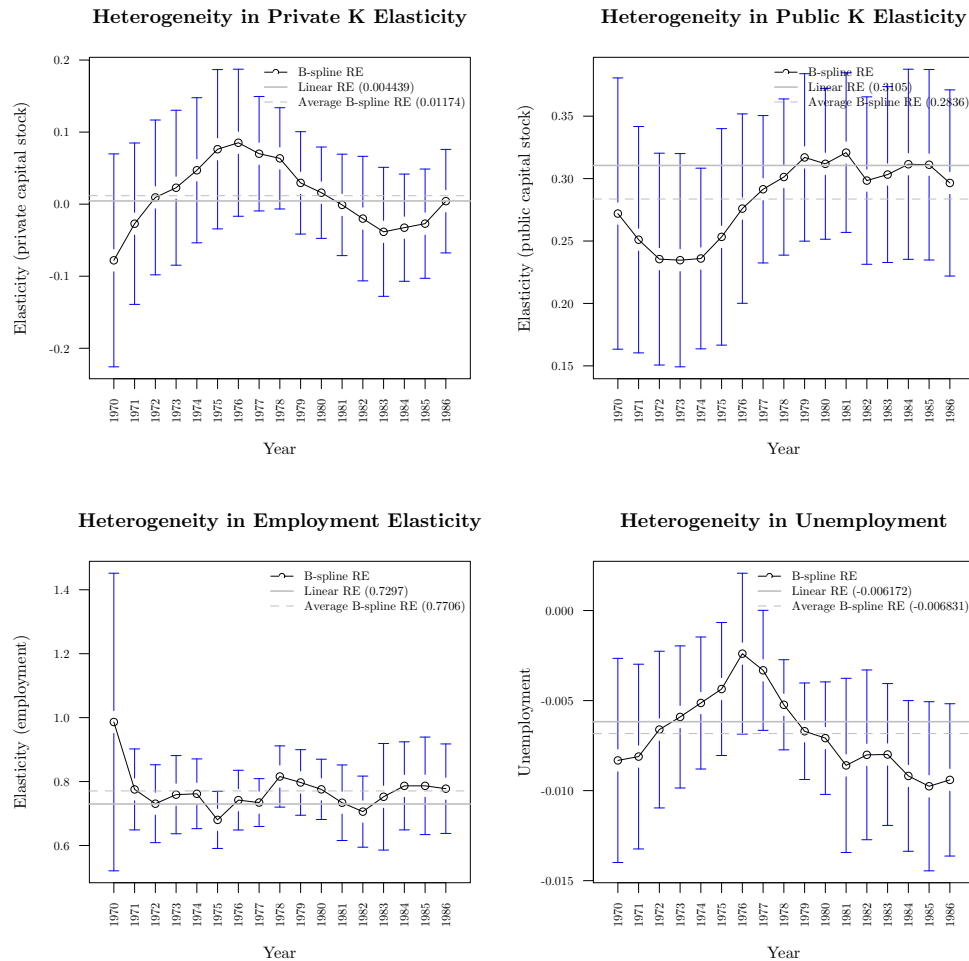


FIGURE B.3. Heterogeneity in B-spline RE elasticity estimates by year (solid line with dots is the average taken over state elasticities by year, vertical bars show 95% confidence intervals). Horizontal lines represent Linear RE estimates and average of mean B-spline RE estimates taken over both year and state.

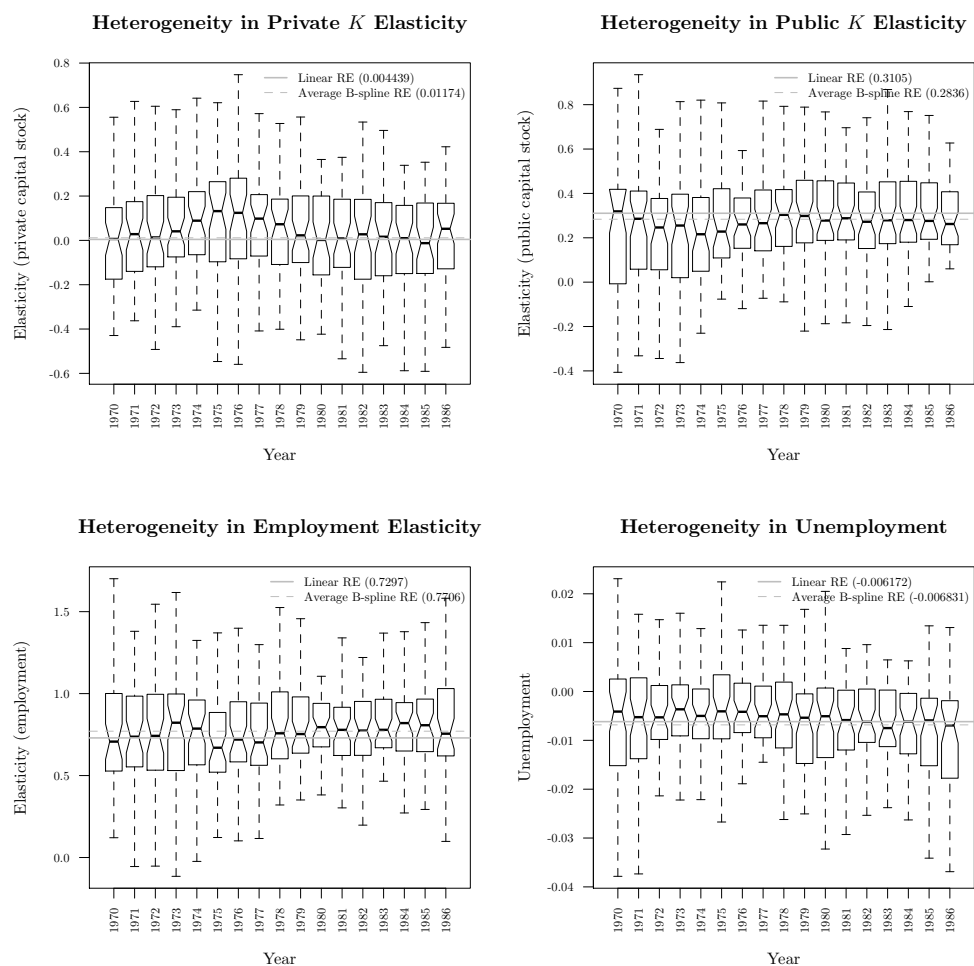


FIGURE B.4. Heterogeneity in B-spline RE elasticity estimates by year (boxplots). Horizontal lines represent Linear RE estimates and average of mean B-spline RE estimates taken over both year and state.

**B.3.  $AIC_c$  Cross Validation (degree  $\geq 0$ ).** Figures B.5 and B.6 present results for the case where  $AIC_c$  cross-validation is used and where the degree of the polynomial can be zero, thereby allowing for the removal of variables from the resulting estimate. Figure B.5 presents the mean elasticities along with asymptotic 95% confidence intervals, while Figure B.6 presents boxplot summaries which highlight the median and interquartile ranges.

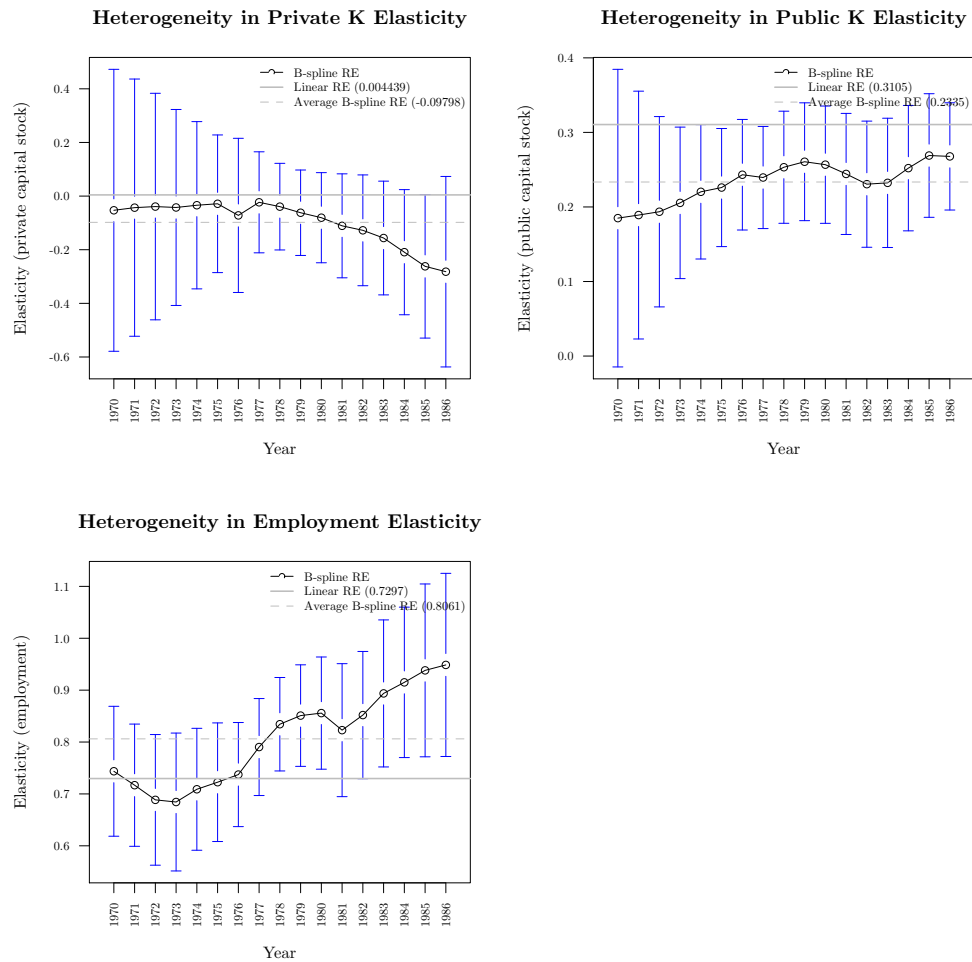


FIGURE B.5. Heterogeneity in B-spline RE elasticity estimates by year (solid line with dots is the average taken over state elasticities by year, vertical bars show 95% confidence intervals). Horizontal lines represent Linear RE estimates and average of mean B-spline RE estimates taken over both year and state.

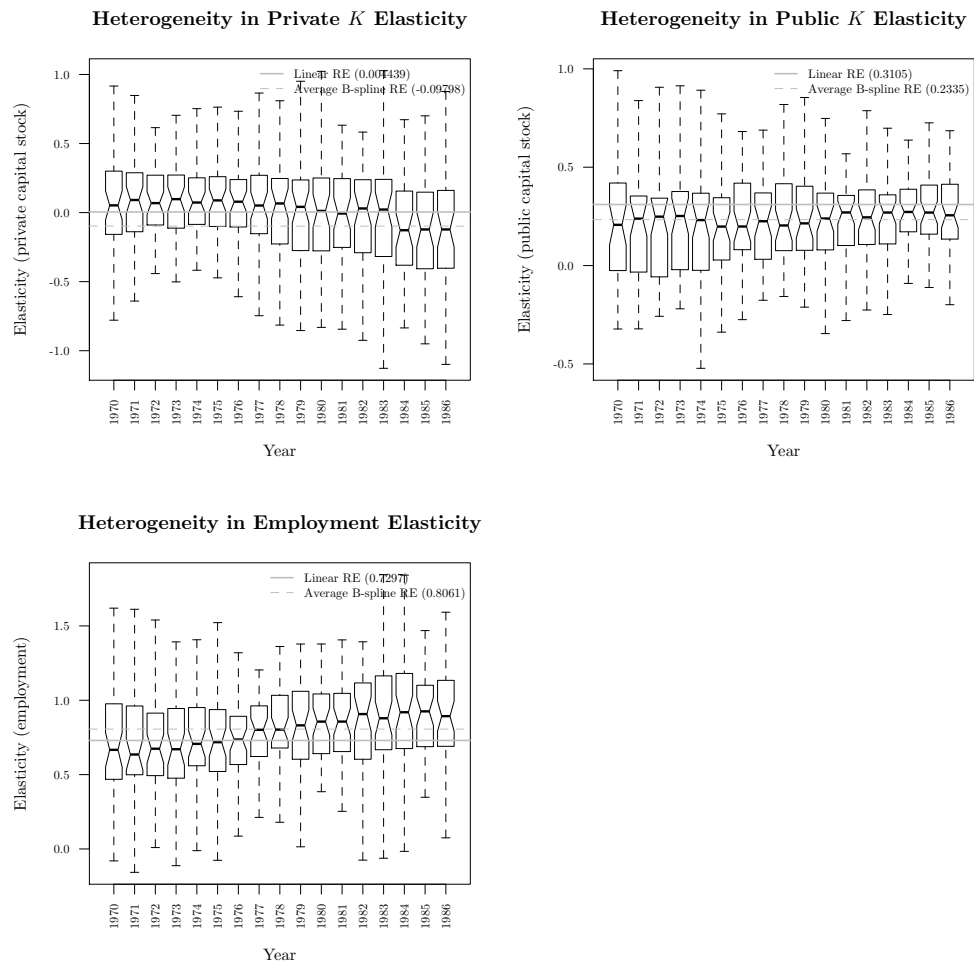


FIGURE B.6. Heterogeneity in B-spline RE elasticity estimates by year (boxplots). Horizontal lines represent Linear RE estimates and average of mean B-spline RE estimates taken over both year and state.

**B.4.  $AIC_c$  Cross Validation, All Variables Forced to Be Relevant (degree  $\geq 1$ ).** Figures B.7 and B.8 present results for the case where  $AIC_c$  cross-validation is used and where the degree of the polynomial is  $\geq 1$ , thereby forcing all variables to be included in the resulting estimate. Figure B.7 presents the mean elasticities along with asymptotic 95% confidence intervals, while Figure B.8 presents boxplot summaries which highlight the median and interquartile ranges.

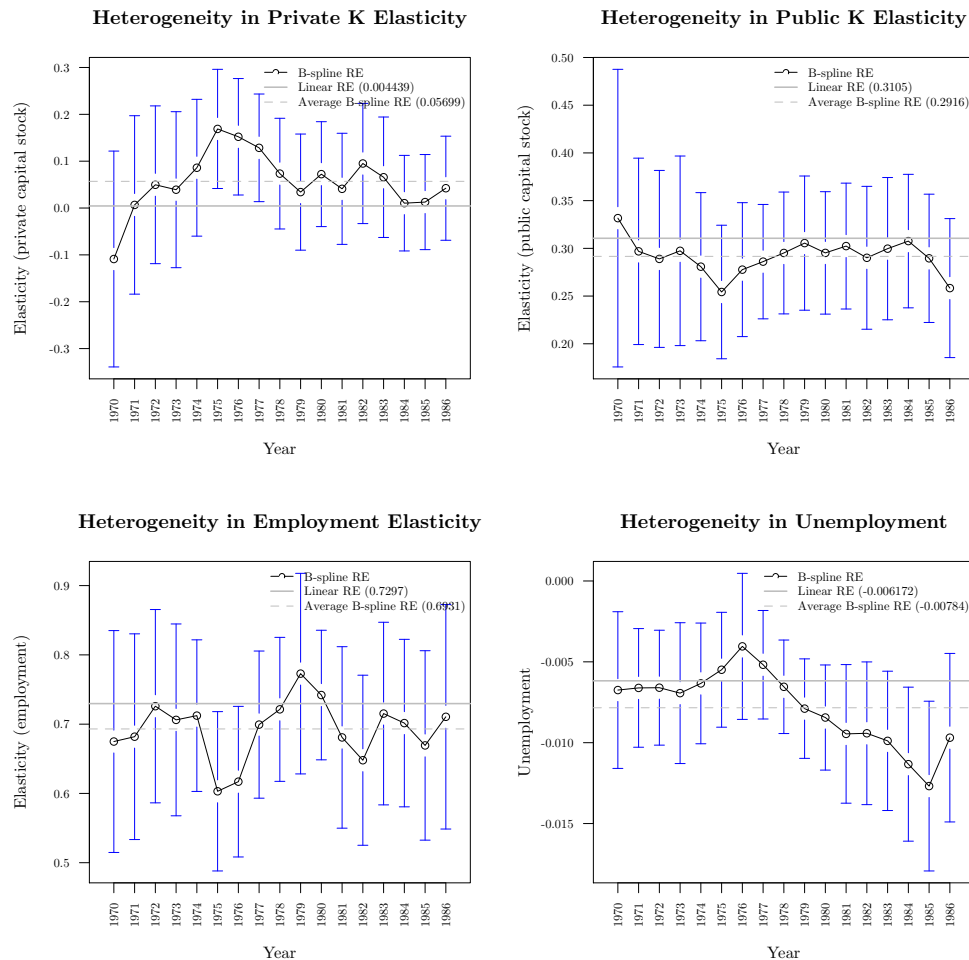


FIGURE B.7. Heterogeneity in B-spline RE elasticity estimates by year (solid line with dots is the average taken over state elasticities by year, vertical bars show 95% confidence intervals). Horizontal lines represent Linear RE estimates and average of mean B-spline RE estimates taken over both year and state.

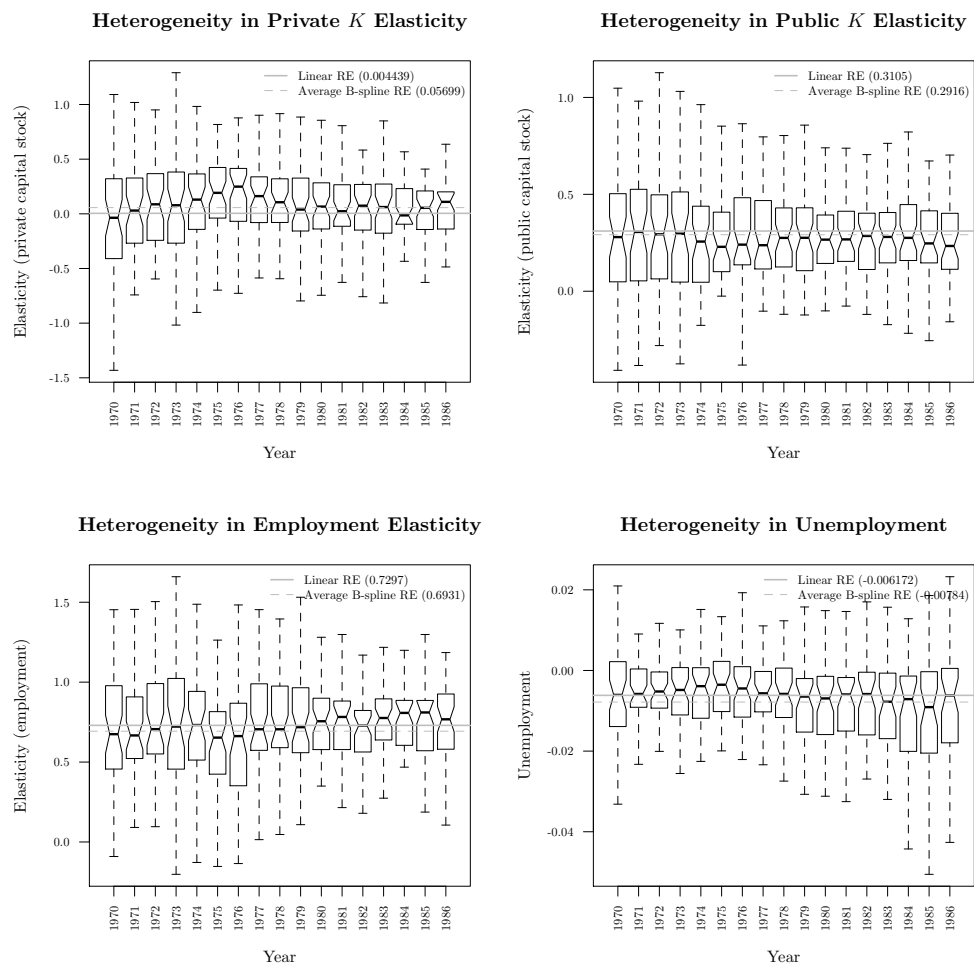


FIGURE B.8. Heterogeneity in B-spline RE elasticity estimates by year (boxplots). Horizontal lines represent Linear RE estimates and average of mean B-spline RE estimates taken over both year and state.

**B.5. Individual Model Summaries.** Tables B.9 and B.10 present results for the pooled and random effects linear parametric specifications.

TABLE B.9. Linear Parametric Pooled OLS Model

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	1.6433	0.0576	28.54	0.0000
log(pcap)	0.1550	0.0172	9.04	0.0000
log(pc)	0.3092	0.0103	30.10	0.0000
log(emp)	0.5939	0.0137	43.20	0.0000
unemp	-0.0067	0.0014	-4.75	0.0000

TABLE B.10. Linear Parametric Random Effects Model

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	2.1354	0.1335	16.00	0.0000
log(pcap)	0.0044	0.0234	0.19	0.8497
log(pc)	0.3106	0.0198	15.68	0.0000
log(emp)	0.7297	0.0249	29.28	0.0000
unemp	-0.0062	0.0009	-6.80	0.0000

(Shujie Ma) DEPARTMENT OF STATISTICS, UNIVERSITY OF CALIFORNIA-RIVERSIDE, SHUJIE.MA@UCR.EDU

(Jeffrey S. Racine) DEPARTMENT OF ECONOMICS AND GRADUATE PROGRAM IN STATISTICS, MCMASTER UNIVERSITY, RACINEJ@MCMASTER.CA; DEPARTMENT OF ECONOMICS AND FINANCE, LA TROBE UNIVERSITY; INFO-METRICS INSTITUTE, AMERICAN UNIVERSITY; RIMINI CENTER FOR ECONOMIC ANALYSIS

(Aman Ullah) DEPARTMENT OF ECONOMICS, UNIVERSITY OF CALIFORNIA RIVERSIDE, AMAN.ULLAH@UCR.EDU