10-2014

Additive nonparametric regression in the presence of endogenous regressors

Deniz OZABACI
University of Connecticut - Storrs

Daniel J. HENDERSON
University of Alabama - Tuscaloosa

Liangjun SU
Singapore Management University, ljsu@smu.edu.sg

Citation

OZABACI, Deniz; HENDERSON, Daniel J.; and SU, Liangjun. Additive nonparametric regression in the presence of endogenous regressors. (2014). *Journal of Business and Economic Statistics*. 32, (4), 555-575.
Available at: https://ink.library.smu.edu.sg/soe_research/1634

Published in Journal of Business & Economic Statistics, 2014, Volume 32, Issue 4, 555-575.
http://dx.doi.org/10.1080/07350015.2014.917590

Additive Nonparametric Regression in the Presence of Endogenous Regressors

Deniz Ozabaci*
Department of Economics
State University of New York at Binghamton

Daniel J. Henderson†
Department of Economics, Finance and Legal Studies
University of Alabama

Liangjun Su‡
School of Economics
Singapore Management University

March 22, 2013

Abstract

In this paper we present an oracle efficient estimator for a structural equation model under full additivity constraint. We propose estimators for both the conditional mean and the gradients and they allow us to take care of both curse of dimensionality and oracle efficiency. We show that our estimators are consistent, asymptotically normal and oracle efficient. Monte Carlo simulations support the asymptotic developments.

In the second part of our paper, we apply our model to the question of effects of childcare use on children's cognitive outcomes, to test our estimators in a real life example. With the help of our model we take care of the endogeneity problem using high number of instruments.

Keywords: Additive Regression, Endogeneity, Generated Regressors, Oracle Estimation, Nonparametric, Structural Equation

JEL Codes: C14

---

*Deniz Ozabaci, Department of Economics, State University of New York, Binghamton, NY 13902-6000, (607) 777-2572, Fax: (607) 777-2681, e-mail: dozabac1@binghamton.edu.

†Corresponding author: Daniel J. Henderson, Department of Economics, Finance and Legal Studies, University of Alabama, Tuscaloosa, AL 35487-0224, (205) 348-8991, Fax: (205) 348-0186, e-mail: djhender@cba.ua.edu.

‡Liangjun Su, School of Economics, Singapore Management University, 90 Stamford Road, Singapore, 178903; Tel: (65) 6828-0386; e-mail: ljsu@smu.edu.sg.

1 Introduction

In this paper we consider nonparametric estimation of structural equations. There are various examples in the literature of those who consider endogeneity within the context of nonparametric regression (e.g., Newey and Powell 2003; Newey et al., 1999; Roehing, 1988; Su and Ullah, 2008; Vella, 1991). Different from the aforementioned papers, we are interested in imposing a full additivity constraint on the conditional mean of each equation while at the same time employing kernel methods.

Our work is most similar to the models in Newey et al. (1999) and Su and Ullah (2008). While the assumptions of their models are relatively restrictive as compared to other examples in the literature, their estimators are typically much easier to implement, which is useful for applied work. Newey et al. (1999) discuss the identification strategy as well as propose a series estimator. Su and Ullah (2008) take their general framework and employ fully nonparametric kernel regression via a three-step procedure. While we prefer kernel methods in practice (especially in the presence of discrete covariates), the fully nonparametric estimator suffers from the curse of dimensionality.

Here, differently, we impose additivity constraint on all our stages and we propose a three step estimation procedure for our additively separable nonparametric structural equation model. We employ series estimators for our first two stages where we take the consistent estimates of the partial residuals and replace those to our second stage regression and finally turn to kernel regression in our final step via one stage backfitting. This process gives us an estimator which is free of curse of dimensionality.

In addition to avoiding the curse of dimensionality, our additively separable estimator achieves a high efficiency level compared to the listed nonparametric examples in the literature. We show that our proposed estimators for additive components have the oracle property, in other words they can be estimated at the high efficiency level as if the rest of the smooth functions were known.

In addition to conditional mean estimates, we also provide the gradient estimates as these are of great interest to economists. We show that our gradients are consistent and asymptotically normal. Furthermore, we propose a partially linear extension of our estimators and discuss the related asymptotic properties as well. Finite sample results are also analyzed via a set of Monte Carlo simulations and their results support the asymptotic developments.

In the final part of our paper we aim to analyze the real life data performance of our estimators. We analyze the effects of childcare use on cognitive outcomes of children, taking care of the endogeneity problem. By the help of our model, which is free of curse of dimensionality and which is also oracle efficient, we solve the endogeneity issue using an extensive set of instrumental variables those we adopt from Bernal and Keane (2011). This way we can answer this empirical question via a strong instrument list and use a flexible estimator that will give us more insights regarding the underlying heterogeneity amongst the conditional mean and the partial effect estimates.

Rest of our paper is structured as following. Section 2 describes our methodology, i.e. our model and proposed estimators. Section 3 presents the main asymptotic results those describe asymptotic properties for the listed estimators, and a partially linear extension of these estimators as well. Section 4 describes our application and the data that we use. Section 5 presents and discusses our main estimation results. And Section 6 concludes.

2 Methodology

To proceed, we adopt the following notation. For a real matrix A, we denote its transpose as A', its Frobenius norm as $\|A\|$ $(\equiv [\text{tr}(AA')]^{1/2})$, its spectral norm as $\|A\|_{\text{sp}}$ $(\equiv \sqrt{\lambda_{\max}(A'A)})$, where $\text{tr}(\cdot)$ is the trace operator, $\equiv$ means "is defined as", and $\lambda_{\max}(\cdot)$ denotes the largest eigenvalue of a real symmetric matrix. Note that the two norms are equal when A is a vector and they can be used interchangeably. We use $\lambda_{\min}(\cdot)$ to denote the smallest eigenvalue of a real symmetric matrix. For any function $q(\cdot)$ defined on the real line, we use $\dot{q}(\cdot)$ and $\ddot{q}(\cdot)$ to denote its first and second derivatives, respectively. We use $\overset{D}{\to}$ and $\overset{P}{\to}$ to denote convergence in distribution and probability, respectively.

2.1 Model

We start with the basic set-up of Newey et al. (1999). They consider a triangular system of the following form

$$\begin{cases} Y = g(\mathbf{X}, \mathbf{Z}_1) + \varepsilon, \\ \mathbf{X} = m(\mathbf{Z}_1, \mathbf{Z}_2) + \mathbf{U},\ E(\mathbf{U}|\mathbf{Z}_1, \mathbf{Z}_2) = 0,\ E(\varepsilon|\mathbf{Z}_1, \mathbf{Z}_2, \mathbf{U}) = E(\varepsilon|\mathbf{U}), \end{cases} \tag{2.1}$$

where $\mathbf{X}$ is a $d_x \times 1$ vector of endogenous regressors, $\mathbf{Z}_1 = (Z_{11}, ..., Z_{1d_1})'$ is a $d_1 \times 1$ vector of "included" exogenous regressors, $\mathbf{Z}_2 \equiv (Z_{21}, ..., Z_{2d_2})'$ is a $d_2 \times 1$ vector of "excluded" exogenous regressors, $g(\cdot, \cdot)$ denotes the true unknown structural function of interest, $m \equiv (m_1, ...m_{d_x})'$ is a $d_x \times 1$ vector of smooth functions of the instruments $\mathbf{Z}_1$ and $\mathbf{Z}_2$, and ε and $\mathbf{U} \equiv (U_1, ..., U_{d_x})'$ are error terms. Newey et al. (1999) are interested in estimating g consistently.

Newey et al. (1999) show that g can be identified up to an additive constant under the key identification conditions that $E(\mathbf{U}|\mathbf{Z}_1, \mathbf{Z}_2) = 0$ and $E(\varepsilon|\mathbf{Z}_1, \mathbf{Z}_2, \mathbf{U}) = E(\varepsilon|\mathbf{U})$. If these conditions hold, then

$$\begin{aligned} E(Y|\mathbf{X}, \mathbf{Z}_1, \mathbf{Z}_2, \mathbf{U}) &= g(\mathbf{X}, \mathbf{Z}_1) + E(\varepsilon|\mathbf{X}, \mathbf{Z}_1, \mathbf{Z}_2, \mathbf{U}) \\ &= g(\mathbf{X}, \mathbf{Z}_1) + E(\varepsilon|\mathbf{Z}_1, \mathbf{Z}_2, \mathbf{U}) \\ &= g(\mathbf{X}, \mathbf{Z}_1) + E(\varepsilon|\mathbf{U}). \end{aligned} \tag{2.2}$$

If $\mathbf{U}$ is observed, this is a standard additive nonparametric regression model. However, in practice, $\mathbf{U}$ is not observed and one needs to replace it by a consistent estimate. This motivates Su and Ullah (2008) to consider a three-stage procedure to obtain consistent estimates of g via the technique of local polynomial regression. In the first stage, they regress $\mathbf{X}$ on $(\mathbf{Z}_1, \mathbf{Z}_2)$ via local polynomial regression and obtain the residuals $\hat{\mathbf{U}}$ from this first-stage reduced-form regression. In the second stage, they estimate $E(Y|\mathbf{X}, \mathbf{Z}, \mathbf{U})$ via another local polynomial regression by regressing $\mathbf{Y}$ on $\mathbf{X}$, $\mathbf{Z}_1$, and $\hat{\mathbf{U}}$. In the third stage, they obtain the estimates of $g(\mathbf{x}, \mathbf{z}_1)$ via the method of marginal integration. Unlike previous works in the literature including Newey et al. (1999), Pinkse (2000), and Newey and Powell (2003) that are based upon two-stage series approximations and only establish mean square and uniform convergence, they establish the asymptotic distribution for their three-step local polynomial estimator.

There are two drawbacks associated with the estimate of Su and Ullah (2008). First, it is subject to the notorious "curse of dimensionality" problem in the nonparametric literature. Without any extra restriction, the convergence rate of their second- and third-stage estimators depend on $2d_x + d_1$ and

$d_x + d_1$, respectively, which can be quite slow if either d_x or d_1 is not small. As a result, their estimates may perform badly even for moderately large sample sizes when $d_x + d_1 \geq 3$. Second, their estimator does not have the oracle property which an optimal estimator of the additive component in a nonparametric regression model should exhibit. In this paper we try to address both issues.

To alleviate the curse of dimensionality problem, we propose to impose some structure on $g(\mathbf{X}, \mathbf{Z}_1)$, $E(\varepsilon|\mathbf{U})$, and $m_l(\mathbf{Z}_1, \mathbf{Z}_2)$, where $l = 1, ..., d_x$. Specifically, we assume that $E(\varepsilon) = 0$ and the above nonparametric objects have additive forms:

$$\begin{aligned} g(\mathbf{X}, \mathbf{Z}_1) &= \mu_g + g_1(X_1) + ... + g_{d_x}(X_{d_x}) + g_{d_x+1}(Z_{11}) + ... + g_{d_x+d_1}(Z_{1d_1}), \\ E(\varepsilon|\mathbf{U}) &= \mu_\varepsilon + g_{d_x+d_1+1}(U_1) + ... + g_{2d_x+d_1}(U_{d_x}), \text{ and} \\ m_l(\mathbf{Z}_1, \mathbf{Z}_2) &= \mu_l + m_{l,1}(Z_{11}) + ... + m_{l,d_1}(Z_{1d_1}) + m_{l,d_1+1}(Z_{21}) + ... + m_{l,d}(Z_{2d_2}), \ l = 1, ..., d_x, \end{aligned}$$

where $d = d_1 + d_2$. Consequently, we have

$$\begin{aligned} E(Y|\mathbf{X}, \mathbf{Z}_1, \mathbf{Z}_2, \mathbf{U}) &= \mu + g_1(X_1) + ... + g_{d_x}(X_{d_x}) + g_{d_x+1}(Z_{11}) + ... + g_{d_x+d_1}(Z_{1d_1}) \\ &\quad + g_{d_x+d_1+1}(U_1) + ... + g_{2d_x+d_1}(U_{d_x}) \equiv \bar{g}(\mathbf{X}, \mathbf{Z}_1, \mathbf{U}), \end{aligned} \tag{2.3}$$

where $\mu = \mu_g + \mu_\varepsilon$. Note that the g_j's are not fully identified without further restriction. Depending on the method that is used to estimate the additive components, different identification conditions can be imposed. [1]

Horowitz (2013) reviews methods for estimating nonparametric additive models, including the backfitting method, the marginal integration method, the series method, and the mixture of a series method and a backfitting method to obtain oracle efficiency. It is well known that it is more difficult to study the asymptotic property of the backfitting estimator than the marginal integration estimator, but the latter has a curse of dimensionality problem if one does not impose additivity at the outset of estimation as in conventional kernel methods. Other problems that are associated with the marginal integration estimator include its lack of oracle property and its heavy computational burden. Kim et al. (1999) try to address the latter two problems by proposing a fast instrumental variable (IV) pilot estimator. But they cannot avoid the *curse of dimensionality* problem at all. In fact, their IV pilot estimator depends on the estimation of the density function of the regressors at all data points. In addition, their paper totally ignores the notorious *boundary bias* problem for kernel density estimates and because their IV pilot estimate is not uniformly consistent on the full support, they have to use a *trimming* scheme to obtain the second-stage oracle estimator. To fully overcome the curse of dimensionality problem, Horowitz and Mammen (2004) propose a two-step estimation procedure with series estimation of the nonparametric additive components followed by a backfitting step that turns the series estimates into kernel estimates that are both oracle efficient and free of the curse of dimensionality.

Below we follow the lead of Horowitz and Mammen (2004) and propose a three-stage estimation procedure that is computationally efficient, oracle efficient, and fully overcomes the curse of dimensionality. We shall adopt the following identification restrictions

$$\begin{aligned} g_l(0) &= g_l(x_l)|_{x_l=0} = 0 \text{ for } l = 1, ..., 2d_x + d_1, \text{ and} \\ m_{l,k}(0) &= 0 \text{ for } l = 1, ..., d_x \text{ and } k = 1, 2, ..., d. \end{aligned}$$

[1] For example, for the method of marginal integration, a convenient set of identification conditions would be $E[g_l(X_l)] = 0$ for $l = 1, ..., d_x$, $E[g_{d_x+k}(Z_{1k})] = 0$ for $k = 1, ..., d_1$, $E[g_{d_x+d_1+l}(U_{1l})] = 0$ for $l = 1, ..., d_x$, $E[m_{l,k}(Z_{1k})] = 0$ for $k = 1, ..., d_1$ and $l = 1, ..., d_x$, $E[m_{l,d_1+j}(Z_{2j})] = 0$ for $j = 1, ..., d_2$ and $l = 1, ..., d_x$.

Similar identification conditions are also adopted in Li (2000).

2.2 Estimation

Given a random sample of n observations $\{Y_i, \mathbf{X}_i, \mathbf{Z}_{1i}, \mathbf{Z}_{2i}\}_{i=1}^n$ where $\mathbf{X}_i = (X_{1i}, ..., X_{d_x i})'$, $\mathbf{Z}_{1i} = (Z_{11,i}, ..., Z_{1d_1,i})'$, and $\mathbf{Z}_{2i} = (Z_{21,i}, ..., Z_{2d_2,i})'$, we propose the following three-stage estimation procedure:

1. For $l = 1, ..., d_x$, let $\tilde{\mu}_l$, $\{\tilde{m}_{l,k}(Z_{1k,i}), k = 1, ..., d_1\}$, and $\{\tilde{m}_{l,d_1+j}(Z_{2j,i}), j = 1, ..., d_2\}$, denote the sieve estimates of $\tilde{\mu}_l$, $\{m_{l,k}(Z_{1k,i}), k = 1, ..., d_1\}$, and $\{m_{l,d_1+j}(Z_{2j,i}), j = 1, ..., d_2\}$ in the nonparametric additive regression

$$X_{li} = \mu_l + m_{l,1}(Z_{11,i}) + ... + m_{l,d_1}(Z_{1d_1,i}) + m_{l,d_1+1}(Z_{21,i}) + ... + m_{l,d}(Z_{2d_2,i}) + U_{1i}.$$

   Let $\tilde{U}_{li} \equiv X_{li} - \tilde{\mu}_l - \tilde{m}_{l,1}(Z_{11,i}) - ... - \tilde{m}_{l,d_1}(Z_{1d_1,i}) - \tilde{m}_{l,d_1+1}(Z_{21,i}) - ... - \tilde{m}_{l,d}(Z_{2d_2,i})$ for $l = 1, ..., d_x$ and $i = 1, ..., n$.

2. Estimate μ, $\{g_l(X_{li}), l = 1, ..., d_x\}$, $\{g_{d_x+j}(Z_{1j,i}), j = 1, ..., d_1\}$, $\{g_{d_x+d_1+k}(\tilde{U}_{ki}), k = 1, ..., d_x\}$, in the following additive regression model

$$\begin{aligned} Y_i &= \mu + g_1(X_{1i}) + ... + g_{d_x}(X_{d_x i}) + g_{d_x+1}(Z_{11,i}) + ... + g_{d_x+d_1}(Z_{1d_1,i}) \\ &\quad + g_{d_x+d_1+1}(\tilde{U}_{1i}) + ... + g_{2d_x+d_1}(\tilde{U}_{d_x i}) + \epsilon_i \end{aligned}$$

   by the series method. Here $\epsilon_i = \varepsilon_i + g_{d_x+d_1+1}(U_{1i}) + ... + g_{2d_x+d_1}(U_{d_x i}) - g_{d_x+d_1+1}(\tilde{U}_{1i}) - ... - g_{2d_x+d_1}(\tilde{U}_{d_x i})$ denotes the new error term. Denote the estimates as $\tilde{\mu}$, $\{\tilde{g}_l(X_{li}), l = 1, ..., d_x\}$, $\{\tilde{g}_{d_x+j}(Z_{1j,i}), j = 1, ..., d_1\}$, $\{\tilde{g}_{d_x+d_1+k}(\tilde{U}_{ki}), k = 1, ..., d_x\}$.

3. Estimate $g_1(x_1)$ and its first order derivative by the local linear regression of $\tilde{Y}_{1i} = Y_i - \tilde{\mu} - \tilde{g}_2(X_{2i}) - ... - \tilde{g}_{d_x}(X_{d_x i}) - \tilde{g}_{d_x+1}(Z_{11,i}) - ... - \tilde{g}_{d_x+d_1}(Z_{1d_1,i}) - \tilde{g}_{d_x+d_1+1}(\tilde{U}_{1i}) - ... - \tilde{g}_{2d_x+d_1}(\tilde{U}_{d_x i})$ on X_{1i}. Analogously, one obtains estimates of the other additive components in (2.3) and their first order derivatives.

In comparison with Horowitz and Mammen (2004), the above first stage is new as we have to replace the unobservable U_{li} by their consistent estimates in the second stage. In addition, Horowitz and Mammen (2004) are only interested in the estimation of the nonparametric additive components themselves, while we are also interested in estimating the first order derivatives (gradients).

Alternatively one could follow Kim et al. (1999) and use the kernel estimator in the first two stages. The oracle estimator of Kim et al. (1999) has gained popularity in recent years. For example, Ozabaci and Henderson (2012) obtain the gradients of their estimator for the local-constant case and Martins-Filho and Yang (2007) consider the local-linear version of the oracle estimator, both assuming strictly exogenous regressors. But as mentioned above, using the kernel estimators in the first two stage here have several disadvantages and does not avoid the curse of dimensionality problem at all.

For notational simplicity, let $\mathbf{W} = (\mathbf{X}', \mathbf{Z}_1', \mathbf{U}')'$ and $\mathbf{w} = (\mathbf{x}', \mathbf{z}_1', \mathbf{u})'$, where, e.g., $\mathbf{u} = (u_1, ..., u_{d_x})'$ denotes a realization of $\mathbf{U}$. We shall use $\mathcal{Z} \equiv \mathcal{Z}_1 \times \mathcal{Z}_2$ and $\mathcal{W} \equiv \mathcal{X} \times \mathcal{Z}_1 \times \mathcal{U}$ to denote the support of $(\mathbf{Z}_1, \mathbf{Z}_2)$ and $\mathbf{W}$, respectively. Let $\{p_l(\cdot), l = 1, 2, ...\}$ denote a sequence of basis functions that can approximate any square-integrable function very well (to be precise later). Let $\kappa_1 = \kappa_1(n)$ and $\kappa = \kappa(n)$ be some integers such that $\kappa_1, \kappa \to \infty$ as $n \to \infty$. Let $p^{\kappa_1}(v) \equiv [p_1(v), ...p_{\kappa_1}(v)]'$. Define

$$\begin{aligned} P^{\kappa_1}(\mathbf{z}_1, \mathbf{z}_2) &\equiv \left[1,\ p^{\kappa_1}(z_{11})', ..., p^{\kappa_1}(z_{1d_1})', p^{\kappa_1}(z_{21})', ..., p^{\kappa_1}(z_{2d_2})'\right]', \\ \Phi^{\kappa}(\mathbf{w}) &\equiv \left[1,\ p^{\kappa}(x_1)', ..., p^{\kappa}(x_{d_x})', p^{\kappa}(z_{11})', ..., p^{\kappa_1}(z_{1d_1})', p^{\kappa}(u_1)', ..., p^{\kappa}(u_{d_x})'\right]'. \end{aligned}$$

For each $(\mathbf{z}_1, \mathbf{z}_2) \in \mathcal{Z}$, we approximate $m_l(\mathbf{z}_1, \mathbf{z}_2)$ and $\bar{g}(\mathbf{w})$ by $P^{\kappa_1}(\mathbf{z}_1, \mathbf{z}_2)' \alpha_l$ and $\Phi^{\kappa}(\mathbf{w})' \beta$, respectively, for $l = 1, ..., d_x$, where $\alpha_l \equiv (\mu_l, \alpha'_{l,1}, ..., \alpha'_{l,d})'$ and $\beta = (\mu, \beta'_1, ..., \beta'_{2d_x+d_1})'$ are $(1 + d\kappa_1) \times 1$ and $(1 + (2d_x + d_1)\kappa) \times 1$ vectors of unknown parameters to be estimated. Here, each $\alpha_{l,k}$, $k = 1, ..., d$, is a $\kappa_1 \times 1$ vector; each β_j, $j = 1, ..., 2d_x + d_1$, is a $\kappa \times 1$ vector. Let $\mathbb{S}_{1k}$ and $\mathbb{S}_k$ denote $\kappa_1 \times (1 + d\kappa_1)$ and $\kappa \times (1 + (2d_x + d_1)\kappa)$ selection matrices, respectively, such that $\mathbb{S}_{1k}\alpha_l = \alpha_{l,k}$ and $\mathbb{S}_k \beta_l = \beta_l$.

To obtain the first stage estimators of the m_l's, let $\tilde{\alpha}_l \equiv (\tilde{\mu}_l, \tilde{\alpha}'_{l,1}, ..., \tilde{\alpha}'_{l,d})'$ be the solution to $\min_{\alpha_l} n^{-1} \times \sum_{i=1}^n \left[X_{li} - P^{\kappa_1}(\mathbf{Z}_{1i}, \mathbf{Z}_{2i})' \alpha_l\right]^2$. The series estimator of $m_l(\mathbf{z})$ is given by

$$\begin{aligned} \tilde{m}_l(\mathbf{z}_1, \mathbf{z}_2) &= P^{\kappa_1}(\mathbf{z}_1, \mathbf{z}_2)' \tilde{\alpha}_l \\ &= P^{\kappa_1}(\mathbf{z}_1, \mathbf{z}_2) \left[n^{-1} \sum_{i=1}^n P^{\kappa_1}(\mathbf{Z}_{1i}, \mathbf{Z}_{2i}) P^{\kappa_1}(\mathbf{Z}_{1i}, \mathbf{Z}_{2i})'\right]^{-} n^{-1} \sum_{i=1}^n P^{\kappa_1}(\mathbf{Z}_{1i}, \mathbf{Z}_{2i}) X_{li} \\ &= \tilde{\mu}_l + \sum_{k=1}^{d_1} \tilde{m}_{l,k}(z_{1k}) + \sum_{j=1}^{d_2} \tilde{m}_{l,d_1+j}(z_{2j}) \end{aligned}$$

where A^- denotes the Moore-Penrose generalized inverse of A, $\tilde{m}_{l,k}(z_{1k}) = p^{\kappa_1}(z_{1k})' \tilde{\alpha}_{l,k}$ is a series estimator of $m_{l,k}(z_{1k})$ for $k = 1, ..., d_1$, and $\tilde{m}_{l,d_1+j}(z_{2j}) = p^{\kappa_1}(z_{2j})' \tilde{\alpha}_{l,d_1+j}$ is a series estimator of $m_{l,d_1+j}(z_{2j})$ for $j = 1, ..., d_2$.

To obtain the second stage estimators of the g_l's, let $\tilde{\beta} \equiv (\tilde{\mu}, \tilde{\beta}'_1, ..., \tilde{\beta}'_{2d_x+d_1})'$ be a solution to $\min_\beta n^{-1} \sum_{i=1}^n \left[Y_i - P^{\kappa}\left(\tilde{\mathbf{W}}_i\right)' \beta\right]^2$, where $\tilde{\mathbf{W}}_i = \left(\mathbf{X}'_i, \mathbf{Z}'_{1i}, \tilde{\mathbf{U}}'_i\right)'$ and $\tilde{\mathbf{U}}_i = (\tilde{U}_{1i}, ..., \tilde{U}_{d_x i})'$. The series estimator of $\bar{g}(\mathbf{w})$ is given by

$$\tilde{\bar{g}}(\mathbf{w}) = P^{\kappa}(\mathbf{w})' \tilde{\beta} = \tilde{\mu} + \sum_{l=1}^{d_x} \tilde{g}_l(x_l) + \sum_{k=1}^{d_1} \tilde{g}_{d_x+k}(z_{1k}) + \sum_{j=1}^{d_x} \tilde{g}_{d_x+d_1+k}(u_j).$$

Let $\beta_1(x_1) \equiv [g_1(x_1), \dot{g}_1(x_1)]'$. We use $\hat{\beta}_1(x_1) \equiv [\hat{g}_1(x_1), \hat{\dot{g}}_1(x_1)]'$ to denote the local linear estimate of $\beta_1(x_1)$ in the third stage by using the kernel function $K(\cdot)$ and bandwidth h. Let $\tilde{\mathbf{Y}}_1 \equiv (\tilde{Y}_{11}, ..., \tilde{Y}_{1n})'$, $X^*_{1i}(x_1) \equiv [1, X_{1i} - x_1]$, $\mathbb{X}_1(x_1) \equiv [X^*_{11}(x_1), ..., X^*_{1n}(x_1)]'$, and $\mathbb{K}_{x_1} \equiv$diag$(K_{1x_1}, ..., K_{nx_1})$ where $K_{ix_1} \equiv K_h(X_{1i} - x_1)$ and $K_h(u) \equiv K(u/h)/h$. Then

$$\hat{\beta}_1(x_1) = \left[\mathbb{X}_1(x_1)' \mathbb{K}_{x_1} \mathbb{X}_1(x_1)\right]^{-1} \mathbb{X}_1(x_1)' \mathbb{K}_{x_1} \mathbb{X}_1(x_1) \tilde{\mathbf{Y}}_1.$$

Below we will study the asymptotic properties of $\hat{\beta}_1(x_1)$ via the study of asymptotic expansion for $\tilde{\beta}$.

3 Main Asymptotic Results

In this section we first provide assumptions that are used to prove the main results and then study the asymptotic properties of the proposed estimators. We also discuss the possible extension to partially linear additive models.

3.1 Assumptions

A real-valued function q on the real line is said to satisfy a Hölder condition with exponent $r \in [0,1]$ if there is c_q such that $|q(v)-q(\tilde{v})| \le c_q|v-\tilde{v}|^r$ for all v and $\tilde{v}$ on the support of q. q is said to be γ-smooth, $\gamma = r+m$, if it is m-times continuously differentiable on $\mathcal{U}$ and its mth derivative, $\partial^m q$, satisfies a Hölder condition with exponent r. The γ-smooth class of functions are popular in econometrics because a γ-smooth function can be approximated well by various linear sieves; see, e.g., Chen (2007). For any scalar function q on the real line that has r derivatives and support $\mathcal{S}$, let $|q|_r \equiv \max_{s\le r}\sup_{v\in\mathcal{S}}|\partial^s q(v)|$.

Let $\mathcal{X}_l$ and $\mathcal{U}_l$ denote the support of X_l and U_l, respectively, for $l=1,...,d_x$. Let $\mathcal{Z}_{sk}$ denote the support of Z_{sk} for $k=1,...,d_s$ and $s=1,2$. We shall use Y_i, $\mathbf{W}_i \equiv (\mathbf{X}_i', \mathbf{Z}_{1i}, \mathbf{U}_i')'$, $\mathbf{Z}_{2i}$, and U_{li} to denote the ith random observation of Y, $\mathbf{W}$, $\mathbf{Z}_2$, and U_l, respectively. Let $Q_{PP} \equiv E[P^{\kappa_1}(\mathbf{Z}_1,\mathbf{Z}_2)P^{\kappa_1}(\mathbf{Z}_1,\mathbf{Z}_2)']$, $Q_{\Phi\Phi} \equiv E[\Phi^{\kappa}(\mathbf{W})\Phi^{\kappa}(\mathbf{W})']$, and $Q_{PP,U_l} = E[P^{\kappa_1}(\mathbf{Z}_1,\mathbf{Z}_2)P^{\kappa_1}(\mathbf{Z}_1,\mathbf{Z}_2)'U_l^2]$ for $l=1,...,d_x$. We make the following set of basic assumptions.

Assumption A1. (i) $\{(Y_i,\mathbf{X}_i,\mathbf{Z}_{1i},\mathbf{Z}_{2i}), i=1,...,n\}$ are an IID random sample.

(ii) The supports $\mathcal{W}$ and $\mathcal{Z}$ of $\mathbf{W}_i$ and $(\mathbf{Z}_{1i},\mathbf{Z}_{2i})$ are compact.

(iii) The distributions of $\mathbf{W}_i$ and $(\mathbf{Z}_{1i},\mathbf{Z}_{2i})$ are absolutely continuous with respect to the Lebesgue measure.

Assumption A2.(i) For every κ_1 that is sufficiently large, there exist $\underline{c}_1$ and $\bar{c}_1$ such that $0<\underline{c}_1\le\lambda_{\min}(Q_{PP})\le\lambda_{\max}(Q_{PP})\le\bar{c}_1<\infty$, and $\lambda_{\max}(Q_{PP,U_l})\le\bar{c}_1<\infty$ for $l=1,...,d_x$.

(ii) For every κ that is sufficiently large, there exist $\underline{c}_2$ and $\bar{c}_2$ such that $0<\underline{c}_2\le\lambda_{\min}(Q_{\Phi\Phi})\le\lambda_{\max}(Q_{\Phi\Phi})\le\bar{c}_2<\infty$.

(iii) The functions $\{m_{l,k}(\cdot),\ l=1,...,d,k=1,...,d\}$ and $\{g_j(\cdot),\ j=2d_x+d_1\}$ belong to the class of γ-smooth functions with $\gamma\ge 2$.

(iv) There exist $\alpha_{l,k}$'s such that $\sup_{z\in\mathcal{Z}_{1k}}|m_{l,k}(z)-p^{\kappa_1}(z)'\alpha_{l,k}| = O(\kappa_1^{-\gamma})$ for $l=1,...,d_x$ and $k=1,...,d_1$, $\sup_{z\in\mathcal{Z}_{2k}}|m_{l,d_1+k}(z)-p^{\kappa_1}(z)'\alpha_{l,d_1+k}| = O(\kappa_1^{-\gamma})$ for $l=1,...,d_x$ and $k=1,...,d_2$.

(v) There exist β_l's such that $\sup_{x\in\mathcal{X}_l}|g_l(x)-p^{\kappa}(x)'\beta_l| = O(\kappa^{-\gamma})$ for $l=1,...,d_x$, $\sup_{z\in\mathcal{Z}_{1l}}|g_{d_z+k}(\cdot)-p^{\kappa}(z)'\beta_{d_z+k}| = O(\kappa^{-\gamma})$ for $k=1,...,d_1$, and $|g_{d_x+d_1+l}(\cdot)-p^{\kappa}(\cdot)'\beta_{d_x+d_1+l}|_1 = O(\kappa^{-\gamma})$ for $l=1,...,d_x$.

(vi) The set of basis functions, $\{p_j(\cdot), j=1,2,...\}$, are twice continuously differentiable everywhere on the support of U_{li} for $l=1,...,d_x$. $\max_{1\le l\le d_x}\max_{0\le s\le r}\sup_{u_l\in\mathcal{U}_l}\|\partial^s p^{\kappa}(u_l)\|\le\varsigma_{r\kappa}$ for $r=0,1,2$.

Assumption A3. (i) The probability density functions (PDF) of any two elements in $\mathbf{W}_i$ are bounded, bounded away from zero, and twice continuously differentiable.

(ii) Let $e_i \equiv Y_i-\bar{g}(\mathbf{X}_i,\mathbf{Z}_{1i},\mathbf{U}_i)$ and $\sigma_i^2\equiv\sigma^2(\mathbf{X}_i,\mathbf{Z}_{1i},\mathbf{Z}_{2i},\mathbf{U}_i)\equiv E\left(e_i^2|\mathbf{X}_i,\mathbf{Z}_{1i},\mathbf{Z}_{2i},\mathbf{U}_i\right)$. Let $Q_{sk,pp}\equiv E[p^{\kappa_1}(Z_{sk,i})p^{\kappa_1}(Z_{sk,i})'\sigma_i^2]$ for $k=1,...,d_s$ and $s=1,2$. The largest eigenvalue of $Q_{sk,pp}$ is bounded uniformly in κ_1.

Assumption A4. The kernel function $K(\cdot)$ is a PDF that is symmetric, bounded, and has compact support $[-c_K,c_K]$. It satisfies the Lipschitz condition $|K(v_1)-K(v_2)|\le C_K|v_1-v_2|$ for all $v_1,v_2\in[-c_K,c_K]$.

Assumption A5.

(i) $\kappa_1\le\kappa$. As $n\to\infty$, $\kappa_1\to\infty$, $\kappa^3/n\to 0$, and $\tau_n\to c_1\in[0,\infty)$, where $\tau_n\equiv\left(\kappa^{1/2}\varsigma_{0\kappa}+\varsigma_{1\kappa}\right)\nu_{1n}+\varsigma_{0\kappa}\varsigma_{2\kappa}\nu_{1n}^2$, $\nu_{1n}\equiv\kappa_1^{1/2}/n^{1/2}+\kappa_1^{-\gamma}$ and $\nu_n\equiv\kappa^{1/2}/n^{1/2}+\kappa^{-\gamma}$.

(ii) As $n \to \infty$, $h \to 0$, $nh^3 \log n \to \infty$, $nh\kappa^{-2\gamma} \to 0$, $\tau_n \nu_{1n} = o(n^{-1/2}h^{-1/2})$, and $[h^{1/2}\varsigma_{1\kappa}(1+n^{1/2}\kappa_1^{-\gamma}) + \varsigma_{2\kappa} n^{1/2} h^{1/2} \nu_{1n}^2](\nu_n + \nu_{1n}) \to 0$.

Assumptions A1(i)-(ii) impose IID sampling and compactness on the support of the exogenous independent variables. Either assumption can be relaxed at lengthy arguments; see, e.g., Su and Jin (2012) who allow for both weakly dependent data and infinite support for their regressors. A1(iii) requires that the variables in $\mathbf{W}_i$ and $(\mathbf{Z}_{1i}, \mathbf{Z}_{2i})$ be continuously valued, which is standard in the literature on sieve estimation. The extension to allow for both continuous and discrete variables is possible but will not be pursued in this paper.

Assumption A2(i)-(ii) ensure the existence and nonsingularity of the covariance matrix of the asymptotic form of the first two stage estimators. They are standard in the literature; see, e.g., Newey (1997), Li (2000), and Horowitz and Mammen (2004). Note that all of these authors assume that the conditional variances of the error terms given the exogenous regressors are uniformly bounded, in which case the second part of A1(i) becomes redundant. A2(iii) imposes smoothness conditions on the relevant functions and A2(iv)-(v) quantifies the approximation error for γ-smooth functions. These conditions are satisfied, for example, for polynomials, splines, and wavelets. A2(vi) is needed for the application of Taylor expansions. It is well known that $\varsigma_{r\kappa} = O\left(\kappa^{r+1/2}\right)$ and $O\left(\kappa^{2r+1}\right)$ for B-splines and power series, respectively; see Newey (1997). The rate at which splines uniformly approximate a function is the same as that for power series, so that the uniform convergence rate for splines is faster than power series. In addition, the low multicollinearity of B-splines and recursive formula for calculation also leads to computational advantages; see Powell (1981, Chapter 19) and Schumaker (2007, Chapter 4). For these reasons, B-splines are widely used in the literature.

Assumptions A3(i)-(ii) and A4 are needed for the establishment of the asymptotic property of the third stage estimator. A3(ii) is redundant under Assumption A2(i) if one assumes that the conditional variances of e_i's given $(\mathbf{X}_i, \mathbf{Z}_{1i}, \mathbf{Z}_{2i}, \mathbf{U}_i)$ are uniformly bounded. A4 is standard for local linear regression; see Fan and Gijbels (1996) and Masry (1996). The compact support condition is convenient for the demonstration of the uniform convergence rate in Theorem 3.2 below. It can be removed at the cost of some lengthy arguments; see, e.g., Hansen (2008). In particular, the Gaussian kernel can be applied. Assumptions A5(i)-(ii) specify conditions on κ_1, κ, and h. Note that we allow the use of different series approximation terms in the first and second stage estimation, which allows us to see clearly the effect of the first stage estimates on the second stage estimates. The first condition (namely, $\kappa_1 \le \kappa$) in A5(i) is needed for the proof of a technical lemma (see Lemma A.5(iii)) in the appendix, and it can be removed at the cost of some additional assumption on the basis functions. The terms that are associated with ν_{1n} arise because of the use of the nonparametrically generated regressors in the second stage series estimation. The appearance of $\log n$ is due to the goal to establish some uniform consistency result in Theorem 3.2 below and it can be replaced by 1 if one is only interested in the pointwise result. In the case where $\varsigma_{r\kappa} = O\left(\kappa^{r+1/2}\right)$ in Assumption A2(vi), $\tau_n = O\left(\kappa^{3/2}\nu_{1n} + \kappa^3 \nu_{1n}^2\right)$. In practice, we recommend setting $\kappa_1 = \kappa$. These restrictions, in conjunction with the condition $\gamma \ge 2$, imply that the conditions in Assumption A5 can be greatly simplified as follows:

Assumption A5*.

(i) As $n \to \infty$, $\kappa \to \infty$, $\kappa^4/n \to c_1 \in [0, \infty)$.

(ii) As $n \to \infty$, $h \to 0$, $nh^3 \log n \to \infty$, $nh\kappa^{-2\gamma} \to 0$, and $n^{-1}h\kappa^5 \to 0$.

3.2 Asymptotic properties

In this section we state two theorems that given the main results of the paper. Even though several results are available in the literature on nonparametric or semiparametric regressions with nonparametrically generated regressors [see, e.g., Mammen et al. (2012) and Hahn and Ridder (2013) for recent contributions], none of them can be directly applied to our framework. In particular, Hahn and Ridder (2013) study the asymptotic distribution of three-step estimators of a *finite-dimensional* parameter vector where the second step consists of one or more nonparametric generated regressions on a regressor that is estimated in the first step. In sharp contrast, our third-stage estimator is also a nonparametric estimator. Under fairly general conditions, Mammen et al. (2012) focus on two-stage nonparametric regression where the first stage can be kernel or series estimation while the second stage is a local linear estimation. In principle, one can treat our second and third stage estimation as their first and second stage estimation, respectively, and then apply their results to our case. But their results are built upon high-level assumptions and are usually not optimal. For this reason, we derive the asymptotic properties of our three-stage estimators under some primitive conditions specified in the preceding section.

The asymptotic properties of the second stage series estimator $\tilde{\beta}$ are reported in the following theorem.

Theorem 3.1 *Suppose that Assumptions A.1-A.5(i) hold. Then*

(i) $\tilde{\beta}-\beta = Q_{\Phi\Phi}^{-1}n^{-1}\sum_{i=1}^{n}\Phi_i e_i + Q_{\Phi\Phi}^{-1}n^{-1}\sum_{i=1}^{n}\Phi_i\left[\bar{g}\left(X_i,Z_{1i},U_i\right)-\Phi_i'\beta\right] - Q_{\Phi\Phi}^{-1}n^{-1}\sum_{i=1}^{n}\Phi_i\sum_{l=1}^{d_x}\dot{g}_{d_x+d_1+l}\left(U_{li}\right)\left(\tilde{U}_{li}-U_{li}\right)+\mathbf{R}_{n,\beta}$;

(ii) $\left\|\tilde{\beta}-\beta\right\| = O_P\left(\nu_n+\nu_{1n}\right);$

(iii) $\sup_{\mathbf{w}\in\mathcal{W}}\left|\tilde{\bar{g}}\left(\mathbf{w}\right)-\bar{g}\left(\mathbf{w}\right)\right| = O_P\left[\varsigma_{0\kappa}\left(\nu_n+\nu_{1n}\right)\right];$

where $\left\|\mathbf{R}_{n,\beta}\right\| = \tau_n O_P\left(\nu_n+\nu_{1n}\right),$ *and* $\nu_{1n},\ \nu_n,$ *and* τ_n *are defined in Assumption A.5*(i).

To appreciate the effect of the first stage series estimation on the second stage series estimation, let $\bar{\beta}$ denote a series estimator of β by using $\mathbf{U}_i$ together as $(\mathbf{X}_i,\mathbf{Z}_{1i})$ as the regressors. Then it is standard to show that

$$\begin{aligned}\bar{\beta}-\beta &= Q_{\Phi\Phi}^{-1}n^{-1}\sum_{i=1}^{n}\Phi_i e_i + Q_{\Phi\Phi}^{-1}n^{-1}\sum_{i=1}^{n}\Phi_i\left[\bar{g}\left(X_i,Z_{1i},U_i\right)-\Phi_i'\beta\right]+\bar{\mathbf{R}}_{n,\beta},\ \text{and}\\ \left\|\bar{\beta}-\beta\right\| &= O_P\left(\nu_n\right)\end{aligned}$$

where $\left\|\bar{\mathbf{R}}_{n,\beta}\right\| = O_P(\kappa n^{-1/2}\nu_n) = o\left(\nu_n\right)$. The third term on the right hand side of the expression in Theorem 3.1(i) signifies the asymptotically non-negligible dominant effect of the first stage estimation on the second stage estimation.

With Theorem 3.1, it is straightforward to show the asymptotic distribution of our three-stage estimator of $g_1\left(x_1\right)$ and its gradient.

Theorem 3.2 *Let* $H \equiv$*diag*$(1,h)$. *Suppose that Assumptions A.1-A.5 hold. Then*

(i) *(Normality)*

$$\sqrt{nh}H\left[\hat{\beta}_1\left(x_1\right)-\beta_1\left(x_1\right)-b_1\left(x_1\right)\right]\overset{D}{\to}N\left(0,\Omega_1\left(x_1\right)\right),$$

where $b_1\left(x_1\right)\equiv\begin{pmatrix}\frac{\upsilon_{21}}{2}h^2\ddot{g}_1\left(x_1\right)\\ 0\end{pmatrix}$, $\Omega_1\left(x_1\right)\equiv\begin{pmatrix}\sigma^2(x_1)/f_{X_1}\left(x_1\right) & 0\\ 0 & \upsilon_{22}\sigma^2(x_1)/\left[\upsilon_{21}^2 f_{X_1}\left(x_1\right)\right]\end{pmatrix}$, $\sigma^2(x_1)$ $\equiv E\left(e_i^2|X_{11,i}=x_1\right),$ *and* $\upsilon_{ij}\equiv\int v^i k(v)^j dv,\ i,\ j=0,1,2.$

(ii) (Uniform consistency) Suppose that $Q_{\Phi\Phi,e} \equiv E\left(\Phi_i \Phi_i' e_i^2\right)$ has bounded maximum eigenvalue. Then

$$\sup_{x_1 \in \mathcal{X}_1} \left\| H\left[\hat{\beta}_1(x_1) - \beta_1(x_1)\right] \right\| = O_P\left((nh \log n)^{-1/2} + h^2 \right).$$

Theorem 3.2(i) indicates that our three-step estimator of $\beta_1(x_1) = [g_1(x_1), \dot{g}_1(x_1)]'$ has the asymptotic oracle property. Asymptotically, the asymptotic distribution of local linear estimator of $\beta_1(x_1)$ is not affected by random sampling errors in the first two-stage estimators. In fact, the three-step estimator of $\beta_1(x_1)$ has the same asymptotic distribution that one could have if the other components in $\bar{g}(x, z_1, u)$ were known and a local linear procedure were used to estimate $\beta_1(x_1)$. Theorem 3.2(ii) gives the uniform convergence rate for $\hat{\beta}_1(x_1)$. Similar properties can be established for the local linear estimators of other components of $\bar{g}(x, z_1, u)$. In addition, following the standard exercise in the nonparametric kernel literature, one can also demonstrate that these estimators are asymptotically independently distributed.

3.3 Extension to partially linear additive models

In this section we consider a slight extension of the model in (2.1) to the following partially linear functional coefficient model

$$\begin{cases} Y = g(\mathbf{X}, \mathbf{Z}_1) + \theta' \mathbf{V} + \varepsilon, \\ \mathbf{X} = m(\mathbf{Z}_1, \mathbf{Z}_2) + \Psi \mathbf{V} + \mathbf{U},\ E(\mathbf{U}|\mathbf{Z}_1, \mathbf{Z}_2, \mathbf{V}) = 0,\ E(\varepsilon|\mathbf{Z}_1, \mathbf{Z}_2, \mathbf{U}, \mathbf{V}) = E(\varepsilon|\mathbf{U}), E(\varepsilon) = 0, \end{cases} \tag{3.1}$$

where Y, $\mathbf{X}$, $\mathbf{Z}_1$, $\mathbf{Z}_2$, $\mathbf{Z}$, and ε are defined as above, $\mathbf{V}$ is a $k \times 1$ vector of exogenous variables, θ is a $k \times 1$ parameter vector, and $\Psi = [\psi_1', ..., \psi_{d_x}']'$ is a $d_x \times k$ matrix of parameters in the reduced form regression for $\mathbf{X}$. To avoid the curse of dimensionality, we continue to assume that $m(\mathbf{Z}_1, \mathbf{Z}_2)$, $g(\mathbf{X}, \mathbf{Z}_1)$, and $E(\varepsilon|\mathbf{U})$ have the additive forms given in Section 2.1.

We remark that the results developed in previous sections extend straightforwardly to the model specified in (3.1). Note that

$$E(Y|\mathbf{X}, \mathbf{Z}_1, \mathbf{Z}_2, \mathbf{U}, \mathbf{V}) = g(\mathbf{X}, \mathbf{Z}_1) + E(\varepsilon|\mathbf{U}) + \theta'\mathbf{V} = \bar{g}(\mathbf{X}, \mathbf{Z_1}, \mathbf{U}) + \theta'\mathbf{V}, \text{ and} \tag{3.2}$$

$$E(\mathbf{X}|\mathbf{Z}_1, \mathbf{Z}_2, \mathbf{V}) = m(\mathbf{Z}_1, \mathbf{Z}_2) + \Psi\mathbf{V}. \tag{3.3}$$

Given a random sample $\{(Y_i, \mathbf{X}_i, \mathbf{Z}_{1i}, \mathbf{Z}_{2i}, \mathbf{V}_i), i = 1, ..., n\}$, we can continue to adopt the three-step procedure outlined in Section 2.2 to estimate the above model. First, we choose $(\alpha_l,\ \psi_l)$ to minimize $n^{-1} \sum_{i=1}^n \left[X_{li} - P^{\kappa_1}(\mathbf{Z}_{1i}, \mathbf{Z}_{2i})' \alpha_l - \mathbf{V}_i'\psi_l\right]^2$. Let $(\tilde{\alpha}_l, \tilde{\psi}_l)$ denote the solution. The series estimator of $m_l(\mathbf{z}_1, \mathbf{z}_2)$ is given by $\tilde{m}_l(\mathbf{z}_1, \mathbf{z}_2) = P^{\kappa_1}(\mathbf{z}_1, \mathbf{z}_2)' \tilde{\alpha}_l$. Define the residuals $\tilde{U}_{li} = X_{li} - \tilde{m}_l(\mathbf{Z}_{1i}, \mathbf{Z}_{2i}) - \tilde{\psi}_l'\mathbf{V}_i$. Let $\tilde{\mathbf{U}}_i = (\tilde{U}_{1i}, ..., \tilde{U}_{d_x i})'$, $\tilde{\mathbf{W}}_i = \left(\mathbf{X}_i', \mathbf{Z}_{1i}', \tilde{\mathbf{U}}_i'\right)'$, and $P^\kappa\left(\tilde{\mathbf{W}}_i\right)$ be defined as before. Second, we choose $(\beta,\ \theta)$ to minimize $n^{-1} \sum_{i=1}^n \left[Y_i - P^\kappa\left(\tilde{\mathbf{W}}_i\right)' \beta - \mathbf{V}_i'\theta\right]^2$. Let $\tilde{\beta} \equiv (\tilde{\mu}, \tilde{\beta}_1', ..., \tilde{\beta}_{2d_x+d_1}')'$ and $\tilde{\theta}$ denote the solution. Define $\tilde{Y}_{1i} = Y_i - \tilde{\beta}_{(-1)}' P^\kappa\left(\tilde{\mathbf{W}}_i\right) - \tilde{\theta}'\mathbf{V}_i$ where $\tilde{\beta}_{(-1)}$ is defined as $\tilde{\beta}$ with its component $\tilde{\beta}_1$ being replaced by a $\kappa \times 1$ vector of zeros. Third, we estimate $g_1(x_1)$ and its first order derivative by regressing $\tilde{Y}_{1i}$ on X_{1i} via the local linear procedure. Let $\hat{\beta}_1(x_1)$ denote the estimate of $\beta_1(x_1)$ and via the local linear fitting.

It is well known that the finite dimensional parameter vectors ψ_l's and θ can be estimated at the parametric $\sqrt{n}$-rate and the appearance of the linear components in (3.1) won't affect the asymptotic properties of $\tilde{\beta}$ and $\hat{\beta}_1(x_1)$. To conserve space, we do not repeat the arguments here.

4 Application

5 Results

6 Conclusion

Appendix

A Proof of the Results in Section 3

For notational simplicity, let $\mathbf{Z}_i \equiv (\mathbf{Z}_{1i}', \mathbf{Z}_{2i}')'$, $P_i \equiv P^{\kappa_1}(\mathbf{Z}_i)$, $\Phi_i = \Phi^{\kappa}(\mathbf{W}_i)$, and $\tilde{\Phi}_i = \Phi^{\kappa}(\tilde{\mathbf{W}}_i)$. Then $Q_{PP} \equiv E(P_i P_i')$ and $Q_{\Phi\Phi} \equiv E(\Phi_i \Phi_i')$. Let $Q_{n,PP} \equiv n^{-1}\sum_{i=1}^n P_i P_i'$, $Q_{n,\Phi\Phi} \equiv n^{-1}\sum_{i=1}^n \Phi_i \Phi_i'$, and $\tilde{Q}_{n,\Phi\Phi} \equiv n^{-1}\sum_{i=1}^n \tilde{\Phi}_i \tilde{\Phi}_i'$. By Lemmas A.1$(ii)$ and(v) and Lemma A.4(iv) below, $Q_{n,PP}$, $Q_{n,\Phi\Phi}$ and $\tilde{Q}_{n,\Phi\Phi}$ are invertible with probability approaching 1 (w.p.a.1) so that in large samples we can replace the generalized inverse $Q_{n,PP}^-$, $Q_{n,\Phi\Phi}^-$ and $\tilde{Q}_{n,\Phi\Phi}^-$ by $Q_{n,PP}^{-1}$, $Q_{n,\Phi\Phi}^{-1}$ and $\tilde{Q}_{n,\Phi\Phi}^{-1}$, respectively. Recall $\nu_{1n} \equiv \kappa_1^{1/2}/n^{1/2} + \kappa_1^{-\gamma}$ and $\nu_n \equiv \kappa^{1/2}/n^{1/2} + \kappa^{-\gamma}$.

Lemma A.1 *Suppose that Assumptions A1 and A2(i)-(ii) and (vi) hold. Then*
(i) $\|Q_{n,PP} - Q_{PP}\|^2 = O_P(\kappa_1^2/n)$;
(ii) $\lambda_{\min}(Q_{n,pp}) = \lambda_{\min}(Q_{PP}) + o_P(1)$ *and* $\lambda_{\max}(Q_{n,pp}) = \lambda_{\max}(Q_{PP}) + o_P(1)$;
(iii) $\left\|Q_{n,PP}^{-1} - Q_{PP}^{-1}\right\|_{sp} = O_P(\kappa_1/n^{1/2})$;
(iv) $\|Q_{n,\Phi\Phi} - Q_{\Phi\Phi}\|^2 = O_P(\kappa^2/n)$;
(v) $\lambda_{\min}(Q_{n,\Phi\Phi}) = \lambda_{\min}(Q_{\Phi\Phi}) + o_P(1)$ *and* $\lambda_{\max}(Q_{n,\Phi\Phi}) = \lambda_{\max}(Q_{\Phi\Phi}) + o_P(1)$.

Proof. By straightforward moment calculations, we can show that $E\|Q_{n,PP} - Q_{PP}\|^2 = O(\kappa_1^2/n)$ under Assumption A1(i)-(ii) and A2(vi). Then (i) follows from Markov inequality. By Weyl inequality [e.g., Bernstein (2005, Theorem 8.4.11)] and the fact that $\lambda_{\max}(A) \leq \|A\|$ for any symmetric matrix A (as $|\lambda_{\max}(A)|^2 = \lambda_{\max}(AA) \leq \|A\|^2$), we have

$$\begin{aligned}\lambda_{\min}(Q_{n,PP}) &\leq \lambda_{\min}(Q_{PP}) + \lambda_{\max}(Q_{n,PP} - Q_{PP}) \\ &\leq \lambda_{\min}(Q_{PP}) + \|Q_{n,PP} - Q_{PP}\| = \lambda_{\min}(Q_{PP}) + o_P(1).\end{aligned}$$

Similarly,

$$\begin{aligned}\lambda_{\min}(Q_{n,PP}) &\geq \lambda_{\min}(Q_{PP}) + \lambda_{\min}(Q_{n,PP} - Q_{PP}) \\ &\geq \lambda_{\min}(Q_{PP}) - \|Q_{n,PP} - Q_{PP}\| = \lambda_{\min}(Q_{\kappa_1}) - o_P(1).\end{aligned}$$

Analogously, we can prove the second part of (ii). Thus (ii) follows. By the submultiplicative property of the spectral norm, (i)-(ii), and Assumption A2(i)

$$\begin{aligned}\left\|Q_{n,PP}^{-1} - Q_{PP}^{-1}\right\|_{\text{sp}} &= \left\|Q_{n,PP}^{-1}(Q_{PP} - Q_{n,PP})Q_{PP}^{-1}\right\|_{\text{sp}} \leq \left\|Q_{n,PP}^{-1}\right\|_{\text{sp}}\|Q_{PP} - Q_{n,PP}\|_{\text{sp}}\left\|Q_{PP}^{-1}\right\|_{\text{sp}} \\ &= O_P(1)\,O_P\left(\kappa_1/n^{1/2}\right)O_P(1) = O_P\left(\kappa_1/n^{1/2}\right),\end{aligned}$$

where we use the fact that $\left\|Q_{n,PP}^{-1}\right\|_{\text{sp}} = [\lambda_{\min}(Q_{n,PP})]^{-1} = [\lambda_{\min}(Q_{PP}) + o_P(1)]^{-1} = O_P(1)$ by (ii) and Assumption A2(i). Then (iii) follows. The proof of (iv)-(v) is analogous to that of (i)-(ii) and thus omitted. ■

Lemma A.2 *Let* $\xi_{nl} \equiv n^{-1}\sum_{i=1}^n P_i U_{li}$ *and* $\zeta_{nl} \equiv n^{-1}\sum_{i=1}^n P_i[m_l(\mathbf{Z}_i) - P_i'\alpha_l]$ *for* $l = 1, ..., d_x$. *Suppose that Assumptions A1-A2 hold. Then*
(i) $\|\xi_{nl}\|^2 = O_P(\kappa_1/n)$;
(iii) $\|\zeta_{nl}\|^2 = O_P(\kappa_1^{-2\gamma})$;
(iii) $\tilde{\alpha}_l - \alpha_l = Q_{\kappa_1}^{-1}n^{-1}\sum_{i=1}^n P_i U_{li} + Q_{\kappa_1}^{-1}n^{-1}\sum_{i=1}^n P_i[m_l(\mathbf{Z}_i) - P_i'\alpha_l] + r_{nl}$;
where $\|r_{nl}\| = O_P(\kappa_1/n + \kappa_1^{-\gamma+1/2}/n^{1/2})$ *and* $l = 1, ..., d_x$.

Proof. (i) By Assumption A1(i) and A2(i), $E\|\xi_{nl}\|^2 = n^{-2}\mathrm{tr}\{\sum_{i=1}^n E(P_iP_i'U_{li}^2)\} \leq n^{-1}(1+d\kappa_1)$ $\lambda_{\max}(Q_{PP,U_l}) = O(\kappa_1/n)$. Then $\|\xi_{nl}\|^2 = O_P(\kappa_1/n)$ by Markov inequality.

(ii) By the facts that $\|a\|_{\mathrm{sp}}^2 = \|a\|^2$ for any vector a, $|a'b| \leq \|a\|\,\|b\|$ for any two conformable vectors a and b, and that $\varkappa' A\varkappa \leq \lambda_{\max}(A)\|\varkappa\|^2$ for any p.s.d. matrix A and conformable vector $\varkappa$, Cauchy-Schwarz inequality, Lemma A.1(ii) and Assumptions A2(iv), we have

$$\begin{aligned}
\|\zeta_{nl}\|^2 &= \|\zeta_{nl}\|_{\mathrm{sp}}^2 = \lambda_{\max}(\zeta_{nl}\zeta_{nl}') = \max_{\|\varkappa\|=1} n^{-2}\sum_{i=1}^n\sum_{j=1}^n \varkappa' P_iP_j'\varkappa\left[m_l(\mathbf{Z}_i) - P_i'\alpha_l\right]\left[m_l(\mathbf{Z}_j) - P_j'\alpha_l\right] \\
&\leq \max_{\|\varkappa\|=1}\left\{n^{-1}\sum_{i=1}^n\left\{\varkappa' P_iP_i'\varkappa\left[m_l(\mathbf{Z}_i) - P_i'\alpha_l\right]^2\right\}^{1/2}\right\}^2 \\
&\leq O_P(\kappa_1^{-2\gamma})\max_{\|\varkappa\|=1}\left\{n^{-1}\sum_{i=1}^n \varkappa' P_iP_i'\varkappa\right\} \leq O_P(\kappa_1^{-2\gamma})\lambda_{\max}(Q_{n,PP}) = O_P(\kappa_1^{-2\gamma}).
\end{aligned}$$

(iii) Noting that $X_{li} = m_l(\mathbf{Z}_i) + U_{li} = P_i'\alpha_l + U_{li} + [m_l(\mathbf{Z}_i) - P_i'\alpha_l]$, by Lemma A.1$(ii)$, w.p.a.1 we have

$$\begin{aligned}
\tilde{\alpha}_l - \alpha_l &= \left(\sum_{i=1}^n P_iP_i'\right)^{-}\sum_{i=1}^n P_iX_{li} - \alpha_l \\
&= Q_{n,PP}^{-1}n^{-1}\sum_{i=1}^n P_iU_{li} + Q_{n,PP}^{-1}n^{-1}\sum_{i=1}^n P_i\left[m_l(\mathbf{Z}_i) - P_i'\alpha_l\right] \\
&= Q_{n,PP}^{-1}\xi_{nl} + Q_{n,PP}^{-1}\zeta_{nl} \equiv a_{1l} + a_{2l}, \text{ say.}
\end{aligned} \tag{A.1}$$

Note that $a_{1l} = Q_{\kappa_1}^{-1}\xi_{nl} + r_{1nl}$ where $r_{1nl} = \left(Q_{n,PP}^{-1} - Q_{PP}^{-1}\right)\xi_{nl}$ satisfies that

$$\begin{aligned}
\|r_{1l}\| &\leq = \left\{\mathrm{tr}\left[\left(Q_{n,PP}^{-1} - Q_{PP}^{-1}\right)\xi_{nl}\xi_{nl}'\left(Q_{n,PP}^{-1} - Q_{PP}^{-1}\right)\right]\right\}^{1/2} \\
&\leq \|\xi_{nl}\|_{\mathrm{sp}}\left\|Q_{n,PP}^{-1} - Q_{PP}^{-1}\right\| = O_P(\kappa_1^{1/2}/n^{1/2})O_P(\kappa_1^{1/2}/n^{1/2}) = O_P(\kappa_1/n)
\end{aligned}$$

by Lemmas A.1(iii) and A.2(i). For a_{2l}, we have $a_{2l} = Q_{\kappa_1}^{-1}\zeta_{nl} + r_{2nl}$ where $r_{2nl} = \left(Q_{n\kappa_1}^{-1} - Q_{n\kappa_1}^{-1}\right)\zeta_{nl}$ satisfies that

$$\|r_{2l}\| \leq \|\zeta_{nl}\|_{\mathrm{sp}}\left\|Q_{n,PP}^{-1} - Q_{PP}^{-1}\right\| = O_P(\kappa_1^{-\gamma})O_P(\kappa_1^{1/2}/n^{1/2}) = O_P(\kappa_1^{-\gamma+1/2}/n^{1/2})$$

by Lemmas A.1(iii) and A.2(ii). The result follows. ■

Lemma A.3 *Suppose that Assumptions A1-A3 hold. Then for* $l = 1, ..., d_x$,

(i) $n^{-1}\sum_{i=1}^n\left(\tilde{U}_{li} - U_{li}\right)^2\left[\sigma_i^2\right]^r = O_P\left(\nu_{1n}^2\right)$ *for* $r = 0, 1$;

(ii) $n^{-1}\sum_{i=1}^n\left(\tilde{U}_{li} - U_{li}\right)^2\|\Phi_i\|^r = O_P\left(\varsigma_{0\kappa}^r\nu_{1n}^2\right)$ *for* $r = 1, 2$;

(iii) $n^{-1}\sum_{i=1}^n\left\|p^\kappa\left(\tilde{U}_{li}\right) - p^\kappa\left(U_{li}\right)\right\|^2 = O_P\left(\varsigma_{1\kappa}^2\nu_{1n}^2\right)$;

(iv) $\left\|n^{-1}\sum_{i=1}^n\left[p^\kappa\left(\tilde{U}_{li}\right) - p^\kappa\left(U_{li}\right)\right]\Phi_i'\right\| = O_P\left(\kappa^{1/2}\varsigma_{0\kappa}\nu_{1n} + \varsigma_{0\kappa}\varsigma_{2\kappa}\nu_{1n}^2\right)$;

(v) $\left\|n^{-1}\sum_{i=1}^n\left[p^\kappa\left(\tilde{U}_{li}\right) - p^\kappa\left(U_{li}\right)\right]e_i\right\| = O_P\left(n^{-1/2}\varsigma_{1\kappa}\nu_{1n}\right)$.

Proof. (i) We only prove the $r=1$ case as the proof of the other case is almost identical. By the definition of $\tilde{U}_{li}$ and (A.1), we can decompose $\tilde{U}_{li}-U_{li}=[X_{li}-\tilde{m}_l(\mathbf{Z}_i)]-U_{li}$ as follows

$$\begin{aligned}
\tilde{U}_{li}-U_{li} &= (\mu_l-\tilde{\mu}_l)+\sum_{k=1}^{d_1}[m_{l,k}(Z_{1k,i})-\tilde{m}_{l,k}(Z_{1k,i})]+\sum_{k=1}^{d_2}[m_{l,d_1+k}(Z_{2k,i})-\tilde{m}_{l,d_1+k}(Z_{2k,i})] \\
&= -(\tilde{\mu}_l-\mu_l)-\sum_{k=1}^{d_1}p^{\kappa_1}(Z_{1k,i})'\mathbb{S}_{1k}a_{1l}-\sum_{k=1}^{d_2}p^{\kappa_1}(Z_{2k,i})'\mathbb{S}_{1,d_1+k}a_{1l} \\
&\quad -\sum_{k=1}^{d_1}p^{\kappa_1}(Z_{1k,i})'\mathbb{S}_{1k}a_{2l}-\sum_{k=1}^{d_2}p^{\kappa_1}(Z_{2k,i})'\mathbb{S}_{1,d_1+k}a_{2l} \\
&\equiv -u_{1l,i}-u_{2l,i}-u_{3l,i}-u_{4l,i}-u_{5l,i},\ \text{say.} \qquad \text{(A.2)}
\end{aligned}$$

Then by Cauchy-Schwarz inequality, $n^{-1}\sum_{i=1}^n(\tilde{U}_{li}-U_{li})^2\sigma_i^2\leq 5\sum_{s=1}^5 n^{-1}\sum_{i=1}^n u_{sl,i}^2\sigma_i^2\equiv 5\sum_{s=1}^5 V_{nl,s}$, say. Apparently, $V_{nl,1}=O_P\left(n^{-1}\right)$ as $\tilde{\mu}_l-\mu_l=O_P\left(n^{-1/2}\right)$.

$$\begin{aligned}
V_{nl,2} &= n^{-1}\sum_{i=1}^n\left(\sum_{k=1}^{d_1}p^{\kappa_1}(Z_{1k,i})'\mathbb{S}_{1k}a_{1l}\right)^2\sigma_i^2 \\
&\leq d_1\sum_{k=1}^{d_1}n^{-1}\sum_{i=1}^n\left(p^{\kappa_1}(Z_{1k,i})'\mathbb{S}_{1k}a_{1l}\right)^2\sigma_i^2=d_1\sum_{k=1}^{d_1}\operatorname{tr}\left(\mathbb{S}_{1k}a_{1l}a_{1l}'\mathbb{S}_{1k}'Q_{n1k,pp}\right) \\
&\leq d_1\sum_{k=1}^{d_1}\lambda_{\max}\left(Q_{n1k,pp}\right)\operatorname{tr}\left(a_{1l}a_{1l}'\mathbb{S}_{1k}'\mathbb{S}_{1k}\right)\leq d_1\sum_{k=1}^{d_1}\lambda_{\max}\left(Q_{n1k,pp}\right)\left\|\mathbb{S}_{1k}\right\|_{\text{sp}}^2\left\|a_{1l}\right\|^2.
\end{aligned}$$

where $Q_{n1k,pp}=n^{-1}\sum_{i=1}^n p^{\kappa_1}(Z_{1k,i})\,p^{\kappa_1}(Z_{1k,i})'\sigma_i^2$ such that $\lambda_{\max}(Q_{n1k,pp})=O_P(1)$ by Assumption A3(ii) and arguments analogous to those used in the proof of Lemma A.1(ii). In addition, $\left\|\mathbb{S}_{1k}\right\|_{\text{sp}}^2=\lambda_{\max}\left(\mathbb{S}_{1k}\mathbb{S}_{1k}'\right)=1$, and $\left\|a_{1l}\right\|^2\leq\left\|Q_{n,PP}^{-1}\right\|_{\text{sp}}^2\left\|\xi_{nl}\right\|^2=O_P(1)\,O_P(\kappa_1/n)=O_P(\kappa_1/n)$ by Lemma A.1(iii) and A.2(i) and Assumption A2(i). It follows that $V_{nl,2}=O_P(1)\times 1\times O_P(\kappa_1/n)=O_P(\kappa_1/n)$. Similarly, using the fact that $\left\|a_{2l}\right\|^2\leq\left\|Q_{n,PP}^{-1}\right\|_{\text{sp}}^2\left\|\zeta_{nl}\right\|^2=O_P(1)\,O_P(\kappa_1^{-2\gamma})$, we have

$$\begin{aligned}
V_{nl,4} &= n^{-1}\sum_{i=1}^n\left(\sum_{k=1}^{d_1}p^{\kappa_1}(Z_{1k,i})'\mathbb{S}_{1k}a_{2l}\right)^2\sigma_i^2\leq d_1\sum_{k=1}^{d_1}\lambda_{\max}\left(Q_{n1k,pp}\right)\left\|\mathbb{S}_{1k}\right\|_{\text{sp}}^2\operatorname{tr}\left(a_{2l}a_{2l}'\right) \\
&= O_P(1)\times 1\times O_P(\kappa_1^{-2\gamma})=O_P(\kappa_1^{-2\gamma}).
\end{aligned}$$

By the same token, $V_{nl,3}=O_P(\kappa_1 n^{-1})$ and $V_{nl,5}=O_P(\kappa_1^{-2\gamma})$.

(ii) The result follows from (i) and the fact that $\max_{1\leq i\leq n}\left\|\Phi_i\right\|=O_P\left(\varsigma_{0\kappa}\right)$ under Assumption A2(vi).

(iii) By Assumption A2(vi), Taylor expansion, and (i),

$$\begin{aligned}
n^{-1}\sum_{i=1}^n\left\|p^\kappa\left(\tilde{U}_{li}\right)-p^\kappa\left(U_{li}\right)\right\|^2 &= n^{-1}\sum_{i=1}^n\left\|\dot{p}^\kappa\left(U_{li}^\dagger\right)\left(\tilde{U}_{li}-U_{li}\right)\right\|^2 \\
&\leq O\left(\varsigma_{1\kappa}^2\right)n^{-1}\sum_{i=1}^n\left(\tilde{U}_{li}-U_{li}\right)^2=O_P\left(\varsigma_{1\kappa}^2\nu_{1n}^2\right),
\end{aligned}$$

where $U_{li}^\dagger$ lies between $\tilde{U}_{li}$ and U_{li}.

(iv) By Assumption A2(vi), Taylor expansion, and triangle inequality, $\left\|n^{-1}\sum_{i=1}^n\left[p^\kappa\left(\tilde{U}_{li}\right)-p^\kappa\left(U_{li}\right)\right]\Phi_i'\right\|_{\text{sp}}$ is bounded by

$$\left\|n^{-1}\sum_{i=1}^n\dot{p}^\kappa\left(U_{li}\right)\Phi_i'\left(\tilde{U}_{li}-U_{li}\right)\right\|_{\text{sp}}+\frac{1}{2}\left\|n^{-1}\sum_{i=1}^n\ddot{p}^\kappa\left(U_{li}^\ddagger\right)\Phi_i'\left(\tilde{U}_{li}-U_{li}\right)^2\right\|_{\text{sp}}\equiv T_{nl,1}+T_{nl,2},$$

where $U_{li}^{\ddagger}$ lies between $\tilde{U}_{li}$ and U_{li}. By triangle and Cauchy-Schwarz inequalities and (ii),

$$\begin{aligned} T_{nl,1} &\leq n^{-1}\sum_{i=1}^{n}\|\dot{p}^{\kappa}(U_{li})\|_{\mathrm{sp}}\left\|\Phi_i'\left(\tilde{U}_{li}-U_{li}\right)\right\|_{\mathrm{sp}} \\ &\leq \left\{n^{-1}\sum_{i=1}^{n}\|\dot{p}^{\kappa}(U_{li})\|^2\right\}^{1/2}\left\{n^{-1}\sum_{i=1}^{n}\|\Phi_i\|^2\left|\tilde{U}_{li}-U_{li}\right|^2\right\}^{1/2} \\ &= O_P\left(\kappa^{1/2}\right)O_P\left(\varsigma_{0\kappa}\nu_{1n}\right)=O_P\left(\kappa^{1/2}\varsigma_{0\kappa}\nu_{1n}\right). \end{aligned}$$

By triangle inequality and (i), $T_{nl,2}\leq O\left(\varsigma_{0\kappa}\varsigma_{2\kappa}\right)n^{-1}\sum_{i=1}^{n}(\tilde{U}_{li}-U_{li})^2=O_P(\varsigma_{0\kappa}\varsigma_{2\kappa}\nu_{1n}^2)$. Then (iv) follows.

(v) Let $\Gamma_{nl}\equiv[p^{\kappa}(\tilde{U}_{l1})-p^{\kappa}(U_{l1}),...,[p^{\kappa}(\tilde{U}_{ln})-p^{\kappa}(U_{ln})]]'$ and $\mathbf{e}=(e_1,...,e_n)'$. Then we can write $n^{-1}\sum_{i=1}^{n}[p^{\kappa}(\tilde{U}_{li})-p^{\kappa}(U_{li})]e_i$ as $n^{-1}\Gamma_{nl}'\mathbf{e}$. Let $\mathbb{D}_n\equiv\{(\mathbf{X}_i,\mathbf{Z}_i,\mathbf{U}_i)\}_{i=1}^{n}$. By the law of iterated expectations, Taylor expansion, Assumptions A1(i), A3(ii) and A2(vi), and (i)

$$\begin{aligned} E\left\{\left\|n^{-1}\Gamma_{nl}'\mathbf{e}\right\|^2|\mathbb{D}_n\right\} &= n^{-2}E\left[\mathrm{tr}\left(\Gamma_n'\mathbf{e}\mathbf{e}'\Gamma_n\right)\right]=n^{-2}E\left[\mathrm{tr}\left(\Gamma_n'E\left(\mathbf{e}\mathbf{e}'|\mathbb{D}_n\right)\Gamma_n\right)\right] \\ &= n^{-2}\sum_{i=1}^{n}[p^{\kappa}(\tilde{U}_{li})-p^{\kappa}(U_{li})]^2\sigma_i^2 \\ &\leq O_P\left(\varsigma_{1\kappa}\right)n^{-2}\sum_{i=1}^{n}\left(\tilde{U}_{li}-U_{li}\right)^2\sigma_i^2=O_P\left(n^{-1}\varsigma_{1\kappa}^2\nu_{1n}^2\right). \end{aligned}$$

It follows that $\left\|n^{-1}\Gamma_{nl}'\mathbf{e}\right\|=O_P\left(n^{-1/2}\varsigma_{1\kappa}\nu_{1n}\right)$ by the conditional Chebyshev inequality. ■

Lemma A.4 *Suppose Assumptions A1-A3 hold. Then*
(i) $n^{-1}\sum_{i=1}^{n}\left\|\tilde{\Phi}_i-\Phi_i\right\|^2=O_P\left(\varsigma_{1\kappa}^2\nu_{1n}^2\right)$;
(ii) $\left\|n^{-1}\sum_{i=1}^{n}\left(\tilde{\Phi}_i-\Phi_i\right)\Phi_i'\right\|_{sp}=O_P\left(\kappa^{1/2}\varsigma_{1\kappa}\nu_{1n}\right)$;
(iii) $\left\|\tilde{Q}_{n,\Phi\Phi}-Q_{n,\Phi\Phi}\right\|_{sp}=O_P\left(\kappa^{1/2}\varsigma_{0\kappa}\nu_{1n}+\varsigma_{0\kappa}\varsigma_{2\kappa}\nu_{1n}^2\right)$;
(iv) $\left\|\tilde{Q}_{n,\Phi\Phi}^{-1}-Q_{\Phi\Phi}^{-1}\right\|_{sp}=O_P\left(\kappa^{1/2}\varsigma_{0\kappa}\nu_{1n}+\varsigma_{0\kappa}\varsigma_{2\kappa}\nu_{1n}^2\right)$;
(v) $\left\|n^{-1}\sum_{i=1}^{n}\left(\tilde{\Phi}_i-\Phi_i\right)e_i\right\|=O_P\left(n^{-1/2}\varsigma_{1\kappa}\nu_{1n}\right)$;
(vi) $n^{-1}\sum_{i=1}^{n}\left(\tilde{\Phi}_i-\Phi_i\right)[\bar{g}(X_i,Z_{1i},U_i)-\Phi_i'\beta]=O_P\left(\kappa_1^{-\gamma}\varsigma_{1\kappa}\nu_{1n}\right)$.

Proof. (i) Noting that $n^{-1}\sum_{i=1}^{n}\left\|\tilde{\Phi}_i-\Phi_i\right\|^2=\sum_{l=1}^{d_x}n^{-1}\sum_{i=1}^{n}\left\|p^{\kappa}\left(\tilde{U}_{li}\right)-p^{\kappa}(U_{li})\right\|^2$, the result follows from Lemma A.3(iii).

(ii) Noting that $\left\|n^{-1}\sum_{i=1}^{n}\left(\tilde{\Phi}_i-\Phi_i\right)\Phi_i'\right\|^2=\sum_{l=1}^{d_x}\left\|n^{-1}\sum_{i=1}^{n}\left[p^{\kappa}\left(\tilde{U}_{li}\right)-p^{\kappa}(U_{li})\right]\Phi_i'\right\|^2$, the result follows from Lemma A.3(iv).

(iii) Noting that $\tilde{Q}_{n,\Phi\Phi}-Q_{n,\Phi\Phi}=n^{-1}\sum_{i=1}^{n}(\tilde{\Phi}_i\tilde{\Phi}_i'-\Phi_i\Phi_i')=n^{-1}\sum_{i=1}^{n}(\tilde{\Phi}_i-\Phi_i)(\tilde{\Phi}_i-\Phi_i)'+n^{-1}\sum_{i=1}^{n}(\tilde{\Phi}_i-\Phi_i)\Phi_i'+n^{-1}\sum_{i=1}^{n}\Phi_i(\tilde{\Phi}_i-\Phi_i)'$, the result follows from (i)-(ii) and the triangle inequality.

(iv) By the triangle inequality $\left\|\tilde{Q}_{n,\Phi\Phi}^{-1}-Q_{\Phi\Phi}^{-1}\right\|_{\mathrm{sp}}\leq\left\|\tilde{Q}_{n,\Phi\Phi}^{-1}-Q_{n,\Phi\Phi}^{-1}\right\|_{\mathrm{sp}}+\left\|Q_{n,\Phi\Phi}^{-1}-Q_{\Phi\Phi}^{-1}\right\|_{\mathrm{sp}}$. Arguments like those used in the proof of Lemma A.1(ii) show that $\left\|\tilde{Q}_{n,\Phi\Phi}^{-1}\right\|_{\mathrm{sp}}=\left[\lambda_{\min}\left(\tilde{Q}_{n,\Phi\Phi}\right)\right]^{-1}=\left[\lambda_{\min}\left(Q_{\Phi\Phi}\right)+o_P(1)\right]^{-1}=O_P(1)$ where the second equality follows from (iii) and Lemma A.1(ii). By the submultiplicative property of the spectral norm and (iii),

$$\begin{aligned}\left\|\tilde{Q}_{n,\Phi\Phi}^{-1}-Q_{n,\Phi\Phi}^{-1}\right\|_{\text{sp}} &= \left\|\tilde{Q}_{n,\Phi\Phi}^{-1}\left(\tilde{Q}_{n,\Phi\Phi}-\tilde{Q}_{n,\Phi\Phi}\right)Q_{n,\Phi\Phi}^{-1}\right\|_{\text{sp}} \\ &\leq \left\|\tilde{Q}_{n,\Phi\Phi}^{-1}\right\|_{\text{sp}}\left\|\tilde{Q}_{n,\Phi\Phi}-\tilde{Q}_{n,\Phi\Phi}\right\|_{\text{sp}}\left\|Q_{n,\Phi\Phi}^{-1}\right\|_{\text{sp}} \\ &= O_P\left(\kappa^{1/2}\varsigma_{0\kappa}\nu_{1n}+\varsigma_{0\kappa}\varsigma_{2\kappa}\nu_{1n}^2\right).\end{aligned}$$

Similarly, $\left\|Q_{n,\Phi\Phi}^{-1}-Q_{\Phi\Phi}^{-1}\right\|_{\text{sp}}=O_P\left(\kappa/n^{1/2}\right)$ by Lemma A.1(iii). It follows that $\left\|\tilde{Q}_{n,\Phi\Phi}^{-1}-Q_{\Phi\Phi}^{-1}\right\|_{\text{sp}}=O_P\left(\kappa^{1/2}\varsigma_{0\kappa}\nu_{1n}+\varsigma_{0\kappa}\varsigma_{2\kappa}\nu_{1n}^2\right)$.

(v) Noting that $\left\|n^{-1}\sum_{i=1}^n\left(\tilde{\Phi}_i-\Phi_i\right)e_i\right\|^2=\sum_{l=1}^{d_x}\left\|n^{-1}\sum_{i=1}^n\left[p^\kappa\left(\tilde{U}_{li}\right)-p^\kappa\left(U_{li}\right)\right]e_i\right\|^2$, the result follows from Lemma A.3(v).

(vi) Let $\delta_i\equiv\bar{g}\left(X_i,Z_{1i},U_i\right)-\Phi_i'\beta$. By triangle inequality, Assumption A2(v), Jensen inequality and (i), we have $\left\|n^{-1}\sum_{i=1}^n\left(\tilde{\Phi}_i-\Phi_i\right)\delta_i\right\|\leq O_P\left(\kappa^{-\gamma}\right)n^{-1}\sum_{i=1}^n\left\|\tilde{\Phi}_i-\Phi_i\right\|=O_P\left(\kappa^{-\gamma}\right)O_P\left(\varsigma_{1\kappa}\nu_{1n}\right)$ $=O_P\left(\kappa^{-\gamma}\varsigma_{1\kappa}\nu_{1n}\right)$. ■

Lemma A.5 *Let* $\xi_n\equiv n^{-1}\sum_{i=1}^n\Phi_ie_i$ *and* $\zeta_n\equiv n^{-1}\sum_{i=1}^n\Phi_i\left[\bar{g}\left(X_i,Z_{1i},U_i\right)-\Phi_i'\beta\right]$. *Suppose Assumptions A1-A3 hold. Then*

(i) $\|\xi_n\|=O_P(\kappa^{1/2}/n^{1/2})$;

(iii) $\|\zeta_n\|=O_P(\kappa^{-\gamma})$;

(iii) $\left\|Q_{\Phi\Phi}^{-1}n^{-1}\sum_{i=1}^n\Phi_i\sum_{l=1}^{d_x}\dot{p}^\kappa\left(U_{li}\right)'\beta_{d_x+d_1+l}\left(\tilde{U}_{li}-U_{li}\right)\right\|=O_P\left(\nu_{1n}\right)$ *for* $l=1,...,d_x$.

Proof. The proof of (i)-(ii) is analogous to that of Lemma A.2 (i)-(ii), respectively. Noting that $\left\|Q_{\Phi\Phi}^{-1}\right\|_{\text{sp}}=O(1)$ by Assumption A2(ii), we can prove (iii) by showing that $\|T_{nl}\|=O_P(\nu_{1n})$, where $T_{nl}=n^{-1}\sum_{i=1}^n\Phi_i\delta_{li}(\tilde{U}_{li}-U_{li})$ where $\delta_{li}=\dot{p}^\kappa\left(U_{li}\right)'\beta_{d_x+d_1+l}$. By triangle inequality and Assumptions A1(ii) and A2(iii) and (v)

$$\begin{aligned}c_{\delta_l} &\equiv \max_{1\leq i\leq n}\|\delta_{li}\|\leq\sup_{u_l\in\mathcal{U}_l}\left\|\dot{g}_{d_x+d_1+l}\left(u_l\right)-\dot{p}^\kappa\left(u_l\right)'\beta_{d_x+d_1+l}\right\|+\sup_{u_l\in\mathcal{U}_l}\left\|\dot{g}_{d_x+d_1+l}\left(u_l\right)\right\| \\ &= O\left(\kappa^{-\gamma}\right)+O(1)=O(1).\end{aligned}$$

By (A.2), $T_{nl}=n^{-1}\sum_{i=1}^n\Phi_i\delta_{li}(\tilde{U}_{li}-U_{li})=\sum_{s=1}^5n^{-1}\sum_{i=1}^n\Phi_i\delta_{li}u_{sl,i}=\sum_{s=1}^5T_{nl,s}$, say.

Let $\eta_{nlk}\equiv n^{-1}\sum_{i=1}^n\delta_{li}\Phi_ip^{\kappa_1}\left(Z_{1k,i}\right)'$ and $\bar{\eta}_{lk}=E\left(\eta_{nlk}\right)$. Then $\|\eta_{nlk}-\bar{\eta}_{lk}\|=O_P((\kappa\kappa_1/n)^{1/2})$ by Chebyshev inequality and

$$\|\bar{\eta}_k\|_{\text{sp}}^2=\left\|E\left[\delta_{li}\Phi_ip^{\kappa_1}\left(Z_{1k,i}\right)'\right]\right\|_{\text{sp}}^2\leq c_\delta^2\lambda_{\max}(M)=O(1),$$

where $M\equiv E\left[\Phi_ip^{\kappa_1}\left(Z_{1k,i}\right)'\right]E\left[p^{\kappa_1}\left(Z_{1k,i}\right)\Phi_i'\right]$, and we use the fact that M has bounded largest eigenvalue. To see the last point, first note that for $\kappa_1\leq\kappa$, $E\left[\Phi_ip^{\kappa_1}\left(Z_{1k,i}\right)'\right]$ is a submatrix of $A\equiv E\left(\Phi_i\Phi_i'\right)$ which has bounded largest eigenvalue. Partition A as follows

$$A=\begin{bmatrix}A_{11} & A_{12} & A_{13}\\ A_{21} & A_{22} & A_{23}\\ A_{31} & A_{32} & A_{33}\end{bmatrix}$$

where $A_{ij}=A_{ji}'$ for i, $j=1,2,3$, and $E\left[\Phi_ip^{\kappa_1}\left(Z_{1k,i}\right)'\right]=\begin{bmatrix}A_{12}' & A_{22} & A_{32}'\end{bmatrix}'$. Then

$$M=\begin{bmatrix}A_{12}A_{12}' & A_{12}A_{22} & A_{12}A_{32}'\\ A_{22}A_{12}' & A_{22}A_{22} & A_{22}A_{32}'\\ A_{32}A_{12}' & A_{32}A_{22} & A_{32}A_{32}'\end{bmatrix}.$$

By Thompson and Freede (1970, Theorem 2), $\lambda_{\max}(M) \leq \lambda_{\max}(A_{12}A_{12}') + \lambda_{\max}(A_{22}A_{22}') + \lambda_{\max}(A_{32}A_{32}')$. By Fact 8.9.3 in Bernstein (2005), the positive definiteness of A ensures that both $A_{12}A_{12}'$ and $A_{32}A_{32}'$ have finite maximum eigenvalues as both $\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$ and $\begin{bmatrix} A_{22} & A_{23} \\ A_{32} & A_{33} \end{bmatrix}$ are also positive definite. In addition, $\lambda_{\max}(A_{22}A_{22}) = [\lambda_{\max}(A_{22})]^2$ is finite as A has bounded maximum eigenvalue. It follows that $\lambda_{\max}(M) = O(1)$. Consequently, $\|\eta_{nlk}\| = O_P(1 + (\kappa\kappa_1/n)^{1/2}) = O_P(1)$.

Analogously, noting that 1 is the first element of Φ_i, we can show that $\left\|n^{-1}\sum_{i=1}^n \Phi_i\delta_i\right\|_{\text{sp}} = O_P(1 + (\kappa/n)^{1/2}) = O_P(1)$. It follows that

$$\begin{aligned} \|T_{nl,1}\| &= \left\|n^{-1}\sum_{i=1}^n \Phi_i\delta_{li}\right\|_{\text{sp}} |\tilde{\mu}_l - \mu_l| = O_P(1)\, O_P\left(n^{-1/2}\right) = O_P\left(n^{-1/2}\right), \\ \|T_{nl,2} + T_{nl,4}\| &\leq \sum_{k=1}^{d_1} \|\eta_{nlk}\| \, \|\mathbb{S}_{1k}\|_{\text{sp}} (\|a_{1l}\| + \|a_{2l}\|) = O_P(1)\, O(1)\, O(\nu_{1n}) = O(\nu_{1n}), \end{aligned}$$

and $\|T_{nl,3} + T_{nl,5}\| = O(\nu_{1n})$ by the same token. Thus we have shown that $\|T_{nl}\| = O_P(\nu_{1n})$. ■

Lemma A.6 *Let $c \equiv (c_1, c_2)'$ be an arbitrary 2×1 nonrandom vector such that $\|c\| = 1$. Suppose that Assumptions A1-A5 hold. Then for $l = 2, ..., d_x$*

(i) $S_{2nl}(x_1) \equiv n^{-1/2}h^{1/2}\sum_{i=1}^n K_{ix_1}c'H^{-1}X_{1i}^*(x_1)(\tilde{U}_{li} - U_{li})\dot{p}^\kappa(U_{li}) = h^{1/2}\varsigma_{1\kappa}O_P(1 + n^{1/2}\kappa_1^{-\gamma})$ *uniformly in x_1;*

(ii) $S_{2nl}(x_1) \equiv n^{-1/2}h^{1/2}\sum_{i=1}^n K_{ix_1}\left|c'H^{-1}X_{1i}^*(x_1)\right|(\tilde{U}_{li} - U_{ki})^2 = n^{1/2}h^{1/2}O_P(\nu_{1n}^2)$ *uniformly in x_1.*

Proof. *(i)* Let $\eta_{nl}(x_1) \equiv n^{-1}\sum_{i=1}^n K_{ix}c'H^{-1}X_{1i}^*(x_1)\,p^\kappa(X_{li})$ and $\bar{\eta}_l(x_1) \equiv E[\eta_{nl}(x_1)]$. By straightforward moment calculations and Chebyshev inequality, we have $\eta_{nl}(x_1) = \bar{\eta}_l(x_1) + r_{\eta l}(x_1)$ where $\|r_{\eta l}(x_1)\| = O_P(\kappa^{1/2}n^{-1/2}h^{-1/2})$. In fact, $\sup_{x_1\in\mathcal{X}_1}\|r_{\eta l}(x_1)\| = O_P(\kappa^{1/2}(nh/\log n)^{-1/2})$ with a simple application of Bernstein inequality for independent observations [see, e.g., Serfling (1980, p. 95)].[2] Note that for $l = 2, ..., d_x$,

$$\begin{aligned} \bar{\eta}_l(x_1) &= E\left[K_h(X_{1i} - x_1)\, c'H^{-1}X_{1i}^*(x_1)\, p^\kappa(X_{li})\right] \\ &= \int K(v)(c_1 + c_2 v)\, p^\kappa(x_l)\, f_{1l}\left(x_1 + h^{1/2}v, x_l\right) dv dx_l \\ &= c_1\int f_{1l}(x_1, x_l)\, p^\kappa(x_l)\, dx_l + c_1\int K(v)\left[f_{1l}(x_1 + hv, x_l) - f_{1l}(x_1, x_l)\right] p^\kappa(x_l)\, dx_l \\ &\quad + c_2\int K(v)\, v\left[f_{1l}(x_1 + hv, x_l) - f_{1l}(x_1, x_l)\right] dv p^\kappa(x_l)\, dx_l \\ &\equiv c_1\bar{\eta}_{1l}(x_1) + c_1\bar{\eta}_{2l}(x_1) + c_2\bar{\eta}_{3l}(x_1). \end{aligned}$$

As in Horowitz and Mammen (2004, p. 2435), in view of the fact that the components of $\bar{\eta}_{1l}(x_1)$ are the Fourier coefficients of a function that is bounded uniformly over $\mathcal{X}_1$, we have $\sup_{x_1\in\mathcal{X}_1}\|\bar{\eta}_{1l}(x_1)\|^2 = O(1)$. In addition, using Assumptions A2(v) and A3(i), we can readily show that $\sup_{x_1\in\mathcal{X}_1}\|\bar{\eta}_{2l}(x_1)\| = O_P\left(\kappa^{1/2}h^2\right)$ and $\sup_{x_1\in\mathcal{X}_1}\|\bar{\eta}_{3l}(x_1)\| = O_P\left(\kappa^{1/2}h\right)$. It follows that $\sup_{x_1\in\mathcal{X}_1}\|\bar{\eta}_l(x_1)\| = O_P\left(1 + \kappa^{1/2}h\right) = O_P(1)$ under Assumption A5(ii) and $\sup_{x_1\in\mathcal{X}_1}\|\eta_{nl}(x_1)\| = O_P(1)$.

By (A.2), $S_{1nl}(x_1) = -\sum_{s=1}^5 n^{-1/2}h^{1/2}\sum_{i=1}^n K_{ix_1}c'H^{-1}X_{1i}^*(x_1)\,\dot{p}^\kappa(U_{li})\,u_{sl,i} \equiv -\sum_{s=1}^5 S_{1nl,s}(x_1)$, say. Noting that $S_{1nl,1}(x_1) = n^{1/2}h^{1/2}\eta_{nl}(x_1)(\tilde{\mu}_l - \mu_l)$, we have

$$\sup_{x_1\in\mathcal{X}_1}\|S_{1nl,1}(x_1)\| \leq n^{1/2}h^{1/2}\sup_{x_1\in\mathcal{X}_1}\|\eta_{nl}(x_1)\| \; |\tilde{\mu}_l - \mu_l| = n^{1/2}h^{1/2}O_P(1)\, O_P\left(n^{-1/2}\right) = o_P(1).$$

[2] The proof of Lemma 7 in Horowitz and Mammen (2004) contains various errors as they ignore the fact that κ is diverging to infinity as $n \to \infty$.

Next, note that $S_{1nl,2}(x_1) = \sum_{k=1}^{d_1} S_{1nl,2k}(x_1)$ where $S_{1nl,2k}(x_1) = n^{-1/2}h^{1/2}\sum_{i=1}^{n} K_{ix_1}c'H^{-1}X_{1i}^{*}(x_1)$ $\dot{p}^{\kappa}(U_{li})\,p^{\kappa_1}(Z_{1k,i})'\,\mathbb{S}_{1k}a_{1l}$. We decompose $S_{1nl,2k}$ as follows:

$$
\begin{aligned}
S_{1nl,2k}(x_1) &= n^{-1/2}h^{1/2}\sum_{i=1}^{n} K_{ix_1}c'H^{-1}X_{1i}^{*}\dot{p}^{\kappa}(U_{li})\,p^{\kappa_1}(Z_{1k,i})'\,\mathbb{S}_{1k}Q_{n,PP}^{-1}\xi_{nl} \\
&= n^{1/2}h^{1/2}\psi_{nkl}(x_1)\,\mathbb{S}_{1k}Q_{PP}^{-1}\xi_{nl} + n^{1/2}h^{1/2}\psi_{nkl}(x_1)\,\mathbb{S}_{1k}\left(Q_{n,PP}^{-1} - Q_{PP}^{-1}\right)\xi_{nl} \\
&\equiv S_{1nl,2k1}(x_1) + S_{1nl,2k2}(x_1)\,, \text{ say.}
\end{aligned}
$$

where $\psi_{nkl}(x_1) \equiv n^{-1}\sum_{i=1}^{n} K_{ix_1}c'H^{-1}X_{1i}^{*}(x_1)\,\dot{p}^{\kappa}(U_{li})\,p^{\kappa_1}(Z_{1k,i})'$. Let $\bar{\psi}_{kl}(x_1) \equiv E[\psi_{nkl}(x_1)]$. As in the analysis of $\eta_{nl}(x_1)$, we can show that $\sup_{x_1\in\mathcal{X}_1}\left\|\bar{\psi}_{kl}(x_1)\right\|_{\rm sp} = O_P(\varsigma_{1\kappa})$ and $\sup_{x_1\in\mathcal{X}_1}\left\|\psi_{nkl}(x_1) - \bar{\psi}_{kl}(x_1)\right\|_{\rm sp}$ $\equiv O_P((\kappa_1\kappa\log n/n)^{-1/2})$. It follows that $\sup_{x_1\in\mathcal{X}_1}\|\psi_{nkl}(x_1)\|_{\rm sp} = O_P(\varsigma_{1\kappa}+(\kappa_1\kappa\log n/n)^{-1/2}) = O_P(\varsigma_{1\kappa})$ under Assumption A5(i). Then following the analysis of $B_{nl,1}(x_1)$ in the proof of Theorem 3.2, we can show that $\|S_{1nl,2k1}(x_1)\| = O_P\left(h^{1/2}\varsigma_{1\kappa}\right)$ uniformly in x_1. In addition,

$$
\begin{aligned}
\sup_{x_1\in\mathcal{X}_1}\|S_{1nl,2k2}(x_1)\| &\leq n^{1/2}h^{1/2}\sup_{x_1\in\mathcal{X}_1}\|\psi_{nkl}(x_1)\|_{\rm sp}\|\mathbb{S}_{1k}\|_{\rm sp}\left\|Q_{n,PP}^{-1} - Q_{PP}^{-1}\right\|_{\rm sp}\|\xi_{nl}\| \\
&= n^{1/2}h^{1/2}O_P(\varsigma_{1\kappa})\,O(1)\,O_P(\kappa_1 n^{-1/2})O_P(\kappa_1^{1/2}n^{-1/2}) = O_P(\varsigma_{1\kappa}\kappa_1^{3/2}n^{-1/2}h^{1/2}).
\end{aligned}
$$

It follows that $\sup_{x_1\in\mathcal{X}_1}\|S_{1nl,2k}(x_1)\| = O_P(h^{1/2}\varsigma_{1\kappa}) + O_P(\varsigma_{1\kappa}\kappa_1^{3/2}n^{-1/2}h^{1/2}) = O_P(h^{1/2}\varsigma_{1\kappa})$ under Assumption A5(i). Analogously,

$$
\begin{aligned}
\sup_{x_1\in\mathcal{X}_1}\|S_{1nl,4}(x_1)\| &\leq \sum_{k=1}^{d_1} n^{1/2}h^{1/2}\sup_{x_1\in\mathcal{X}_1}\|\psi_{nkl}(x_1)\|\,\|\mathbb{S}_{1k}\|_{\rm sp}\|a_{2l}\| \\
&= n^{1/2}h^{1/2}O_P(\varsigma_{1\kappa})O_P(\kappa_1^{-\gamma}) = O_P(n^{1/2}h^{1/2}\varsigma_{1\kappa}\kappa_1^{-\gamma}).
\end{aligned}
$$

By the same token, we can show that $\sup_{x_1\in\mathcal{X}_1}\|S_{1nl,3}(x_1)\| = O_P\left(h^{1/2}\varsigma_{1\kappa}\right)$ and $\sup_{x_1\in\mathcal{X}_1}\|S_{1nl,5}(x_1)\| =$ $O_P(n^{1/2}h^{1/2}\varsigma_{1\kappa}\kappa_1^{-\gamma})$. It follows that $\sup_{x_1\in\mathcal{X}_1}\|S_{1n}(x_1)\| = h^{1/2}\varsigma_{1\kappa}O_P(1 + n^{1/2}\kappa_1^{-\gamma})$.

(ii) By (A.2) and Cauchy-Schwarz inequality

$$
S_{2nl}(x_1) \leq 5\sum_{s=1}^{5} n^{-1/2}h^{1/2}\sum_{i=1}^{n} K_{ix_1}\left|c'H^{-1}X_{1i}^{*}(x_1)\right| u_{sl,i}^2 \equiv 5\sum_{s=1}^{5} S_{2nl,s}, \text{ say.}
$$

It is easy to show that $\sup_{x_1\in\mathcal{X}_1} S_{2nl,1}(x_1) = O_P(n^{-1/2}h^{1/2})$. Note that $S_{2nl,2}(x_1) = n^{-1/2}h^{1/2}\sum_{i=1}^{n} K_{ix_1}$ $\times\left|c'H^{-1}X_{1i}^{*}(x_1)\right| u_{2l,i}^2 \leq d_1\sum_{k=1}^{d_1} S_{2nl,2k}(x_1)$, where

$$
\begin{aligned}
S_{2nl,2k}(x_1) &= n^{-1/2}h^{1/2}\sum_{i=1}^{n} K_{ix_1}\left|c'H^{-1}X_{1i}^{*}(x_1)\right| p^{\kappa_1}(Z_{1k,i})'\,\mathbb{S}_{1k}a_{1l}a_{1l}'\mathbb{S}_{1k}'p^{\kappa_1}(Z_{1k,i}) \\
&= n^{1/2}h^{1/2}\mathrm{tr}\left(\mathbb{S}_{1k}a_{1l}a_{1l}'\mathbb{S}_{1k}'v_{nlk}(x_1)\right)
\end{aligned}
$$

and $v_{nlk}(x_1) \equiv n^{-1}\sum_{i=1}^{n} K_{ix_1}\left|c'H^{-1}X_{1i}^{*}\right| p^{\kappa_1}(Z_{1k,i})\,p^{\kappa_1}(Z_{1k,i})'$. As in the analysis of $\eta_{nl}(x_1)$, we can show that $\sup_{x_1\in\mathcal{X}_1}\|v_{nlk}(x_1)\|_{\rm sp} = O_P(1)$. By the fact $\mathrm{tr}(AB) \leq \lambda_{\max}(A)\mathrm{tr}(B)$ and $\|B\|_{\rm sp} = \lambda_{\max}(B)$ for any symmetric matrix A and conformable positive-semidefinite matrix B.

$$
\begin{aligned}
S_{2nl,2k}(x_1) &\leq n^{1/2}h^{1/2}\mathrm{tr}\left(\mathbb{S}_{1k}a_{1l}a_{1l}'\mathbb{S}_{1k}'\right)\|v_{nlk}(x_1)\|_{\rm sp} = n^{1/2}h^{1/2}\|\mathbb{S}_{1k}a_{1l}\|_{\rm sp}^2\|v_{nlk}(x_1)\|_{\rm sp} \\
&\leq n^{1/2}h^{1/2}\|\mathbb{S}_{1k}\|_{\rm sp}^2\|a_{1l}\|_{\rm sp}^2\|v_{nlk}(x_1)\|_{\rm sp} \\
&= n^{1/2}h^{1/2}O(1)\,O_P\left(\kappa_1 n^{-1}\right)O_P(1) = O_P\left(\kappa_1 n^{-1/2}h^{1/2}\right) \text{ uniformly in } x_1.
\end{aligned}
$$

It follows that $\sup_{x_1 \in \mathcal{X}_1} S_{2nl,2}(x_1) = O_P\left(\kappa_1 n^{-1/2} h^{1/2}\right)$. Similarly, uniformly in x_1

$$\begin{aligned} S_{2nl,4}(x_1) &\leq n^{1/2} h^{1/2} \|\mathbb{S}_{1k}\|_{\text{sp}}^2 \|a_{2l}\|_{\text{sp}}^2 \|v_{nlk}(x_1)\|_{\text{sp}} \\ &= n^{1/2} h^{1/2} O(1) O_P(\kappa_1^{-2\gamma}) O_P(1) = O_P\left(n^{1/2} \kappa_1^{-2\gamma} h^{1/2}\right). \end{aligned}$$

By the same token, $S_{2nl,3}(x_1) = O_P(\kappa_1 n^{-1/2} h^{1/2})$ and $S_{2nl,5}(x_1) = O_P(n^{1/2} \kappa_1^{-2\gamma} h^{1/2})$ uniformly in x_1. Consequently, $\sup_{x_1 \in \mathcal{X}_1} S_{2nl}(x_1) = n^{1/2} h^{1/2} O_P(\nu_{1n}^2)$. ■

Proof of Theorem 3.1. (i) Noting that $Y_i = \bar{g}(X_i, Z_{1i}, U_i) + e_i = \tilde{\Phi}_i' \beta + e_i + \left[\bar{g}(X_i, Z_{1i}, U_i) - \tilde{\Phi}_i' \beta\right]$, we have

$$\begin{aligned} \tilde{\beta} - \beta &= \tilde{Q}_{n,\Phi\Phi}^{-1} n^{-1} \sum_{i=1}^n \tilde{\Phi}_i Y_i - \beta = \tilde{Q}_{n,\Phi\Phi}^{-1} n^{-1} \sum_{i=1}^n \tilde{\Phi}_i e_i + \tilde{Q}_{n,\Phi\Phi}^{-1} n^{-1} \sum_{i=1}^n \tilde{\Phi}_i \left[\bar{g}(X_i, Z_{1i}, U_i) - \tilde{\Phi}_i' \beta\right] \\ &= \tilde{Q}_{n,\Phi\Phi}^{-1} \xi_n + \tilde{Q}_{n,\Phi\Phi}^{-1} \zeta_n + \tilde{Q}_{n,\Phi\Phi}^{-1} n^{-1} \sum_{i=1}^n \Phi_i \left(\Phi_i - \tilde{\Phi}_i\right)' \beta + \tilde{Q}_{n,\Phi\Phi}^{-1} n^{-1} \sum_{i=1}^n \left(\tilde{\Phi}_i - \Phi_i\right) e_i \\ &\quad + \tilde{Q}_{n,\Phi\Phi}^{-1} n^{-1} \sum_{i=1}^n \left(\tilde{\Phi}_i - \Phi_i\right) [\bar{g}(X_i, Z_{1i}, U_i) - \Phi_i' \beta] - \tilde{Q}_{n,\Phi\Phi}^{-1} n^{-1} \sum_{i=1}^n \left(\tilde{\Phi}_i - \Phi_i\right) \left(\tilde{\Phi}_i - \Phi_i\right)' \beta \\ &\equiv b_{1n} + b_{2n} + b_{3n} + b_{4n} + b_{5n} - b_{6n}, \text{ say.} \end{aligned}$$

Note that $b_{1n} = Q_{\Phi\Phi}^{-1} \xi_n + r_{1n}$, where $r_{1n} = \left(\tilde{Q}_{n,\Phi\Phi}^{-1} - Q_{\Phi\Phi}^{-1}\right) \xi_n$ satisfies $\|r_{1n}\| \leq \left\|\tilde{Q}_{n,\Phi\Phi}^{-1} - Q_{\Phi\Phi}^{-1}\right\|_{\text{sp}} \|\xi_n\|_{\text{sp}} = O_P\left[(\kappa^{1/2} \varsigma_{0\kappa} \nu_{1n} + \varsigma_{0\kappa} \varsigma_{2\kappa} \nu_{1n}^2) \kappa^{1/2} n^{-1/2}\right]$ by Lemmas A.4(iv) and A.5(i). Similarly, $b_{2n} = Q_{\Phi\Phi}^{-1} \zeta_n + r_{2n}$, where $r_{2n} = \left(\tilde{Q}_{n,\Phi\Phi}^{-1} - Q_{\Phi\Phi}^{-1}\right) \zeta_n$ satisfies $\|r_{2n}\| \leq \left\|\tilde{Q}_{n,\Phi\Phi}^{-1} - Q_{\Phi\Phi}^{-1}\right\| \|\zeta_n\|_{\text{sp}} = O_P\left[(\kappa^{1/2} \varsigma_{0\kappa} \nu_{1n} + \varsigma_{0\kappa} \varsigma_{2\kappa} \nu_{1n}^2) \kappa^{-\gamma}\right]$ by Lemmas A.4(iv) and A.5(ii). Next, we decompose b_{3n} as follows:

$$b_{3n} = Q_{\Phi\Phi}^{-1} n^{-1} \sum_{i=1}^n \Phi_i \left(\Phi_i - \tilde{\Phi}_i\right)' \beta + \left(\tilde{Q}_{n,\Phi\Phi}^{-1} - Q_{\Phi\Phi}^{-1}\right) n^{-1} \sum_{i=1}^n \Phi_i \left(\Phi_i - \tilde{\Phi}_i\right)' \beta \equiv b_{3n,1} + b_{3n,2}, \text{ say.}$$

We further decompose $b_{3n,1}$ as follows:

$$\begin{aligned} b_{3n,1} &= -Q_{\Phi\Phi}^{-1} n^{-1} \sum_{i=1}^n \Phi_i \sum_{l=1}^{d_x} \left[p^\kappa\left(\tilde{U}_{li}\right) - p^\kappa(U_{li})\right]' \beta_{d_x+d_1+l} \\ &= \sum_{l=1}^{d_x} Q_{\Phi\Phi}^{-1} n^{-1} \sum_{i=1}^n \Phi_i \dot{p}^\kappa\left(U_{li}^\dagger\right)' \beta_{d_x+d_1+l} \left(U_{li} - \tilde{U}_{li}\right) \\ &= \sum_{l=1}^{d_x} Q_{\Phi\Phi}^{-1} n^{-1} \sum_{i=1}^n \Phi_i \dot{g}_{d_x+d_1+l}(U_{li}) \left(U_{li} - \tilde{U}_{li}\right) \\ &\quad + \sum_{l=1}^{d_x} Q_{\Phi\Phi}^{-1} n^{-1} \sum_{i=1}^n \Phi_i \left[\dot{g}_{d_x+d_1+l}\left(U_{li}^\dagger\right) - \dot{g}_{d_x+d_1+l}(U_{li})\right] \left(U_{li} - \tilde{U}_{li}\right) \\ &\quad + \sum_{l=1}^{d_x} Q_{\Phi\Phi}^{-1} n^{-1} \sum_{i=1}^n \Phi_i \left[\dot{p}^\kappa\left(U_{li}^\dagger\right)' \beta_{d_x+d_1+l} - \dot{g}_{d_x+d_1+l}\left(U_{li}^\dagger\right)\right] \left(U_{li} - \tilde{U}_{li}\right) \\ &\equiv \sum_{l=1}^{d_x} b_{3n,11l} + \sum_{l=1}^{d_x} b_{3n,12l} + \sum_{l=1}^{d_x} b_{3n,13l}, \text{ say,} \end{aligned}$$

where $U_{li}^{\dagger}$ lies between $\tilde{U}_{li}$ and U_{li}. Noting that $|\dot{g}_{d_x+d_1+l}\left(U_{li}^{\dagger}\right) - \dot{g}_{d_x+d_1+l}\left(U_{li}\right)| \leq c_{\dot{g}}|\tilde{U}_{li} - U_{li}|$ where $c_{\dot{g}} = \max_{1\leq l\leq d_x}\max_{u_l\in\mathcal{U}_l}|\dot{g}_{d_x+d_1+l}\left(u_l\right)| = O\left(1\right)$ by Assumptions A1(ii) and A2(iii),

$$\|b_{3n,12l}\| \leq c_{\dot{g}}n^{-1}\sum_{i=1}^{n}\|\Phi_i\|\left(U_{li} - \tilde{U}_{li}\right)^2 = \varsigma_{0\kappa}O_P\left(\nu_{1n}^2\right)$$

by Lemma A.3(i). By Assumption A2(ii), Cauchy-Schwarz inequality, and Lemma A.3(i)

$$\begin{aligned}
\|b_{3n,13l}\| &\leq O\left(\kappa^{-\gamma}\right)\left\|Q_{\Phi\Phi}^{-1}\right\|_{\text{sp}} n^{-1}\sum_{i=1}^{n}\|\Phi_i\|\left|U_{li} - \tilde{U}_{li}\right| \\
&\leq O\left(\kappa^{-\gamma}\right)\left\|Q_{\Phi\Phi}^{-1}\right\|_{\text{sp}}\left\{n^{-1}\sum_{i=1}^{n}\|\Phi_i\|^2\right\}^{1/2}\left\{n^{-1}\sum_{i=1}^{n}\left(U_{li} - \tilde{U}_{li}\right)^2\right\}^{1/2} \\
&= O\left(\kappa^{-\gamma}\right)O\left(1\right)O\left(\kappa^{1/2}\right)O_P\left(\nu_{1n}\right) = \kappa^{-\gamma+1/2}O_P\left(\nu_{1n}\right).
\end{aligned}$$

By Lemma A.5(iii), $\|b_{3n,11l}\| = O_P(\nu_{1n})$ which dominates both $\|b_{3n,12l}\|$ and $\|b_{3n,13l}\|$. Thus $\|b_{3n,2}\| \leq \left\|\tilde{Q}_{n,\Phi\Phi}^{-1} - Q_{\Phi\Phi}^{-1}\right\|_{\text{sp}} O_P(\nu_{1n}) = O_P[(\kappa^{1/2}\varsigma_{0\kappa}\nu_{1n}+\varsigma_{0\kappa}\varsigma_{2\kappa}\nu_{1n}^2)\nu_{1n}]$. It follows that $b_{3n} = \sum_{l=1}^{d_x}Q_{\Phi\Phi}^{-1}n^{-1}\sum_{i=1}^{n}\Phi_i$ $\times\dot{g}_{d_x+d_1+l}\left(U_{li}\right)(U_{li} - \tilde{U}_{li}) + \bar{b}_{3n}$, where $\left\|\bar{b}_{3n}\right\| = O_P[(\kappa^{1/2}\varsigma_{0\kappa}\nu_{1n} + \varsigma_{0\kappa}\varsigma_{2\kappa}\nu_{1n}^2)\nu_{1n}]$. By Lemmas A.4$(v)$-$(vi)$, $\|b_{4n}\| = O_P\left(n^{-1/2}\varsigma_{1\kappa}\nu_{1n}\right)$, and

$$\|b_{5n}\| \leq \left\|\tilde{Q}_{n,\Phi\Phi}^{-1}\right\|_{\text{sp}}\left\|n^{-1}\sum_{i=1}^{n}\left(\tilde{\Phi}_i - \Phi_i\right)\left[\bar{g}\left(X_i, Z_{1i}, U_i\right) - \Phi_i'\beta\right]\right\| = O_P\left(\kappa^{-\gamma}\varsigma_{1\kappa}\nu_{1n}\right),$$

where we use the fact that $\left\|\tilde{Q}_{n,\Phi\Phi}^{-1}\right\|_{\text{sp}} \leq \left\|\tilde{Q}_{n,\Phi\Phi}^{-1} - Q_{\Phi\Phi}^{-1}\right\|_{\text{sp}} + \left\|Q_{\Phi\Phi}^{-1}\right\|_{\text{sp}} = o_P\left(1\right) + O\left(1\right) = O_P\left(1\right)$. For b_{6n}, we have by Taylor expansion and triangle inequality that

$$\begin{aligned}
\|b_{6n}\| &\leq \sum_{l=1}^{d_x}\left\|\tilde{Q}_{n,\Phi\Phi}^{-1}n^{-1}\sum_{i=1}^{n}\left(\tilde{\Phi}_i - \Phi_i\right)\left[p^{\kappa}\left(\tilde{U}_{li}\right) - p^{\kappa}\left(U_{li}\right)\right]'\beta_{d_x+d_1+l}\right\| \\
&= \sum_{l=1}^{d_x}\left\|\tilde{Q}_{n,\Phi\Phi}^{-1}n^{-1}\sum_{i=1}^{n}\left(\tilde{\Phi}_i - \Phi_i\right)p^{\kappa}\left(U_{li}^{\dagger}\right)'\beta_{d_x+d_1+l}\left(\tilde{U}_{li} - U_{li}\right)\right\| \\
&\leq \sum_{l=1}^{d_x}\left\|\tilde{Q}_{n,\Phi\Phi}^{-1}\right\|_{\text{sp}}\left\|n^{-1}\sum_{i=1}^{n}\left(\tilde{\Phi}_i - \Phi_i\right)\dot{g}_{d_x+d_1+l}\left(U_{li}^{\dagger}\right)\left(\tilde{U}_{li} - U_{li}\right)\right\| \\
&\quad + \sum_{l=1}^{d_x}\left\|\tilde{Q}_{n,\Phi\Phi}^{-1}\right\|_{\text{sp}}\left\|n^{-1}\sum_{i=1}^{n}\left(\tilde{\Phi}_i - \Phi_i\right)\left[p^{\kappa}\left(U_{li}^{\dagger}\right)'\beta_{d_x+d_1+l} - \dot{g}_{d_x+d_1+l}\left(U_{li}^{\dagger}\right)\right]\left(\tilde{U}_{li} - U_{li}\right)\right\| \\
&\equiv \sum_{l=1}^{d_x}b_{6nl,1} + \sum_{l=1}^{d_x}b_{6nl,2}, \text{ say.}
\end{aligned}$$

By the triangle inequality, Lemmas A.3(i) and A.4(i),

$$\begin{aligned}
b_{6nl,1} &\leq c_{\dot{g}}\left\|\tilde{Q}_{n,\Phi\Phi}^{-1}\right\|_{\text{sp}}\left\{n^{-1}\sum_{i=1}^{n}\left\|\tilde{\Phi}_i - \Phi_i\right\|^2\right\}^{1/2}\left\{n^{-1}\sum_{i=1}^{n}\left(\tilde{U}_{li} - U_{li}\right)^2\right\}^{1/2} \\
&= O_P(1)O_P\left(\varsigma_{1\kappa}\nu_{1n}\right)O_P\left(\nu_{1n}\right) = O_P\left(\varsigma_{1\kappa}\nu_{1n}^2\right).
\end{aligned}$$

Similarly, we can show that $b_{6nl,2} = \kappa^{-\gamma}O_P\left(\varsigma_{1\kappa}\nu_{1n}^2\right)$ by Assumption A2(v) and Lemmas A.3(i) and A.4(i). It follows that $\|b_{6n}\| = O_P\left(\varsigma_{1\kappa}\nu_{1n}^2\right)$. Combining the above results yield the conclusion in (i).

(ii) Noting that $\left\|Q_{\Phi\Phi}^{-1}\xi_n\right\| \leq \left\|Q_{\Phi\Phi}^{-1}\right\|_{\rm sp}\|\xi_n\| = O_P\left(\kappa^{1/2}/n^{1/2}\right)$ and $||Q_{\Phi\Phi}^{-1}\zeta_n|| \leq \left\|Q_{\Phi\Phi}^{-1}\right\|_{\rm sp}\|\zeta_n\| = O_P\left(\kappa^{-\gamma}\right)$ by Lemmas A.5(i)-(ii), the result in part (ii) follows from part (i), Lemma A.4, and the fact that $\|\mathbf{R}_{n,\beta}\| = O_P(\nu_{1n})$ under Assumption A5(i)

(iii) By (ii) and Assumptions A2(v) , $\sup_{\mathbf{w}\in\mathcal{W}}\left|\widetilde{g}(\mathbf{w}) - \bar{g}(\mathbf{w})\right| = \sup_{\mathbf{w}\in\mathcal{W}}|\Phi(\mathbf{w})'(\tilde{\beta}-\beta) + [\beta'\Phi(\mathbf{w}) - \bar{g}(\mathbf{w})]| \leq \sup_{\mathbf{w}\in\mathcal{W}}\|\Phi(\mathbf{w})\|\left\|\tilde{\beta}-\beta\right\| + \sup_{w\in\mathcal{W}}|\beta'\Phi(\mathbf{w}) - \bar{g}(\mathbf{w})| = O_P[\varsigma_{0\kappa}(\nu_n+\nu_{1n})]$ as the second term is $O(\nu_n)$. ■

Proof of Theorem 3.2.

Let $Y_{1i} \equiv Y_i - \mu - g_2(X_{2i}) - ... - g_{d_x}(X_{d_x i}) - g_{d_x+1}(Z_{11,i}) - ... - g_{d_x+d_1}(Z_{1d_1,i}) - g_{d_x+d_1+1}(U_{1i}) - ... - g_{2d_x+d_1}(U_{d_x i})$, and $\mathbf{Y}_1 \equiv (Y_{11},...,Y_{1n})'$. Using the notation defined at the end of section 2.2, we have

$$\begin{aligned} H\hat{\beta}_1(x_1) &= \left[H^{-1}\mathbb{X}_1(x_1)'\mathbb{K}_{x_1}\mathbb{X}_1(x_1)H^{-1}\right]^{-1}H^{-1}\mathbb{X}_1(x_1)'\mathbb{K}_{x_1}\mathbb{X}_1(x_1)\mathbf{Y}_1 \\ &\quad + \left[H^{-1}\mathbb{X}_1(x_1)'\mathbb{K}_{x_1}\mathbb{X}_1(x_1)H^{-1}\right]^{-1}H^{-1}\mathbb{X}_1(x_1)\mathbb{K}_{x_1}(\tilde{\mathbf{Y}}_1 - \mathbf{Y}_1) \\ &\equiv J_{1n}(x_1) + J_{2n}(x_1), \text{ say.} \end{aligned}$$

By standard results in local linear regressions [e.g., Masry (1996) and Hansen (2008)], $n^{-1}H^{-1}\mathbb{X}_1(x_1)'\mathbb{K}_{x_1}\mathbb{X}_1(x_1)H^{-1} = f_{X_1}(x_1)\begin{pmatrix} 1 & 0 \\ 0 & \int u^2K(u)\,du\end{pmatrix} + o_P(1)$ uniformly in x_1, $n^{1/2}h^{1/2}[J_{1n}(x_1) - b_1(x_1)] \xrightarrow{D} N(0,\Omega_1(x_1))$, and $\sup_{x_1\in\mathcal{X}_1}\|J_{1n}(x_1)\| = O_P\left((nh\log n)^{-1/2} + h^2\right)$, where $b_1(x_1)$ and $\Omega_1(x_1)$ are defined in Theorem 3.2. It suffices to prove the theorem by showing that $n^{-1/2}h^{1/2}H^{-1}\mathbb{X}_1(x_1)'\mathbb{K}_{x_1}(\tilde{\mathbf{Y}}_1 - \mathbf{Y}_1) = o_P(1)$ uniformly in x_1.[3]

We make the following decomposition:

$$\begin{aligned} (n/h)^{-1/2}H^{-1}\mathbb{X}_1(x_1)\mathbb{K}_{x_1}(\mathbf{Y}_1 - \tilde{\mathbf{Y}}_1) &= n^{-1/2}h^{1/2}\sum_{i=1}^n K_{ix_1}H^{-1}X_{1i}^*(x_1)\left(Y_{1i} - \tilde{Y}_{1i}\right) \\ &= \sqrt{n}(\tilde{\mu}-\mu)n^{-1}h^{1/2}\sum_{i=1}^n K_{ix_1}H^{-1}X_{1i}^*(x_1) \\ &\quad + \sum_{l=2}^{d_x} n^{-1/2}h^{1/2}\sum_{i=1}^n K_{ix_1}H^{-1}X_{1i}^*(x_1)[\tilde{g}_l(X_{li}) - g_l(X_{li})] \\ &\quad + \sum_{j=1}^{d_1} n^{-1/2}h^{1/2}\sum_{i=1}^n K_{ix_1}H^{-1}X_{1i}^*(x_1)[\tilde{g}_{d_x+j}(Z_{1j,i}) - g_{d_x+j}(Z_{1j,i})] \\ &\quad + \sum_{l=1}^{d_x} n^{-1/2}h^{1/2}\sum_{i=1}^n K_{ix_1}H^{-1}X_{1i}^*(x_1)\left[\tilde{g}_{d_x+d_1+l}(\tilde{U}_{li}) - g_{d_x+d_1+l}(U_{li})\right] \\ &\equiv A_n(x_1) + \sum_{l=2}^{d_x}B_{nl}(x_1) + \sum_{j=1}^{d_1}C_{nj}(x_1) + \sum_{l=1}^{d_x}D_{nl}(x_1). \end{aligned}$$

We prove the first part of the theorem by showing that $(i1)$ $A_n(x_1) = o_P(1)$, $(i2)$ $B_{nl}(x_1) = o_P(1)$ for $l = 2,...,d_x$, $(i3)$ $C_{nj}(x_1) = o_P(1)$ for $j = 1,...,d_1$, and $(i4)$ $D_{nl}(x_1) = o_P(1)$ for $l = 1,...,d_x$, all uniformly in x_1.

$(i1)$ holds by noticing that $\sqrt{n}(\tilde{\mu}-\mu) = O_P(1)$ and $n^{-1}\sum_{i=1}^n K_{ix_1}H^{-1}X_{1i}^*(x_1) = O_P(1)$ uniformly in x_1. Let $c \equiv (c_1,c_2)'$ be an arbitrary 2×1 nonrandom vector such that $\|c\| = 1$. Recall that $\eta_{nl}(x_1) \equiv$

[3] For part (i) of Theorem 3.2 we only need the pointwise result to hold.

$n^{-1}\sum_{i=1}^{n}K_{ix}c'H^{-1}X_{1i}^{*}(x_1)\,p^{\kappa}(X_{li})$. For $(i2)$, we make the following decomposition

$$\begin{aligned}
c'B_{nl}(x_1) &= n^{-1/2}h^{1/2}\sum_{i=1}^{n}K_{ix_1}c'H^{-1}X_{1i}^{*}(x_1)\,p^{\kappa}(X_{li})'\,\mathbb{S}_l\left(\tilde{\beta}-\beta\right)\\
&\quad+n^{-1/2}h^{1/2}\sum_{i=1}^{n}K_{ix_1}c'H^{-1}X_{1i}^{*}(x_1)\left[p^{\kappa}(X_{li})'\,\mathbb{S}_l\beta-g_l(X_{li})\right]\\
&= n^{1/2}h^{1/2}\eta_{nl}(x_1)'\,\mathbb{S}_lQ_{\Phi\Phi}^{-1}\xi_n+n^{1/2}h^{1/2}\eta_{nl}(x_1)\,\mathbb{S}_lQ_{\Phi\Phi}^{-1}\zeta_n\\
&\quad-n^{-1/2}h^{1/2}\eta_{nl}(x_1)'\,\mathbb{S}_lQ_{\Phi\Phi}^{-1}\sum_{j=1}^{n}\Phi_j\sum_{k=1}^{d_x}\delta_{kj}\left(\tilde{U}_{kj}-U_{kj}\right)+n^{-1/2}h^{1/2}\eta_{nl}(x_1)'\,\mathbb{S}_l\mathbf{R}_{n,\beta}\\
&\quad+n^{-1/2}h^{1/2}\sum_{i=1}^{n}K_{ix_1}c'H^{-1}X_{1i}^{*}(x_1)\left[p^{\kappa}(X_{li})'\,\mathbb{S}_l\beta-g_l(X_{li})\right]\\
&\equiv B_{nl,1}(x_1)+B_{nl,2}(x_1)-B_{nl,3}(x_1)+B_{nl,4}(x_1)+B_{nl,5}(x_1),
\end{aligned}$$

where recall $\delta_{kj}\equiv\dot{p}^{\kappa}(U_{kj})'\,\beta_{d_x+d_1+k}$, $\xi_n\equiv n^{-1}\sum_{j=1}^{n}\Phi_je_j$, and $\zeta_n\equiv n^{-1}\sum_{j=1}^{n}\Phi_j\left[\bar{g}(X_j,Z_{1j},U_j)-\beta'\Phi_j\right]$. Let $\bar{\eta}_l(x_1)\equiv E[\eta_{nl}(x_1)]$ and $r_{\eta l}(x_1)=\eta_{nl}(x_1)-\bar{\eta}_l(x_1)$. By the proof of Lemma A.6(i), $\|r_{\eta l}(x_1)\|=O_P(\kappa^{1/2}(nh/\log n)^{-1/2})$, $\|\bar{\eta}_l(x_1)\|=O_P\left(1+\kappa^{1/2}h\right)$, and $\|\eta_{nl}(x_1)\|=O_P(1)$ uniformly in x_1. Note that

$$B_{nl,1}(x_1)=n^{1/2}h^{1/2}\bar{\eta}_l(x_1)'\,\mathbb{S}_lQ_{\Phi\Phi}^{-1}\xi_n+n^{1/2}h^{1/2}r_{\eta l}(x_1)'\,\mathbb{S}_lQ_{\Phi\Phi}^{-1}\xi_n\equiv B_{nl,11}(x_1)+B_{nl,12}(x_1),\ \text{say.}$$

Noting that

$$\begin{aligned}
E\left[B_{nl,11}^2(x_1)\right] &= h\bar{\eta}_l(x_1)'\,\mathbb{S}_lQ_{\Phi\Phi}^{-1}E\left(\Phi_j\Phi_j'e_j^2\right)P_{\Phi\Phi}^{-1}\mathbb{S}_l'\bar{\eta}_l(x_1)\\
&\leq h\lambda_{\max}\left(E\left(\Phi_j\Phi_j'e_j^2\right)\right)\left[\lambda_{\min}(P_{\Phi\Phi})\right]^{-2}\lambda_{\max}(\mathbb{S}_l\mathbb{S}_l')\,\|\bar{\eta}_l(x_1)\|^2\\
&= hO(1)\,O_P(1)\,O_P(1)=O_P(h),
\end{aligned}$$

we have $|B_{nl,11}(x_1)|=O_P\left(h^{1/2}\right)$ for each $x_1\in\mathcal{X}_1$. Let $\breve{\eta}_l(x_1)\equiv Q_{\Phi\Phi}^{-1}\mathbb{S}_l'\bar{\eta}_l(x_1)$. Then we can write $\bar{\eta}_l(x_1)'\,\mathbb{S}_lQ_{\Phi\Phi}^{-1}\xi_n$ as $n^{-1}\sum_{i=1}^{n}\breve{\eta}_l(x_1)'\,\Phi_ie_i$. Noting that $E[\breve{\eta}_l(x_1)'\,\Phi_ie_i]=0$ and $E\left[\breve{\eta}_l(x_1)'\,\Phi_ie_i\right]^2=\breve{\eta}_l(x_1)'\,E\left(\Phi_i\Phi_i'e_i^2\right)\breve{\eta}_l(x_1)\leq\lambda_{\max}(Q_{\Phi\Phi,e})\left\|Q_{\Phi\Phi}^{-1}\right\|_{\rm sp}^2\sup_{x_1\in\mathcal{X}_1}\|\bar{\eta}_l(x_1)\|=O(1)$, we can readily divide $\mathcal{X}_l$ into intervals of appropriate length and apply Bernstein inequality to show that $\bar{\eta}_l(x_1)'\,\mathbb{S}_lQ_{\Phi\Phi}^{-1}\xi_n=O_P\left((n/\log n)^{-1/2}\right)$. Consequently,

$$\sup_{x_1\in\mathcal{X}_1}|B_{nl,11}(x_1)|=n^{1/2}h^{1/2}O_P\left((n/\log n)^{-1/2}\right)=O_P\left((h/\log n)^{-1/2}\right)=o_P(1).$$

For $B_{nl,12}(x_1)$, we have by Lemma A.5(i)

$$\begin{aligned}
\sup_{x_1\in\mathcal{X}_1}\|B_{nl,12}(x_1)\| &\leq n^{1/2}h^{1/2}\sup_{x_1\in\mathcal{X}_1}\|r_{\eta l}(x_1)\|\,\|\mathbb{S}_l\|_{\rm sp}\left\|Q_{\Phi\Phi}^{-1}\right\|_{\rm sp}\|\xi_n\|\\
&= n^{1/2}h^{1/2}O_P(\kappa^{1/2}(nh/\log n)^{-1/2})O(1)\,O_P(1)\,O_P(\kappa^{1/2}n^{-1/2})\\
&= O_P(\kappa(n/\log n)^{-1/2})=o_P(1).
\end{aligned}$$

It follows that $\sup_{x_1\in\mathcal{X}_1}|B_{nl,1}(x_1)|=o_P(1)$. By Lemma A.5$(ii)$ and Assumptions A2(ii) and (v) and A5,

$$\begin{aligned}
\sup_{x_1\in\mathcal{X}_1}|B_{nl,2}(x_1)| &= \sup_{x_1\in\mathcal{X}_1}n^{1/2}h^{1/2}\left|\eta_{nl}(x_1)'\,\mathbb{S}_lQ_{\Phi\Phi}^{-1}\zeta_n\right|\leq n^{1/2}h^{1/2}\sup_{x_1\in\mathcal{X}_1}\|\eta_{nl}(x_1)\|\left\|Q_{\Phi\Phi}^{-1}\right\|_{\rm sp}\|\mathbb{S}_l\|_{\rm sp}\|\zeta_n\|\\
&= n^{1/2}h^{1/2}O_P(1)\,O(1)\,O(1)O_P(\kappa^{-\gamma})=o_P(1),\\
\sup_{x_1\in\mathcal{X}_1}|B_{nl,4}(x_1)| &= \sup_{x_1\in\mathcal{X}_1}n^{1/2}h^{1/2}\left|\eta_{nl}(x_1)'\,\mathbb{S}_l\mathbf{R}_{n,\beta}\right|\leq n^{1/2}h^{1/2}\sup_{x_1\in\mathcal{X}_1}\|\eta_{nl}(x_1)\|\,\|\mathbb{S}_l\|_{\rm sp}\|\mathbf{R}_{n,\beta}\|\\
&= n^{1/2}h^{1/2}O_P(1)\,O(1)\,o_P\left(n^{-1/2}h^{-1/2}\right)=o_P(1),
\end{aligned}$$

and

$$\sup_{x_1\in\mathcal{X}_1}|B_{nl,5}(x_1)|\leq O\left(\kappa^{-\gamma}\right)n^{1/2}h^{1/2}\left\{\sup_{x_1\in\mathcal{X}_1}n^{-1}\sum_{i=1}^{n}K_{ix_1}\left|c'H^{-1}X_{1i}^{*}(x_1)\right|\right\}=O_P\left(n^{1/2}h^{1/2}\kappa^{-\gamma}\right)=o_P(1).$$

For $B_{nl,3}(x_1)$, we have $B_{nl,3}(x_1)=\sum_{k=1}^{d_x}B_{nl,3k}(x_1)$ where $B_{nl,3k}(x_1)=n^{-1/2}h^{1/2}\eta_{nl}(x_1)'\mathbb{S}_lQ_{\Phi\Phi}^{-1}\sum_{j=1}^{n}\Phi_j\delta_{kj}(\tilde{U}_{kj}-U_{kj})$. Using (A.2), we make the following decomposition

$$B_{nl,3k}(x_1)=-\sum_{s=1}^{5}n^{-1/2}h^{1/2}\eta_{nl}(x_1)'\mathbb{S}_lQ_{\Phi\Phi}^{-1}\sum_{j=1}^{n}\Phi_j\delta_{kj}u_{sk,j}\equiv-\sum_{s=1}^{5}B_{nl,3ks}(x_1),\text{ say.}$$

First, noting that δ_{kj} is uniformly bounded, we can show $\left\|n^{-1}\sum_{j=1}^{n}\Phi_j\delta_{kj}\right\|_{\text{sp}}=O_P(1)$ using arguments similar to those used in the proof of Lemma A.5(iii). It follows that

$$\begin{aligned}\sup_{x_1\in\mathcal{X}_1}|B_{nl,3k1}(x_1)| &\leq h^{1/2}\sup_{x_1\in\mathcal{X}_1}\|\eta_{nl}(x_1)\|\,\|\mathbb{S}_l\|_{\text{sp}}\left\|Q_{\Phi\Phi}^{-1}\right\|_{\text{sp}}\left\|n^{-1}\sum_{j=1}^{n}\Phi_j\delta_{kj}\right\|_{\text{sp}}n^{1/2}|\mu_k-\tilde{\mu}_k|\\ &= h^{1/2}O_P(1)\,O(1)\,O(1)\,O_P(1)\,O(1)\,O_P(1)=o_P(1).\end{aligned}$$

Now we decompose $B_{nl,3k2}(x_1)$ as follows:

$$\begin{aligned}B_{nl,3k2}(x_1) &= \sum_{m=1}^{d_1}n^{-1/2}h^{1/2}\bar{\eta}_l(x_1)'\mathbb{S}_lQ_{\Phi\Phi}^{-1}\sum_{j=1}^{n}\delta_{kj}\Phi_jp^{\kappa_1}(Z_{1m,j})'\mathbb{S}_{1m}a_{1k}\\ &\quad+\sum_{m=1}^{d_1}n^{-1/2}h^{1/2}r_{\eta l}(x_1)'\mathbb{S}_lQ_{\Phi\Phi}^{-1}\sum_{j=1}^{n}\delta_{kj}\Phi_jp^{\kappa_1}(Z_{1m,j})'\mathbb{S}_{1m}a_{1k}\\ &\equiv B_{nl,3k2}^{(1)}(x_1)+B_{nl,3k2}^{(2)}(x_1),\text{ say.}\end{aligned}$$

Let $\varphi_{nlkm}(x_1)=\bar{\eta}_l(x_1)'\mathbb{S}_lQ_{\Phi\Phi}^{-1}n^{-1}\sum_{j=1}^{n}\delta_{kj}\Phi_jp^{\kappa_1}(Z_{1m,j})$ and $\bar{\varphi}_{lkm}(x_1)=E[\varphi_{nlkm}(x_1)]$. Arguments like those used to study $\eta_{nl}(x_1)$ in the proof of Lemma A.6(i) show that $\|\bar{\varphi}_{lkm}(x_1)\|=O(\|\bar{\eta}_l(x_1)\|)=O\left(1+\kappa^{1/2}h\right)=O(1)$ under Assumption A5(ii) and $\|\varphi_{nlkm}(x_1)-E[\varphi_{nlkm}(x_1)]\|=\|\bar{\eta}_l(x_1)\|\,O_P((\kappa^{1/2}\log n/n)^{-1/2})=O_P((\kappa^{1/2}\log n/n)^{-1/2})$ uniformly in x_1. We further decompose $B_{nl,3k2}^{(1)}(x_1)$ as follows

$$\begin{aligned}B_{nl,3k2}^{(1)}(x_1) &= \sum_{m=1}^{d_1}n^{-1/2}h^{1/2}\bar{\eta}_l(x_1)'\mathbb{S}_lQ_{\Phi\Phi}^{-1}\sum_{j=1}^{n}\delta_{kj}\Phi_jp^{\kappa_1}(Z_{1m,j})'\mathbb{S}_{1m}Q_{n,PP}^{-1}\xi_{nk}\\ &= \sum_{m=1}^{d_1}n^{1/2}h^{1/2}\bar{\varphi}_{lkm}(x_1)'\mathbb{S}_{1m}Q_{PP}^{-1}\xi_{nk}+\sum_{m=1}^{d_1}n^{1/2}h^{1/2}\bar{\varphi}_{lkm}(x_1)'\mathbb{S}_{1m}(Q_{n,PP}^{-1}-Q_{PP}^{-1})\xi_{nk}\\ &\quad+\sum_{m=1}^{d_1}n^{1/2}h^{1/2}r_{nlkm}(x_1)'\mathbb{S}_{1m}Q_{n,PP}^{-1}\xi_{nk}\\ &\equiv B_{nl,3k2}^{(1,1)}(x_1)+B_{nl,3k2}^{(1,2)}(x_1)+B_{nl,3k2}^{(1,3)}(x_1).\end{aligned}$$

Following the analysis of $B_{nl,11}(x_1)$, we can show that $\sup_{x_1\in\mathcal{X}_1}\left|B_{nl,3k2}^{(1,1)}(x_1)\right|=O_P\left((h/\log n)^{1/2}\right)$. In addition,

$$\begin{aligned}\sup_{x_1\in\mathcal{X}_1}\left|B_{nl,3k2}^{(1,2)}(x_1)\right| &\leq n^{1/2}h^{1/2}\sup_{x_1\in\mathcal{X}_1}\sum_{m=1}^{d_1}\|\bar{\varphi}_{lkm}(x_1)\|\,\|\mathbb{S}_{1m}\|_{\text{sp}}\left\|Q_{n,PP}^{-1}-Q_{PP}^{-1}\right\|_{\text{sp}}\|\xi_{nk}\|\\ &= n^{1/2}h^{1/2}O_P(1)O(1)\,O_P\left(\kappa_1n^{-1/2}\right)O_P(\kappa_1^{1/2}n^{-1/2})=o_P(1),\end{aligned}$$

and

$$\sup_{x_1\in\mathcal{X}_1}\left|B^{(1,3)}_{nl,3k2}(x_1)\right| \leq n^{1/2}h^{1/2}\sup_{x_1\in\mathcal{X}_1}\sum_{m=1}^{d_1}\|r_{nlkm}(x_1)\|\,\|\mathbb{S}_{1m}\|_{\rm sp}\left\|Q^{-1}_{n,PP}\right\|_{\rm sp}\|\xi_{nk}\|_{\rm sp}$$
$$= n^{1/2}h^{1/2}O_P((\kappa^{1/2}\log n/n)^{-1/2})O(1)\,O_P(1)\,O_P\left(\kappa_1^{1/2}n^{-1/2}\right)=o_P(1).$$

It follows that $\sup_{x_1\in\mathcal{X}_1}\left|B^{(1)}_{nl,3k2}(x_1)\right|=o_P(1)$. For $B^{(2)}_{nl,3k2}(x_1)$, we have

$$\sup_{x_1\in\mathcal{X}_1}\left|B^{(2)}_{nl,3k2}(x_1)\right| \leq n^{1/2}h^{1/2}\sup_{x_1\in\mathcal{X}_1}\|r_{\eta l}(x_1)\|\,\|\mathbb{S}_l\|_{\rm sp}\left\|Q^{-1}_{\Phi\Phi}\right\|_{\rm sp}\sum_{m=1}^{d_1}\|t_{nkm}\|_{\rm sp}\|\mathbb{S}_{1m}\|_{\rm sp}\|a_{1k}\|$$
$$= n^{1/2}h^{1/2}O_P\left((\kappa^{1/2}\log n/n)^{-1/2}\right)O(1)\,O_P(1)\,O(1)\,O_P\left(\kappa_1^{1/2}n^{-1/2}\right)=o_P(1),$$

where $t_{nkm}\equiv n^{-1}\sum_{j=1}^n\delta_{kj}\Phi_j p^{\kappa_1}(Z_{1m,j})'$, we use the fact that $\|t_{nkm}\|_{\rm sp}=O_P(1)$ by following similar arguments to those used in the proof of Lemma A.5(iii) and noticing that δ_{kj} is uniformly bounded. Consequently we have shown that $\sup_{x_1\in\mathcal{X}_1}|B_{nl,3k2}(x_1)|=o_P(1)$. Analogously,

$$\sup_{x_1\in\mathcal{X}_1}|B_{nl,3k4}(x_1)| \leq n^{1/2}h^{1/2}\sup_{x_1\in\mathcal{X}_1}\|\eta_{nl}(x_1)\|_{\rm sp}\|\mathbb{S}_l\|_{\rm sp}\left\|Q^{-1}_{\Phi\Phi}\right\|_{\rm sp}\sum_{m=1}^{d_1}\|t_{nkm}\|_{\rm sp}\|\mathbb{S}_{1m}\|_{\rm sp}\|a_{2k}\|$$
$$= n^{1/2}h^{1/2}O_P(1)O(1)\,O_P(1)\,O_P(1)\,O(1)\,O_P\left(\kappa_1^{-\gamma}\right)=o_P(1).$$

By the same token, we can show that $B_{nl,3k3}(x_1)=o_P(1)$ and $B_{nl,3k}(x_1)=o_P(1)$ uniformly in x_1. It follows that $\sup_{x_1\in\mathcal{X}_1}\|B_{nl,3k}(x_1)\|=o_P(1)$ for $k=1,...,d_x$. Analogously, we can show that $(i3)$: $\sup_{x_1\in\mathcal{X}_1}\|C_{nj}(x_1)\|=o_P(1)$ for $j=1,...,d_1$.

Now we show $(i4)$. We make the following decomposition

$$c'D_{nl}(x_1) = n^{-1/2}h^{1/2}\sum_{i=1}^n K_{ix_1}c'H^{-1}X^*_{1i}(x_1)\left[\tilde{g}_{d_x+d_1+l}(\tilde{U}_{li})-g_{d_x+d_1+l}(\tilde{U}_{li})\right]$$
$$+n^{-1/2}h^{1/2}\sum_{i=1}^n K_{ix_1}c'H^{-1}X^*_{1i}(x_1)\left[g_{d_x+d_1+l}(\tilde{U}_{li})-g_{d_x+d_1+k}(U_{li})\right]$$
$$\equiv D_{nl,1}(x_1)+D_{nl,2}(x_1),\ \text{say}.$$

In view of the fact that $\tilde{g}_{d_x+d_1+l}(\tilde{U}_{li})-g_{d_x+d_1+l}(\tilde{U}_{li})=p^{\kappa}(\tilde{U}_{li})'\mathbb{S}_{d_x+d_1+k}(\tilde{\beta}-\beta)+\left[p^{\kappa}(\tilde{U}_{li})'\beta_l-g_{d_x+d_1+l}(\tilde{U}_{li})\right]$, we continue to decompose $D_{nl,1}(x_1)$ as follows:

$$D_{nl,1}(x_1) = n^{-1/2}h^{1/2}\sum_{i=1}^n K_{ix_1}c'H^{-1}X^*_{1i}(x_1)\,p^{\kappa}(U_{li})'\,\mathbb{S}_{d_x+d_1+l}\left(\tilde{\beta}-\beta\right)$$
$$+n^{-1/2}h^{1/2}\sum_{i=1}^n K_{ix_1}c'H^{-1}X^*_{1i}(x_1)\left[p^{\kappa}(\tilde{U}_{li})-p^{\kappa}(U_{li})\right]'\mathbb{S}_{d_x+d_1+l}\left(\tilde{\beta}-\beta\right)$$
$$-n^{-1/2}h^{1/2}\sum_{i=1}^n K_{ix_1}c'H^{-1}X^*_{1i}(x_1)\left[g_{d_x+d_1+l}(\tilde{U}_{li})-p^{\kappa}(\tilde{U}_{li})'\beta_l\right]$$
$$\equiv D_{nl,11}(x_1)+D_{nl,12}(x_1)+D_{nl,13}(x_1),\ \text{say}.$$

Analogous to the analysis of $B_{nl,1}(x_1)$, we can readily show that $\sup_{x_1\in\mathcal{X}_1}|D_{nl,11}(x_1)| = o_P(1)$. For $D_{nl,12}(x_1)$, by Taylor expansion,

$$\begin{aligned} D_{nl,12}(x_1) &= n^{-1/2}h^{1/2}\sum_{i=1}^{n}K_{ix_1}c'H^{-1}X_{1i}^*(x_1)(\tilde{U}_{li}-U_{li})\dot{p}^{\kappa}(U_{li})'\left(\tilde{\beta}_l-\beta_l\right) \\ &\quad +\frac{1}{2}n^{-1/2}h^{1/2}\sum_{i=1}^{n}K_{ix_1}c'H^{-1}X_{1i}^*(x_1)(\tilde{U}_{li}-U_{li})^2\ddot{p}^{\kappa}\left(U_{li}^{\ddagger}\right)'\left(\tilde{\beta}_l-\beta_l\right) \\ &\equiv D_{nl,121}(x_1)+\frac{1}{2}D_{nl,122}(x_1),\text{ say,} \end{aligned}$$

where $U_{li}^{\ddagger}$ lies between $\tilde{U}_{li}$ and U_{li}. By Theorem 3.1 and Lemmas A.6(i)-(ii), $\sup_{x_1\in\mathcal{X}_1}|D_{nl,121}(x_1)| = h^{1/2}\varsigma_{1\kappa}\,O_P(1+n^{1/2}\kappa_1^{-\gamma})O_P(\nu_n+\nu_{1n}) = o_P(1)$, and

$$\begin{aligned} \sup_{x_1\in\mathcal{X}_1}|D_{nl,122}(x_1)| &\le \varsigma_{2\kappa}\sup_{x_1\in\mathcal{X}_1}\left\{n^{-1/2}h^{1/2}\sum_{i=1}^{n}K_{ix_1}c'H^{-1}X_{1i}^*(x_1)(\tilde{U}_{li}-U_{li})^2\right\}\left\|\tilde{\beta}_l-\beta_l\right\| \\ &= \varsigma_{2\kappa}n^{1/2}h^{1/2}O_P(\kappa_1 n^{-1}+\kappa_1^{-2\gamma})O_P(\nu_n+\nu_{1n}) = o_P(1). \end{aligned}$$

In addition, $\sup_{x_1\in\mathcal{X}_1}\|D_{nl,13}(x_1)\| \le n^{1/2}h^{1/2}O(\kappa^{-\gamma})\sup_{x_1\in\mathcal{X}_1}n^{-1}\sum_{i=1}^{n}K_{ix_1}\left\|H^{-1}X_{1i}^*(x_1)\right\| = O_P(n^{1/2}h^{1/2}\kappa^{-\gamma}) = o_P(1)$. It follows that $\sup_{x_1\in\mathcal{X}_1}|D_{nl,1}(x_1)| = o_P(1)$.

By Taylor expansion,

$$\begin{aligned} D_{nl,2}(x_1) &= n^{-1/2}h^{1/2}\sum_{i=1}^{n}K_{ix_1}c'H^{-1}X_{1i}^*(x_1)\dot{g}(U_{li})\left(\tilde{U}_{li}-U_{li}\right) \\ &\quad +n^{-1/2}h^{1/2}\sum_{i=1}^{n}K_{ix_1}c'H^{-1}X_{1i}^*(x_1)\ddot{g}_{d_x+d_1+l}\left(U_{li}^{\ddagger}\right)\left(\tilde{U}_{li}-U_{li}\right)^2 \\ &\equiv D_{nl,21}(x_1)+D_{nl,22}(x_1). \end{aligned}$$

Arguments like those used to study $B_{nl,3}(x_1)$ show that $\sup_{x_1\in\mathcal{X}_1}|D_{nl,21}(x_1)| = o_P(1)$. By Lemma A.6($ii$), $\sup_{x_1\in\mathcal{X}_1}|D_{nl,22}(x_2)| \le c_{\ddot{g}}\sup_{x_1\in\mathcal{X}_1}\{n^{-1/2}h^{1/2}\sum_{i=1}^{n}K_{ix_1}\left|c'H^{-1}X_{1i}^*(x_1)\right|(\tilde{U}_{li}-U_{ki})^2\} = n^{1/2}h^{1/2}O_P(\nu_{1n}^2) = o_P(1)$, where $c_{\ddot{g}} = \sup_{u_l\in\mathcal{U}_l}\ddot{g}_{d_x+d_1+l}(u_l) = O(1)$. ■

References

Bernstein, D. S., 2005. *Matrix Mathematics: Theory, Facts, and Formulas with Application to Linear Systems Theory*. Princeton University Press, Princeton.

Chen, X., 2007. Large sample sieve estimation of semi-nonparametric models. In J. J. Heckman and E. Leamer (eds.), *Handbook of Econometrics*, 6B (Chapter 76), 5549-5632. New York: Elsevier Science.

Fan, J., Gijbels, I., 1996. *Local Polynomial Modelling and Its Applications*. Chapman & Hall, London.

Hahn, J., Ridder, G., 2013. Asymptotic variance of semiparametric estimators with generated regressors. *Econometrica* 81, 315-340.

Hansen, B. E., 2008. Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory* 24, 726-748.

Horowitz, J. L., 2013. Nonparametric additive models. In A. Ullah, J. Racine, and L. Su (eds.), *Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*, forthcoming, Oxford University Press, Oxford.

Horowitz, J. L., Mammen, E., 2004. Nonparametric estimation of an additive model with a link function. *Annals of Statistics* 32, 2412-2443.

Kim, W., Linton O. B., Hengartner N. W., 1999. A computationally efficient estimator for additive nonparametric regression with bootstrap confidence intervals. *Journal of Computational and Graphical Statistics* 8, 278-297.

Li, Q., 2000. Efficient estimation of additive partially linear models. *International Economic Review* 41, 1073-1092.

Li, Q., Racine, J., 2007. *Nonparametric Econometrics: Theory and Practice.* Princeton University Press: Princeton.

Mammen, E., Rothe, C., Schienle, M., 2012. Nonparametric regression with nonparametrically generated covariates. *Annals of Statistics* 40, 1132-1170.

Martins-Filho, C., Yang, K., 2007. Finite sample performance of kernel-based regression methods for nonparametric additive models under common bandwidth selection criterion. *Journal of Nonparametric Statistics* 19, 23-62.

Masry, E., 1996. Multivariate local polynomial regression for time series: uniform strong consistency rates. *Journal of Time Series Analysis,* 17, 571-599.

Newey, W. K., 1997. Convergence rates and asymptotic normality for series estimators. Journal of Econometrics 79, 147-168.

Newey, W. K., Powell, J. L., 2003. Instrumental variable estimation of nonparametric models. *Econometrica* 71, 1565-1578.

Newey, W. K., Powell, J. L., Vella, F., 1999. Nonparametric estimation of triangular simultaneous equation models. *Econometrica* 67, 565-603.

Ozabaci, D., Henderson, D. J. 2012. Gradients via oracle estimation for additive nonparametric regression with application to returns to schooling. Working paper, State University of New York at Binghamton.

Pagan, A., Ullah, A. 1999. *Nonparametric Econometrics.* Cambridge University Press, Cambridge.

Pinkse, J., 2000. Nonparametric two-step regression estimation when regressors and error are dependent. *Canadian Journal of Statistics* 28, 289-300.

Powell, M. J. D., 1981. *Approximation Theory and Methods.* Cambridge University Press, Cambridge.

Schumaker, L. L., 2007. *Spline Functions: Basic Theory*, 3rd ed. Cambridge University Press, Cambridge.

Serfling, R. J., 1980. *Approximation Theorems of Mathematical Statistics.* John Wiley & Sons, New York.

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

Su, L., Jin, S., 2012. Sieve estimation of panel data models with cross section dependence. *Journal of Econometrics* 169, 34-47.

Su, L., Ullah, A. 2008. Local polynomial estimation of nonparametric simultaneous equations models. *Journal of Econometrics* 144, 193-218.

Thompson, R., Freede, 1974. Eigenvalues of partitioned Hermitian matrices. *Bulletin Australian Mathematical Society* 3, 23-37

Yang, L., Sperlich, S., Härdle, W. 2003. Derivative estimation and testing in generalized additive models. *Journal of Statistical Planning and Inference.* 115: 521:542.