International Journal of Statistics and Management System
Vol. 4 No. 1 (January-June, 2019)

Statistical inference for seemingly unrelated varying-coefficient nonparametric regression models*

Jinhong You,† Bin Zhou‡ and Haibo Zhou§

Received: 13th March 2018 Revised: 24th April 2018 Accepted: 10th October 2018

Abstract

This paper is concerned with the statistical inference of seemingly unrelated (SU) varying-coefficient nonparametric regression models. We propose an estimation for the unknown coefficient functions, which is an extension of the two-stage procedure proposed by Linton al. [19] in the longitudinal data framework where they focused on purely nonparametric regression. We show the resulted estimators are asymptotically normal and more efficient than those based on only the individual regression equation even when the error covariance matrix is homogeneous. Another focus of this paper is to extend the generalized likelihood ratio technique developed by Fan, Zhang and Zhang [7] for testing the goodness of fit of models to the setting of SU regression. A wild block bootstrap based method is used to compute p-value of the test. Some simulation studies are given in support of the asymptotics. A real data set from an ongoing environmental epidemiologic study is used to illustrate the proposed procedures.

Key words and phrases: seemingly unrelated, varying-coefficient, two-stage, goodness of fit.

†*Mailing Address:* Department of Statistics, Shanghai University of Finance and Economics, Shanghai, P.R. China 200433. *E–mail:* `jyou@incresearch.com`

‡*Mailing Address:* Department of Management and Insurance, East China Normal University, Shanghai, P.R. China 20062. *E–mail:* `bzhou@stats.ecnu.cn`

§*Mailing Address:* Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7400, USA. *E–mail:* `zhou@bios.unc.edu`

1 Introduction. Seemingly unrelated (SU) regression models proposed by Zellner [36, 37] have broad applicability in the analysis of behaviors in biology, economics, education and social science. These models take account of the fact that subtle interactions often may be present between individual statistical relationships when each of these relationships is being used to model some aspect of behavior (Srivastava and Giles [29]). Many works in the literature have contributed to analyze the SU regression models, which include for example Kakwani [14], Srivastava and Giles [29], Rocke [27], Neudecker and Windmeijer [23], Mandy and Martins [22], Kurata [16], Hougaard [12], Liu [20], Ng [24], Kalbfleisch and Prentice [15]), He and Lawless [11], Carroll et al. [3] and so on.

All results mentioned above for SU regression models are established in the setting of parametric regressions, mainly linear regression. However, in practice the regression functions are usually unknown, so that parametric models may be misspecified and hence be inadequate to capture the underlying relationships between response variables and their associated covariates, which may cause large bias of modeling. In order to reduce the modeling bias, recently Smith and Kohn [28], Wang, Guo and Brown [33] and Welsh and Yee [34] proposed SU nonparametric regression models. The immediate advantage of the nonparametric regression models is that no prior information on model structure is assumed. Further, they may provide useful insight for further parametric fitting. However, a purely nonparametric method is hampered by its serious drawbacks, such as the curse of dimensionality, difficulty of interpretation, and lack of extrapolation capability. Like other nonparametric regression model, the SU nonparametric regression model suffers from the well-known "curse of dimensionality" when the covariates are multidimensional, that is the optimal rate of convergence decreases with the increase of the covariate's dimensionality. In addition, it is hard to describe, interpret, and understand the estimated regression surface when the dimension is more than two. In order to avoid these shortcomings of purely nonparametric regression models some authors including Lang, Adebayo and Fahrmeir [17], Lang et al. [18], Poirer, Koop and Tobias [25] have studied SU semiparametric regression models and SU structural nonparametric regression models. Specially, Lang, Adebayo and Fahrmeir [17], Poirer, Koop and Tobias [25] focused on SU partially linear regression model, and Lang et al. [18] on SU additive nonparametric regression model. Another important structural nonparametric regression model is the varying-coefficient regression model proposed by Hastie and Tibshirani [10], which allows appreciable flexibility on the structure of fitted model without suffering from the "the curse of dimensionality". The varying-coefficient regression model is a useful extension of thresholding models in Tong [31]. It also appears natural in the longitudinal data analysis where one wishes to explore the extent to which covariates affect response changing over time. See for example Wu, Chiang and Hoover [35] and Fan and Zhang [6] for novel applications of the model to longitudinal data. The varying-coefficient models are also useful for analyzing functional types of data. See Ramsay and Silverman [26] and Brumback and Rice [1] for details. In this paper we propose the following SU varying-coefficient regression model.

$$Y_{ij} = X_{ij}^1 \alpha_{1j}(U_{ij}) + \ldots + X_{ij}^{p_j} \alpha_{p_j j}(U_{ij}) + \varepsilon_{ij} \text{ for } i = 1, \ldots, n \text{ and } j = 1, \ldots, J \tag{1}$$

where Y_{ij}'s are responses, $\alpha_{sj}(\cdot)$'s are unknown functions, X_{ij}^s and U_{ij}'s are design points, ε_{ij} are errors with $E(\varepsilon_{ij}) = 0$ and $E(\varepsilon_{ij_1}\varepsilon_{ij_2}) = \sigma^2_{j_1 j_2}$ for $j, j_1, j_2 = 1, \ldots, J$. Further, $E(\varepsilon_{i_1 j_1}\varepsilon_{i_2 j_2}) = 0$ when $i_1 \neq i_2$. Our research is motivated by a recent environmental epidemiology study, the Collaborative Prenatal Projects (CPP) (Gray et al. [9]). The investigators are interested in assessing the relationship of the women's PCB exposure and their children's hearing of left ear and right ear. One hand, the investigators doubt that the women's PCB exposure has the same effect on their children's left ear hearing and right ear hearing, and the effect may be nonlinear. On the other hand, they believe that there exist correlation between left ear hearing and right ear hearing.

We propose a two-stage local polynomial estimation for the unknown coefficient functions in model (1), which is an extension of the procedure developed by Linton et al. [19] in the longitudinal data framework. Linton et al. [19] focused on the purely nonparametric regression. We show the resulted estimators are asymptotically normal and more efficient than those only based on the individual regression equation even when the error covariance matrix is homogeneous. Therefore, our estimation is also asymptotically more efficient than the weighted local polynomial estimations proposed by Welsh and Yee [34] in the setting of purely SU nonparametric regression when the error covariance matrix is homogeneous.

Another important statistical question in fitting model (1) is if there exists a parametric structure for some $\alpha_{sj}(\cdot)$. This amounts to testing if $\alpha_{sj}(\cdot)$ is in a certain parametric form. A testing procedure is proposed based on the comparison of the sum of residual squares under the null and alternative models. This testing procedure is an extension of the generalized likelihood ratio technique proposed by Fan, Zhang and Zhang [7] to the setting of SU regression. A wild block bootstrap method is used to find the null distribution of the test statistic.

The layout of the remainder of this paper is as follows. In Section 2 we present the two-stage estimation for model (1). Section 3 establishes the asymptotic properties of the resulted estimators. A wild block bootstrap based test is proposed in Section 4. Some simulation studies are conducted in Section 5. A real data from an ongoing environmental epidemiologic study is analyzed in Section 6. Section 7 concludes. The proof of the main results are collected in Appendix.

2 Two-stage estimation. Throughout this paper we will assume $\sigma^2_{j_1j_2}$'s are known. If $\sigma^2_{j_1j_2}$'s are unknown we can estimate them root-n consistently, which is faster than the rates available in nonparametric regression. Therefore, for purposes of asymptotic theory we may assume without loss of generality that $\sigma^2_{j_1j_2}$'s are known. The root-n consistent estimates of $\sigma^2_{j_1j_2}$ can be obtained by computing the residuals which are based on only the individual regression equation. If the errors of model (1) are heterogeneous and their covariance matrix is a smoothing function of covariate U, we can estimate it by applying kernel smoothing to the residuals. We believe the subsequent results still hold although the kernel smoothing covariance matrix estimator is not root-n consistent.

The two-stage local polynomial procedure consists of two steps.

1. First, based on the individual regression equation we fit the functions $\boldsymbol{\alpha}_j(\cdot) = (\alpha_{1j}(\cdot), \ldots, \alpha_{p_jj}(\cdot))^T$ using local linear method.

2. At the second step, we construct a linear transformation of $(\mathbf{Y}_1^T, \ldots, \mathbf{Y}_J^T)^T$ that has the same mean as $(\mathbf{Y}_1^T, \ldots, \mathbf{Y}_J^T)^T$ and diagonal covariance matrix, and then apply the local linear method to this transformation where $\mathbf{Y}_j = (Y_{1j}, \ldots, Y_{nj})^T$.

This two-stage procedure was proposed by Linton et al. [19] in the longitudinal data framework. They focused on the purely nonparametric regression. We will show that this two-stage procedure works even for SU structural nonparametric regression models such as the SU varying-coefficient regression model.

2.1 First-stage estimation. In the first stage, based on the individual regression equation we estimate the unknown coefficient functions of model (1) by local polynomial (linear) smoother. According to Fan and Gijbels [5], the local polynomial smoother has attractive properties. For example, it reduces the bias of the Nadaraya-Watson estimators and the variance of the Gasser-Müller estimator, and adapts automatically to the boundary of design points.

For fixed j, suppose that $(Y_{ij}, \mathbf{X}_{ij}^T, U_{ij})_{i=1}^n$ is a sample from model (1.1), namely they satisfy

$$Y_{ij} = \mathbf{X}_{ij}^T \boldsymbol{\alpha}_j(U_{ij}) + \varepsilon_{ij}, \quad i = 1, \ldots, n \tag{2}$$

where $\mathbf{X}_{ij} = (X_{ij}^1, \ldots, X_{ij}^{p_j})^T$ and $\boldsymbol{\alpha}_j(\cdot) = (\alpha_{1j}(\cdot), \ldots, \alpha_{p_j j}(\cdot))^T$. Now, apply a local linear regression technique to estimate the coefficient functions $\{\alpha_{sj}(\cdot), s = 1, \ldots, p_j\}$ in model (2) as Cai, Fan and Yao [2]. For U in a small neighborhood of u, one can approximate $\alpha_{sj}(U)$ locally by a linear function

$$\alpha_{sj}(U) \approx \alpha_{sj}(u) + \alpha'_{sj}(u)(U - u) \equiv a_{sj} + b_{sj}(U - u), \quad s = 1, \ldots, p_j$$

where $\alpha'_{sj}(u) = \partial\alpha_{sj}(u)/\partial u$. This leads to the following weighted local least-squares problem: find $\{(a_{sj}, b_{sj}), s = 1, \ldots, p_j\}$ to minimize

$$\sum_{i=1}^n \left[Y_{ij} - \sum_{s=1}^{p_j} \{a_{sj} + b_{sj}(U_{ij} - u)\} X_{ij}^s \right]^2 K_h(U_{ij} - u), \tag{3}$$

where $K(\cdot)$ is a kernel function, h is a bandwidth and $K_h(\cdot) = K(\cdot/h)/h$. The solution to problem (3) is given by

$$\left\{ \hat{a}_{1j}(u), \ldots, \hat{a}_{p_j j}(u), h\hat{b}_{1j}(u), \ldots, h\hat{b}_{p_j j}(u) \right\}^T = \left(\mathbf{D}_{ju}^T \mathbf{W}_{ju} \mathbf{D}_{ju} \right)^{-1} \mathbf{D}_{ju}^T \mathbf{W}_{ju} \mathbf{Y}_j. \tag{4}$$

where

$$\mathbf{D}_{ju} = \begin{pmatrix} \mathbf{X}_{1j}^T & \frac{U_{1j}-u}{h} \mathbf{X}_{1j}^T \\ \vdots & \vdots \\ \mathbf{X}_{nj}^T & \frac{U_{nj}-u}{h} \mathbf{X}_{nj}^T \end{pmatrix}, \quad \text{and} \quad \mathbf{W}_{ju} = \operatorname{diag}(K_h(U_{1j} - u), \ldots, K_h(U_{nj} - u)).$$

Thus, an estimator for $\boldsymbol{\Psi}_j(u) = (\alpha_{1j}(u), \ldots, \alpha_{p_j j}(u), \alpha'_{1j}(u), \ldots, \alpha'_{p_j j}(u))$ has the form

$$\hat{\boldsymbol{\Psi}}_j(u) = \mathbf{H}^{-1} \left(\mathbf{D}_{ju}^T \mathbf{W}_{ju} \mathbf{D}_{ju} \right)^{-1} \mathbf{D}_{ju}^T \mathbf{W}_{ju} \mathbf{Y}_j.$$

where $\mathbf{H} = \operatorname{diag}(1, h) \otimes \mathbf{I}_{p_j}$, $\mathbf{I}_{p_j}$ is a $p_j \times p_j$ identity matrix and $\otimes$ is the Kronecker product. Especially, an estimator of $\alpha_{sj}(u)$ is

$$\hat{\alpha}_{sj}(u) = \mathbf{e}_{s,2p_j}^T \mathbf{H}^{-1} \left(\mathbf{D}_{ju}^T \mathbf{W}_{ju} \mathbf{D}_{ju} \right)^{-1} \mathbf{D}_{ju}^T \mathbf{W}_{ju} \mathbf{Y}_j,$$

where $\mathbf{e}_{s,2p_j}$ is a $2p_j$-column vector with 1 in the sth position and zeros elsewhere. The estimator $\hat{\boldsymbol{\Psi}}_j(\cdot)$ or $\hat{\boldsymbol{\alpha}}_j(\cdot) = (\hat{\alpha}_{1j}(\cdot), \ldots, \hat{\alpha}_{p_j j}(\cdot))^T$ is the well-known local linear estimator. Since the correlation between the equations is not considered in the estimators $\hat{\boldsymbol{\Psi}}_j(u)$ and $\hat{\boldsymbol{\alpha}}_j(u)$ one can not expect that they are asymptotically efficient. In the following, we will propose an improved estimation.

2.2 Two-stage estimation. Obviously, the correlation of the equations which comprise the structure of the model (1), and the form of the associated disturbance variance covariance matrix, introduce additional information over the available when the individual equations are considered separately. This suggests that treating model (1) as a collection of separate relationships will be suboptimal when drawing inferences about the model's unknown coefficient functions. Indeed, as we shall see, in general the sharpness of these inferences may be improved by taking into account of the correlation inherent in model (1), rather than ignoring it. Recognizing this point motivates us to consider the two-stage nonparametric estimation.

Let $\mathbf{\Sigma} = (\sigma^2_{j_1j_2})^J_{j_1,j_2=1}$ and $\mathbf{\Phi} = \left\{(\text{diag}(\mathbf{\Sigma}^{-\frac{1}{2}}))^{-1}(\mathbf{\Sigma}^{-\frac{1}{2}} - \text{diag}(\mathbf{\Sigma}^{-\frac{1}{2}}))\right\} \otimes \mathbf{I}_n$ where $\mathbf{I}_n$ is an $n \times n$ identity matrix. For any function $(m_{11}(\cdot), \dots, m_{p_11}(\cdot), \dots, m_{p_JJ}(\cdot))^T$, define

$$\begin{pmatrix} \mathbf{Z}_1(m_{11}, \dots, m_{p_11}, \dots, m_{p_JJ}) \\ \vdots \\ \mathbf{Z}_J(m_{11}, \dots, m_{p_11}, \dots, m_{p_JJ}) \end{pmatrix} = \begin{pmatrix} \mathbf{Y}_1 \\ \vdots \\ \mathbf{Y}_J \end{pmatrix} + \mathbf{\Phi} \begin{pmatrix} \mathbf{Y}_1 - \mathbf{G}_1 \\ \vdots \\ \mathbf{Y}_J - \mathbf{G}_J \end{pmatrix}$$

with $\mathbf{G}_j = (\mathbf{X}^T_{1j}\mathbf{m}_j(U_{1j}), \dots, \mathbf{X}^T_{nj}\mathbf{m}_j(U_{nj}))^T$ and $\mathbf{m}_j(\cdot) = (m_{1j}(\cdot), \dots, m_{p_jj}(\cdot))^T$. Note that

$$\begin{pmatrix} \mathbf{Z}_1(\alpha_{11}, \dots, \alpha_{p_11}, \dots, \alpha_{p_JJ}) \\ \vdots \\ \mathbf{Z}_J(\alpha_{11}, \dots, \alpha_{p_11}, \dots, \alpha_{p_JJ}) \end{pmatrix} = \begin{pmatrix} \mathbf{M}_1 \\ \vdots \\ \mathbf{M}_J \end{pmatrix} + \mathbf{\Phi} \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \vdots \\ \boldsymbol{\varepsilon}_J \end{pmatrix}$$

and

$$\text{Cov}\left\{ \mathbf{\Phi} \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \vdots \\ \boldsymbol{\varepsilon}_J \end{pmatrix} \right\} = (\text{diag}(\mathbf{\Sigma}^{-\frac{1}{2}}))^{-2} \otimes \mathbf{I}_n$$

where $\mathbf{M}_j = (\mathbf{X}^T_{1j}\boldsymbol{\alpha}_j(U_{1j}), \dots, \mathbf{X}^T_{nj}\boldsymbol{\alpha}_j(U_{nj}))^T$ and $\boldsymbol{\varepsilon}_j = (\varepsilon_{1j}, \dots, \varepsilon_{nj})^T$.

This implies that $\mathbf{Z}_{j_1}(\alpha_{11}, \dots, \alpha_{p_11}, \dots, \alpha_{p_JJ})$ and $\mathbf{Z}_{j_2}(\alpha_{11}, \dots, \alpha_{p_11}, \dots, \alpha_{p_JJ})$ are uncorrelated for $j_1 \neq j_2$. Further, if we denote

$$\mathbf{\Omega} = (\omega_{j_1j_2})^J_{j_1,j_2=1} = \{\text{diag}(\mathbf{\Sigma}^{-\frac{1}{2}})\}^{-1}(\mathbf{\Sigma}^{-\frac{1}{2}} - \text{diag}(\mathbf{\Sigma}^{-\frac{1}{2}}))$$

we have

$$\mathbf{Z}_j(\alpha_{11}, \dots, \alpha_{p_11}, \dots, \alpha_{p_JJ}) = \mathbf{Y}_j + \sum_{j_1=1}^{J} \omega_{jj_1}(\mathbf{Y}_{j_1} - \mathbf{M}_{j_1}).$$

Thus, the two-stage local linear estimator of $\mathbf{\Psi}_j(u)$ is

$$\hat{\mathbf{\Psi}}_j^{TS}(u) = \mathbf{H}^{*-1}\left(\mathbf{D}_{ju}^{*T}\mathbf{W}_{ju}^{*}\mathbf{D}_{ju}^{*}\right)^{-1}\mathbf{D}_{ju}^{*T}\mathbf{W}_{ju}^{*}\left\{\mathbf{Y}_j + \sum_{j_1=1}^{J}\omega_{jj_1}(\mathbf{Y}_{j_1} - \mathbf{M}_{j_1})\right\}.$$

where $\mathbf{H}^*, \mathbf{D}_{ju}^*$ and $\mathbf{W}_{ju}^*$ have the same definitions as $\mathbf{H}, \mathbf{D}_{ju}$ and $\mathbf{W}_{ju}$ except that h is replaced by h^*. However, $\mathbf{M}_j$'s of the right side of $\hat{\mathbf{\Psi}}_j^{TS}(u)$ are unknown. Therefore, we replace $\mathbf{M}_j$ by the estimator from the first stage and it results in the following feasible estimator

$$\hat{\mathbf{\Psi}}_j^{TS}(u) = \mathbf{H}^{*-1}\left(\mathbf{D}_{ju}^{*T}\mathbf{W}_{ju}^{*}\mathbf{D}_{ju}^{*}\right)^{-1}\mathbf{D}_{ju}^{*T}\mathbf{W}_{ju}^{*}\left\{\mathbf{Y}_j + \sum_{j_1=1}^{J}\omega_{jj_1}(\mathbf{Y}_{j_1} - \hat{\mathbf{M}}_{j_1})\right\}.$$

where $\hat{\mathbf{M}}_j = (\mathbf{X}_{1j}^T\hat{\boldsymbol{\alpha}}_j(U_{1j}), \ldots, \mathbf{X}_{nj}^T\hat{\boldsymbol{\alpha}}_j(U_{nj}))^T$. Especially, a two-stage local linear estimator of $\alpha_{sj}(u)$ is

$$\hat{\alpha}_{sj}^{TS}(u) = \mathbf{e}_{s,2p_j}^T\mathbf{H}^{*-1}\left(\mathbf{D}_{ju}^{*T}\mathbf{W}_{ju}^{*}\mathbf{D}_{ju}\right)^{-1}\mathbf{D}_{ju}^{T}\mathbf{W}_{ju}^{*}\left\{\mathbf{Y}_j + \sum_{j_1=1}^{J}\omega_{jj_1}(\mathbf{Y}_{j_1} - \hat{\mathbf{M}}_{j_1})\right\}.$$

Remark 2.1. *The proposed two-stage estimation is easy to be extended to the case of the error covariance matrix being heterogeneous. Let* $Cov(\boldsymbol{\varepsilon}_i^*) = \mathbf{\Sigma}_i = (\sigma_{ij_1j_2}^2)_{j_1,j_2}^J$ *and* $(\omega_{ij_1j_2})_{j_1,j_2=1}^J = \{diag(\mathbf{\Sigma}_i^{-\frac{1}{2}})(\mathbf{\Sigma}_i^{-\frac{1}{2}} - diag(\mathbf{\Sigma}_i^{-\frac{1}{2}}))\}$ *where* $\boldsymbol{\varepsilon}_i^* = (\varepsilon_{i1}, \ldots, \varepsilon_{iJ})^T$*. Then the feasible two-stage local linear estimators for* $\mathbf{\Psi}_j(\cdot)$ *and* $\alpha_{sj}(\cdot)$ *are, respectively*

$$\hat{\mathbf{\Psi}}_j^{TS}(u) = \mathbf{H}^{*-1}\left(\mathbf{D}_{ju}^{*T}\mathbf{W}_{ju}^{*}\mathbf{D}_{ju}^{*}\right)^{-1}\mathbf{D}_{ju}^{*T}\mathbf{W}_{ju}^{*}\left\{\mathbf{Y}_j + \sum_{j_1=1}^{J} diag(\omega_{1jj_1}, \ldots, \omega_{njj_1})(\mathbf{Y}_{j_1} - \hat{\mathbf{M}}_{j_1})\right\}.$$

and

$$\hat{\alpha}_{sj}^{TS}(u) = \mathbf{e}_{s,2p_j}^T\mathbf{H}^{*-1}\left(\mathbf{D}_{ju}^{*T}\mathbf{W}_{ju}^{*}\mathbf{D}_{ju}\right)^{-1}\mathbf{D}_{ju}^{T}\mathbf{W}_{ju}^{*}\left\{\mathbf{Y}_j + \sum_{j_1=1}^{J} diag(\omega_{1jj_1}, \ldots, \omega_{njj_1})(\mathbf{Y}_{j_1} - \hat{\mathbf{M}}_{j_1})\right\}.$$

In the next section, we will show the two-stage estimators are asymptotically normal and more efficient than those based on only individual equation.

3 Asymptotic properties of two-stage estimators. In order to establish the asymptotic properties of $\hat{\mathbf{\Psi}}_j^{TS}(\cdot)$ and $\hat{\alpha}_{sj}^{TS}(u)$ we make the following assumptions. These assumptions, while look a little bit lengthy, are actually quite mild and can be easily satisfied.

Assumption 1. *For fixed j, $(\mathbf{X}_{ij}^T, U_{ij}, \varepsilon_{ij})^T$ are i.i.d random vectors. U_{1j} has a bounded support $\mathcal{U}_j$ and its density function $f_j(\cdot)$ is Lipschitz continuous and bounded away from 0 on its support. In addition, there exist a constant c such that $f_{j,j_1}(u_j, u_{j_1}) < c$ where $f_{j,j_1}(\cdot,\cdot)$ is the joint density function of U_{ij} and U_{ij_1}.*
Assumption 2. *The $p\times p$ matrix $\mathbf{\Gamma}_j(u) = E(\mathbf{X}_{1j}\mathbf{X}_{1j}^T | U_{1j} = u)$ is positive definite for each $u \in \mathcal{U}_j$.*
Assumption 3. *There is an $s > 2$ such that $E\varepsilon_{1j}^{2s} < \infty$, $E||\mathbf{X}_{1j}||^{2s} < \infty$ for $j = 1,\ldots,J$.*
Assumption 4. *$\{\alpha_{sj}(\cdot), s = 1,\ldots,p_j, j = 1,\ldots,J\}$ have the continuous second derivatives in $u \in \mathcal{U}_j$.*
Assumption 5. *The function $K(\cdot)$ is a density function with compact support and the bandwidths h and h^* satisfy $nh^8 \to 0$, $nh^2/(\log n)^2 \to \infty$, $nh^{*8} \to 0$, $nh^{*2}/(\log n)^2 \to \infty$ and $h^*/h \to 0$ as $n \to \infty$.*

Remark 3.1. *Throughout this paper, we assume that $\boldsymbol{\alpha}_j(\cdot)$'s have the same smoothness for different j. This is just for simplicity. Actually, the proposed two-stage local linear estimation can easily adopt the case that $\boldsymbol{\alpha}_j$'s have different smoothness. In addition, for fixed j, when $\alpha_{sj}(\cdot)$'s have the different smoothness for different s, the estimation developed by Fan and Zhang [6] may be useful here.*

Let
$$\mu_j = \int_{-\infty}^{\infty} u^j K(u)du, \qquad \nu_j = \int_{-\infty}^{\infty} u^j K^2(u)du.$$

The following theorem shows that $\hat{\mathbf{\Psi}}_j^{TS}(\cdot)$ is asymptotically normal.

Theorem 3.1. *Suppose that Assumptions 1 to 5 hold. Then it holds that*

$$\sqrt{nh^*}\left[\mathbf{H}^*\left\{\hat{\mathbf{\Psi}}_j^{TS}(u) - \mathbf{\Psi}_j(u)\right\} - \frac{h^{*2}}{2}\frac{1}{\mu_2-\mu_1^2}\begin{pmatrix}(\mu_2-\mu_1\mu_3)\boldsymbol{\alpha}_j''(u)\\(\mu_3-\mu_1\mu_2)\boldsymbol{\alpha}_j''(u)\end{pmatrix}\right] \xrightarrow{D} N(0, \mathbf{\Sigma}_j^{TS})$$

as $n \to \infty$, where $\boldsymbol{\alpha}_j''(u) = (\alpha_{1j}''(u),\ldots,\alpha_{p_jj}''(u))^T$ with $\alpha_{sj}''(u) = \partial^2\alpha_{sj}(u)/\partial u^2$, "$\xrightarrow{D}$" denotes convergence in distribution, and

$$\begin{aligned}\mathbf{\Sigma}_j^{TS} &= \frac{((diag(\mathbf{\Sigma}^{-\frac{1}{2}}))^{-2})_{jj}}{f_j(u)}\mathbf{\Gamma}_j^{-1}(u)\\ &\otimes\frac{1}{(\mu_2-\mu_1^2)^2}\begin{pmatrix}\mu_2^2\nu_0 - 2\mu_1\mu_2\nu_1 + \mu_1^2\nu_2 & (\mu_1^2+\mu_2)\nu_1 - \mu_1\mu_2\nu_0 - \mu_1\nu_2\\ (\mu_1^2+\mu_2)\nu_1 - \mu_1\mu_2\nu_0 - \mu_1\nu_2 & \nu_2 - \mu_1(2\nu_1+\mu_1\nu_0)\end{pmatrix}.\end{aligned}$$

Corollary 1 *Under conditions of Theorem 3.1, we have*

$$\sqrt{nh^*}\left\{\hat{\boldsymbol{\alpha}}_j^{TS}(u) - \boldsymbol{\alpha}_j(u) - \frac{h^{*2}}{2}\frac{\mu_2^2 - \mu_1\mu_3}{\mu_2 - \mu_1^2}\boldsymbol{\alpha}_j''(u)\right\} \xrightarrow{D} N(0, \boldsymbol{\Sigma}_{\boldsymbol{\alpha}_j}^{TS}) \text{ as } n \to \infty$$

where $\hat{\boldsymbol{\alpha}}_j^{TS}(u) = (\hat{\alpha}_{1j}^{TS}(u), \ldots, \hat{\alpha}_{p_j j}^{TS}(u))^T$, $\boldsymbol{\alpha}_j(u) = (\alpha_{1j}(u), \ldots, \alpha_{p_j j}(u))^T$ *and*

$$\boldsymbol{\Sigma}_{\boldsymbol{\alpha}_j}^{TS} = \frac{((\mathrm{diag}(\boldsymbol{\Sigma}^{-\frac{1}{2}}))^{-2})_{jj}(c_0^2\nu_0 + 2c_0c_1\nu_1 + c_1^2\nu_2)}{f_j(u)}\boldsymbol{\Gamma}_j^{-1}(u)$$

with $c_0 = \mu_2/(\mu_2 - \mu_1^2)$ *and* $c_1 = -\mu_1/(\mu_2 - \mu_1^2)$.

Remark 3.2. *It is easy to see that under Assumptions 1 to 5 we have*

$$\sqrt{nh}\left\{\hat{\boldsymbol{\alpha}}_j(u) - \boldsymbol{\alpha}_j(u) - \frac{h^2}{2}\frac{\mu_2^2 - \mu_1\mu_3}{\mu_2 - \mu_1^2}\boldsymbol{\alpha}_j''(u)\right\} \xrightarrow{D} N(0, \boldsymbol{\Sigma}_{\boldsymbol{\alpha}_j}) \quad \text{as } n \to \infty$$

where $\hat{\boldsymbol{\alpha}}_j(u) = (\hat{\alpha}_{1j}(u), \ldots, \hat{\alpha}_{p_j j}(u))^T$ *and*

$$\boldsymbol{\Sigma}_{\boldsymbol{\alpha}_j} = \frac{\sigma_{jj}^2(c_0^2\nu_0 + 2c_0c_1\nu_1 + c_1^2\nu_2)}{f_j(u)}\boldsymbol{\Gamma}_j^{-1}(u)$$

By the same argument as in Appendix 5.2 of Linton et al. [19] we can show that

$$((diag(\boldsymbol{\Sigma}^{-\frac{1}{2}}))^{-2})_{jj} \le \sigma_{jj}^2.$$

Therefore, the two-stage estimator $\hat{\boldsymbol{\alpha}}_j^{TS}(\cdot)$ *is asymptotically more efficient than those based on only the individual regression equation.*

In addition, when the error covariance matrix is heterogeneous we have

$$\sqrt{nh^*}\left\{\hat{\boldsymbol{\alpha}}_j^{TS}(u) - \boldsymbol{\alpha}_j(u) - \frac{h^{*2}}{2}\frac{\mu_2^2 - \mu_1\mu_3}{\mu_2 - \mu_1^2}\boldsymbol{\alpha}_j''(u)\right\} \xrightarrow{D} N(0, \boldsymbol{\Sigma}_{\boldsymbol{\alpha}_j}^{TS*}) \quad \text{as } n \to \infty$$

and

$$\sqrt{nh}\left\{\hat{\boldsymbol{\alpha}}_j(u) - \boldsymbol{\alpha}_j(u) - \frac{h^2}{2}\frac{\mu_2^2 - \mu_1\mu_3}{\mu_2 - \mu_1^2}\boldsymbol{\alpha}_j''(u)\right\} \xrightarrow{D} N(0, \boldsymbol{\Sigma}_{\boldsymbol{\alpha}_j}^{*}) \quad \text{as } n \to \infty$$

where

$$\boldsymbol{\Sigma}_{\boldsymbol{\alpha}_j}^{TS*} = \frac{\{\lim_{n\to\infty} n^{-1}\sum_{i=1}^n((diag(\boldsymbol{\Sigma}_i^{-\frac{1}{2}}))^{-2})_{jj}\}(c_0^2\nu_0 + 2c_0c_1\nu_1 + c_1^2\nu_2)}{f_j(u)}\boldsymbol{\Gamma}_j^{-1}(u)$$

and

$$\boldsymbol{\Sigma}_{\boldsymbol{\alpha}_j}^{*} = \frac{\{\lim_{n\to\infty} n^{-1}\sum_{i=1}^n \sigma_{ijj}^2\}(c_0^2\nu_0 + 2c_0c_1\nu_1 + c_1^2\nu_2)}{f_j(u)}\boldsymbol{\Gamma}_j^{-1}(u)$$

Obviously, $\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}((diag(\mathbf{\Sigma}_i^{-\frac{1}{2}}))^{-2})_{jj} \leq \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\sigma_{ijj}^2$*. Therefore, when the error covariance matrix is heterogeneous the two-stage local linear estimator* $\hat{\boldsymbol{\alpha}}_j^{TS}(\cdot)$ *is still asymptotically more efficient than those based on only the individual regression equation.*

In order to apply Theorem 3.1 or Corollary 1 to make statistical inference for $\hat{\mathbf{\Psi}}_j^{TS}(\cdot)$ or $\hat{\boldsymbol{\alpha}}_j^{TS}(\cdot)$, a consistent estimator of $\mathbf{\Sigma}_j^{TS}$ or $\mathbf{\Sigma}_{\boldsymbol{\alpha}_j}^{TS}$ is needed. It is easy to see that ν_j and μ_j are known constants. In addition, we can construct root-n consistent estimators of $\sigma_{j_1 j_2}^2$ based on estimated residuals. Therefore, in order to construct the consistent estimators for $\mathbf{\Sigma}_j^{TS}$ and $\mathbf{\Sigma}_{\boldsymbol{\alpha}_j}^{TS}$ we just provide a consistent estimator for $\mathbf{\Gamma}_j(u)f_j(u)$. Define

$$\hat{\mathbf{\Gamma}}_j(u) = \frac{1}{nh^*}\sum_{i=1}^{n} K_{h^*}(U_{ij}-u)\mathbf{X}_{ij}\mathbf{X}_{ij}^T\ .$$

The following theorem shows that $\hat{\mathbf{\Gamma}}_j(u)$ is a consistent estimator of $\mathbf{\Gamma}_j(u)f_j(u)$.

Theorem 3.2. *Suppose that Assumptions 1,2 and 5 hold. Then it holds that*

$$\hat{\mathbf{\Gamma}}_j(u) \to_p \mathbf{\Gamma}_j(u)f_j(u) \quad as\ n\to\infty.$$

4 Wild block bootstrap based goodness of fit test. To test whether model (1) holds with a specified parametric form for the varying-coefficient functions such as an SU linear regression model, we extend the generalized likelihood technique in Fan, Zhang and Zhang [7] to the current setting. Applying the the generalized likelihood ratio technique to test the goodness of fit of non-SU varying-coefficient regressions is considered by Cai, Fan and Yao [2]. The null hypothesis is

$$H_0 : \alpha_{sj}(u) = a_{sj}(u, \boldsymbol{\theta}_{sj}), \quad s = 1, \ldots, p_j, j = 1, \ldots, J, \tag{5}$$

where $a_{sj}(\cdot, \boldsymbol{\theta}_{sj})$ is a given family of functions indexed by an unknown parameter vector $\boldsymbol{\theta}_{sj}$. Let $\hat{\boldsymbol{\theta}}_{sj}$ be a consistent estimator of $\boldsymbol{\theta}_{sj}$ and denote $\hat{\mathbf{a}}_j(U_{ij}) = (\hat{a}_{1j}(U_{ij}, \hat{\boldsymbol{\theta}}_{1j}), \ldots, \hat{a}_{p_j j}(U_{ij}, \hat{\boldsymbol{\theta}}_{p_j j}))^T$. The residual sum of squares under the null hypothesis is

$$\mathrm{RSS}_0 = (nJ)^{-1}\sum_{j=1}^{J}\sum_{i=1}^{n}(Y_{ij} - \mathbf{X}_{ij}^T\hat{\mathbf{a}}_j(U_{ij}))^2,$$

and the residual sum of squares corresponding to model (1.1) is

$$\mathrm{RSS}_1 = (nJ)^{-1} \sum_{j=1}^{J} \sum_{i=1}^{n} (Y_{ij} - \mathbf{X}_{ij}^T \hat{\boldsymbol{\alpha}}_j^{TS}(U_{ij}))^2.$$

Our test statistic is defined as

$$Q_n = (\mathrm{RSS}_0 - \mathrm{RSS}_1)/\mathrm{RSS}_1 = \mathrm{RSS}_0/\mathrm{RSS}_1 - 1,$$

for which we have the following theorem.

Theorem 4.1. *Suppose that Assumptions 1 to 5 hold. Then under H_0, $Q_n \to 0$ in probability as $n \to \infty$. Otherwise, if $\inf_{\boldsymbol{\theta}_{sj}} \left\{ \int_{\mathcal{U}_j} (\alpha_{sj}(u) - a_{sj}(u, \boldsymbol{\theta}_{sj}))^2 du \right\}^2 > 0$ for some $s = 1, \ldots, p_j, j = 1, \ldots, p$, then there exists a constant $\delta > 0$ such that $Q_n > \delta$ with probability approaching one as $n \to \infty$.*

Theorem 4.1 suggests that we should reject the null hypothesis (5) for large values of Q_n. However, the distribution of Q_n is hard to obtain. Inspired by Stute, Gonzá lez and Presedo [30], Godfrey and Tremayne [8] and Ioannidis and Peel [13] among others who have successfully used the wild bootstrap method to calculate the p-values of their tests and note that there exists correlation between ε_{ij_1} and ε_{ij_2} even when $j_1 \neq j_2$, we develop a wild block bootstrap procedure to compute the p-value of Q_n.

Follow the steps below to implement our goodness-of-fit test:

1. Fit the SU varying-coefficient regression model (1) as described in Section 2 and calculate Q_n and the estimated pseudo residuals $\{\hat{\varepsilon}_{ij}\}_{i=1}^n$, where

$$\hat{\varepsilon}_{ij} = Y_{ij} - \mathbf{X}_{ij}^T \hat{\boldsymbol{\alpha}}_j^{TS}(U_{ij}), \quad i = 1, \ldots, n, j = 1, \ldots, J.$$

2. Generate a sequence of i.i.d. random variables $\{\tau_i\}_{i=1}^n$ from a symmetric distribution function $F(\cdot)$ such that $E\tau_1 = 0$, $E\tau_1^2 = 1$ and $E|\tau_1|^3 < \infty$. We stress that $F(\cdot)$ is chosen independently of the given regression model.
3. Set $Y_{ij}^* = \mathbf{X}_{ij}^T \hat{\boldsymbol{\alpha}}_j^{TS}(U_{ij}) + \hat{\varepsilon}_{ij}\tau_i$, $i = 1, \ldots, n$ and calculate the bootstrap test statistic Q_n^* based on the sample $\{Y_{ij}^*, \mathbf{X}_{ij}^T, U_{ij}, i = 1, \ldots, n, j = 1, \ldots, J\}$.
4. Repeat Step 2 and Step 3 a large number of times to generate a bootstrap distribution of Q_n^*.
5. Reject the null hypothesis H_0 at level α if Q_n is greater than the $100(1-\alpha)\%$ quantile of the bootstrap distribution of Q_n^*.

The p-value of our test is simply the relative frequency of the event $\{Q_n^* \geq Q_n\}$ in the replications of the bootstrap sampling.

5 Some simulation studies. In this section we carry out some simulation studies to demonstrate the finite sample performance of the proposed procedures. The data are generated from the following SU varying-coefficient nonparametric regression model

$$y_{ij} = x_{ij}\alpha_j(u_{ij}) + \varepsilon_{ij}, \quad i = 1, \ldots, n, \quad j = 1, 2,$$

where $u_{i1} \sim U(0,1)$, $u_{i2} \sim U(0,1)$, $x_{i1} \sim N(1, 0.25^2)$, $x_{i2} \sim N(1, 0.25^2)$, $\alpha_1(u_{i1}) = 2\sin(2\pi u_{i1})$, $\alpha_2(u_{i2}) = 2\cos(3\pi u_{i2})$, $\varepsilon_{i1} \sim N(0,1)$, $\varepsilon_{i2} \sim N(0,1)$ and $E(\varepsilon_{i1}\varepsilon_{i2}) = \sigma_{12}^2$. We take $\sigma_{11}^2 = 0.5$ and 0.9 with respect to different degrees of correlation.

Samples of size $n = 50$, 100, 200 and 400 are drawn repeatedly. In each case the number of simulated realizations is $1,000$. We take the Gaussian kernel.

The estimators $\hat{\alpha}_1(\cdot)$, $\hat{\alpha}_2(\cdot)$, and $\hat{\alpha}_1^{TS}(\cdot)$, $\hat{\alpha}_2^{TS}(\cdot)$ are assessed via the Square-Root of Averaged Squared Errors (RASE):

$$\text{RASE} = \left[n^{-1}\sum_{i=1}^{n}\{\tilde{\alpha}_1(u_{i1}) - \alpha_1(u_{i1})\}^2\right]^{\frac{1}{2}} + \left[n^{-1}\sum_{i=1}^{n}\{\tilde{\alpha}_2(u_{i2}) - \tilde{\alpha}_2(u_{i2})\}^2\right]^{\frac{1}{2}}.$$

where $\tilde{\alpha}_j(\cdot)$ is an estimator of $\alpha_j(\cdot)$. The results are summarized in the following Table 1.

Table 1: The finite sample performance of the estimators for the coefficient functions.

			$n = 50$	$n = 100$	$n = 200$	n=400
$\sigma_{12}^2 = 0.5$	RASE1	sm	0.8601	0.6685	0.4706	0.3623
		std	0.1345	0.0945	0.0945	0.0480
	RASE2	sm	0.7999	0.6199	0.4320	0.3388
		std	0.1233	0.0857	0.0590	0.0363
$\sigma_{12}^2 = 0.9$	RASE1	sm	0.8537	0.6147	0.4668	0.3645
		std	0.1463	0.0919	0.0655	0.0405
	RASE2	sm	0.6092	0.4226	0.3032	0.0405
		std	0.1161	0.0818	0.0580	0.0308

Note: RASE1 means the RASE of the estimators which only based on individual equations and RASE2 means the RASE of the proposed estimators.

From Table 1 we can see that the two-stage estimators perform much better than the one based on only the individual equation, especially when σ_{12}^2 is big. For example, when $n = 100$ and $\sigma_{12}^2 = 0.5$ the RASE of $\hat{\alpha}_1(\cdot)$ and $\hat{\alpha}_2(\cdot)$ is 0.6685, and that of $\hat{\alpha}_1^{TS}(\cdot)$ and $\hat{\alpha}_2^{TS}(\cdot)$

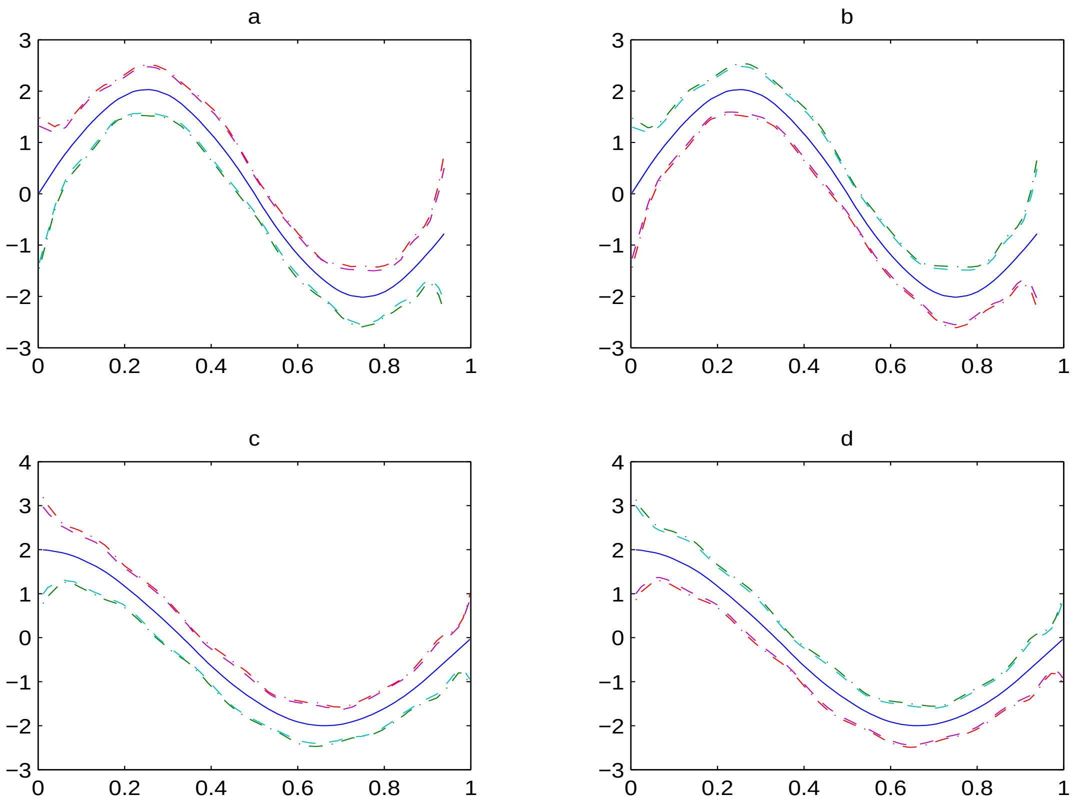

Figure 1: $\sigma_{12}^2 = 0.5$. Plots (a) and (b) are the confidence bands for $\alpha_1(\cdot)$. Plots (c) and (d) are the confidence bands for $\alpha_2(\cdot)$. Plots (a) and (c) are the 95% Monte Carlo simulation confidence bands. Plots (b) and (d) are the 95% asymptotic confidence bands. Dash-dotted curves: the confidence bands based on the usual local linear estimator. Dashed curves: the confidence bands based on two-stage estimators. Solid curves: the true curves.

is 0.6199. However, when $n = 100$ and $\sigma_{12}^2 = 0.9$ the RASE of $\hat{\alpha}_1(\cdot)$ and $\hat{\alpha}_2(\cdot)$ is 0.6147, and that of $\hat{\alpha}_1^{TS}(\cdot)$ and $\hat{\alpha}_2^{TS}(\cdot)$ is 0.4226.

We also plot the 2.5% and 97.5% quantiles from the Monte Carlo (MC) simulation and the 95% asymptotic pointwise confidence bands using the asymptotic results established in Section 3 in Figures 1 and 2. From them we can see that the asymptotic confidence band is very close to the quantile confidence band.

To demonstrate the power of the proposed bootstrap test, we consider the following null hypothesis: $H_0 : \alpha_j(u) = \theta_j$, for all $j = 1, 2$, namely an SU linear regression model, against the alternative $H_1 : \alpha_j(u) \neq \theta_j$, for at least one j.

The power function is evaluated under a sequence of the alternative models indexed by

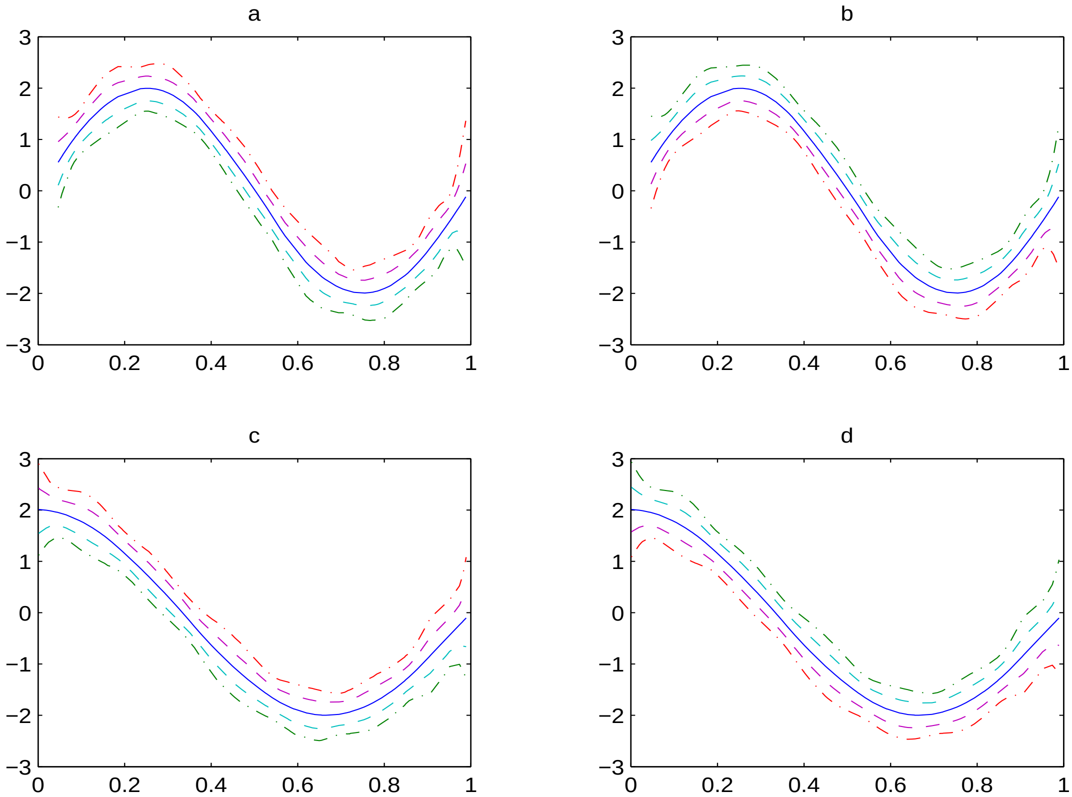

Figure 2: $\sigma_{12}^2 = 0.9$. Plots (a) and (b) are the confidence bands for $\alpha_1(\cdot)$. Plots (c) and (d) are the confidence bands for $\alpha_2(\cdot)$. Plots (a) and (c) are the 95% Monte Carlo simulation confidence bands. Plots (b) and (d) are the 95% asymptotic confidence bands. Dash-dotted curves: the confidence bands based on the usual local linear estimator. Dashed curves: the confidence bands based on two-stage estimators. Solid curves: the true curves.

c:

$$H_1 : \alpha_j(u) = \bar{\alpha}_j^0 + c(\alpha_j^0(u) - \bar{\alpha}_j), \qquad j = 1, 2, \quad 0 \leq c \leq 1$$

where $\alpha_1^0(u) = 2\sin(2\pi u)$, $\alpha_2^0(u) = 2\cos(3\pi u)$ and $\bar{\alpha}_j$ is the average height of $\alpha_j^0(u)$. We apply the goodness of fit test described in last section in a simulation with sample size being 500. We take $\tau_i = -(\sqrt{5}-1)/2$ with probability $(\sqrt{5}+1)/(2\sqrt{5})$ and $\tau_i = (\sqrt{5}+1)/2$ with probability $1 - (\sqrt{5}+1)/(2\sqrt{5})$. For each realization, we repeat bootstrap sampling 500 samples. Figure 3 plots the simulated power function against c.

When the null hypothesis holds the sizes are very significantly close to the nominal level 5%. This demonstrates that bootstrap estimate of the null distribution is approximately correct. Meanwhile, from Figure 3, the power function shows that our test is indeed

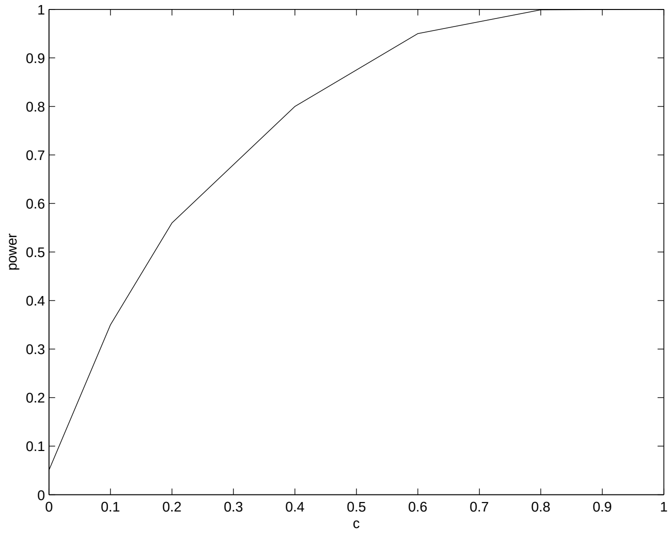

Figure 3: The power curves of testing the goodness of fit of model in which $n = 100, \sigma_{12}^2 = 0.5$.

powerful.

6 An application. We apply the proposed method to analyze a data set from the Collaborative Perinatal Project (CPP). CPP is a prospectively designed study to provide precise data for studies of a wide variety of neurological outcome and birth detects (Gray et al. [9]. Subjects were enrolled through 12 university affiliated medical clinics, with the centers contributing unequal numbers of subjects. In all, 55,908 pregnancies were registered, representing the experience of about 44,000 women. The children born during the study were followed for various outcomes for up to 8 years. One of the hypotheses is that the PCB levels are related to performance on the hearing level for children at 7 years of age, taking children's gender into account. The PCBs were measured by analyzing the third trimester blood serum specimens that have been preserved from mothers in the CPP study. We use the average of hearing level at 1,000, 2,000 and 4,000 for left and right ear of each child at 7 years of age as the outcome variable and the PCB, gender as the exposure variables. The following model is used.

$$\text{Left ear hearing} = \alpha_{11}(\text{PCB}) + \text{Gender} \times \alpha_{12}(\text{PCB}) + \varepsilon_L$$

$$\text{Right ear hearing} = \alpha_{21}(\text{PCB}) + \text{Gender} \times \alpha_{22}(\text{PCB}) + \varepsilon_R$$

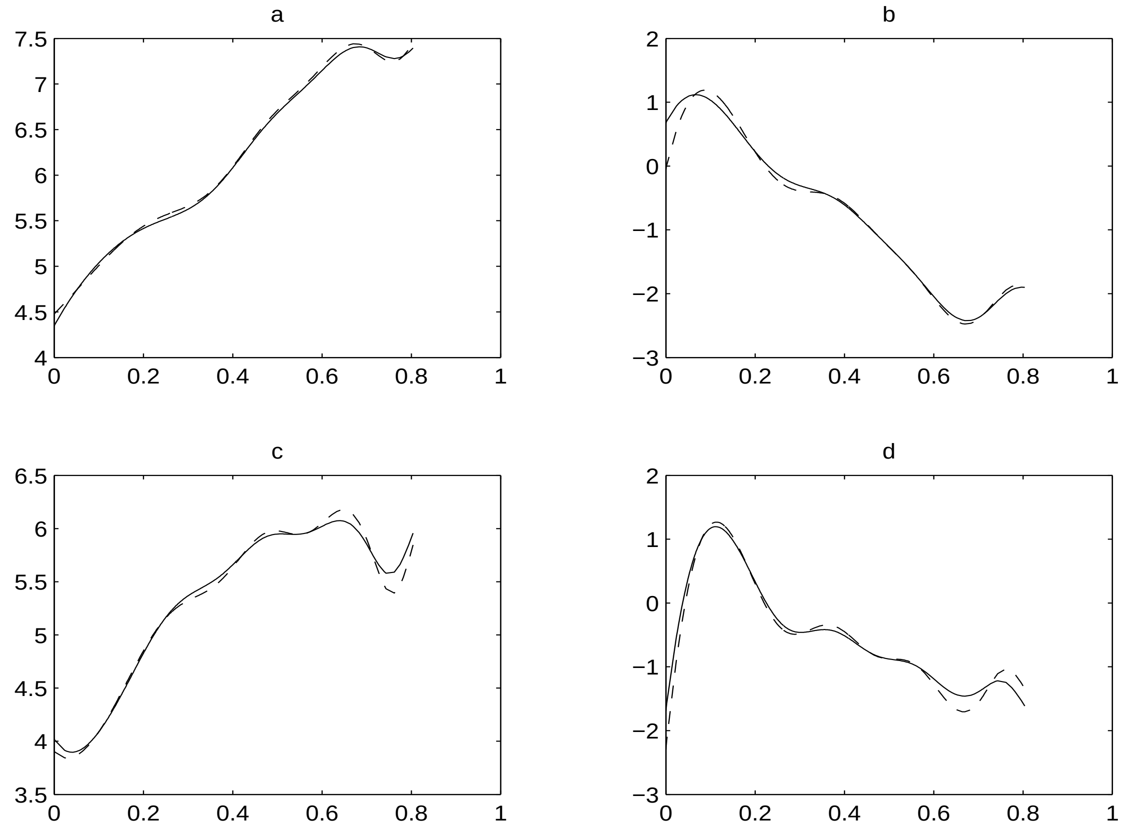

Figure 4: Application to CPP data. Plots (a) and (b) are for left ear. Plots (c) and (d) are for right ear. Plots (a) and (c) are estimators of intercept coefficient function. Plots (b) and (d) are estimators of the gender coefficient function. Dashed curves: two-stage estimators. Solid curves: local linear estimator.

For left ear fitting the error variance is 14.5803, for right ear fitting the error variance is 20.7763 and the correlation is 0.8593. Obviously, compared with error variances, the correlation is strong. By the test, we reject the hypothesis: $\alpha_{sj}(\cdot)$ = constant for $s = 1, 2$ and $j = 1, 2$. The local linear estimators which neglect the correlation between the equations and the proposed two-stage estimators of $\alpha_{11}(\cdot), \alpha_{12}(\cdot)$, and $\alpha_{21}(\cdot), \alpha_{22}(\cdot)$, are shown in Figure 4 and the pointwise standard deviations of these estimators are shown in Figure 5.

From these figures we can see the PCB have the same effecting pattern on left and right ears. However, the degrees of effecting are different. In addition, we can see that the two-stage estimator has smaller pointwise error variance than the usual local linear estimator.

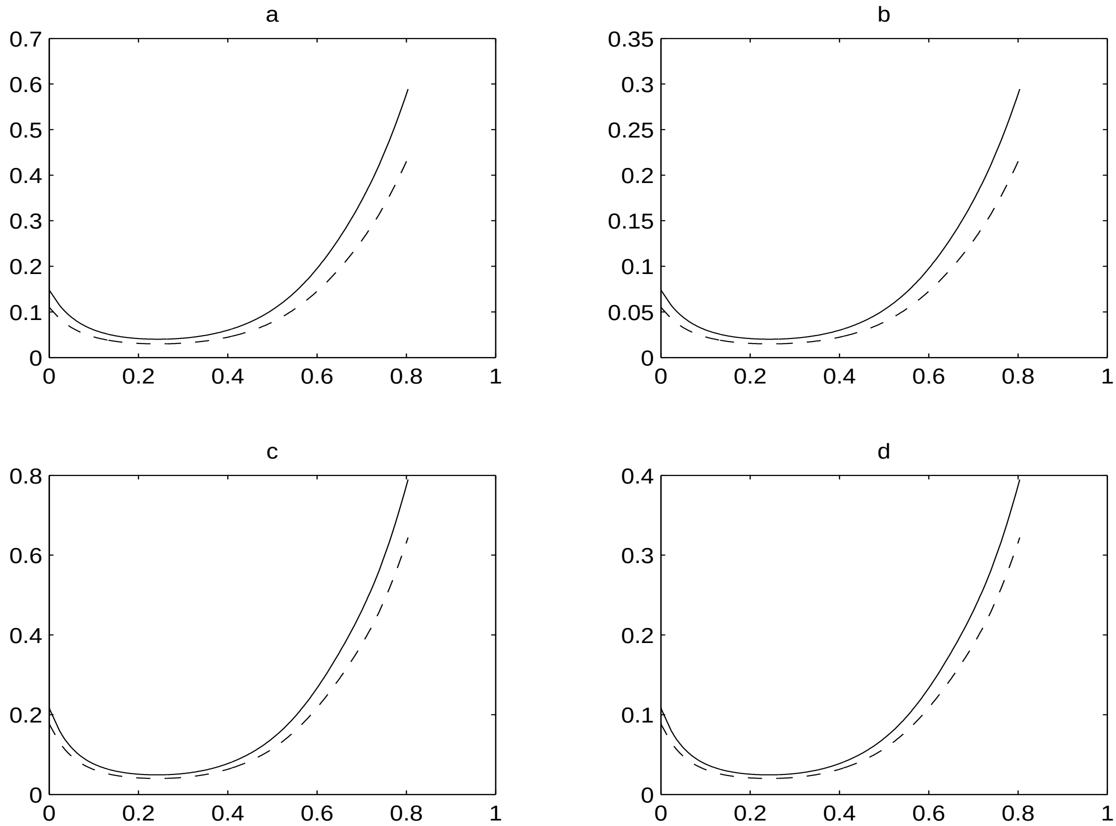

Figure 5: Application to CPP data. Plots (a) and (b) are for left ear. Plots (c) and (d) are for right ear. Plots (a) and (c) are the pointwise error variances of the estimators for intercept coefficient function. Plots (b) and (d) are the pointwise error variances of estimators for the gender coefficient function. Dashed curves: the pointwise error variances of the two-stage estimators. Solid curves: the pointwise error variances of local linear estimator.

7 Concluding remarks. In this paper, we have investigated the estimating problem of seemingly unrelated varying-coefficient nonparametric regression models. We proposed a two-stage local polynomial procedure to estimate the unknown coefficient functions. It was shown that the resulted estimators are asymptotically normal and more efficient than those based on only the individual regression equation when the error covariance matrix is homogeneous or heterogeneous. A statistic is also proposed to test the goodness of fit of models.

For our results, we need an assumption that $\varepsilon_{i_1j_1}$ and $\varepsilon_{i_2j_2}$ are independent for $i_1 \neq i_2$. This assumption may be not true sometimes. For instances, crop yields from units in the same plot observed time points would generally be correlated at each time point as well exhibiting time correlation. Responses of individuals belonging to to the same group tend to be correlated within the group as well as along time axis. In this situations, it would be desirable to consider a seemingly unrelated varying-coefficient regression model which permits correlation within each block or group and also incorporates dependence between the observation vectors at different time points. For this setting, how to improve the ordinary local polynomial estimation is still an open problem.

8 Appendix: proofs of main results. In order to prove the main results we first present a lemma.

Lemma 8.1. *Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be i.i.d random vectors, where the Y_i's are scalar random variables. Further assume that $E|Y_i|^4 < \infty$ and $\sup_x \int |y|^4 f(x,y)dy < \infty$, where f denotes the joint density of (X, Y). Let K be a bounded positive function with a bounded support, and satisfies Lipschitz's condition. Then if $nh^8 \to 0$ and $nh^2/(\log n)^2 \to \infty$, it holds that*

$$\sup_x \left| \frac{1}{n} \sum_{i=1}^{n} [K_h(X_i - X)Y_i - E\{K_h(X_i - X)Y_i\}] \right| = O_p\left(\left\{ \frac{\log(1/h)}{nh} \right\}^{\frac{1}{2}} \right).$$

The proof of Lemma 8.1 follows immediately from the result of Mack and Silverman [21].

Proof of Theorem 3.1 According to the definition of $\hat{\mathbf{\Psi}}_j^{TS}(u)$ it holds that

$$\begin{aligned} \mathbf{H}^* \hat{\mathbf{\Psi}}_j^{TS}(u) &= (\mathbf{D}_{ju}^{*T} \mathbf{W}_{ju}^* \mathbf{D}_{ju}^*)^{-1} \mathbf{D}_{ju}^{*T} \mathbf{W}_{ju}^* (\boldsymbol{\varepsilon}_j + \sum_{j_1=1}^{J} \omega_{jj_1} \boldsymbol{\varepsilon}_{j_1}) \\ &\quad + (\mathbf{D}_{ju}^{*T} \mathbf{W}_{ju}^* \mathbf{D}_{ju}^*)^{-1} \mathbf{D}_{ju}^{*T} \mathbf{W}_{ju}^* \sum_{j_1=1}^{J} \omega_{jj_1} (\hat{\mathbf{M}}_j - \mathbf{M}_j) \\ &\quad + (\mathbf{D}_{ju}^{*T} \mathbf{W}_{ju}^* \mathbf{D}_{ju}^*)^{-1} \mathbf{D}_{ju}^{*T} \mathbf{W}_{ju}^* \mathbf{M}_j = J_1 + J_2 + J_3, \quad \text{say} \end{aligned}$$

where $\hat{\mathbf{M}}_j = (\mathbf{X}_{1j}^T \hat{\boldsymbol{\alpha}}_j(U_{1j}), \ldots, \mathbf{X}_{nj}^T \hat{\boldsymbol{\alpha}}_j(U_{nj}))^T$ and $\mathbf{M}_j = (\mathbf{X}_{1j}^T \boldsymbol{\alpha}_j(U_{1j}), \ldots, \mathbf{X}_{nj}^T \boldsymbol{\alpha}_j(U_{nj}))^T$. For J_2, we have

$$J_2 = (\mathbf{D}_{ju}^{*T} \mathbf{W}_{ju}^* \mathbf{D}_{ju}^*)^{-1} \begin{pmatrix} \sum_{i=1}^{n} K_{h^*}(U_{ij} - u) \mathbf{X}_{ij} \sum_{j_1=1}^{J} \omega_{jj_1} \mathbf{X}_{ij_1}^T (\hat{\boldsymbol{\alpha}}_{j_1}(U_{ij_1}) - \boldsymbol{\alpha}_{j_1}(U_{ij_1})) \\ \sum_{i=1}^{n} \frac{U_{ij} - u}{h^*} K_{h^*}(U_{ij} - u) \mathbf{X}_{ij} \sum_{j_1=1}^{J} \omega_{jj_1} \mathbf{X}_{ij_1}^T (\hat{\boldsymbol{\alpha}}_{j_1}(U_{ij_1}) - \boldsymbol{\alpha}_{j_1}(U_{ij_1})) \end{pmatrix}.$$

Let $\mathbf{S}_j(u) = \begin{pmatrix} \mathbf{\Gamma}_j(u) & \mu_1\mathbf{\Gamma}_j(u) \\ \mu_1\mathbf{\Gamma}_j(u) & \mu_2\mathbf{\Gamma}_j(u) \end{pmatrix}$. Then by Cai, Fan and Yao [2] it holds that

$$\begin{aligned}
& \hat{\boldsymbol{\alpha}}_j(U_{ij}) - \boldsymbol{\alpha}_j(U_{ij}) \\
= \; & (\mathbf{I}_p, \mathbf{0}_{p\times p}) f_j(U_{ij}) \left\{\mathbf{S}_j(U_{ij})\right\}^{-1} \begin{pmatrix} \frac{1}{n}\sum_{i_1=1}^n \mathbf{X}_{i_1j} K_h(U_{i_1j} - U_{ij})\varepsilon_{ij} \\ \frac{1}{n}\sum_{i_1=1}^n \mathbf{X}_{i_1j}\left(\frac{U_{i_1j}-U_{ij}}{h}\right) K_h(U_{i_1j} - U_{ij})\varepsilon_{ij} \end{pmatrix} \\
& + \frac{h^2}{2}\left\{\mathbf{S}_j(U_{ij})\right\}^{-1} \begin{pmatrix} \mu_2\mathbf{\Gamma}_j(U_{ij}) \\ \mu_3\mathbf{\Gamma}_j(U_{ij}) \end{pmatrix} \boldsymbol{\alpha}_j^{''}(U_{ij}) + o_p(h^2) \\
= \; & \frac{\mu_2\left\{\mathbf{\Gamma}_j(U_{ij})\right\}^{-1}}{f_j(U_{ij})(\mu_2-\mu_1^2)} \frac{1}{n}\sum_{i_1=1}^n \mathbf{X}_{i_1j} K_h(U_{i_1j} - U_{ij})\varepsilon_i \\
& - \frac{\mu_1\left\{\mathbf{\Gamma}_j(U_{ij})\right\}^{-1}}{f_j(U_{ij})(\mu_2-\mu_1^2)} \frac{1}{n}\sum_{i_1=1}^n \mathbf{X}_{i_1j}\left(\frac{U_{i_1j}-U_{ij}}{h}\right) K_h(U_{i_1j} - U_{ij})\varepsilon_i \\
& + \frac{h^2}{2}\frac{\mu^2 - \mu_1\mu_3}{\mu_2-\mu_1^2}\boldsymbol{\alpha}_j^{''}(U_{ij}) + o_p(h^2) = I_1 + I_2 + I_3 + o(h^2), \quad \text{say.}
\end{aligned}$$

Denote $\mathbf{\Gamma}^*(U_{ij}) = [\mu_2\left\{\mathbf{\Gamma}_j(U_{ij})\right\}^{-1}]/\{f_j(U_{ij})(\mu_2-\mu_1^2)\}$. We have

$$\begin{aligned}
& \frac{1}{n}\sum_{i=1}^n K_{h^*}(U_{ij}-u)\mathbf{X}_{ij}\sum_{j_1=1}^J \omega_{jj_1}\mathbf{X}_{ij_1}^T I_1 \\
= \; & \frac{1}{n}\sum_{j_1=1}^J \omega_{jj_1}\sum_{i_1=1}^n \varepsilon_{i_1j_1}\frac{1}{n}\sum_{i=1}^n K_{h^*}(U_{ij}-u)\mathbf{X}_{ij}\mathbf{X}_{ij_1}^T\mathbf{\Gamma}^*(U_{ij_1})\mathbf{X}_{i_1j_1}K_h(U_{i_1j_1}-U_{ij_1}) \\
= \; & \frac{1}{n^2}\sum_{j_1=1}^J \omega_{jj_1}\sum_{i_1=1}^n \varepsilon_{i_1j_1}K_{h^*}(U_{i_1j}-u)\mathbf{X}_{i_1j}\mathbf{X}_{i_1j_1}^T\mathbf{\Gamma}^*(U_{i_1j_1})\mathbf{X}_{i_1j_1}K_h(0) \\
& + \frac{1}{n}\sum_{j_1=1}^J \omega_{jj_1}\sum_{i_1=1}^n \varepsilon_{i_1j_1}\frac{1}{n}\sum_{i\neq i_1}^n K_{h^*}(U_{ij}-u)\mathbf{X}_{ij}\mathbf{X}_{ij_1}^T\mathbf{\Gamma}^*(U_{ij_1})\mathbf{X}_{i_1j_1}K_h(U_{i_1j_1}-U_{ij_1}) \\
= \; & I_4 + I_5, \quad \text{say.}
\end{aligned}$$

By Lemma 8.1 it is easy to see that $I_4 = o_p(n^{-1/2})$. On the other hand, obviously $\varepsilon_{i_1j_1}$ and $n^{-1}\sum_{i=1}^n K_{h^*}(U_{ij}-u)\mathbf{X}_{ij}\mathbf{X}_{ij_1}^T\mathbf{\Gamma}^*(U_{ij_1})\mathbf{X}_{i_1j_1}K_h(U_{i_1j_1}-U_{ij_1})$ are independent. Moreover, it holds that

$$\begin{aligned}
& E\left|\left|\frac{1}{n}\sum_{i=1}^n K_{h^*}(U_{ij}-u)\mathbf{X}_{ij}\mathbf{X}_{ij_1}^T\mathbf{\Gamma}^*(U_{ij_1})\mathbf{X}_{i_1j_1}K_h(U_{i_1j_1}-U_{ij_1})\right|\right| \\
\leq \; & \int\int\int f_{jj_1}(u_1,u_2)f_{j_1}(u_3)E\left(\left|\left|\mathbf{X}_{1j}\mathbf{X}_{1j_1}^T\mathbf{\Gamma}^*(u_{j_1})\right|\right| |U_{1j}=u_1, U_{1j_1}=u_2\right)E(||\mathbf{X}_{1j_1}|| \,|U_{1j_1}=u_3) \\
& \cdot K_{h^*}(u_1-u)K_h(u_2-u_3)du_1du_2du_3 = O(1)
\end{aligned}$$

Therefore,

$$\frac{1}{n}\sum_{i=1}^{n} K_{h^*}(U_{ij}-u)\mathbf{X}_{ij}\sum_{j_1=1}^{J}\omega_{jj_1}\mathbf{X}_{ij_1}^{T} I_1 = O_p(n^{-\frac{1}{2}}).$$

Moreover, combining Lemma 8.1 it is easy to see that

$$\frac{1}{n}\sum_{i=1}^{n} K_{h^*}(U_{ij}-u)\mathbf{X}_{ij}\sum_{j_1=1}^{J}\omega_{jj_1}\mathbf{X}_{ij_1}^{T}(I_2+I_3+o_p(h^2)) = O_p(h^2).$$

This implies that

$$\frac{1}{n}\sum_{i=1}^{n} K_{h^*}(U_{ij}-u)\mathbf{X}_{ij}\sum_{j_1=1}^{J}\omega_{jj_1}\mathbf{X}_{ij_1}^{T}(\hat{\boldsymbol{\alpha}}_{j_1}(U_{ij_1})-\boldsymbol{\alpha}_{j_1}(U_{ij_1})) = o_p(h^{*2}+1/\sqrt{nh^*}).$$

By the same argument, we can show that

$$\frac{1}{n}\sum_{i=1}^{n}\frac{U_{ij}-u}{h^*} K_{h^*}(U_{ij}-u)\mathbf{X}_{ij}\sum_{j_1=1}^{J}\omega_{jj_1}(\hat{\boldsymbol{\alpha}}_{j_1}(U_{ij_1})-\boldsymbol{\alpha}_{j_1}(U_{ij_1})) = o_p(h^{*2}+1/\sqrt{nh^*}).$$

Note that

$$\mathbf{D}_{ju}^{*T}\mathbf{W}_{ju}^{*}\mathbf{D}_{ju}^{*} = \begin{pmatrix} \sum_{i=1}^{n} K_{h^*}(U_{ij}-u) & \sum_{i=1}^{n}\left(\frac{U_{ij}-u}{h^*}\right) K_{h^*}(U_{ij}-u) \\ \sum_{i=1}^{n}\left(\frac{U_{ij}-u}{h^*}\right) K_{h^*}(U_{ij}-u) & \sum_{i=1}^{n}\left(\frac{U_{ij}-u}{h^*}\right)^2 K_{h^*}(U_{ij}-u) \end{pmatrix}.$$

Each element of the above matrix is in the form of kernel regression. By Lemma 8.1 it holds that

$$\frac{1}{n}\mathbf{D}_{ju}^{*T}\mathbf{W}_{ju}^{*}\mathbf{D}_{ju}^{*} = f_j(u)\boldsymbol{\Gamma}_j(u)\otimes\begin{pmatrix}\mu_0 & \mu_1\\ \mu_1 & \mu_2\end{pmatrix}\cdot O_p\left[1+\left\{\frac{\log(1/h^*)}{nh^*}\right\}^{\frac{1}{2}}\right].$$

Therefore, we have $J_2 = o_p\{h^{*2}+1/\sqrt{nh^*}\}$.

Since the coefficient functions $\alpha_{sj}(u)$ $(s=1,2,\cdots,p_j)$ are smooth in the neighborhood of $|U_{ij}-u|<h$, by Taylor's expansion,

$$\mathbf{X}_{ij}^{T}\boldsymbol{\alpha}_j(U_{ij}) = \mathbf{X}_{ij}^{T}\boldsymbol{\alpha}_j(u)+(U_{ij}-u)\mathbf{X}_{ij}^{T}\boldsymbol{\alpha}_j'(u)+\frac{h^{*2}}{2}\left(\frac{U_{ij}-u}{h^*}\right)^2\mathbf{X}_{ij}^{T}\boldsymbol{\alpha}_j''(u)+o(h^{*2})$$

where $\boldsymbol{\alpha}_j'(u)$ and $\boldsymbol{\alpha}_j''(u)$ are the vectors consisting of the first and the second derivatives of the functions $\alpha_{sj}(\cdot)$. Thus, we have

$$J_3 = \begin{pmatrix}\boldsymbol{\alpha}_j(u)\\ h^*\boldsymbol{\alpha}_j'(u)\end{pmatrix}+\frac{h^{*2}}{2}\left(\mathbf{D}_{ju}^{*T}\mathbf{W}_{ju}^{*}\mathbf{D}_{ju}^{*}\right)^{-1}\mathbf{D}_{ju}^{*T}\mathbf{W}_{ju}^{*}\mathbf{A}_{ju}\mathbf{X}_j^{T}\boldsymbol{\alpha}_j''(u)+o(h^{*2})$$

where

$$\mathbf{A}_{ju} = \operatorname{diag}\left\{\left(\frac{U_{1j}-u}{h^*}\right)^2 \cdots, \left(\frac{U_{nj}-u}{h^*}\right)^2\right\}.$$

By Lemma 8.1

$$\frac{1}{n}\mathbf{D}_{ju}^{*T}\mathbf{W}_u^*\mathbf{A}_{ju}\mathbf{X}_j^T\boldsymbol{\alpha}_j''(u) = f_j(u)\boldsymbol{\Gamma}_j(u)\boldsymbol{\alpha}_j''(u) \otimes [\mu_2, \mu_3]^T\{1 + o_p(1)\}.$$

By applying the fact

$$(\mathbf{A} + a\mathbf{B})^{-1} = \mathbf{A}^{-1} - a\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1} + O(a^2) \quad \text{as } a \to 0$$

to $\left(\mathbf{D}_{ju}^{*T}\mathbf{W}_{ju}^*\mathbf{D}_{ju}^*\right)^{-1}$ we have

$$\frac{h^{*2}}{2}\left(\mathbf{D}_{ju}^{*T}\mathbf{W}_{ju}^*\mathbf{D}_{ju}^*\right)^{-1}\mathbf{D}_{ju}^{*T}\mathbf{W}_{ju}^*\mathbf{A}_{ju}\mathbf{X}_j^T\boldsymbol{\alpha}_j''(u) = \frac{1}{\mu_2 - \mu_1^2}\begin{pmatrix} (\mu_2 - \mu_1\mu_3)\boldsymbol{\alpha}_j''(u) \\ (\mu_3 - \mu_1\mu_2)\boldsymbol{\alpha}_j''(u) \end{pmatrix}\{1 + o_p(1)\}.$$

Further, write

$$J_1 = (\mathbf{D}_{ju}^{*T}\mathbf{W}_{ju}^*\mathbf{D}_{ju}^*)^{-1}J_{12}$$

where $J_{12} = \mathbf{D}_{ju}^{*T}\mathbf{W}_{ju}^*(\boldsymbol{\varepsilon}_j + \sum_{j_1=1}^J \omega_{j_1 j}\boldsymbol{\varepsilon}_{j_1})$. For any $2p_j \times 1$ nonzero vector $\boldsymbol{\iota} = (\iota_1, \ldots, \iota_{2p_j})^T$ it holds

$$\frac{\sqrt{h^*}}{\sqrt{n}}\boldsymbol{\iota}^T J_{12} = \frac{\sqrt{h^*}}{\sqrt{n}}\sum_{i=1}^n\left\{\sum_{s=1}^{p_j}\iota_s X_{ijs}K_{h^*}(U_{ij}-u)\varepsilon_{ij} + \sum_{s=1}^{p_j}\iota_{j+p_j}X_{ijs}K_{h^*}(U_{ij}-u)\left(\frac{U_{ij}-u^*}{h}\right)\varepsilon_{ij}\right\}.$$

It is easy to see that $\sum_{s=1}^{p_j}\iota_s X_{ijs}K_{h^*}(U_{ij}-u)\varepsilon_{ij} + \sum_{s=1}^{p_j}\iota_{j+p_j}X_{ijs}K_{h^*}(U_{ij}-u)\left(\frac{U_{ij}-u^*}{h}\right)\varepsilon_{ij}$ are i.i.d random variables with mean zero. Therefore, we have that

$$\begin{aligned}
\operatorname{Var}\left(\frac{\sqrt{h^*}}{\sqrt{n}}\boldsymbol{\iota}^T J_{12}\right) &= \frac{h^*((\operatorname{diag}(\boldsymbol{\Sigma}^{-\frac{1}{2}}))^{-2})_{jj}}{n}\sum_{i=1}^n K_{h^*}^2(U_{ij}-u)\boldsymbol{\iota}^{*T}\mathbf{X}_{ij}\mathbf{X}_{ij}^T\boldsymbol{\iota}^* \\
&+ \frac{h^*((\operatorname{diag}(\boldsymbol{\Sigma}^{-\frac{1}{2}}))^{-2})_{jj}}{n}\sum_{i=1}^n\left(\frac{U_{ij}-u}{h^*}\right)^2 K_{h^*}^2(U_{ij}-u)\boldsymbol{\iota}^{**T}\mathbf{X}_{ij}\mathbf{X}_{ij}^T\boldsymbol{\iota}^{**} \\
&+ \frac{h^*((\operatorname{diag}(\boldsymbol{\Sigma}^{-\frac{1}{2}}))^{-2})_{jj}}{n}\sum_{i=1}^n\left(\frac{U_{ij}-u}{h^*}\right)K_{h^*}^2(U_{ij}-u)\boldsymbol{\iota}^{*T}\mathbf{X}_{ij}\mathbf{X}_{ij}^T\boldsymbol{\iota}^{**} \\
&+ \frac{h^*((\operatorname{diag}(\boldsymbol{\Sigma}^{-\frac{1}{2}}))^{-2})_{jj}}{n}\sum_{i=1}^n\left(\frac{U_{ij}-u}{h^*}\right)K_{h^*}^2(U_{ij}-u)\boldsymbol{\iota}^{**T}\mathbf{X}_{ij}\mathbf{X}_{ij}^T\boldsymbol{\iota}^* = Q_1 + Q_2 + Q_3 + Q_4, \text{ say}
\end{aligned}$$

where $\boldsymbol{\iota}^*$ is the first p_j components of $\boldsymbol{\iota}$ and $\boldsymbol{\iota}^{**}$ is the last p_j components of $\boldsymbol{\iota}$. It is easy to see that

$$Q_1 \to_p f_j(u)((\operatorname{diag}(\boldsymbol{\Sigma}^{-\frac{1}{2}}))^{-2})_{jj}\nu_0\boldsymbol{\iota}^{*T}\boldsymbol{\Gamma}_j(u)\boldsymbol{\iota}^*, \; Q_2 \to_p f_j(u)((\operatorname{diag}(\boldsymbol{\Sigma}^{-\frac{1}{2}}))^{-2})_{jj}\nu_2\boldsymbol{\iota}^{**T}\boldsymbol{\Gamma}_j(u)\boldsymbol{\iota}^{**}$$

$Q_3 \to_p f_j(u)((\mathrm{diag}(\boldsymbol{\Sigma}^{-\frac{1}{2}}))^{-2})_{jj}\nu_1\boldsymbol{\iota}^{*T}\boldsymbol{\Gamma}_j(u)\boldsymbol{\iota}^{**}$, $Q_4 \to_p f_j(u)((\mathrm{diag}(\boldsymbol{\Sigma}^{-\frac{1}{2}}))^{-2})_{jj}\nu_1\boldsymbol{\iota}^{**T}\boldsymbol{\Gamma}_j(u)\boldsymbol{\iota}^{*}$.

Therefore, by the central limit theorem we have that

$$\frac{\sqrt{nh^*}}{n}J_{12} \xrightarrow{D} N(0, \boldsymbol{\Sigma}_1) \quad \text{as } n \to \infty$$

$$\boldsymbol{\Sigma}_1 = ((\mathrm{diag}(\boldsymbol{\Sigma}^{-\frac{1}{2}}))^{-2})_{jj} f_j(u)\boldsymbol{\Gamma}_j(u) \otimes \begin{pmatrix} \nu_0 & \nu_1 \\ \nu_1 & \nu_2 \end{pmatrix}.$$

Above all, it holds

$$\sqrt{nh^*}\left[\mathbf{H}^*\left\{\hat{\boldsymbol{\Psi}}_j^{TS}(u) - \boldsymbol{\Psi}_j(u)\right\} - \frac{h^{*2}}{2}\frac{1}{\mu_2-\mu_1^2}\begin{pmatrix} (\mu_2-\mu_1\mu_3)\boldsymbol{\alpha}_j''(u) \\ (\mu_3-\mu_1\mu_2)\boldsymbol{\alpha}_j''(u) \end{pmatrix}\right] \xrightarrow{D} N(0, \boldsymbol{\Sigma}_j^{TS})$$

as $n \to \infty$, where

$$\begin{aligned}
\boldsymbol{\Sigma}_j^{TS} &= \left\{f_j(u)\boldsymbol{\Gamma}_j(u) \otimes \begin{pmatrix} \mu_0 & \mu_1 \\ \mu_1 & \mu_2 \end{pmatrix}\right\}^{-1} ((\mathrm{diag}(\boldsymbol{\Sigma}^{-\frac{1}{2}}))^{-2})_{jj} f_j(u)\boldsymbol{\Gamma}_j(u) \\
&\quad \otimes \begin{pmatrix} \nu_0 & \nu_1 \\ \nu_1 & \nu_2 \end{pmatrix}\left\{f_j(u)\boldsymbol{\Gamma}_j(u) \otimes \begin{pmatrix} \mu_0 & \mu_1 \\ \mu_1 & \mu_2 \end{pmatrix}\right\}^{-1} \\
&= \frac{((\mathrm{diag}(\boldsymbol{\Sigma}^{-\frac{1}{2}}))^{-2})_{jj}}{f_j(u)}\boldsymbol{\Gamma}_j^{-1}(u) \otimes \\
&\quad \frac{1}{(\mu_2-\mu_1^2)^2}\begin{pmatrix} \mu_2^2\nu_0 - 2\mu_1\mu_2\nu_1 + \mu_1^2\nu_2 & (\mu_1^2+\mu_2)\nu_1 - \mu_1\mu_2\nu_0 - \mu_1\nu_2 \\ (\mu_1^2+\mu_2)\nu_1 - \mu_1\mu_2\nu_0 - \mu_1\nu_2 & \nu_2 - \mu_1(2\nu_1 + \mu_1\nu_0) \end{pmatrix}.
\end{aligned}$$

This implies that Theorem 3.1 holds.

Proof of Theorem 3.2 Applying Lemma 8.1 the proof of Theorem 3.2 is trivial.

Proof of Theorem 4.1 Under the null hypothesis, it is easy to see that

$$\begin{aligned}
\mathrm{RSS}_0 - \mathrm{RSS}_1 &= (nJ)^{-1}\sum_{j=1}^{J}\sum_{i=1}^{n}\left\{\mathbf{X}_{ij}^T(\mathbf{a}_j(U_{ij}) - \hat{\mathbf{a}}_j(U_{ij}))\right\}^2 \\
&- (nJ)^{-1}\sum_{j=1}^{J}\sum_{i=1}^{n}\left\{\mathbf{X}_{ij}^T(\mathbf{a}_j(U_{ij}) - \hat{\boldsymbol{\alpha}}_j^{TS}(U_{ij}))\right\}^2 + o_p(1) = K_1 + o_p(1) \text{ say.}
\end{aligned}$$

Applying the consistency of $\hat{\boldsymbol{\theta}}_{sj}$ and $\hat{\boldsymbol{\alpha}}_j^{TS}(\cdot)$ we can show that $K_1 \to_p 0$ as $n \to \infty$. This implies $\mathrm{RSS}_0 - \mathrm{RSS}_1 \to_p 0$ under the null hypothesis. Therefore, in order to prove the first result it suffices to show that RSS_1 is bounded away from zero and infinity. According to the definition and Theorem 3 we can show that

$$\mathrm{RSS}_1 \to_p J^{-1} \sum_{j=1}^{J} \sigma_{jj}^2 > 0 \quad \text{as} \quad n \to \infty.$$

The second result can be proved in the same way.

REFERENCES

[1] Brumback, B. and Rice, J. (1998). Smoothing spline models for the analysis of nested and crossed samples of curves (with discussion). *J. Amer. Statist. Assoc.* **93** 961–976.

[2] Cai, Z.; Fan, J. and Yao, Q. (2000). Functional-coefficient regression models for nonlinear time series. *J. Amer. Statist. Assoc.* **95** 941–956.

[3] Carroll, R. J.; Midthune, D.; Freedman, L. S. and Kipnis, V. (2006). Seemingly unrelated measurement error models, with application to nutritional epidemiology. *Biometrics* **62** 75–84.

[4] Fan, J. and Gijbels, I. (1995). Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation. *J. Royal Statist. Soc. B.* **57** 371–394.

[5] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modeling and its Applications.* Chapman and Hall, London.

[6] Fan, J. and Zhang, J. (2000). Two-step estimation of functional linear models with applications to longitudinal data. *J. R. Stat. Soc. Ser. B* **62** 303–322.

[7] Fan, J.; Zhang, C. and Zhang, J. (2001). Generalized likelihood ratio statistics and Wilks phenomenon. *Ann. Statist.* **29** 153–193.

[8] Godfrey, L.G. and Tremayne, A.R. (2005). The wild bootstrap and heteroskedasticity-robust tests for serial correlation in dynamic regression models. *Comput. Statist. Data Anal.* **49** 377–395.

[9] Gray, K.A.; Klebanoff, M.A.; Brock, J.W.; Zhou, H.; Darden, R.; Needham, L. and and Longnecker, M. P. (2005). In Utero exposure to background levels of Polychlorinated Biphenls and cognitive functioning among school-aged children. *Am J Epidemiology.* **162** 17–26.

[10] Hastie, T. J. and Tibshirani, R. (1993). varying-coefficient models. *J. R. Stat. Soc. Ser. B.* **55** 757–796.

[11] He, W. and Lawless, J. F. (2005). Bivariate location-scale models for regression analysis, with applications to lifetime data. *J. R. Stat. Soc. Ser. B.* **67** 63–78.

[12] Hougaard, P. (2000). *Analysis of Multivariate Survival Data.* Springer, New York.

[13] Ioannidis, C. and Peel, D. A. (2005). Testing for market efficiency in gambling markets when the errors are non-normal and heteroskedastic an application of the wild bootstrap. *Econom. Lett.* **87** 221–226.

[14] Kakwani, N. C. (1974). A note on the efficiency of the Zellner's seemingly unrelated regressions estimator. *Ann. Inst. Statist. Math.* **26** 361–362.

[15] Kalbfleisch, J. D. and Prentice, R.L. (2002). *The Statistical Analysis of Failure Time Data, 2nd Edition.* John Wiley, New York.

[16] Kurata, H. (1999). On the efficiencies of several generalized least squares estimators in a seemingly unrelated regression model and a heteroscedastic model. *J. Multivariate Anal.* **70** 86–94.

[17] Lang, S.; Adebayo, S. and Fahremir, L. (2002). Bayesian semiparametric seemingly unrelated regression. In: Härdle, W. and Rönz, B. (eds.): *Proceedings in Computational Statistics*, 195–200, Physika-Verlag, Heidelberg.

[18] Lang, S.; Adebayo, S.; Fahremir, L. and Steiner, W. (2003). Bayesian geoadditive seemingly unrealted regression. *Compu. Statist.*, **18**, 263–292.

[19] Linton, O. B.; Mammen, E.; Lin, X. and Carroll, R. J. (2004). Correlation and marginal longitudinal kernel nonparametric regression. *Lecture Notes in Statist. 179.* Proceedings of the Second Seattle Symposium in Biostatistics. **23–33**. Springer, New York.

[20] Liu, A. (2002). Efficient estimation of two seemingly unrelated regression equations. *J. Multivariate Anal.* **82** 445–456.

[21] Mack, Y. P. and Silverman, B. W. (1982). Weak and strong uniform consistency of kernel regression estimates. *Probab. Theory Related Fields.* **61** 405–415.

[22] Mandy, D. M. and Martins-Filho, C. (1993). Seemingly unrelated regressions under additive heteroscedasticity: theory and share equation applications. *J. Econometrics.* **58** 315–346.

[23] Neudecker, H. and Windmeijer, F. A. G. (1991). R^2 in seemingly unrelated regression equations. *Statist. Neerlandica* **45** 405–411.

[24] Ng, V. M. (2002). Robust Bayesian inference for seemingly unrelated regressions with elliptical errors. *J. Multivariate Anal.* **83** 409–414.

[25] Poirer, D.; Koop, G. and Tobias, J. (2005). Semiparametric Bayesian inference in multiple equation models. *J. Applied Econom.* **20** 723–747.

[26] Ramsay, J. O. and Silverman, B. W. (1997). *Functional Data Analysis.* Springer-Verlag, New York.

[27] Rocke, D. M. (1989). Bootstrap Bartlett adjustment in seemingly unrelated regression. *J. Amer. Statist. Assoc.* **84** 598–601.

[28] Smith, M. and Kohn, R. (2000). Nonparametric seemingly unrelated regression. *J. Econom.* **98** 257–281.

[29] Srivastava, V. K. and Giles, D. E. A. (1987). *Seemingly unrelated regression equations models. Estimation and Inference.* CRC Press, New York,

[30] Stute, W.; Gonzalez M. W. and Presedo, Q. M. (1998). Bootstrap approximations in model checks for regression. *J. Amer. Statist. Assoc.* **93** 141–149.

[31] Tong, H. (1990). *Nonlinear Time Series: A Dynamical System Approach.* Oxford University Press, Oxford.

[32] Wang, N. (2003). Marginal nonparametric kernel regression accounting for within-subject correlation. *Biometrika* **90** 43–52.

[33] Wang, Y. D.; Guo, W. S. and Brown, B. (2000). Spline smoothing for bivariate data with application between hormones. *Statistica Sinica.* **10** 377–397.

[34] Welsh, A. H. and Yee, T. W. (2005). Local regression for vector responses. *J. Stat. Plann. Infer.* **135** 307–331.

[35] Wu, C. O.; Chiang, C. T. and Hoover, D. R. (1998). Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data. *J. Amer. Statist. Assoc.* **93** 1388–1402.

[36] Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *J. Amer. Statist. Assoc.* **57** 348–368.

[37] Zellner, A. (1963). Estimators for seemingly unrelated equations: some exact finite sample results. *J. Amer. Statist. Assoc.* **58** 977–992.