Econometric Theory, **30**, 2014, 407–435.
doi:10.1017/S0266466613000352

ASYMPTOTIC THEORY IN FIXED EFFECTS PANEL DATA SEEMINGLY UNRELATED PARTIALLY LINEAR REGRESSION MODELS

JINHONG YOU
Shanghai University of Finance and Economics and Key Laboratory of Mathematical Economics

XIAN ZHOU
Macquarie University

This paper deals with statistical inference for the fixed effects panel data seemingly unrelated partially linear regression model. The model naturally extends the traditional fixed effects panel data regression model to allow for semiparametric effects. Multiple regression equations are permitted, and the model includes the aggregated partially linear model as a special case. A weighted profile least squares estimator for the parametric components is proposed and shown to be asymptotically more efficient than those neglecting the contemporaneous correlation. Furthermore, a weighted two-stage estimator for the nonparametric components is also devised and shown to be asymptotically more efficient than those based on individual regression equations. The asymptotic normality is established for estimators of both parametric and nonparametric components. The finite-sample performance of the proposed methods is evaluated by simulation studies.

1. INTRODUCTION

Econometric models often involve jointly estimated equation systems. Examples include demand equations across different sectors and industries or regions, cost share equations for a translog cost function, etc. (see Srivastava and Giles, 1987). The seemingly unrelated (SU) regression estimation approach proposed by Zellner (1962) uses the correlation between disturbances in different equations to improve the efficiency of joint estimation. There are many applications of the SU regression procedure in both time series and cross section data areas—see, e.g., Zellner (1962, 1963), Kakwani (1974), Gallant (1975), Ghazal (1986),

You's research was supported by grants from the National Natural Science Foundation of China (NSFC) (11071154). In addition, the research is partially supported by leading Academic Discipline Program, 211 Project for Shanghai University of Finance and Economics (the third phase) and project B803. We are truly grateful to Professor Piet de Jong of Macquarie University for carefully editing this paper. Address correspondence to Jinhong You, School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai 200433, P. R. China. e-mail: jinhongyou@gmail.com.

© Cambridge University Press 2013

https://doi.org/10.1017/S0266466613000352 Published online by Cambridge University Press

Robinson (1988), Rocke (1989), Neudecker and Windmeijer (1991), Mandy and Martins-Filho (1993), Kurata (1999), Smith and Kohn (2000), Wang, Guo, and Brown (2000), Liu (2002), Ng (2002), Koop, Poirer, and Tobias (2005), and Welsh and Yee (2006).

The analysis of panel data has attracted considerable research interest in economics. Examples of panel data are abundant in a number of fields; see Baltagi (1995) for many interesting examples. According to Baltagi (1995), the widespread attention enjoyed by panel data models is amply justified by many attractive features they offer, most notably their allowance for observation heterogeneity, permitting more efficient estimation of parameters. Avery (1977) proposed a panel data SU regression model combining the SU regression approach with panel data modeling. The combined model was motivated by an application of estimating a system of consumer demand equations on a set of panel data, to take the advantages of the SU regression approach and panel data modeling. Many authors have subsequently studied the statistical inference and applications of the panel data SU regression model. Baltagi (1980) took into account the contemporaneous correlation among equations in addition to the one-way error structure to construct efficient estimators for the regression coefficients of the panel data SU regression model. Verbon (1980) applied this model to a set of four labor demand equations with data from the Netherlands on 18 industries over 10 semiannual periods during 1972–1979. He extended the error component specification to allow for heteroskedasticity and applied the Lagrange multiplier test of Breusch and Pagan (1979) to test for heteroskedasticity. Beierlein, Dunn, and McConnon (1981) used the panel data SU regression model with a two-way error component disturbance to estimate the demand for electricity and natural gas in the northeastern United States. Wan, Griffiths, and Anderson (1992) applied this model with heteroskedastic disturbances to estimate the rice, maize, and wheat production in China. Carroll, Dong, Larry, and Victor (2006) applied this model to analyze the nutritional data with measurement errors on some of the covariates. Further applications and references can be found in the comprehensive monograph of Baltagi (1995).

Most estimation results on panel data SU regression models are established in the setting of parametric regressions, mainly linear regression. Parametric specifications are often subject to significant modeling biases. Zhou, Xu, and You (2011) studied the estimation of nonparametric panel data SU regressions and proposed efficient weighted estimation. Nonparametric regression models do not require detailed information on model structure and often provide useful insight for further parametric fitting.

Nonparametric methods, however, can have serious drawbacks, including the "curse of dimensionality," lack of extrapolation capability, and difficulty of interpretation. The panel data SU nonparametric regression model shares these drawbacks including the curse of dimensionality when the covariates are multidimensional, causing the optimal rate of convergence to decrease with an increasing dimensionality of covariates (see Stone, 1986). In addition, it is

https://doi.org/10.1017/S0266466613000352 Published online by Cambridge University Press

difficult to describe, interpret, and understand the estimated regression surface when the dimension exceeds two.

To circumvent these shortcomings, this paper proposes a panel data SU semiparametric regression model, namely, the *panel data SU partially linear regression model*:

$$Y_{sit} = \mu_{si} + \mathbf{X}_{sit}^{\tau}\boldsymbol{\beta}_s + g_s(U_{sit}) + \nu_{sit}, \quad s = 1, \ldots, M, \quad i = 1, \ldots, k, \quad t = 1, \ldots, T, \tag{1.1}$$

where Y_{sit} is the measurement of the sth response variable on the ith individual at the tth time point, $(\mathbf{X}_{sit}^{\tau}, U_{sit})$ are the corresponding explanatory variables, $\boldsymbol{\beta}_s$ is an unknown $p_s \times 1$ parameter vector, $g_s(\cdot)$ is an unknown function, μ_{si} is the individual effect of the ith individual about the sth response variable, ν_{sit} is the random error, and the superscript τ is the notation for transpose.

Panel data models are often divided into two classes based on the relationship between the explanatory variables and individual effects. First is the random effects model where individual effects are random, the explanatory variables and individual effects are uncorrelated, and individual effects are random errors. Second is the fixed effects model where individual effects are fixed or correlated with the explanatory variables and treated as unknown nuisances. The fixed effects specification has the advantage of robustness compared to the random effects specification (e.g., Baltagi, 1995; Entorf, 1997; Hahn, 1999; Abrevaya, 1999; Ahn and Schmidt, 2000; Hahn, 2001; Horowitz and Lee, 2004). However, the analysis of the fixed effects panel data is more challenging because of increasing number of parameters with the sample size, yielding the famous Neyman and Scott problem (cf. Neyman and Scott, 1948). In this paper, we allow μ_{si} to correlate with $(\mathbf{X}_{sit}^{\tau}, U_{sit})^{\tau}$ with an unknown correlation structure. Hence (1.1) is assumed a fixed effects model.

Semiparametric regression helps solve complex scientific problems. Model (1.1) includes many well-known regression models. When $M = 1$, the model reduces to the partially linear in-slide model, independently proposed by Baltagi and Li (2002) and Fan, Peng, and Huang (2005) and applied, by the latter, to the normalization of microarray data. When $\mu_{s_1 i} = \mu_{s_2 i}$ and $\mathbb{E}(\nu_{s_1 it}\nu_{s_2 it}) \equiv 0$ for all $1 \le s_1 \ne s_2 \le M$, model (1.1) reduces to the aggregated partially linear in-slide model, which was studied by Fan et al. (2005). When $\mu_{si} \equiv 0$ for all $s = 1, \ldots, M$ and $T = 1$, model (1.1) becomes the SU partially linear model mentioned by Robinson (1988). When $g_s(\cdot) \equiv 0$ for all $s = 1, \ldots, M$, i.e., the nonlinear components are removed, model (1.1) becomes the well-known panel data SU linear model, which has been widely applied in economics; see the literature mentioned earlier.

Throughout this paper assume, in (1.1), that k is large and T and M are small. This is typical in labor or consumer panel data studies (see, e.g., Baltagi, 1995). Because the number of nuisance parameters μ_{si} $(s = 1, \ldots, M, i = 1, \ldots, k)$ grows with k, estimating the parametric and nonparametric components requires care. In the context of $\mu_{s_1 i} = \mu_{s_2 i}$ and $\mathbb{E}(\nu_{s_1 it}\nu_{s_2 it}) \equiv 0$ for all $1 \le s_1 \ne s_2 \le M$

https://doi.org/10.1017/S0266466613000352 Published online by Cambridge University Press

(the so-called aggregated partially linear in-slide model), Fan et al. (2005) proposed an aggregated profile least squares estimator for the parametric components by combining the local polynomial with least squares techniques and established their asymptotic normality. No results are available on the nonparametric components other than some description of their estimation. This makes it difficult to select bandwidth and hinders statistical inference for the nonparametric components.

Open questions in the setting of model (1.1) include (i) whether the contemporaneous correlation among equations should be taken into account to construct an efficient estimator for the parametric components (ii) whether it is possible to construct an estimator for the nonparametric components with easily derived asymptotic properties, and (iii) whether the nonparametric component estimator should take into account the contemporaneous correlation among equations to improve the efficiency. These problems are addressed in the following sections. Moreover, a weighted two-stage estimator for the nonparametric components is proposed, and its asymptotic normality is derived. Taking into account the contemporaneous correlation among equations is shown to increase the efficiencies of the estimators for both parametric and nonparametric components.

The construction of efficient estimators of the parametric and nonparametric components relies on the estimates of the contemporaneous correlation among equations of model (1.1), which in turn requires pilot estimates of the parametric and nonparametric components. We take the profile least squares estimators proposed by Fan et al. (2005) as our pilot estimators. For ease of understanding, the profile least squares estimators are described in Section 2. Section 3 investigates the estimating problems of the error variance and contemporaneous correlation. A weighted profile least squares estimator for the parametric components is proposed in Section 4, and its asymptotic normality is derived. Section 5 describes a weighted two-stage estimator for the nonparametric components and provides the corresponding asymptotic properties. Simulation studies are reported in Section 6, followed by a concluding discussions in Section 7. Proofs of the main results are presented in the Appendix.

2. PROFILE LEAST SQUARES ESTIMATION

The profile least squares (PLS) estimators proposed by Fan et al. (2005) are taken as our pilot estimators of the parametric and nonparametric components. Based on a random sample $\{Y_{sit}, \mathbf{X}_{sit}, U_{sit} : s = 1, \ldots, M; i = 1, \ldots, k; t = 1, \ldots, T\}$ from model (1.1), the PLS estimation proceeds as follows. For fixed s write (1.1) as

$$Y_{sit} - \mu_{si} - \mathbf{X}_{sit}^{\tau}\boldsymbol{\beta}_s = g_s(U_{sit}) + v_{sit}, \quad i = 1, \ldots, k, \quad s = 1, \ldots, T. \tag{2.1}$$

This transforms the partially linear model into the usual nonparametric model. Applying the well-known local linear regression technique in a small neighborhood of u, $g_s(U_{sit})$ can be locally approximated by a linear function

https://doi.org/10.1017/S0266466613000352 Published online by Cambridge University Press

$g_s(U_{sit}) \approx g_s(u) + g_s^{(\prime)}(u)(u - U_{sit}) \equiv a_s + b_s(u - U_{sit})$, where $g_s^{(\prime)}(u) = \partial g_s(\cdot)/\partial u$. This leads to local least squares to find $\{(a_s, b_s)\}$ that minimize

$$\sum_{i=1}^{k}\sum_{t=1}^{T}\left[Y_{sit} - \mu_{si} - \mathbf{X}_{sit}^{\tau}\boldsymbol{\beta}_s - a_s - b_s(U_{sit} - u)\right]^2 K_{h_s}(U_{sit} - u), \tag{2.2}$$

where $K_{h_s}(\cdot) = K(\cdot/h_s)/h_s$ with a kernel function $K(\cdot)$ and bandwidth h_s. The solution is given by $\left(\hat{a}_s(u), h_s\hat{b}_s(u)\right)^{\tau} = \left(\mathbf{D}_{su}^{\tau}\mathbf{W}_{su}\mathbf{D}_{su}\right)^{-1}\mathbf{D}_{su}^{\tau}\mathbf{W}_{su}\left(\mathbf{Y}_s - \mathbf{B}\boldsymbol{\mu}_s - \mathbf{X}_s\boldsymbol{\beta}_s\right)$, where $\mathbf{B} = \mathbf{I}_k \otimes \mathbf{1}_T$ with $\otimes$ the Kronecker product,

$$\mathbf{D}_{su} = \begin{pmatrix} 1 & \cdots & 1 & \cdots & 1 \\ \dfrac{U_{s11} - u}{h_s} & \cdots & \dfrac{U_{s1T} - u}{h_s} & \cdots & \dfrac{U_{skT} - u}{h_s} \end{pmatrix}^{\tau},$$

and

$$\mathbf{W}_{su} = \operatorname{diag}(K_{h_s}(U_{s11} - u), \ldots, K_{h_s}(U_{s1T} - u), \ldots, K_{h_s}(U_{skT} - u)).$$

Substituting $\hat{a}_s(u)$ into model (2.1) yields the following estimator of $\boldsymbol{\beta}$:

$$\hat{\boldsymbol{\beta}}_s = \left\{\hat{\mathbf{X}}_s^{\tau}(\mathbf{I}_{kT} - \mathbf{P}_{\hat{\mathbf{B}}_s})\hat{\mathbf{X}}_s\right\}^{-1}\hat{\mathbf{X}}_s^{\tau}(\mathbf{I}_{kT} - \mathbf{P}_{\hat{\mathbf{B}}_s})\hat{\mathbf{Y}}_s,$$

where $\hat{\mathbf{X}}_s = (\hat{\mathbf{X}}_{s11}, \ldots, \hat{\mathbf{X}}_{s1T}, \ldots, \hat{\mathbf{X}}_{skT})^{\tau} = (\mathbf{I}_{kT} - \mathbf{S}_s)\mathbf{X}_s$, $\hat{\mathbf{B}}_s = \left(\hat{\mathbf{B}}_{s11}, \ldots, \hat{\mathbf{B}}_{s1T}, \ldots, \hat{\mathbf{B}}_{skT}\right)^{\tau} = \left(\mathbf{I}_{kT} - \mathbf{S}_s\right)\mathbf{B}$, $\hat{\mathbf{Y}}_s = \left(\hat{Y}_{s1j}, \ldots, \hat{Y}_{s1T}, \ldots, \hat{Y}_{skT}\right)^{\tau} = \left(\mathbf{I}_{kT} - \mathbf{S}_s\right)\mathbf{Y}_s$, $\mathbf{P}_{\hat{\mathbf{B}}_s} = \hat{\mathbf{B}}_s\left(\hat{\mathbf{B}}_s^{\tau}\hat{\mathbf{B}}_s\right)^{-1}\hat{\mathbf{B}}_s^{\tau}$, $\mathbf{X}_s = \left(\mathbf{X}_{s11}, \ldots, \mathbf{X}_{s1T}, \ldots, \mathbf{X}_{skT}\right)^{\tau}$, $\mathbf{Y}_s = \left(Y_{s1j}, \ldots, Y_{s1T}, \ldots, Y_{skT}\right)^{\tau}$, and

$$\mathbf{S}_s = \begin{pmatrix} (1\,0)\left(\mathbf{D}_{sU_{s11}}^{\tau}\mathbf{W}_{sU_{s11}}\mathbf{D}_{sU_{s11}}\right)^{-1}\mathbf{D}_{sU_{s11}}^{\tau}\mathbf{W}_{sU_{s11}} \\ \vdots \\ (1\,0)\left(\mathbf{D}_{sU_{s1T}}^{\tau}\mathbf{W}_{sU_{s1T}}\mathbf{D}_{sU_{s1T}}\right)^{-1}\mathbf{D}_{sU_{s1T}}^{\tau}\mathbf{W}_{sU_{s1T}} \\ \vdots \\ (1\,0)\left(\mathbf{D}_{sU_{skT}}^{\tau}\mathbf{W}_{sU_{skT}}\mathbf{D}_{sU_{skT}}\right)^{-1}\mathbf{D}_{sU_{skT}}^{\tau}\mathbf{W}_{sU_{skT}} \end{pmatrix}.$$

Further, the nonparametric estimator of $g_s(\cdot)$ has the form

$$\begin{aligned}\hat{\mathbf{G}}_s &= \left(\hat{g}_s(U_{s11}), \ldots, \hat{g}_s(U_{s1T}), \ldots, \hat{g}_s(U_{skT})\right)^{\tau} \\ &= \left\{\mathbf{I}_{kT} - (\mathbf{I}_{kT} - \mathbf{S}_s\overline{\mathbf{P}}_{\mathbf{B}})^{-1}(\mathbf{I}_{kT} - \mathbf{S}_s)\right\}\left(\mathbf{Y}_s - \mathbf{X}_s\hat{\boldsymbol{\beta}}_s\right),\end{aligned}$$

where $\overline{\mathbf{P}}_{\mathbf{B}} = \mathbf{B}(\mathbf{I}_{kT} - \boldsymbol{\iota}_{kT}\boldsymbol{\iota}_{kT}^{\tau}/kT)(\mathbf{B}^{\tau}\mathbf{B})^{-1}\mathbf{B}^{\tau}$ with $\mathbf{B}$ defined as before and $\boldsymbol{\iota}_{kT}$ is the kT-vector with all elements equal to 1.

Let $\mathbf{h}_{st}(u) = \mathbb{E}(\mathbf{X}_{sit}|U_{sit} = u)$, $\boldsymbol{\Pi}_{sit} = \mathbf{X}_{sit} - \mathbf{h}(U_{sit})$, $\mathbf{P}_{\mathbf{B}} = \mathbf{B}(\mathbf{B}^{\tau}\mathbf{B})^{-1}\mathbf{B}^{\tau}$, and

$$\mathbf{D}_{ss} = \mathbb{E}\left(\sum_{t=1}^{T}\boldsymbol{\Pi}_{sit}\boldsymbol{\Pi}_{sit}^{\tau}\right) - \frac{1}{T}\mathbb{E}\left(\sum_{t_1=1}^{T}\boldsymbol{\Pi}_{sit_1}\sum_{t_2=1}^{T}\boldsymbol{\Pi}_{sit_2}^{\tau}\right).$$

https://doi.org/10.1017/S0266466613000352 Published online by Cambridge University Press

The asymptotic properties of $\hat{\boldsymbol{\beta}}_s$ and $\hat{\mathbf{G}}_s$ are derived under the following assumptions.

Assumption 1. For fixed s, the function, $\mathrm{g}_s(\cdot)$ has bounded second derivatives.

Assumption 2. For fixed s, $\left(\mathbf{X}_{si1}^{\tau}, \ldots, \mathbf{X}_{siT}^{\tau}, U_{si1}, \ldots, U_{siT}\right)^{\tau}$ are independent and identically distributed (i.i.d.) over $i = 1, \ldots, k$. In addition, $\mathbf{D}_{ss}$ is nonsingular, and each component of $\mathbf{h}(u)$ satisfies the Lipschitz condition in u.

Assumption 3. For fixed s, the random variable U_{sit} has a bounded support $\mathcal{U}_s$. Its density function $f_s(\cdot)$ is Lipschitz continuous and bounded away from 0 on $\mathcal{U}_s$.

Assumption 4. The kernel function $K(\cdot)$ is a density function with a compact support.

Assumption 5. For fixed s, the bandwidth h_s satisfies $kh_s^8 \to 0$ and $kh_s^2/(\log k)^2 \to \infty$ as $k \to \infty$.

Assumption 6. For fixed s, each component of $\left(\mathbf{X}_{si1}^{\tau}, \ldots, \mathbf{X}_{siT}^{\tau}\right)^{\tau}$ is bounded in probability.

Assumption 7. For fixed s, ν_{sit} are i.i.d. over $i = 1, \ldots, k$ and $t = 1, \ldots, T$ with mean zero and variance $\sigma_{\nu ss}^2$. In addition, $\sigma_{\nu s_1 s_2}^2 = \mathbb{E}\left(\nu_{s_1 11}\nu_{s_2 11}\right) \neq 0$ and $\mathbb{E}\left(\nu_{s_1 11}^2\nu_{s_2 11}^2\right) < \infty$ for $s_1, s_2 = 1, \ldots, M$. Further, $\nu_{s_1 \mathrm{it}}$ are independent of $\left(\mathbf{X}_{s_2 \mathrm{it}}^{\tau}, \mathrm{U}_{s_2 \mathrm{it}}\right)^{\tau}$ for all $s_1, s_2 = 1, \ldots, M$, $i = 1, \ldots, k$ and $t = 1, \ldots, T$, and each μ_{sit} has a zero mean and finite variance.

Remark 1. Assumptions 1–7 are not the weakest possible conditions, but they are imposed to facilitate proofs. These assumptions were also made by Fan et al. (2005). Assumption 5 gives a range of the bandwidth h_j. Obviously, the usual optimal bandwidth of order $\mathrm{k}^{-1/5}$ falls into this range. Assumption 6 is needed to prove $k^{-1}\hat{\mathbf{X}}_s^{\tau}\left(\mathbf{I}_{kT} - \mathbf{P}_{\hat{\mathbf{B}}_s}\right)\hat{\mathbf{X}}_s \to_p \mathbf{D}_{ss}$ as $k \to \infty$; see the proof of Lemma A.2 in Fan et al. (2005). Assumption 7 shows the usual restrictions for random errors and identifiability conditions in model (1.1). These identifiability conditions are the same as those in Baltagi and Li (2002), Fan et al. (2005), Su and Ullah (2006), and Henderson, Carroll, and Li (2008). Note that our results could be extended to the scenario that ν_{sit}, and $\left(\mathbf{X}_{s_2 it}^{\tau}, \mathrm{U}_{s_2 it}\right)^{\tau}$ are mean independent in the sense that $\mathbb{E}\left[\nu_{s_1 it}|(\mathbf{X}_{s_2 it}^{\tau}, \mathrm{U}_{s_2 it})^{\tau}\right] = 0$.

The following asymptotic properties of $\hat{\boldsymbol{\beta}}_s$ and $\hat{\mathbf{G}}_s$ are established by Fan et al. (2005).

LEMMA 1. *Under Assumptions 1–7, we have* $\sqrt{k}\left(\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s\right) \xrightarrow{D} N\left(\mathbf{0}, \mathbf{D}_{ss}^{-1}\right)$ *as* $k \to \infty$ *and*

$$\mathbf{MSE}\left\{\hat{g}_s(u)|U_{s11}, \ldots, U_{s1T}, \ldots, U_{skT}\right\}$$
$$= \frac{1}{kT}\sum_{i=1}^{k}\sum_{t=1}^{T}\mathbb{E}\left\{\left(\hat{g}_s(U_{sit}) - g_s(U_{sit})\right)^2|U_{s11}, \ldots, U_{s1T}, \ldots, U_{skT}\right\}$$

https://doi.org/10.1017/S0266466613000352 Published online by Cambridge University Press

$$\leq \sqrt{T}\left(\sqrt{T}-1\right)^2 \left(\frac{\left(\alpha_2^2-\alpha_1\alpha_3\right)h_s^4}{4(\alpha_2-\alpha_1^2)}\mathbb{E}[g_s''(U_{sit})]^2 \right.$$
$$\left. \times\left\{ \frac{(T-1)\sigma_{\nu ss}^2|\mathcal{U}_s|}{T^2}+e\frac{(T-1)^2\sigma_{\nu ss}^2}{T^2} \right\}\frac{\delta_0}{kTh_s} \right)$$
$$+ o_p\left(h_s^4+\frac{1}{kTh_s}\right),$$

where $\mathbf{D}_{ss}$ *is defined in Assumption 2,* $\alpha_i=\int u^i K(u)du$, $\delta_i=\int u^i K^2(u)du$, *and* $|\mathcal{U}_s|$ *is the length of* $\mathcal{U}_s$.

Because $\hat{\boldsymbol{\beta}}_s$ and $\hat{\mathbf{G}}_s$ do not take into account the contemporaneous correlation among equations, they may not be asymptotically efficient. By Lemma 1, however, they are consistent estimators of $\boldsymbol{\beta}_s$ and $\mathbf{G}_s$. Therefore, we can use $\hat{\boldsymbol{\beta}}_s$ and $\hat{\mathbf{G}}_s$ as a basis to estimate residuals and then the contemporaneous correlation and error variances. These will be shown in the next section.

3. ESTIMATING ERROR VARIANCE AND CONTEMPORANEOUS CORRELATION

The error variance is $\sigma_{\nu s_1 s_1}^2$, and the contemporaneous correlation is given by $\sigma_{\nu s_1 s_2}^2/(\sigma_{\nu s_1 s_1}\sigma_{\nu s_2 s_2})$ for $s_1\neq s_2$. Consistent estimation of $\sigma_{\nu s_1 s_2}^2$ is required for the construction of efficient estimators of the parametric and nonparametric components, the design of model-based tests, and the selection of the smoothing parameters. This section outlines such consistent estimation.

Let $\boldsymbol{\varepsilon}_s=\left(\varepsilon_{s1},\ldots,\varepsilon_{s1T},\ldots,\varepsilon_{skT}\right)^\tau$. Because $\mathbb{E}\left[\boldsymbol{\varepsilon}_{s_1}^\tau(\mathbf{I}_{kT}-\mathbf{P}_\mathbf{B})\boldsymbol{\varepsilon}_{s_2}\right]=(T-1)/T\sigma_{\nu s_1 s_2}^2$, a natural estimator of $\sigma_{\nu s_1 s_2}^2$ is

$$\hat{\sigma}_{\nu s_1 s_2}^2=\frac{1}{k(T-1)}\left(\hat{\mathbf{Y}}_{s_1}-\hat{\mathbf{X}}_{s_1}\hat{\boldsymbol{\beta}}_{s_1}-\hat{\mathbf{G}}_{s_1}\right)^\tau(\mathbf{I}_n-\mathbf{P}_\mathbf{B})\left(\hat{\mathbf{Y}}_{s_2}-\hat{\mathbf{X}}_{s_2}\hat{\boldsymbol{\beta}}_{s_2}-\hat{\mathbf{G}}_{s_2}\right),$$

where $\mathbf{P}_\mathbf{B}=\mathbf{B}(\mathbf{B}^\tau\mathbf{B})^{-1}\mathbf{B}$ with $\mathbf{B}=\mathbf{I}_k\otimes\mathbf{1}_T$.

The following theorem summarizes the asymptotic normality of $\hat{\sigma}_{\nu s_1 s_2}^2$.

THEOREM 1. *Under Assumptions 1–7,*

$$\sqrt{k}\left(\hat{\sigma}_{\nu s_1 s_2}^2-\sigma_{\nu s_1 s_2}^2\right)\xrightarrow{D}N\left(0,\kappa_{s_1 s_2}\right)\quad as\quad k\to\infty,$$

where

$$\kappa_{s_1 s_2}=\frac{1}{T}\left\{\mathbb{E}\left(\nu_{s_1 11}^2\nu_{s_2 11}^2\right)+\frac{3-T}{T-1}\sigma_{\nu s_1 s_1}^2\sigma_{\nu s_2 s_2}^2\right\}.$$

To apply Theorem 1 for making statistical inference for $\sigma_{\nu s_1 s_2}^2$, a consistent estimator of $\kappa_{s_1 s_2}$ is required. This is given by

https://doi.org/10.1017/S0266466613000352 Published online by Cambridge University Press

$$\hat{\kappa}_{s_1s_2} = \left(1+\frac{6}{T^2}-\frac{3}{T^3}-\frac{4}{T}\right)^{-1} \times \left[\frac{1}{kT}\sum_{i=1}^{k}\sum_{t=1}^{T}\hat{\psi}_{s_1it}^2\hat{\psi}_{s_2it}^2 - \frac{T-1}{T^2}\left(6-\frac{9}{T}\right)\hat{\sigma}_{\nu s_1s_1}^2\hat{\sigma}_{\nu s_2s_2}^2\right] + \frac{3-T}{T-1}\hat{\sigma}_{\nu s_1s_1}^2\hat{\sigma}_{\nu s_2s_2}^2,$$

where $\left(\hat{\psi}_{s11},\ldots,\hat{\psi}_{s1T},\ldots,\hat{\psi}_{skT}\right)^\tau = \hat{\Psi}_s = \left(\mathbf{I}_{kT}-\mathbf{P}_{\mathbf{B}}\right)\left(\mathbf{Y}_s-\mathbf{X}_s\hat{\boldsymbol{\beta}}_s-\hat{\mathbf{G}}_s\right)$.

The next theorem shows the consistency of $\hat{\kappa}_{s_1s_2}$.

THEOREM 2. *Under Assumptions 1–7,* $\hat{\kappa}_{s_1s_2} \to_p \kappa_{s_1s_2}$ *as* $k \to \infty$.

4. WEIGHTED PROFILE LEAST SQUARES ESTIMATION

The contemporaneous correlation among the equations is crucial to the structure of SU regression models and the form of the associated disturbance variance-covariance matrix. It allows the model to contain more information than individual equations taken separately, and it is required for efficient statistical inference on the parameters in the SU regression model (see, e.g., Srivastava and Giles, 1987). We now consider weighted profile least squares (WPLS) estimation using the estimated contemporaneous correlation among the equations.

Write $\boldsymbol{\nu} = \left(\boldsymbol{\nu}_1^\tau,\ldots,\boldsymbol{\nu}_M^\tau\right)^\tau$ and $\Omega = \left(\sigma_{\nu s_1s_2}^2\right)_{s_1,s_2=1}^M$. Then

$$\mathrm{Cov}(\boldsymbol{\nu}) = \mathrm{Cov}\left\{\left(\boldsymbol{\nu}_1^\tau,\ldots,\boldsymbol{\nu}_M^\tau\right)^\tau\right\} = \left(\sigma_{\nu s_1s_2}^2\right)_{s_1,s_2=1}^M \otimes \mathbf{I}_{kT} = \Omega\otimes\mathbf{I}_{kT}, \quad \text{say.}$$

The following weighted profile least square estimator of $\boldsymbol{\beta} = \left(\boldsymbol{\beta}_1^\tau,\ldots,\boldsymbol{\beta}_M^\tau\right)^\tau$ is proposed:

$$\begin{aligned}\hat{\boldsymbol{\beta}}^W &= \left(\hat{\boldsymbol{\beta}}_1^{W\tau},\ldots,\hat{\boldsymbol{\beta}}_M^{W\tau}\right)^\tau \\ &= \left[\hat{\mathbf{X}}^\tau\left(\mathbf{I}_{kTM}-\mathbf{P}_{\hat{\mathbf{B}}}\right)\left(\hat{\Omega}^{-1}\otimes\mathbf{I}_{kT}\right)\left(\mathbf{I}_{kTM}-\mathbf{P}_{\hat{\mathbf{B}}}\right)\hat{\mathbf{X}}\right]^{-1} \\ &\quad\times\hat{\mathbf{X}}^\tau\left(\mathbf{I}_{kTM}-\mathbf{P}_{\hat{\mathbf{B}}}\right)\left(\hat{\Omega}^{-1}\otimes\mathbf{I}_{kT}\right)\left(\mathbf{I}_{kTM}-\mathbf{P}_{\hat{\mathbf{B}}}\right)\hat{\mathbf{Y}},\end{aligned}$$

where $\mathbf{P}_{\hat{\mathbf{B}}} = \mathrm{blockdiag}\left(\hat{\mathbf{B}}_1\left(\hat{\mathbf{B}}_1^\tau\hat{\mathbf{B}}_1\right)^{-1}\hat{\mathbf{B}}_1^\tau,\ldots,\hat{\mathbf{B}}_M\left(\hat{\mathbf{B}}_M^\tau\hat{\mathbf{B}}_M\right)^{-1}\hat{\mathbf{B}}_M^\tau\right)$, $\hat{\mathbf{X}} = \mathrm{blockdiag}\left(\hat{\mathbf{X}}_1,\ldots,\hat{\mathbf{X}}_M\right)$, $\hat{\Omega} = \left(\hat{\sigma}_{\nu s_1s_2}^2\right)_{s_1,s_2=1}^M$, $\hat{\sigma}_{\nu s_1s_2}^2$ is the estimator of $\sigma_{\nu s_1s_2}^2$ proposed in the last section, $\hat{\mathbf{Y}} = \left(\hat{\mathbf{Y}}_1^\tau,\ldots,\hat{\mathbf{Y}}_M^\tau\right)^\tau$, and $\hat{\mathbf{X}}_s$, $\hat{\mathbf{Y}}_s$, and $\hat{\mathbf{B}}_s$ are defined in Section 2.

Let $\Sigma_1 = \left(\sigma_\nu^{2s_1s_2}\mathbf{D}_{s_1s_2}\right)_{s_1,s_2=1}^M$, where $\sigma_\nu^{2s_1s_2}$ is the (s_1,s_2) entry of Ω^{-1} and

$$\mathbf{D}_{s_1s_2} = \mathbb{E}\left(\sum_{t=1}^{T}\boldsymbol{\Pi}_{s_1it}\boldsymbol{\Pi}_{s_2it}^\tau\right) - \frac{1}{T}\mathbb{E}\left(\sum_{t_1=1}^{T}\boldsymbol{\Pi}_{s_1it_1}\sum_{t_2=1}^{T}\boldsymbol{\Pi}_{s_2it_2}^\tau\right).$$

https://doi.org/10.1017/S0266466613000352 Published online by Cambridge University Press

THEOREM 3. *Under Assumptions 1–7,*

$$\sqrt{k}\left(\hat{\boldsymbol{\beta}}^{W}-\boldsymbol{\beta}\right)\xrightarrow{D}N\left(\mathbf{0},\Sigma_1^{-1}\right)\quad as\ k\to\infty.$$

Let $\hat{\boldsymbol{\beta}}=\left(\hat{\boldsymbol{\beta}}_1^{\tau},\ldots,\hat{\boldsymbol{\beta}}_M^{\tau}\right)^{\tau}$, where $\hat{\boldsymbol{\beta}}_j$'s are defined in Section 2. By the same argument as proving the first result of Theorem 2 in Fan et al. (2005), we have the asymptotic covariance matrix of $\sqrt{k}\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)$ as

$$\Sigma_2=\text{blockdiag}\left(\mathbf{D}_{11}^{-1},\ldots,\mathbf{D}_{MM}^{-1}\right)\Sigma_3\text{blockdiag}\left(\mathbf{D}_{11}^{-1},\ldots,\mathbf{D}_{MM}^{-1}\right),$$

where $\Sigma_3=\left(\sigma_{\nu s_1 s_2}^2\mathbf{D}_{s_1 s_2}\right)_{s_1,s_2=1}^{M}$. According to Lemma 2 in the Appendix,

$$\lim_{k\to\infty}k\left\{\hat{\mathbf{X}}^{\tau}\left(\mathbf{I}_{kTM}-\mathbf{P}_{\hat{\mathbf{B}}}\right)\left(\Omega^{-1}\otimes\mathbf{I}_{kT}\right)\left(\mathbf{I}_{kTM}-\mathbf{P}_{\hat{\mathbf{B}}}\right)\hat{\mathbf{X}}\right\}^{-1}=\Sigma_1^{-1}.$$

Similarly,

$$\lim_{k\to\infty}k\left\{\hat{\mathbf{X}}^{\tau}\left(\mathbf{I}_{kTM}-\mathbf{P}_{\hat{\mathbf{B}}}\right)\hat{\mathbf{X}}\right\}^{-1}\hat{\mathbf{X}}^{\tau}\left(\mathbf{I}_{kTM}-\mathbf{P}_{\hat{\mathbf{B}}}\right)\left(\Omega\otimes\mathbf{I}_{kT}\right)\left(\mathbf{I}_{kTM}-\mathbf{P}_{\hat{\mathbf{B}}}\right)$$
$$\hat{\mathbf{X}}\left\{\hat{\mathbf{X}}^{\tau}\left(\mathbf{I}_{kTM}-\mathbf{P}_{\hat{\mathbf{B}}}\right)\hat{\mathbf{X}}\right\}^{-1}=\Sigma_2.$$

By the fact that

$$\begin{aligned}
&\left\{\hat{\mathbf{X}}^{\tau}\left(\mathbf{I}_{kTM}-\mathbf{P}_{\hat{\mathbf{B}}}\right)\hat{\mathbf{X}}\right\}^{-1}\hat{\mathbf{X}}^{\tau}\left(\mathbf{I}_{kTM}-\mathbf{P}_{\hat{\mathbf{B}}}\right)\left(\Omega\otimes\mathbf{I}_{kT}\right)\left(\mathbf{I}_{kTM}-\mathbf{P}_{\hat{\mathbf{B}}}\right)\\
&\quad\times\hat{\mathbf{X}}\left\{\hat{\mathbf{X}}^{\tau}\left(\mathbf{I}_{kTM}-\mathbf{P}_{\hat{\mathbf{B}}}\right)\hat{\mathbf{X}}\right\}^{-1}\\
&\quad-\left\{\hat{\mathbf{X}}^{\tau}\left(\mathbf{I}_{kTM}-\mathbf{P}_{\hat{\mathbf{B}}}\right)\left(\Omega^{-1}\otimes\mathbf{I}_{kT}\right)\left(\mathbf{I}_{kTM}-\mathbf{P}_{\hat{\mathbf{B}}}\right)\hat{\mathbf{X}}\right\}^{-1}\\
&=\left\{\hat{\mathbf{X}}^{\tau}\left(\mathbf{I}_{kTM}-\mathbf{P}_{\hat{\mathbf{B}}}\right)\hat{\mathbf{X}}\right\}^{-1}\hat{\mathbf{X}}^{\tau}\left(\mathbf{I}_{kTM}-\mathbf{P}_{\hat{\mathbf{B}}}\right)\left(\Omega^{1/2}\otimes\mathbf{I}_{kT}\right)\\
&\quad\times\left[\mathbf{I}_{kTM}-\left(\Omega^{-1/2}\otimes\mathbf{I}_{kT}\right)\left(\mathbf{I}_{kTM}-\mathbf{P}_{\hat{\mathbf{B}}}\right)\hat{\mathbf{X}}\right.\\
&\quad\times\left\{\hat{\mathbf{X}}^{\tau}\left(\mathbf{I}_{kTM}-\mathbf{P}_{\hat{\mathbf{B}}}\right)\left(\Omega^{-1}\otimes\mathbf{I}_{kT}\right)\left(\mathbf{I}_{kTM}-\mathbf{P}_{\hat{\mathbf{B}}}\right)\hat{\mathbf{X}}\right\}^{-1}\\
&\quad\left.\times\mathbf{X}^{\tau}\left(\mathbf{I}_{kTM}-\mathbf{P}_{\hat{\mathbf{B}}}\right)\left(\Omega^{-1/2}\otimes\mathbf{I}_{kT}\right)\right]\\
&\quad\times\left(\Omega^{1/2}\otimes\mathbf{I}_{kT}\right)\left(\mathbf{I}_{kTM}-\mathbf{P}_{\hat{\mathbf{B}}}\right)\hat{\mathbf{X}}\left\{\hat{\mathbf{X}}^{\tau}\left(\mathbf{I}_{kTM}-\mathbf{P}_{\hat{\mathbf{B}}}\right)\hat{\mathbf{X}}\right\}^{-1}\geq 0,
\end{aligned}$$

we have

$$\begin{aligned}
&\left\{\hat{\mathbf{X}}^{\tau}\left(\mathbf{I}_{kTM}-\mathbf{P}_{\hat{\mathbf{B}}}\right)\left(\Omega^{-1}\otimes\mathbf{I}_{kT}\right)\left(\mathbf{I}_{kTM}-\mathbf{P}_{\hat{\mathbf{B}}}\right)\hat{\mathbf{X}}\right\}^{-1}\\
&\quad\leq\left\{\hat{\mathbf{X}}^{\tau}\left(\mathbf{I}_{kTM}-\mathbf{P}_{\hat{\mathbf{B}}}\right)\hat{\mathbf{X}}\right\}^{-1}\hat{\mathbf{X}}^{\tau}\left(\mathbf{I}_{kTM}-\mathbf{P}_{\hat{\mathbf{B}}}\right)\left(\Omega\otimes\mathbf{I}_{kT}\right)\left(\mathbf{I}_{kTM}-\mathbf{P}_{\hat{\mathbf{B}}}\right)\\
&\qquad\times\hat{\mathbf{X}}\left\{\hat{\mathbf{X}}^{\tau}\left(\mathbf{I}_{kTM}-\mathbf{P}_{\hat{\mathbf{B}}}\right)\hat{\mathbf{X}}\right\}^{-1}.
\end{aligned}$$

https://doi.org/10.1017/S0266466613000352 Published online by Cambridge University Press

That is, $\Sigma_1^{-1} \leq \Sigma_2$. This shows that the WPLS $\hat{\boldsymbol{\beta}}^W$ is asymptotically more efficient than the unweighted $\hat{\boldsymbol{\beta}}$ that ignores the contemporaneous correlation among the equations.

Remark 2. For $M = 1$, it is easy to see that $\hat{\boldsymbol{\beta}}_1 = \hat{\boldsymbol{\beta}}^W$. Fan et al. (2005) argued that $\hat{\boldsymbol{\beta}}_1$ (or $\hat{\boldsymbol{\beta}}^W$) achieves the semiparametric efficient bound when $M = 1$. For $M > 1$, whether $\hat{\boldsymbol{\beta}}^W$ still achieves the semiparametric efficient bound is still an open problem. We will investigate this in the future.

To make statistical inference based on $\hat{\boldsymbol{\beta}}^W$, a consistent estimator of the asymptotic covariance matrix Σ_1 is needed. This is given by

$$\hat{\Sigma}_1 = k^{-1}\hat{\mathbf{X}}^\tau \left(\mathbf{I}_{kTM} - \mathbf{P}_{\hat{\mathbf{B}}}\right)\left(\hat{\Omega}^{-1} \otimes \mathbf{I}_{kT}\right)\left(\mathbf{I}_{kTM} - \mathbf{P}_{\hat{\mathbf{B}}}\right)\hat{\mathbf{X}}.$$

The next theorem shows that $\hat{\Sigma}_1$ is a consistent estimator of Σ_1.

THEOREM 4. *Under Assumptions 1–7,* $\hat{\Sigma}_1 - \Sigma_1 = o_p(1)$.

We can apply these asymptotic results to construct test statistics and confidence intervals for the parametric components $\boldsymbol{\beta} = \left(\boldsymbol{\beta}_1^\tau, \ldots, \boldsymbol{\beta}_M^\tau\right)^\tau$. For example, the following corollary gives a test for the hypothesis $H_0 : \mathbf{C}\boldsymbol{\beta} = 0$, where $\mathbf{C}$ is a known $d \times (p_1 + \cdots + p_M)$ matrix with rank d.

COROLLARY 1. *If Assumptions 1–7 hold, then under the null hypothesis* $H_0 : \mathbf{C}\boldsymbol{\beta} = 0$,

$$k\left(\mathbf{C}\hat{\boldsymbol{\beta}}^W\right)^\tau \left(\mathbf{C}\hat{\Sigma}_1^{-1}\mathbf{C}^\tau\right)^{-1}\left(\mathbf{C}\hat{\boldsymbol{\beta}}^W\right) \xrightarrow{D} \chi_d^2, \quad as \quad k \to \infty,$$

where χ_d^2 *is the chi-square distribution with degree of freedom* d.

5. WEIGHTED TWO-STAGE ESTIMATION OF NONPARAMETRIC COMPONENTS

The PLS estimator of the nonparametric components proposed by Fan et al. (2005) has two shortcomings. One is the difficulty in deriving its asymptotic distribution, and the other is the difficulty in accounting for the contemporaneous correlation among equations. Therefore, we here propose a new estimator based on a weighted two-stage technique.

If the contemporaneous correlation among equations is ignored, we can construct a two-stage estimator of the nonparametric components as in You and Zhou (2008). For fixed s, subtracting $(T-1)^{-1}\sum_{t_1 \neq t}^{T}\left(Y_{sit_1} - \mathbf{X}_{sit_1}^\tau \hat{\boldsymbol{\beta}}_s - \hat{g}_s(U_{sit_1})\right)$ from

https://doi.org/10.1017/S0266466613000352 Published online by Cambridge University Press

model (1.1) yields

$$Y_{sit} - \frac{1}{T-1}\sum_{t_1\neq t}^{T}\left(Y_{sit_1} - \mathbf{X}_{sit_1}^{\tau}\hat{\boldsymbol{\beta}}_s - \hat{g}_s(U_{sit_1})\right)$$
$$= \mathbf{X}_{sit}^{\tau}\boldsymbol{\beta}_s + g_s(U_{sit}) + v_{sit} + \mu_{si} - \frac{1}{T-1}\sum_{t_1\neq t}^{T}\left(Y_{sit_1} - \mathbf{X}_{sit_1}^{\tau}\hat{\boldsymbol{\beta}}_s - \hat{g}_s(U_{sit_1})\right).$$

According to Lemma 1 in Section 2,

$$\mu_{si} - \frac{1}{T-1}\sum_{t_1\neq t}^{T}\left(Y_{sit_1} - \mathbf{X}_{sit_1}^{\tau}\hat{\boldsymbol{\beta}}_s - \hat{g}_s(U_{sit_1})\right)$$
$$= \frac{1}{T-1}\sum_{t_1\neq t}^{T} v_{sit_1} + O_p\left(h_s^2 + \frac{1}{\sqrt{kh_s}} + \frac{1}{\sqrt{k}}\right).$$

Therefore, if we denote

$$Y_{sit}^* = Y_{sit} - \frac{1}{T-1}\sum_{t_1\neq t}^{T}\left(Y_{sit_1} - \mathbf{X}_{sit_1}^{\tau}\hat{\boldsymbol{\beta}}_s - \hat{g}_s\left(U_{sit_1}\right)\right) \quad \text{and}$$

$$v_{sit}^* = v_{sit} - \frac{1}{T-1}\sum_{t_1\neq t}^{T} v_{sit_1},$$

then

$$Y_{sit}^* - \mathbf{X}_{sit}^{\tau}\hat{\boldsymbol{\beta}}_s = g_s(U_{sit}) + v_{sit}^* + O_p\left(h_s^2 + \frac{1}{\sqrt{kh_s}} + \frac{1}{\sqrt{k}}\right). \tag{5.1}$$

Based on model (5.1), a two-stage estimator of $g_s(\cdot)$ can be constructed by

$$\hat{g}_s^{TS}(u) = (1,0)\left\{\mathbf{D}_{su}^{*\tau}\mathbf{W}_{su}^{*}\mathbf{D}_{su}^{*}\right\}^{-1}\mathbf{D}_{su}^{*\tau}\mathbf{W}_{su}^{*}\left(\mathbf{Y}_s^* - \mathbf{X}_s\hat{\boldsymbol{\beta}}_s^{W}\right),$$

where $\mathbf{D}_{ju}^*$ and $\mathbf{W}_{ju}^*$ have the same definitions as $\mathbf{D}_{ju}$ and $\mathbf{W}_{ju}$, respectively, except that h_s is replaced by h_s^*, and $\mathbf{Y}_s^* = \left(Y_{s11}^*, \ldots, Y_{s1T}^*, \ldots, Y_{skT}^*\right)^{\tau}$.

The following assumption is needed for the asymptotic property of $\hat{g}_s^{TS}(u)$ in the next theorem.

Assumption 8. For fixed s, the bandwidth h_s^* satisfies that $kh_s^{*8} \to 0$ and $kh_s^{*2}/(\log k)^2 \to \infty$ as $k \to \infty$. In addition, $h_s = \mathrm{o}(h_s^*)$.

THEOREM 5. *Under Assumptions 1–8,*

$$\sqrt{kh_s^*}\left[\hat{g}_s^{TS}(u) - g_s(u) - \frac{h_s^{*2}}{2}\frac{\alpha_2^2 - \alpha_1\alpha_3}{\alpha_2 - \alpha_1^2}g_s''(u)\right] \xrightarrow{D} N\left(0, \zeta_s^{TS}(u)\right) \quad \text{as } k \to \infty$$

provided that $f_s(u) \neq 0$, *where*

$$\zeta_s^{TS}(u) = \frac{\sigma_{vss}^2(c_0^2\delta_0 + 2c_0c_1\delta_1 + c_1^2\delta_2)}{f_s(u)}\frac{1}{T-1},$$

$c_1 = \alpha_2/\left(\alpha_2 - \alpha_1^2\right)$, *and* $c_2 = -\alpha_1/\left(\alpha_2 - \alpha_1^2\right)$, *with* α_i *and* δ_i *defined in Lemma 1.*

https://doi.org/10.1017/S0266466613000352 Published online by Cambridge University Press

This $\hat{g}_s^{TS}(u)$, however, may not be asymptotically efficient as it does not take into account the contemporaneous correlation among equations. Consequently, we propose a weighted two-stage estimator for $g_s(\cdot)$ and show it is asymptotically more efficient than $\hat{g}_s^{TS}(u)$. Define

$$\hat{\varepsilon}_{sit}^* = Y_{sit}^* - \mathbf{X}_{sit}^{\tau}\hat{\boldsymbol{\beta}}_s - \hat{g}_s(U_{sit}) \quad \text{and}$$
$$d_{sit} = \hat{\sigma}_{\nu}^{2ss}\left(Y_{sit}^* - \mathbf{X}_{sit}^{T}\hat{\boldsymbol{\beta}}_s^{W}\right) + \sum_{s_1 \neq s}^{M} \hat{\sigma}_{\nu}^{2ss_1} \sum_{t_1=1}^{T} \hat{\varepsilon}_{s_1 it_1}^*.$$

The following weighted two-stage estimate for $g_s(\cdot)$ is proposed:

$$\hat{g}_s^{WTS}(u) = \left(\left(\hat{\sigma}_{\nu}^{2ss}\right)^{-1}, 0\right)\left\{\mathbf{D}_{su}^{*\tau}\mathbf{W}_{su}^*\mathbf{D}_{su}^*\right\}^{-1}\mathbf{D}_{su}^{*\tau}\mathbf{W}_{su}^*\mathbf{d}_s,$$

where $\mathbf{d}_s = (d_{s11}, \ldots, d_{s1T} \ldots, d_{skT})^{\tau}$. This $\hat{g}_s^{WTS}(u)$ has the following asymptotic results.

THEOREM 6. *Under Assumptions 1–8,*

$$\sqrt{kh_s^*}\left[\hat{g}_s^{WTS}(u) - g_s(u) - \frac{h_s^{*2}}{2}\frac{\alpha_2^2 - \alpha_1\alpha_3}{\alpha_2 - \alpha_1^2}g_s''(u)\right] \xrightarrow{D} N\left(0, \zeta_s^{WTS}(u)\right) \quad \text{as } k \to \infty,$$

provided that $f_s(u) \neq 0$, *where* $\zeta_s^{WTS}(u) = \left(c_0^2\delta_0 + 2c_0c_1\delta_1 + c_1^2\delta_2\right) / \left[\sigma_{\nu}^{2ss} f_s(u)(T-1)\right]$.

Remark 5.1. The weighted two-stage estimator $\hat{g}_s^{WTS}$(u) is somewhat like the SU kernel estimator proposed by Wang (2003), which is however focused on longitudinal structure. Recall that σ_{ν}^{2ss} is the (s,s)-entry of Ω^{-1}, $\sigma_{\nu ss}^2$ is the (s,s)-entry of Ω, and Ω is a positive definite matrix. As in Wang (2003), we only need to consider s = 1. If we write $\Omega = \begin{pmatrix} \sigma_{\nu 11}^2 & \Omega_{12} \\ \Omega_{12}^T & \Omega_{22} \end{pmatrix}$, then $\sigma_{\nu}^{211} = \left(\sigma_{\nu 11}^2 - \Omega_{12}\Omega_{22}^{-1}\Omega_{12}^{\tau}\right)^{-1}$. Because Ω_{22}^{-1} is positive definite, $\sigma_{\nu}^{211}\sigma_{\nu 11}^2 \geq 1$. Thus the weighted two-stage estimator $\hat{g}_s^{WTS}(\cdot)$ is asymptotically more efficient than the unweighted $\hat{g}_s^{TS}(\cdot)$.

Selection of the smoothing parameter is is essential in the setting of non/semiparametric regressions. According to the discussion by Fan in Bickel and Kwon (2001), the performance of the parametric component estimators in the semiparametric regression models is not sensitive to the bandwidth, as long as the bandwidth is not so large as to create excessive biases. We also noted this in our simulation studies. Hence, we choose bandwidths h_s^* that are suitable for estimating function $g_s(\cdot)$. By Theorem 6, the asymptotic integrated mean square

https://doi.org/10.1017/S0266466613000352 Published online by Cambridge University Press

error (AIMSE) of $\sqrt{kh_s^*}\left(\hat{g}_s^{WTS}(u) - g_s(u)\right)$ is

$$\text{AIMSE} = \frac{kh_s^{*5}}{4}\left(\frac{\alpha_2^2 - \alpha_1\alpha_3}{\alpha_2 - \alpha_1^2}\right)^2 \int_{\mathcal{U}_s} \left\{g_s''(u)\right\}^2 du + \frac{(c_0^2\delta_0 + 2c_0c_1\delta_1 + c_1^2\delta_2)}{(T-1)\sigma_v^{2ss}} \int_{\mathcal{U}_s} \frac{1}{f_s(u)} du.$$

Minimizing the AIMSE gives the optimal bandwidth

$$h_{sopt}^* = k^{-1/5}\left\{\frac{(c_0^2\delta_0 + 2c_0c_1\delta_1 + c_1^2\delta_2)}{(T-1)\sigma_v^{2ss}} \int_{\mathcal{U}_s} f_s(u)du \left(\frac{\alpha_2 - \alpha_1^2}{\alpha_2^2 - \alpha_1\alpha_3}\right)^2 \times \left[\int_{\mathcal{U}_s} \{g_s''(u)\}^2 du\right]^{-1}\right\}^{-1/5}. \tag{5.2}$$

Hence the optimal convergence rate of the AIMSE for $\hat{g}_s^{WTS}(u)$ is of order $k^{-4/5}$.

Based on $\{U_{sit}, i = 1, \ldots, k, t = 1, \ldots, T\}$, we can estimate $f_s(u)$ by applying the usual density estimation technique. In addition, a consistent estimator of $g_s''(u)$ can be constructed according to Theorem 5 of Horowitz and Mammen (2004). Thus, a plug-in estimator of h_{sopt}^* can be obtained by replacing unknown population quantities on the right-hand side of (5.2) with their estimates.

6. SIMULATION STUDIES

In this section, we report some simulation results on the performance of the estimators proposed in the previous sections. The data are generated from the following fixed effects panel data SU partially linear regression model:

$$y_{sit} = x_{sit1}\beta_{s1} + x_{sit2}\beta_{s2} + g_s(u_{sit}) + \mu_{si} + v_{sit},$$
$$s = 1, 2, \qquad i = 1, \ldots, k, \qquad t = 1, \ldots, T,$$

where $x_{1it1}, x_{2it1} \sim$ i.i.d. $N(0,1)$, $x_{1it2}, x_{2it2} \sim$ i.i.d. Bernoulli(0.45), $u_{1it}, u_{2it} \sim$ i.i.d. $U(0,1)$, $\beta_{11} = 1.5$, $\beta_{12} = 2$, $\beta_{21} = 1$, $\beta_{22} = -2$, $g_1(u) = 2\sin(2\pi u)$, $g_2(u) = 2\cos(2\pi u)$, $\mu_{1i}, \mu_{2i} \sim$ i.i.d $U(-0.5, 0.5)$ and are centralized. In addition, for fixed n, T, and σ_{21}^2, (μ_{1i}, μ_{2i}) are generated only once to get fixed effects, and

$$\begin{bmatrix} v_{1it} \\ v_{2it} \end{bmatrix} \sim \text{i.i.d. } N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{11}^2 & \sigma_{12}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 \end{bmatrix}\right)$$

with $\sigma_{11}^2 = 2$, $\sigma_{22}^2 = 2.25$, and $\sigma_{12}^2 = \sigma_{21}^2 = 1.5$ and 2 to reflect different correlation levels.

Samples of sizes $k = 100, 200, 500$ and $T = 2, 3, 5$ are drawn repeatedly, with 1,000 simulated realizations in each case. The Gaussian kernel is used, and the bandwidth is selected by the method discussed at the end of Section 5.

https://doi.org/10.1017/S0266466613000352 Published online by Cambridge University Press

For a given sample size, the sample means (sms), standard deviations (stds), mean of the variance estimates (ses), and coverage of the 95% nominal confidence intervals (cps) for the estimators of β_{11}, β_{12}, β_{21}, and β_{22} are obtained, and the results are summarized in Tables 1 and 2, which compare the PLS estimator with the WPLS estimator.

TABLE 1. The finite-sample performance of the PLS and WPLS estimators for the parametric components in which $\sigma_{12}^2 = 2$

		$k = 100$			$k = 200$			$k = 500$		
		$T = 2$	$T = 3$	$T = 5$	$T = 2$	$T = 3$	$T = 5$	$T = 2$	$T = 3$	$T = 5$
$\hat{\beta}_{11}$	sm	1.505	1.499	1.501	1.500	1.499	1.499	1.499	1.500	1.506
	std	0.151	0.102	0.071	0.101	0.071	0.049	0.064	0.044	0.029
	se	0.148	0.102	0.071	0.102	0.071	0.050	0.063	0.044	0.030
	cp	0.949	0.953	0.951	0.956	0.953	0.952	0.960	0.956	0.950
$\hat{\beta}_{12}$	sm	1.994	1.993	2.006	2.000	2.002	1.997	2.000	1.996	1.997
	std	0.295	0.200	0.141	0.202	0.140	0.101	0.130	0.095	0.061
	se	0.297	0.205	0.143	0.205	0.143	0.101	0.128	0.090	0.061
	cp	0.949	0.950	0.966	0.954	0.953	0.949	0.943	0.936	0.954
$\hat{\beta}_{21}$	sm	1.005	1.000	0.997	1.007	1.002	0.997	1.000	0.997	0.999
	std	0.162	0.106	0.078	0.114	0.073	0.050	0.066	0.048	0.032
	se	0.157	0.108	0.075	0.108	0.075	0.053	0.067	0.047	0.033
	cp	0.935	0.959	0.941	0.933	0.952	0.958	0.950	0.940	0.961
$\hat{\beta}_{22}$	sm	−1.988	−1.997	−2.003	−2.000	−2.003	−1.997	−2.010	−1.997	−1.998
	std	0.315	0.219	0.151	0.218	0.155	0.107	0.130	0.099	0.066
	se	0.315	0.218	0.152	0.218	0.152	0.107	0.136	0.095	0.067
	cp	0.944	0.945	0.952	0.945	0.944	0.948	0.962	0.942	0.949
$\hat{\beta}_{11}^W$	sm	1.499	1.499	1.500	1.499	1.498	1.500	1.500	1.500	1.501
	std	0.050	0.034	0.024	0.032	0.024	0.016	0.021	0.015	0.010
	se	0.050	0.034	0.023	0.034	0.023	0.016	0.021	0.015	0.010
	cp	0.944	0.952	0.942	0.958	0.942	0.948	0.948	0.946	0.960
$\hat{\beta}_{12}^W$	sm	1.998	2.001	1.999	2.002	1.998	1.998	2.001	1.999	2.001
	std	0.099	0.071	0.047	0.069	0.046	0.034	0.040	0.031	0.020
	se	0.100	0.068	0.048	0.068	0.048	0.033	0.042	0.030	0.021
	cp	0.953	0.939	0.952	0.943	0.959	0.937	0.957	0.944	0.953
$\hat{\beta}_{21}^W$	sm	0.999	0.999	0.999	1.002	1.000	1.000	1.000	0.999	1.001
	std	0.052	0.035	0.025	0.036	0.025	0.017	0.022	0.015	0.010
	se	0.053	0.036	0.025	0.036	0.025	0.017	0.022	0.015	0.011
	cp	0.953	0.956	0.956	0.945	0.950	0.952	0.945	0.951	0.948
$\hat{\beta}_{22}^W$	sm	−2.000	−2.003	−1.999	−2.002	−1.999	−1.999	−1.999	−2.000	−1.999
	std	0.105	0.072	0.050	0.071	0.051	0.034	0.046	0.032	0.020
	se	0.106	0.073	0.050	0.073	0.050	0.035	0.045	0.032	0.020
	cp	0.953	0.961	0.952	0.960	0.940	0.961	0.947	0.940	0.951

https://doi.org/10.1017/S0266466613000352 Published online by Cambridge University Press

TABLE 2. The finite-sample performance of the PLS and WPLS estimators for the parametric components in which $\sigma_{12}^2 = 1.5$

		$k = 100$			$k = 200$			$k = 500$		
		$T = 2$	$T = 3$	$T = 5$	$T = 2$	$T = 3$	$T = 5$	$T = 2$	$T = 3$	$T = 5$
$\hat{\beta}_{11}$	sm	1.506	1.503	1.501	1.503	1.499	1.502	1.503	1.500	1.501
	std	0.149	0.102	0.071	0.102	0.072	0.049	0.063	0.043	0.032
	se	0.148	0.102	0.071	0.102	0.071	0.050	0.063	0.044	0.031
	cp	0.953	0.951	0.955	0.954	0.961	0.951	0.950	0.956	0.937
$\hat{\beta}_{12}$	sm	1.990	1.992	1.994	2.006	1.999	1.998	1.993	2.001	2.001
	std	0.299	0.203	0.137	0.205	0.143	0.097	0.133	0.085	0.065
	se	0.298	0.205	0.143	0.205	0.143	0.100	0.128	0.090	0.063
	cp	0.941	0.947	0.966	0.957	0.946	0.955	0.943	0.967	0.945
$\hat{\beta}_{21}$	sm	1.005	0.996	1.002	0.997	0.999	1.001	0.999	1.000	1.001
	std	0.158	0.106	0.074	0.109	0.075	0.052	0.070	0.046	0.033
	se	0.157	0.108	0.075	0.109	0.075	0.053	0.067	0.047	0.033
	cp	0.946	0.957	0.950	0.951	0.949	0.958	0.944	0.945	0.956
$\hat{\beta}_{22}$	sm	−1.986	−2.003	−2.006	−2.006	−1.994	−1.998	−1.996	−1.996	−1.999
	std	0.305	0.218	0.150	0.210	0.156	0.107	0.133	0.093	0.066
	se	0.315	0.217	0.152	0.218	0.152	0.107	0.136	0.095	0.067
	cp	0.959	0.953	0.945	0.957	0.939	0.945	0.957	0.948	0.954
$\hat{\beta}_{n11}^W$	sm	1.510	1.504	1.500	1.498	1.500	1.502	1.501	1.500	1.500
	std	0.104	0.073	0.051	0.073	0.052	0.036	0.044	0.030	0.023
	se	0.105	0.072	0.050	0.072	0.050	0.035	0.045	0.031	0.022
	cp	0.954	0.946	0.949	0.944	0.942	0.951	0.958	0.953	0.943
$\hat{\beta}_{12}^W$	sm	1.999	2.000	1.997	2.003	1.997	1.999	1.996	1.999	2.001
	std	0.207	0.147	0.101	0.149	0.102	0.074	0.092	0.061	0.045
	se	0.212	0.145	0.101	0.145	0.101	0.071	0.090	0.063	0.045
	cp	0.949	0.936	0.949	0.948	0.946	0.940	0.949	0.955	0.947
$\hat{\beta}_{21}^W$	sm	1.005	0.997	0.999	0.998	0.999	1.001	0.999	0.999	1.000
	std	0.110	0.075	0.051	0.076	0.052	0.036	0.047	0.032	0.023
	se	0.112	0.077	0.053	0.077	0.053	0.037	0.047	0.033	0.023
	cp	0.951	0.956	0.956	0.950	0.958	0.954	0.958	0.945	0.956
$\hat{\beta}_{22}^W$	sm	−1.991	−2.004	−2.004	−2.003	−1.998	−2.001	−2.000	−2.000	−1.999
	std	0.234	0.152	0.107	0.158	0.110	0.076	0.097	0.068	0.046
	se	0.224	0.154	0.108	0.154	0.107	0.075	0.096	0.067	0.047
	cp	0.941	0.960	0.950	0.936	0.942	0.953	0.948	0.948	0.955

Tables 1 and 2 provide the following observations: (i) Both PLS and WPLS estimators are asymptotically unbiased. (ii) The stds of the estimators decrease as the individual size k grows, or as T grows. (iii) The stds of the WPLS estimator are much smaller than those of the PLS estimator, especially when the contemporaneous correlation is strong. (iv) The proposed variance estimator provides

https://doi.org/10.1017/S0266466613000352 Published online by Cambridge University Press

TABLE 3. The finite-sample performance of the two-stage and improved estimators of the nonparametric components

			$k=100$			$k=200$			$k=500$		
			$T=2$	$T=3$	$T=5$	$T=2$	$T=3$	$T=5$	$T=2$	$T=3$	$T=5$
$\sigma_{12}^2=2$	$\mathrm{RASE}(\hat{g}_1^{TS}(\cdot))$	sm	0.138	0.072	0.034	0.076	0.033	0.017	0.028	0.015	0.007
		std	0.065	0.040	0.017	0.040	0.017	0.009	0.015	0.007	0.003
	$\mathrm{RASE}(\hat{g}_2^{TS}(\cdot))$	sm	0.186	0.077	0.041	0.078	0.038	0.020	0.034	0.015	0.007
		std	0.095	0.042	0.023	0.037	0.019	0.010	0.019	0.008	0.004
	$\mathrm{RASE}(\hat{g}_1^{WTS}(\cdot))$	sm	0.017	0.008	0.004	0.009	0.004	0.002	0.003	0.001	0.001
		std	0.009	0.004	0.002	0.005	0.001	0.001	0.001	0.000	0.000
	$\mathrm{RASE}(\hat{g}_2^{WTS}(\cdot))$	sm	0.021	0.009	0.004	0.009	0.004	0.002	0.003	0.001	0.001
		std	0.011	0.004	0.002	0.005	0.002	0.001	0.001	0.000	0.000
$\sigma_{12}^2=1.5$	$\mathrm{RASE}(\hat{g}_1^{TS}(\cdot))$	sm	0.146	0.072	0.034	0.070	0.037	0.016	0.026	0.014	0.007
		std	0.073	0.036	0.018	0.037	0.019	0.008	0.014	0.007	0.004
	$\mathrm{RASE}(\hat{g}_2^{TS}(\cdot))$	sm	0.184	0.077	0.040	0.084	0.037	0.020	0.033	0.015	0.007
		std	0.099	0.038	0.022	0.047	0.017	0.011	0.017	0.008	0.004
	$\mathrm{RASE}(\hat{g}_1^{WTS}(\cdot))$	sm	0.077	0.038	0.017	0.037	0.018	0.008	0.013	0.007	0.003
		std	0.042	0.019	0.008	0.019	0.009	0.005	0.006	0.003	0.001
	$\mathrm{RASE}(\hat{g}_2^{WTS}(\cdot))$	sm	0.091	0.041	0.020	0.040	0.020	0.010	0.016	0.008	0.004
		std	0.048	0.021	0.011	0.020	0.011	0.005	0.009	0.004	0.002

accurate estimates and coverage of the confidence intervals close to the nominal 95% level for the cases studied.

The performance of the estimator $\hat{g}_s(\cdot)$ of the nonparametric component is also assessed via the square-root averaged squared errors (RASE) method:

$$\mathrm{RASE}(\hat{g}_s)=\left[\frac{1}{kT}\sum_{i=1}^{k}\sum_{t=1}^{T}\{\hat{g}_s(U_{sit})-g_s(U_{sit})\}^2\right]^{1/2}.$$

For a given sample size, the sm and std of the RASE are calculated and listed in Table 3, which shows that the weighted two-stage estimator outperforms the original two-stage estimator.

7. CONCLUDING REMARKS

Semiparametric regression models have attracted considerable attention recently as they provide a sensible compromise between the parametric and nonparametric models while retaining their nice features. This paper has investigated the statistical inference of the fixed effects panel data seemingly unrelated partially linear regression model, which naturally extends the traditional fixed effects panel data regression model to semiparametric regressions. It allows multiple regression equations and includes the aggregated partially linear model as a special

https://doi.org/10.1017/S0266466613000352 Published online by Cambridge University Press

case. By taking into account the contemporaneous correlation among equations, and combining the local polynomial techniques with least squares, we proposed a weighted profile least squares estimator for the parametric components and a weighted two-stage estimator for the nonparametric components. The proposed estimators are asymptotically more efficient than those neglecting the contemporaneous correlation.

Interesting topics for further study include the extension of our results to two or more ways error component structure. Recently, Atak, Linton, and Xiao (2011) proposed a semiparametric panel data model for unbalanced data with a stationary β-mixing error process and applied it to climate change in the United Kingdom. Whether and how this new semiparametric panel data model can be extended to fixed effects panel data seemingly unrelated regression models remain open problems.

REFERENCES

Abrevaya, J. (1999) Leapfrog estimation of a fixed-effects model with unknown transformation of the dependent variable. *Journal of Econometrics* 93, 203–228.

Ahn, S.C. & P. Schmidt (2000) Estimation of linear panel data models using GMM. In L. Mátyás (ed.), *Generalized Method of Moments Estimation*. Ch. 8, 211–249. Cambridge University Press.

Atak, A., O. Linton, Z. Xiao (2011) A semiparametric panel model for climate change in the United Kingdom. *Journal of Econometrics* 164, 92–115.

Avery, R.B. (1977) Error components and seemingly unrelated regressions. *Econometrica* 45, 199–209.

Baltagi, B.H. (1980) On seemingly unrelated regressions with error components. *Econometrica* 48, 1547–1551.

Baltagi, B.H. (1995) *Econometric Analysis of Panel Data*. Wiley.

Baltagi, B.H. & D. Li (2002) Series estimation of partially linear panel data models with fixed effects. *Annals of Economics and Finance* 3, 103–116.

Beierlein, J.G., J.W. Dunn & J.C. McConnon Jr. (1981) The demand for electricity and natural gas in the northeastern United States. *Review of Economics and Statistics* 63, 403–408.

Bickel, P.J. & J. Kwon (2001) Inference for semiparametric models: Some questions and an answer. With comments and a rejoinder by the authors. *Statistics Sinica* 11, 863–960.

Breusch, T.S. & A.R. Pagan (1979) A simple test for heteroscedasticity and random coefficient variation. *Econometrica* 47, 1287–1294.

Carroll, R.J., M. Doug, F. Larry & K. Victor (2006) Seemingly unrelated measurement error models, with application to nutritional epidemiology. *Biometrics* 62, 75–84.

Entorf, H. (1997) Random walks with drifts: Nonsense regression and spurious fixed-effect estimation. *Journal of Econometrics* 80, 287–296.

Fan, J., H. Peng & T. Huang (2005) Semilinear high-dimensional model for normalization of microarray data: A theoretical analysis and partial consistency. *Journal of the American Statistical Association* 100, 781–813.

Gallant, R. (1975) Seemingly unrelated nonlinear regressions. *Journal of Econometrics* 3, 35–50.

Ghazal, G.A. (1986) The error of forecast for seemingly unrelated regression (SUR) equations. *Egyptian Statistical Journal* 30, 35–64.

Hahn, J. (1999) How informative is the initial condition in the dynamic panel model with fixed effects? *Journal of Econometrics* 93, 309–326.

Hahn, J. (2001) The information bound of a dynamic panel logit model with fixed effects. *Econometric Theory* 17, 913–932.

https://doi.org/10.1017/S0266466613000352 Published online by Cambridge University Press

Henderson, D.J., R.J. Carroll & Q. Li (2008) Nonparametric estimation and testing of fixed effects panel data models. *Journal of Econometrics* 144, 257–276.

Horowitz, J.L. & S. Lee (2004) Semiparametric estimation of a panel data proportional hazards model with fixed effects. *Journal of Econometrics* 119, 155–198.

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

Kakwani, N.C. (1974) A note on the efficiency of the Zellner's seemingly unrelated regressions estimator. *Annals of the Institute of Statistical Mathematics* 26, 361–362.

Koop, G, Poirer, D. & Tobias, J. (2005) Semiparametric Bayesian inference in multiple equation models. *Journal of Applied Econometrics* 20, 723–747.

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

Liu, A. (2002) Efficient estimation of two seemingly unrelated regression equations. *Journal of Multivariate Analysis* 82, 445–456.

Mack, Y.P. & B.W. Silverman (1982) Weak and strong uniform consistency of kernel regression estimates. *Zoitschcift für Wahrscheinlichkeits theorie und Verwandte Gebiete* 61, 405–415.

Mandy, D.M. & C. Martins-Filho (1993) Seemingly unrelated regressions under additive heteroscedasticity: Theory and share equation applications. *Journal of Econometrics* 58, 315–346.

Neudecker, H. & F.A. Windmeijer (1991) R^2 in seemingly unrelated regression equations. *Statistical Neerlandica* 45, 405–411.

Neyman, J. & E. Scott (1948) Consistent estimates based on partially consistent observations. *Econometrica* 16, 1–32.

Ng, V.M. (2002) Robust Bayesian inference for seemingly unrelated regressions with elliptical errors. *Journal of Multivariate Analysis* 83, 409–414.

Robinson, P. (1988) Root-N-consistent semiparametric regression. *Econometrica* 56, 931–954.

Rocke, D.M. (1989) Bootstrap Bartlett adjustment in seemingly unrelated regression. *Journal of the American Statistical Association* 84, 598–601.

Smith, M. & R. Kohn (2000) Nonparametric seemingly unrelated regression. *Journal of Econometrics* 98, 257–281.

Srivastava, V.K. & E.A. Giles (1987) *Seemingly Unrelated Regression Equations Models. Estimation and Inference.* Statistics: Textbooks and Monographs, 80. Marcel Dekker.

Stone, C.J. (1986) Additive regression and other nonparametric models. *Annals of Statistics* 13, 689–705.

Su, L. & A. Ullah (2006) Profile likelihood estimation of partially linear panel data models with fixed effects. *Economics Letters* 92, 75–81.

Verbon, H.A. (1980) Testing for heteroscedasticity in a model of seemingly unrelated regression equations with variance components (SUREVC). *Economics Letters* 2, 149–153.

Wan, G.H., W.E. Griffiths, & J.R. Anderson (1992) Using panel data to estimate risk effects in seemingly unrelated production functions. *Empirical Economics* 17, 35–49.

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

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

Welsh, A.H. & T.W. Yee (2006) Local regression for vector responses. *Journal of Statical Planning and Inference* 135, 307–331.

You, J.H. & G.M. Chen (2006) Estimation in a semiparametric varying-coefficient partially linear errors-in-variables model. *Journal of Multivariate Analysis* 97, 324–341.

You, J.H. & H. Zhou (2008) A two-stage method for semilinear in-slide regression model. *Journal of Multivariate Analysis* 99, 1610–1634.

Zellner, A. (1962) An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association* 57, 348–368.

https://doi.org/10.1017/S0266466613000352 Published online by Cambridge University Press

Zellner, A. (1963) Estimators for seemingly unrelated equations: Some exact finite sample results. *Journal of the American Statistical Association* 58, 977–992.

Zhou, B., Q.F. Xu & J.H. You (2011) Efficient estimation of seemingly unrelated nonparametric regression models. *Metrika* 73, 121–138.

APPENDIX: Proofs of Main Results

We first present several lemmas that are need to prove the main results.

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

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

where $K_h(\cdot) = K(\cdot/h)/h$.

Lemma 2 follows immediately from Proposition 4 of Mack and Silverman (1982).

LEMMA 3. *Under Assumptions 4 and 5,*

$$\lim_{k\to\infty} \frac{1}{k} \hat{\mathbf{X}}^\tau \left(\mathbf{I}_{kTM} - \mathbf{P}_{\hat{\mathbf{B}}} \right) \left(\Omega^{-1} \otimes \mathbf{I}_{kT} \right) \left(\mathbf{I}_{kTM} - \mathbf{P}_{\hat{\mathbf{B}}} \right) \hat{\mathbf{X}} = \Sigma_1,$$

where Ω *and* Σ_1 *are defined in Section 4.*

Proof. According to the definition of Ω, it is easy to see

$$\begin{aligned} &\frac{1}{k} \hat{\mathbf{X}}^\tau \left(\mathbf{I}_{kTM} - \mathbf{P}_{\hat{\mathbf{B}}} \right) \left(\Omega^{-1} \otimes \mathbf{I}_{kT} \right) \left(\mathbf{I}_{kTM} - \mathbf{P}_{\hat{\mathbf{B}}} \right) \hat{\mathbf{X}} \\ &\quad = \left(\frac{\sigma_\nu^{2s_1 s_2}}{k} \hat{\mathbf{X}}_{s_1}^\tau \left(\mathbf{I}_{kT} - \mathbf{P}_{\hat{\mathbf{B}}_{s_1}} \right) \left(\mathbf{I}_{kT} - \mathbf{P}_{\hat{\mathbf{B}}_{s_2}} \right) \hat{\mathbf{X}}_{s_2} \right)_{s_1, s_2 = 1}^{M}. \end{aligned}$$

Arguments similar to the proof of Theorem 1 in Fan et al. (2005) lead to

$$\frac{\sigma_\nu^{2s_1 s_2}}{k} \hat{\mathbf{X}}_{s_1}^\tau \left(\mathbf{I}_{kT} - \mathbf{P}_{\hat{\mathbf{B}}_{s_1}} \right) \left(\mathbf{I}_{kT} - \mathbf{P}_{\hat{\mathbf{B}}_{s_2}} \right) \hat{\mathbf{X}}_{s_2} \to \sigma_\nu^{2s_1 s_2} \mathbf{D}_{s_1 s_2} \quad \text{as } k \to \infty.$$

This completes the proof. ■

LEMMA 4. *Under Assumptions 1–7,* $\sqrt{k}\left(\tilde{\boldsymbol{\beta}}^W - \boldsymbol{\beta} \right) \xrightarrow{D} N\left(0, \Sigma_{\mathbf{1}}^{-\mathbf{1}}\right)$ *as* $k \to \infty$*, where*

$$\begin{aligned} \tilde{\boldsymbol{\beta}}^W = &\left\{ \hat{\mathbf{X}}^\tau \left(\mathbf{I}_{kTM} - \mathbf{P}_{\hat{\mathbf{B}}} \right) \left(\Omega^{-1} \otimes \mathbf{I}_{kT} \right) \left(\mathbf{I}_{kTM} - \mathbf{P}_{\hat{\mathbf{B}}} \right) \hat{\mathbf{X}} \right\}^{-1} \\ &\hat{\mathbf{X}}^\tau \left(\mathbf{I}_{kTM} - \mathbf{P}_{\hat{\mathbf{B}}} \right) \left(\Omega^{-1} \otimes \mathbf{I}_{kT} \right) \left(\mathbf{I}_{kTM} - \mathbf{P}_{\hat{\mathbf{B}}} \right) \hat{\mathbf{Y}} \end{aligned}$$

with $\hat{\mathbf{X}}$, $\mathbf{P}_{\hat{\mathbf{B}}}$, *and* $\hat{\mathbf{Y}}$ *defined in Section 4.*

https://doi.org/10.1017/S0266466613000352 Published online by Cambridge University Press

Proof. According to the definition of $\widetilde{\boldsymbol{\beta}}^W$, we can decompose $\widetilde{\boldsymbol{\beta}}^W - \boldsymbol{\beta}$ as

$$\begin{aligned}\widetilde{\boldsymbol{\beta}}^W - \boldsymbol{\beta} &= J_1 + J_2 + J_3 \\ &= \left\{\hat{\mathbf{X}}^\tau \left(\mathbf{I}_{kTM} - \mathbf{P}_{\hat{\mathbf{B}}}\right)\left(\Omega^{-1} \otimes \mathbf{I}_{kT}\right)\left(\mathbf{I}_{kTM} - \mathbf{P}_{\hat{\mathbf{B}}}\right)\hat{\mathbf{X}}\right\}^{-1} \\ &\quad \times \left\{\hat{\mathbf{X}}^\tau \left(\mathbf{I}_{kTM} - \mathbf{P}_{\hat{\mathbf{B}}}\right)\left(\Omega^{-1} \otimes \mathbf{I}_{kT}\right)\left(\mathbf{I}_{kTM} - \mathbf{P}_{\hat{\mathbf{B}}}\right)\widetilde{\mathbf{G}}\right. \\ &\quad + \hat{\mathbf{X}}^\tau \left(\mathbf{I}_{kTM} - \mathbf{P}_{\hat{\mathbf{B}}}\right)\left(\Omega^{-1} \otimes \mathbf{I}_{kT}\right)\left(\mathbf{I}_{kTM} - \mathbf{P}_{\hat{\mathbf{B}}}\right)\widetilde{\boldsymbol{\nu}} \\ &\quad \left. + \hat{\mathbf{X}}^\tau \left(\mathbf{I}_{kTM} - \mathbf{P}_{\hat{\mathbf{B}}}\right)\left(\Omega^{-1} \otimes \mathbf{I}_{kT}\right)\left(\mathbf{I}_{kTM} - \mathbf{P}_{\hat{\mathbf{B}}}\right)\boldsymbol{\nu}\right\},\end{aligned}$$

where $\widetilde{\mathbf{G}} = \left(\widetilde{\mathbf{G}}_1^\tau, \ldots, \widetilde{\mathbf{G}}_M^\tau\right)^\tau$, $\widetilde{\mathbf{G}}_s = \left(\widetilde{g}_s(U_{s11}), \ldots, \widetilde{g}_s(U_{s1T}), \ldots, \widetilde{g}_s(U_{skT})\right)^\tau$ with

$$\widetilde{g}_s(U_{sit}) = g(U_{sit}) - (1,0)\left(\mathbf{D}_{sU_{sit}}^\tau \mathbf{W}_{sU_{sit}} \mathbf{D}_{sU_{sit}}\right)^{-1} \mathbf{D}_{sU_{sit}}^\tau \mathbf{W}_{sU_{sit}} \mathbf{G}_s,$$

$\widetilde{\boldsymbol{\nu}} = \left(\widetilde{\boldsymbol{\nu}}_1^\tau, \ldots, \widetilde{\boldsymbol{\nu}}_M^\tau\right)^\tau$, $\widetilde{\boldsymbol{\nu}}_s = \left(\widetilde{\nu}_{s11}, \ldots, \widetilde{\nu}_{s1T}, \ldots, \widetilde{\nu}_{skT}\right)^\tau$ with

$$\widetilde{\nu}_{sit} = -(1,0)\left(\mathbf{D}_{sU_{sit}}^\tau \mathbf{W}_{sU_{sit}} \mathbf{D}_{sU_{sit}}\right)^{-1} \mathbf{D}_{sU_{sit}}^\tau \mathbf{W}_{sU_{sit}} \boldsymbol{\nu}_s,$$

and $\boldsymbol{\nu} = (\boldsymbol{\nu}_1^\tau, \ldots, \boldsymbol{\nu}_M^\tau)^\tau$ with $\boldsymbol{\nu}_s = \left(\nu_{s11}, \ldots, \nu_{s1T}, \ldots, \nu_{skT}\right)^\tau$. Write

$$\begin{aligned}&\frac{1}{k}\hat{\mathbf{X}}^\tau \left(\mathbf{I}_{kTM} - \mathbf{P}_{\hat{\mathbf{B}}}\right)\left(\Omega^{-1} \otimes \mathbf{I}_{kT}\right)\left(\mathbf{I}_{kTM} - \mathbf{P}_{\hat{\mathbf{B}}}\right)\widetilde{\mathbf{G}} \\ &\quad = \left(\sum_{s_2=1}^{M} \frac{\sigma_\nu^{2s_1s_2}}{k} \hat{\mathbf{X}}_{s_1}^\tau \left(\mathbf{I}_{kT} - \mathbf{P}_{\hat{\mathbf{B}}_{s_1}}\right)\left(\mathbf{I}_{kT} - \mathbf{P}_{\hat{\mathbf{B}}_{s_2}}\right)\widetilde{\mathbf{G}}_{s_2}\right)_{s_1=1}^{M}.\end{aligned}$$

By the same arguments as for Theorem 1 of Fan et al. (2005) we obtain

$$\frac{1}{k}\hat{\mathbf{X}}_{s_1}^\tau \left(\mathbf{I}_{kT} - \mathbf{P}_{\hat{\mathbf{B}}_{s_1}}\right)\left(\mathbf{I}_{kT} - \mathbf{P}_{\hat{\mathbf{B}}_{s_2}}\right)\widetilde{\mathbf{G}}_{s_2} = o_p\left(k^{-1/2}\right).$$

It then follows from Lemma 3 that $J_1 = o_p\left(k^{-1/2}\right)$. Similarly, $J_2 = o_p\left(k^{-1/2}\right)$ and

$$\begin{aligned}&\frac{1}{k}\hat{\mathbf{X}}^\tau \left(\mathbf{I}_{kTM} - \mathbf{P}_{\hat{\mathbf{B}}}\right)\left(\Omega^{-1} \otimes \mathbf{I}_{kT}\right)\left(\mathbf{I}_{kTM} - \mathbf{P}_{\hat{\mathbf{B}}}\right)\boldsymbol{\nu} \\ &\quad = \left(\sum_{s_2=1}^{M} \frac{\sigma_\nu^{2s_1s_2}}{k} \boldsymbol{\Pi}_{s_1}^\tau \left(\mathbf{I}_{kT} - \mathbf{P}_{\mathbf{B}}\right)\boldsymbol{\nu}_{s_2}\right)_{s_1=1}^{M} + o_p\left(k^{-1/2}\right),\end{aligned}$$

where $\boldsymbol{\Pi}_s = \left(\boldsymbol{\Pi}_{s11}, \ldots, \boldsymbol{\Pi}_{s1T}, \ldots, \boldsymbol{\Pi}_{skT}\right)^\tau$ with $\boldsymbol{\Pi}_{sit} = \mathbf{X}_{sit} - \mathbb{E}\left(\mathbf{X}_{sit} | U_{sit}\right)$, $s = 1, \ldots, M$, $i = 1, \ldots, k$, and $t = 1, \ldots, T$. As a result, it suffices to prove

$$\sqrt{k}\left(\sum_{s_2=1}^{M} \frac{\sigma_\nu^{2s_1s_2}}{k} \boldsymbol{\Pi}_{j_1}^\tau \left(\mathbf{I}_{kn_0} - \mathbf{P}_{\mathbf{B}}\right)\boldsymbol{\nu}_s\right)_{s_1=1}^{M} \xrightarrow{D} N(\mathbf{0}, \Sigma_1) \quad \text{as} \quad k \to \infty. \tag{A.1}$$

https://doi.org/10.1017/S0266466613000352 Published online by Cambridge University Press

For any nonzero vector $\mathbf{\Lambda} = (\mathbf{\Lambda}_1, \ldots, \mathbf{\Lambda}_M)^\tau$ with $\mathbf{\Lambda}_s = (\Lambda_{s1}, \ldots, \Lambda_{sp_s})^\tau$,

$$\begin{aligned}
&\sqrt{k}\mathbf{\Lambda}^\tau \left(\sum_{s_2=1}^{M} \frac{\sigma_\nu^{2s_1s_2}}{k} \mathbf{\Pi}_{s_1}^\tau (\mathbf{I}_{kT} - \mathbf{P}_\mathbf{B}) \boldsymbol{\nu}_{s_2} \right)_{s_1=1}^{M} \\
&\quad = \frac{1}{\sqrt{k}} \sum_{s_1=1}^{M} \sum_{s_2=1}^{M} \sigma_\nu^{2s_1s_2} \mathbf{\Lambda}_{s_1}^\tau \mathbf{\Pi}_{s_1}^\tau (\mathbf{I}_{kT} - \mathbf{P}_\mathbf{B}) \boldsymbol{\nu}_{s_2} \\
&\quad = \frac{1}{\sqrt{k}} \sum_{s_1=1}^{M} \sum_{s_2=1}^{M} \sigma_\nu^{2s_1s_2} \sum_{m=1}^{p_{s_1}} \sum_{i=1}^{k} \sum_{t=1}^{T} \Lambda_{ms_1} \left(\Pi_{ms_1it} - \frac{1}{T} \sum_{t_1=1}^{T} \Pi_{ms_1it_1} \right) \nu_{s_2it} \\
&\quad = \frac{1}{\sqrt{k}} \sum_{i=1}^{k} \sum_{s_1=1}^{M} \sum_{s_2=1}^{M} \sigma_\nu^{2s_1s_2} \sum_{m=1}^{p_{s_1}} \sum_{t=1}^{T} \Lambda_{ms_1} \left(\Pi_{ms_1it} - \frac{1}{T} \sum_{t_1=1}^{T} \Pi_{ms_1it_1} \right) \nu_{s_2it} \\
&\quad = \frac{1}{\sqrt{k}} \sum_{i=1}^{k} \psi_i, \quad \text{say,}
\end{aligned}$$

where $\mathbf{\Pi}_{msit}$ is the mth entry of Π_{sit}. Note that $\psi_1, \ldots, \psi_k$ are independent random variables with mean zero. It is easy to check that $\{\psi_i\}_{i=1}^k$ satisfy the Linderberg condition. Moreover,

$$\begin{aligned}
&\mathrm{Cov} \left\{ \sqrt{k}\mathbf{\Lambda}^\tau \left(\sum_{s_2=1}^{M} \frac{\sigma_\nu^{2s_1s_2}}{k} \mathbf{\Pi}_{s_1}^\tau (\mathbf{I}_{kT} - \mathbf{P}_\mathbf{B}) \boldsymbol{\nu}_{s_2} \right)_{s_1=1}^{M} \right\} \\
&\quad = \frac{1}{k} \mathbf{\Lambda}^\tau \{\text{blockdiag } (\mathbf{\Pi}_1 (\mathbf{I}_{kT} - \mathbf{P}_\mathbf{B}), \ldots, \mathbf{\Pi}_M (\mathbf{I}_{kT} - \mathbf{P}_\mathbf{B}))\}^\tau (\Omega^{-1} \otimes \mathbf{I}_{kT}) \boldsymbol{\nu}\boldsymbol{\nu}^\tau (\Omega^{-1} \otimes \mathbf{I}_{kT}) \\
&\qquad \times \text{blockdiag } (\mathbf{\Pi}_1 (\mathbf{I}_{kT} - \mathbf{P}_\mathbf{B}), \ldots, \mathbf{\Pi}_M (\mathbf{I}_{kT} - \mathbf{P}_\mathbf{B})) \mathbf{\Lambda} \\
&\quad = \frac{1}{k} \mathbf{\Lambda}^\tau \{\text{blockdiag } (\mathbf{\Pi}_1 (\mathbf{I}_{kT} - \mathbf{P}_\mathbf{B}), \ldots, \mathbf{\Pi}_M (\mathbf{I}_{kT} - \mathbf{P}_\mathbf{B}))\}^\tau (\Omega^{-1} \otimes \mathbf{I}_{kT}) \\
&\qquad \times \text{blockdiag } (\mathbf{\Pi}_1 (\mathbf{I}_{kT} - \mathbf{P}_\mathbf{B}), \ldots, \mathbf{\Pi}_M (\mathbf{I}_{kT} - \mathbf{P}_\mathbf{B})) \mathbf{\Lambda} \\
&\quad = \frac{1}{k} \mathbf{\Lambda}^\tau (\sigma_\nu^{2s_1s_2} \mathbf{\Pi}_{s_1}^\tau (\mathbf{I}_{kT} - \mathbf{P}_\mathbf{B}) \mathbf{\Pi}_{s_2})_{s_1,s_2=1}^{M} \mathbf{\Lambda} \to_p \mathbf{\Lambda}^\tau \mathbf{\Sigma}_1 \mathbf{\Lambda} \quad \text{as } k \to \infty.
\end{aligned}$$

Therefore (A.1) follows from Lemma 3. ■

Proof of Theorem 1. According to the definition, $\hat{\sigma}_{\nu s_1 s_2}^2$ can be decomposed as

$$\begin{aligned}
\hat{\sigma}_{\nu s_1 s_2}^2 &= \frac{1}{k(T-1)} \boldsymbol{\nu}_{s_1}^\tau (\mathbf{I}_{kT} - \mathbf{P}_\mathbf{B}) \boldsymbol{\nu}_{s_2} \\
&\quad + \frac{1}{k(T-1)} \left(\boldsymbol{\beta}_{s_1} - \hat{\boldsymbol{\beta}}_{s_1} + \mathbf{G}_{s_1} - \hat{\mathbf{G}}_{s_1} \right)^\tau \mathbf{X}_{s_1}^\tau (\mathbf{I}_{kT} - \mathbf{P}_\mathbf{B}) \mathbf{X}_{s_2} \left(\boldsymbol{\beta}_{s_2} - \hat{\boldsymbol{\beta}}_{s_2} + \mathbf{G}_{s_2} - \hat{\mathbf{G}}_{s_2} \right) \\
&\quad + \frac{1}{k(T-1)} \left(\boldsymbol{\beta}_{s_1} - \hat{\boldsymbol{\beta}}_{s_1} + \mathbf{G}_{s_1} - \hat{\mathbf{G}}_{s_1} \right)^\tau \mathbf{X}_{s_1}^\tau (\mathbf{I}_{kT} - \mathbf{P}_\mathbf{B}) \boldsymbol{\nu}_{s_2} \\
&\quad + \frac{1}{k(T-1)} \left(\boldsymbol{\beta}_{s_2} - \hat{\boldsymbol{\beta}}_{s_2} + \mathbf{G}_{s_2} - \hat{\mathbf{G}}_{s_2} \right)^\tau \mathbf{X}_{s_2}^\tau (\mathbf{I}_{kT} - \mathbf{P}_\mathbf{B}) \boldsymbol{\nu}_{s_1} \\
&= J_1 + J_2 + J_3 + J_4, \quad \text{say.}
\end{aligned}$$

https://doi.org/10.1017/S0266466613000352 Published online by Cambridge University Press

Write

$$
\begin{aligned}
k(T-1)J_1 &= \boldsymbol{\nu}_{s_1}^{\tau}(\mathbf{I}_{kT}-\mathbf{P}_{\mathbf{B}})\boldsymbol{\nu}_{s_2} = \boldsymbol{\nu}_{s_1}^{\tau}\left(\mathbf{I}_{kT}-\frac{1}{T}\mathbf{I}_k\otimes\boldsymbol{\iota}_T\boldsymbol{\iota}_T^{\tau}\right)\boldsymbol{\nu}_{s_2} \\
&= \sum_{i=1}^{k}\sum_{t=1}^{T}\nu_{s_1it}\nu_{s_2it} - \sum_{i=1}^{k}\frac{1}{T}\left\{\sum_{t=1}^{T}\nu_{s_1it}\right\}\left\{\sum_{t=1}^{T}\nu_{s_2it}\right\} \\
&= \sum_{i=1}^{k}\sum_{t=1}^{T}\left(\nu_{s_1it}-\frac{1}{T}\sum_{t_1=1}^{T}\nu_{s_1it_1}\right)\left(\nu_{s_2it}-\frac{1}{T}\sum_{t_1=1}^{T}\nu_{s_2it_1}\right) \\
&= \sum_{i=1}^{k}\zeta_{s_1s_2i}, \quad \text{say.}
\end{aligned}
$$

The variables $\{\zeta_{s_1s_2i} : i = 1,\ldots,k\}$ are independent random variables with mean

$$
\mathbb{E}[\zeta_{s_1s_2i}] = \sum_{t=1}^{T}\left[\frac{(T-1)^2}{T^2}\mathbb{E}(\nu_{s_1it}\nu_{s_2it}) + \frac{1}{T^2}\sum_{t_1\neq t}^{T}\mathbb{E}(\nu_{s_1it_1}\nu_{s_2it_1})\right] = (T-1)\sigma^2_{\nu s_1s_2}
$$

and variance

$$
\begin{aligned}
\mathrm{Var}\left(\zeta_{s_1s_2i}\right) &= \mathbb{E}\left(\zeta^2_{s_1s_2i}\right) - (\mathbb{E}\zeta_{s_1s_2i})^2 = \mathbb{E}\left[\sum_{t=1}^{T}\nu_{s_1it}\nu_{s_2it}\right]^2 \\
&\quad + \frac{1}{T^2}\mathbb{E}\left[\left(\sum_{t_1=1}^{T}\nu_{s_1it_1}\right)^2\left(\sum_{t_1=1}^{T}\nu_{s_2it_1}\right)^2\right] \\
&\quad - \frac{2}{T}\mathbb{E}\left[\sum_{t=1}^{T}(\nu_{s_1it}\varepsilon_{s_2it})\left(\sum_{t_1=1}^{T}\nu_{s_1it_1}\right)\left(\sum_{t_1=1}^{T}\nu_{s_1it_1}\right)\right] - (T-1)^2\sigma^4_{\nu s_1s_2} \\
&= T\mathbb{E}\left(\nu^2_{s_111}\nu^2_{s_211}\right) + T(T-1)\sigma^4_{\nu s_1s_2} + \frac{1}{T}\mathbb{E}\left(\nu^2_{s_111}\nu^2_{s_211}\right) + \frac{6}{T^2}\frac{T(T-1)}{2}\sigma^4_{\nu s_1s_2} \\
&\quad - \frac{2}{T}\left\{T\mathbb{E}\left(\nu^2_{s_111}\nu^2_{s_211}\right) + T(T-1)\sigma^2_{\nu s_1s_1}\sigma^2_{\nu s_2s_2}\right\} - (T-1)\sigma^2_{\nu s_1s_1}\sigma^2_{\nu s_2s_2} \\
&= \left(T+\frac{1}{T}-2\right)\mathbb{E}\left(\nu^2_{s_111}\nu^2_{s_211}\right) + (T-1)\left(-1+\frac{3}{T}\right)\left(\sigma^2_{\nu s_1s_1}\sigma^2_{\nu s_2s_2}\right).
\end{aligned}
$$

Hence $\mathbb{E}[J_1] = \sigma^2_{\nu s_1s_2}$, and

$$
\mathrm{Var}(J_1) = \frac{1}{k(T-1)^2}\left\{\frac{(T-1)^2}{T}\mathbb{E}\left(\nu^2_{s_111}\nu^2_{s_211}\right) - \frac{(T-1)(T-3)}{T}\left(\sigma^2_{\nu s_1s_1}\sigma^2_{\nu s_2s_2}\right)\right\} = \frac{\kappa_{s_1s_2}}{kT}.
$$

Next, according to Lemma 1,

$$
J_2 \leq \frac{T}{k(T-1)}\left(\boldsymbol{\beta}_{s_1}-\hat{\boldsymbol{\beta}}_{s_1}+\mathbf{G}_{s_1}-\hat{\mathbf{G}}_{s_1}\right)^{\tau}\mathbf{X}_{s_1}^{\tau}\mathbf{X}_{s_2}\left(\boldsymbol{\beta}_{s_2}-\hat{\boldsymbol{\beta}}_{s_2}+\mathbf{G}_{s_2}-\hat{\mathbf{G}}_{s_2}\right) = o_p\left(k^{-1/2}\right).
$$

https://doi.org/10.1017/S0266466613000352 Published online by Cambridge University Press

Furthermore, J_3 can be decomposed as

$$\begin{aligned}J_3 &= \frac{1}{k(T-1)}\sum_{i=1}^{k}\sum_{t=1}^{T}\left(v_{s_1it}-\frac{1}{T}\sum_{t_1=1}^{T}v_{s_1it_1}\right)\\ &\quad\times\left\{\mathbf{X}_{s_2it}^{\tau}\left(\boldsymbol{\beta}_{s_2}-\hat{\boldsymbol{\beta}}_{s_2}\right)-\frac{1}{T}\sum_{t_1=1}^{T}\mathbf{X}_{s_2it_1}^{\tau}\left(\boldsymbol{\beta}_{s_2}-\hat{\boldsymbol{\beta}}_{s_2}\right)\right\}\\ &\quad+\frac{1}{k(T-1)}\sum_{i=1}^{k}\sum_{t=1}^{T}\left(v_{s_1it}-\frac{1}{T}\sum_{t_1=1}^{T}v_{s_1it_1}\right)\left\{g_{s_2}(U_{s_2it})-\hat{g}_{s_2}(U_{s_2it})\right\}\\ &\quad-\frac{1}{k(T-1)}\sum_{i=1}^{k}\sum_{t=1}^{T}\left(v_{isj_1}-\frac{1}{T}\sum_{t=1}^{T}v_{s_1it_1}\right)\frac{1}{T}\sum_{t_1=1}^{T}\left\{g_{s_2}(U_{s_2it_1})-\hat{g}_{s_2}(U_{(s_2it_1)})\right\}\\ &= J_{31}+J_{32}-J_{33},\ \text{ say.}\end{aligned}$$

Because

$$\sum_{i=1}^{k}\sum_{t=1}^{T}\left(v_{s_1it}-\frac{1}{T}\sum_{t=1}^{T}v_{s_1it_1}\right)\mathbf{X}_{s_2it}\quad\text{and}\quad\sum_{i=1}^{k}\sum_{t=1}^{T}\left(v_{s_1it}-\frac{1}{T}\sum_{t=1}^{T}v_{s_1it_1}\right)\frac{1}{T}\sum_{t_1=1}^{T}\mathbf{X}_{s_2it_1}$$

are both of order $O_p(k^{1/2})$, the root-k consistency of $\hat{\boldsymbol{\beta}}_s$ leads to $J_{31}=o_p(k^{-1/2})$.

Let $\mathbf{d}_{it}$ denote a kT-vector with its $\{(i-1)T+t\}$th element being 1 and others 0. According to the definition of $\hat{g}_s(\cdot)$,

$$\begin{aligned}\hat{g}_s(U_{sit}) &= \mathbf{d}_{it}^{\tau}\left\{\mathbf{I}_{kT}-\left(\mathbf{I}_{kT}-\mathbf{S}_s\overline{\mathbf{P}}_{\mathbf{B}}\right)^{-1}\left(\mathbf{I}_{kT}-\mathbf{S}_s\right)\right\}\left(\mathbf{Y}_s-\mathbf{X}_s\hat{\boldsymbol{\beta}}_s\right)\\ &= \mathbf{d}_{it}^{\tau}\left\{\mathbf{I}_{kT}-\left(\mathbf{I}_{kT}-\mathbf{S}_s\overline{\mathbf{P}}_{\mathbf{B}}\right)^{-1}\left(\mathbf{I}_{kT}-\mathbf{S}_s\right)\right\}\mathbf{X}_s\left(\boldsymbol{\beta}_s-\hat{\boldsymbol{\beta}}_s\right)\\ &\quad+\mathbf{d}_{it}^{\tau}\left\{\mathbf{I}_{kT}-\left(\mathbf{I}_{kT}-\mathbf{S}_s\overline{\mathbf{P}}_{\mathbf{B}}\right)^{-1}\left(\mathbf{I}_{kT}-\mathbf{S}_s\right)\right\}\mathbf{G}_s\\ &\quad+\mathbf{d}_{it}^{\tau}\left\{\mathbf{I}_{kT}-\left(\mathbf{I}_{kT}-\mathbf{S}_s\overline{\mathbf{P}}_{\mathbf{B}}\right)^{-1}\left(\mathbf{I}_{kT}-\mathbf{S}_s\right)\right\}\boldsymbol{\nu}_s\\ &\quad+\mathbf{d}_{it}^{\tau}(\mathbf{I}_{kT}-\mathbf{S}_s\overline{\mathbf{P}}_{\mathbf{B}})^{-1}\mathbf{S}_s(\mathbf{I}_{kT}-\overline{\mathbf{P}}_{\mathbf{B}})\mathbf{B}\boldsymbol{\mu}_s.\end{aligned}$$

Because $\sum_{i=1}^{k}\mu_{si}=0$ for $s=1,\ldots,M$, it is easy to see $\left(\mathbf{I}_{kT}-\mathbf{S}_s\overline{\mathbf{P}}_{\mathbf{B}}\right)^{-1}\mathbf{S}_s\left(\mathbf{I}_{kT}-\overline{\mathbf{P}}_{\mathbf{B}}\right)\mathbf{B}\boldsymbol{\mu}_s=0$.

Consequently,

$$\begin{aligned}J_{32} &= \frac{1}{k(T-1)}\sum_{i=1}^{k}\sum_{t=1}^{T}\left(v_{s_1it}-\frac{1}{T}\sum_{t=1}^{T}v_{s_1it_1}\right)\\ &\quad\times\mathbf{d}_{it}^{\tau}\left\{\mathbf{I}_{kT}-(\mathbf{I}_{kT}-\mathbf{S}_s\overline{\mathbf{P}}_{\mathbf{B}})^{-1}(\mathbf{I}_{kT}-\mathbf{S}_s)\right\}\mathbf{X}_s(\boldsymbol{\beta}_s-\hat{\boldsymbol{\beta}}_s)\\ &\quad-\frac{1}{k(T-1)}\sum_{i=1}^{k}\sum_{t=1}^{T}\left(v_{s_1it}-\frac{1}{T}\sum_{t_1=1}^{T}v_{s_1it_1}\right)\mathbf{d}_{it}^{\tau}(\mathbf{I}_{kT}-\mathbf{S}_s\overline{\mathbf{P}}_{\mathbf{B}})^{-1}(\mathbf{I}_{kT}-\mathbf{S}_s)\mathbf{G}_s\\ &\quad+\frac{1}{k(T-1)}\sum_{i=1}^{k}\sum_{t=1}^{T}\left(v_{s_1it}-\frac{1}{T}\sum_{t_1=1}^{T}v_{s_1it_1}\right)\mathbf{d}_{it}^{\tau}\left\{\mathbf{I}_{kT}-(\mathbf{I}_{kT}-\mathbf{S}_s\overline{\mathbf{P}}_{\mathbf{B}})^{-1}(\mathbf{I}_{kT}-\mathbf{S}_s)\right\}\boldsymbol{\nu}_s\\ &= J_{321}+J_{322}+J_{323},\quad\text{ say.}\end{aligned}$$

https://doi.org/10.1017/S0266466613000352 Published online by Cambridge University Press

By the root-k consistency of $\hat{\boldsymbol{\beta}}_s$ and the proof of Theorem 3 of Fan et al. (2005),

$$\begin{aligned} J_{321}^2 &= \left\{\frac{1}{k(T-1)}\right\}^2 \left\| \sum_{i=1}^{k}\sum_{t=1}^{T}\left(\nu_{s_1it} - \frac{1}{T}\sum_{t_1=1}^{T}\nu_{s_1it_1}\right)\right. \\ &\quad \left.\times \left\{\mathbf{I}_{kT} - \mathbf{d}_{it}^{\tau}\left(\mathbf{I}_{kT} - \mathbf{S}_s\overline{\mathbf{P}}_{\mathbf{B}}\right)^{-1}\left(\mathbf{I}_{kT} - \mathbf{S}_s\right)\right\}\right\|^2 \\ &\quad \times \|\mathbf{X}_s(\boldsymbol{\beta}_s - \hat{\boldsymbol{\beta}}_s)\| \\ &= O(k^{-1}) \cdot O_p(h_s^4 + 1/kh_s) = o_p(k^{-1}), \end{aligned}$$

so that $J_{321} = o_p(k^{-1/2})$. Moreover, $J_{322} = o_p(k^{-1/2})$ follows from

$$(\mathbf{I}_{kT} - \mathbf{S}_s)\mathbf{G}_s = \alpha_{2s}/2\left(g_s''(U_{s11}), \ldots, g_s''(U_{s1T}), \ldots, g_s''(U_{skT})\right)^{\tau} h_s^2 + o_p(h_s^2),$$

and $J_{323} = o_p(k^{-1/2})$ follows from the same argument as proving Lemma 4(2) in You and Chen (2006). Thus $J_3 = o_p(k^{-1/2})$. The proof for $J_4 = o_p(k^{-1/2})$ is similar. This completes the proof of Theorem 1. ■

Proof of Theorem 2. Write

$$\begin{aligned} \nabla_{sit} &= \mathbf{X}_{sit}^{\tau}\left(\boldsymbol{\beta}_s - \hat{\boldsymbol{\beta}}_s\right) - \frac{1}{T}\sum_{t_1=1}^{T}\mathbf{X}_{sit_1}^{\tau}\left(\boldsymbol{\beta}_s - \hat{\boldsymbol{\beta}}_s\right) + g_s(U_{sit}) - \hat{g}_s(U_{sit}) \\ &\quad - \frac{1}{T}\sum_{t_1=1}^{T}\left[g_s(U_{(sit_1)}) - \hat{g}_s(U_{sit_1})\right]. \end{aligned}$$

Then

$$\begin{aligned} \frac{1}{kT}\sum_{i=1}^{k}\sum_{t=1}^{T}\hat{\zeta}_{s_1it}^2\hat{\zeta}_{sit_2}^2 &= \frac{1}{kT}\sum_{i=1}^{k}\sum_{t=1}^{T}\left(\nu_{s_1it} - \frac{1}{T}\sum_{t_1=1}^{T}\nu_{s_1it_1}\right)^2\left(\nu_{s_2it} - \frac{1}{T}\sum_{t_1=1}^{T}\nu_{s_2it_1}\right)^2 \\ &\quad + \frac{1}{kT}\sum_{i=1}^{k}\sum_{t=1}^{T}\nabla_{s_1it}^2\nabla_{s_2it}^2 + \frac{4}{kT}\sum_{i=1}^{k}\sum_{t=1}^{T}\left(\nu_{s_1it} - \frac{1}{T}\sum_{t_1=1}^{T}\nu_{s_1it_1}\right)^2 \\ &\quad \times \left(\nu_{s_2it} - \frac{1}{T}\sum_{t_1=1}^{T}\nu_{s_2it_1}\right)\nabla_{s_2it} \\ &\quad + \frac{4}{kT}\sum_{i=1}^{k}\sum_{t=1}^{T}\left(\nu_{s_1it} - \frac{1}{T}\sum_{t_1=1}^{T}\nu_{s_1it_1}\right)^2\nabla_{s_2it}^2 \\ &\quad + \frac{4}{kT}\sum_{i=1}^{k}\sum_{t=1}^{T}\left(\nu_{s_1it} - \frac{1}{T}\sum_{t_1=1}^{T}\nu_{sit_1}\right)\nabla_{s_1it_1}\nabla_{s_2it} \\ &= J_1 + \cdots + J_5, \quad \text{say.} \end{aligned}$$

https://doi.org/10.1017/S0266466613000352 Published online by Cambridge University Press

Calculate

$$
\begin{aligned}
J_1 &= \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\left(\nu_{s_1it}-\frac{1}{T}\sum_{t_1=1}^{T}\nu_{s_1it_1}\right)^2\right]\left(\nu_{s_1it}-\frac{1}{T}\sum_{t_1=1}^{T}\nu_{s_1it_1}\right)^2+o_p(1)\\
&= \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\nu_{s_1it}^2\nu_{s_2it}^2+\frac{1}{T^4}\left(\sum_{t_1=1}^{T}\nu_{s_1it_1}\right)^2\left(\sum_{t_1=1}^{T}\nu_{s_2it_1}\right)^2+\frac{4}{T^2}\nu_{s_1it}^2\left(\sum_{t_1=1}^{T}\nu_{s_2it_1}\right)^2\right.\\
&\qquad +\frac{2}{T^2}\nu_{s_2it}^2\left(\sum_{t_1=1}^{T}\nu_{s_1it_1}\right)^2-\frac{4}{T}\nu_{s_1it}^2\nu_{s_2it}\sum_{t_1=1}^{T}\nu_{s_2it_1}\\
&\qquad \left.-\frac{4}{T^3}\left(\sum_{t_1=1}^{T}\nu_{s_1it_1}\right)^2\left(\sum_{t_1=1}^{T}\nu_{s_2it_1}\right)\nu_{s_2it}\right]\\
&= \mathbb{E}\left[\nu_{s_111}^2\nu_{s_211}^2\right]+\frac{6}{T^3}\sum_{t=1}^{T}\mathbb{E}\left[\nu_{s_11t}^2\left(\sum_{t_1=1}^{T}\nu_{s_21t_1}\right)^2\right]\\
&\quad -\frac{4}{T^2}\sum_{s=1}^{T}\mathbb{E}\left[\nu_{s_1it}^2\nu_{s_2it}^2\sum_{t_1=1}^{T}\nu_{s_2it_1}\right]-\frac{3}{T^4}\mathbb{E}\left(\sum_{t=1}^{T}\nu_{s_11t}\right)^2\left(\sum_{t=1}^{T}\nu_{s_21t}\right)^2\\
&= \left(1+\frac{6}{T^2}-\frac{3}{T^3}-\frac{4}{T}\right)\mathbb{E}\left[\nu_{s_111}^2\nu_{s_211}^2\right]+\frac{(T-1)}{T^2}\left(6-\frac{9}{T}\right)\sigma_{\nu s_1s_1}^2\sigma_{\nu s_2s_2}^2.
\end{aligned}
$$

By Lemma 1, we can show that $\sum_{i=1}^{k}\sum_{t=1}^{T}\nabla_{s_1it}^2\nabla_{s_2it}^2=o_p(1)$. Therefore, the Hölder inequality leads to $J_j=o_p(1)$ for $j=2,\ldots,5$. The theorem then follows. ■

Proof of Theorem 3. Let $\hat{\mathbf{X}}^*=(\mathbf{I}_{kTM}-\mathbf{P}_{\hat{\mathbf{B}}})\hat{\mathbf{X}}$. Then

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}^W-\boldsymbol{\beta} &= \widetilde{\boldsymbol{\beta}}^W-\boldsymbol{\beta}+\left\{\hat{\mathbf{X}}^{*\tau}(\hat{\Omega}^{-1}\otimes\mathbf{I}_{kT})\hat{\mathbf{X}}\right\}^{-1}\left\{\hat{\mathbf{X}}^{*\tau}(\hat{\Omega}^{-1}\otimes\mathbf{I}_{kT})\widetilde{\mathbf{G}}-\hat{\mathbf{X}}^{*\tau}(\Omega^{-1}\otimes\mathbf{I}_{kT})\widetilde{\mathbf{G}}\right\}\\
&\quad +\left[\left\{\hat{\mathbf{X}}^{*\tau}(\hat{\Omega}^{-1}\otimes\mathbf{I}_{kT})\hat{\mathbf{X}}^*\right\}^{-1}-\left\{\hat{\mathbf{X}}^{*\tau}(\Omega^{-1}\otimes\mathbf{I}_{kT})\hat{\mathbf{X}}^*\right\}^{-1}\right]\hat{\mathbf{X}}^{*\tau}(\hat{\Omega}^{-1}\otimes\mathbf{I}_{kT})\widetilde{\mathbf{G}}\\
&\quad +\left\{\hat{\mathbf{X}}^{*\tau}(\hat{\Omega}^{-1}\otimes\mathbf{I}_{kT})\hat{\mathbf{X}}\right\}^{-1}\left\{\hat{\mathbf{X}}^{*\tau}(\hat{\Omega}^{-1}\otimes\mathbf{I}_{kT})\widetilde{\boldsymbol{\nu}}-\hat{\mathbf{X}}^{*\tau}(\Omega^{-1}\otimes\mathbf{I}_{kT})\widetilde{\boldsymbol{\nu}}\right\}\\
&\quad +\left\{\hat{\mathbf{X}}^{*\tau}(\hat{\Omega}^{-1}\otimes\mathbf{I}_{kT})\hat{\mathbf{X}}\right\}^{-1}\left\{\hat{\mathbf{X}}^{*\tau}(\hat{\Omega}^{-1}\otimes\mathbf{I}_{kT})\boldsymbol{\nu}-\hat{\mathbf{X}}^{*\tau}(\Omega^{-1}\otimes\mathbf{I}_{kT})\boldsymbol{\nu}\right\}\\
&\quad +\left[\left\{\hat{\mathbf{X}}^{*\tau}(\hat{\Omega}^{-1}\otimes\mathbf{I}_{kT})\hat{\mathbf{X}}^*\right\}^{-1}-\left\{\hat{\mathbf{X}}^{*\tau}(\Omega^{-1}\otimes\mathbf{I}_{kT})\hat{\mathbf{X}}^*\right\}^{-1}\right]\hat{\mathbf{X}}^{*\tau}(\hat{\Omega}^{-1}\otimes\mathbf{I}_{kT})(\boldsymbol{\nu}-\widetilde{\boldsymbol{\nu}}),
\end{aligned}
$$

where $\widetilde{\mathbf{G}}$ and $\widetilde{\boldsymbol{\nu}}$ are defined in the proof of Lemma 4. According to Lemma 4, it suffices to prove $\widetilde{\boldsymbol{\beta}}^W-\hat{\boldsymbol{\beta}}^W=o_p(k^{-1/2})$. Because $(\mathbf{A}+a\mathbf{B})^{-1}=\mathbf{A}^{-1}-a\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}+O(a^2)$ as

https://doi.org/10.1017/S0266466613000352 Published online by Cambridge University Press

$a \to 0$, Theorem 3 will follow from (A.2)–(A.7):

$$\frac{1}{k}\left\{\hat{\mathbf{X}}^{*\tau}(\hat{\Omega}^{-1}\otimes\mathbf{I}_{kT})\hat{\mathbf{X}}^{*}-\hat{\mathbf{X}}^{*\tau}(\Omega^{-1}\otimes\mathbf{I}_{kT})\hat{\mathbf{X}}^{*}\right\}=o_p(1), \tag{A.2}$$

$$\frac{1}{k}\left\{\hat{\mathbf{X}}^{*\tau}(\hat{\Omega}^{-1}\otimes\mathbf{I}_{kT})\widetilde{\mathbf{G}}-\hat{\mathbf{X}}^{*\tau}(\Omega^{-1}\otimes\mathbf{I}_{kT})\widetilde{\mathbf{G}}\right\}=o_p(k^{-1/2}), \tag{A.3}$$

$$\frac{1}{k}\left\{\hat{\mathbf{X}}^{*\tau}(\hat{\Omega}^{-1}\otimes\mathbf{I}_{kT})\widetilde{\boldsymbol{\nu}}-\hat{\mathbf{X}}^{*\tau}(\Omega^{-1}\otimes\mathbf{I}_{kT})\widetilde{\boldsymbol{\nu}}\right\}=o_p(k^{-1/2}), \tag{A.4}$$

$$\frac{1}{k}\left\{\hat{\mathbf{X}}^{*\tau}(\hat{\Omega}^{-1}\otimes\mathbf{I}_{kT})\boldsymbol{\nu}-\hat{\mathbf{X}}^{*\tau}(\Omega^{-1}\otimes\mathbf{I}_{kT})\boldsymbol{\nu}\right\}=o_p(k^{-1/2}), \tag{A.5}$$

$$\begin{aligned} k\left\{\hat{\mathbf{X}}^{*\tau}(\hat{\Omega}^{-1}\otimes\mathbf{I}_{kT})\hat{\mathbf{X}}\right\}^{-1}&=O_p(1),\\ k\left\{\hat{\mathbf{X}}^{*\tau}(\Omega^{-1}\otimes\mathbf{I}_{kT})\hat{\mathbf{X}}\right\}^{-1}&=O_p(1), \end{aligned} \tag{A.6}$$

$$\begin{aligned} \frac{1}{k}\hat{\mathbf{X}}^{*\tau}\left(\hat{\Omega}^{-1}\otimes\mathbf{I}_{kT}\right)\widetilde{\mathbf{G}}&=o_p\left(k^{-1/2}\right), \quad \text{and}\\ \frac{1}{k}\hat{\mathbf{X}}^{*\tau}\left(\Omega^{-1}\otimes\mathbf{I}_{kT}\right)\widetilde{\boldsymbol{\nu}}&=O_p(k^{-1/2}). \end{aligned} \tag{A.7}$$

It is easy to see that

$$\frac{1}{k}\left\{\hat{\mathbf{X}}^{*\tau}(\hat{\Omega}^{-1}\otimes\mathbf{I}_{kT})\hat{\mathbf{X}}^{*}-\hat{\mathbf{X}}^{*\tau}(\Omega^{-1}\otimes\mathbf{I}_{kT})\hat{\mathbf{X}}^{*}\right\}=\left((\hat{\sigma}_{\nu}^{2s_1s_2}-\sigma_{\nu}^{2s_1s_2})\frac{1}{k}\hat{\mathbf{X}}_{s_1}^{*\tau}\hat{\mathbf{X}}_{s_2}^{*}\right)_{s_1,s_2}^{M}.$$

Thus (A.2) follows from Theorem 1 together with the proof of Lemma 3. By the same arguments combined with the proof of Lemma 4, we can prove (A.3)–(A.5) also. Finally, (A.6)–(A.7) follow from the proof process of Lemma 4 together with Theorem 1. ■

Proof of Theorem 4. The proof of Theorem 4 is straightforward. We here omit the details.

Proof of Theorem 5. Let

$$\Delta_i^s(-t)=\frac{1}{T-1}\sum_{t_1\neq t}^{T}\left(\mathbf{X}_{sit_1}^{\tau}(\boldsymbol{\beta}_s-\hat{\boldsymbol{\beta}}_s)+g_s(U_{sit_1})-\hat{g}_s(U_{sit_1})\right)$$

for $i=1,\ldots,k$, $t=1,\ldots,T$, and $s=1,\ldots,M$.

According to the definition of $\hat{g}_s^{TS}(u)$,

$$\begin{aligned} \hat{g}_s^{TS}(u)&=(1,0)(\mathbf{D}_{su}^{\tau}\mathbf{W}_{su}^{*}\mathbf{D}_{su}^{*})^{-1}\mathbf{D}_{su}^{*\tau}\mathbf{W}_{su}^{*}\\ &\quad\times\left\{\boldsymbol{\nu}_s-\left(\frac{1}{T-1}\sum_{t_1\neq 1}^{T}\nu_{s1t_1},\ldots,\frac{1}{T-1}\sum_{t_1\neq T}^{T}\nu_{s1t_1},\ldots,\frac{1}{T-1}\sum_{t_1\neq T}^{T}\nu_{skt_1}\right)^{\tau}\right\}\\ &\quad+(1,0)(\mathbf{D}_{su}^{*\tau}\mathbf{W}_{su}^{*}\mathbf{D}_{su}^{*})^{-1}\mathbf{D}_{su}^{*\tau}\mathbf{W}_{su}^{*}\left(\Delta_1^s(-1),\ldots,\Delta_1^s(-T),\ldots,\Delta_k^s(-T)\right)^{\tau}\\ &\quad+(1,0)(\mathbf{D}_{su}^{*\tau}\mathbf{W}_{su}^{*}\mathbf{D}_{su}^{*})^{-1}\mathbf{D}_{su}^{*\tau}\mathbf{W}_{su}^{*}\mathbf{X}_s(\boldsymbol{\beta}_s-\hat{\boldsymbol{\beta}}_s^{W})\\ &\quad+(1,0)(\mathbf{D}_{su}^{*\tau}\mathbf{W}_{su}^{*}\mathbf{D}_{su}^{*})^{-1}\mathbf{D}_{su}^{*\tau}\mathbf{W}_{su}^{*}\mathbf{G}_s\\ &=J_1+J_2+J_3+J_4, \quad \text{say.} \end{aligned}$$

https://doi.org/10.1017/S0266466613000352 Published online by Cambridge University Press

It is easy to see that

$$\mathbf{D}_{su}^{*\tau}\mathbf{W}_{su}^{*}\mathbf{D}_{su}^{*}=\begin{pmatrix}\sum_{i=1}^{k}\sum_{t=1}^{T}K_{h_s^*}(U_{sit}-u) & \sum_{i=1}^{k}\sum_{t=1}^{T}\left(\frac{U_{sit}-u}{h_s^*}\right)K_{h_s^*}(U_{sit}-u)\\ \sum_{i=1}^{k}\sum_{t=1}^{T}\left(\frac{U_{sit}-u}{h_s^*}\right)K_{h_s^*}(U_{sit}-u) & \sum_{i=1}^{k}\sum_{t=1}^{T}\left(\frac{U_{sit}-u}{h_s^*}\right)^2K_{h_s^*}(U_{sit}-u)\end{pmatrix}.$$

Each element of the preceding matrix is in the form of kernel regression. By Lemma 2,

$$\mathbf{D}_{su}^{*\tau}\mathbf{W}_{su}^{*}\mathbf{D}_{su}=kTf_s(u)\otimes\begin{pmatrix}1&0\\0&\alpha_2\end{pmatrix}\left[1+\left\{\frac{\log(1/h_s^*)}{kh_s^*}\right\}^{1/2}\right]$$

holds uniformly in $\mathcal{U}_s$. The same argument leads to

$$\mathbf{D}_{su}^{*\tau}\mathbf{W}_{su}^{*}\mathbf{X}_s=kTf_s(u)\mathbb{E}(\boldsymbol{\iota}_{p_s}^{\tau}\mathbf{X}_{s11}|U_{s11}=u)\otimes\begin{pmatrix}1&0\\0&\alpha_2\end{pmatrix}\left[1+\left\{\frac{\log(1/h_s^*)}{kh_s^*}\right\}^{1/2}\right].$$

Combining these with the fact $\|\boldsymbol{\beta}_s-\hat{\boldsymbol{\beta}}_s\|=O_p(k^{-1/2})$, we get $J_3=O_p(k^{-1/2})$. Moreover, let

$$\Delta_i^{s(1)}(-t)=\frac{1}{T-1}\sum_{t_1\neq t}^{T}\mathbf{X}_{sit_1}^{\tau}(\boldsymbol{\beta}_s-\hat{\boldsymbol{\beta}}_s)\quad\text{and}$$

$$\Delta_i^{s(2)}(-t)=\frac{1}{T-1}\sum_{t_1\neq t}^{T}(g_s(U_{sit_1})-\hat{g}_s(U_{sit_1}))$$

for $i=1,\ldots,k, t=1,\ldots,T$, and $s=1,\ldots,M$. Then

$$\begin{aligned}J_2&=(1,0)(\mathbf{D}_{su}^{*\tau}\mathbf{W}_{su}^{*}\mathbf{D}_{su}^{*})^{-1}\begin{pmatrix}\sum_{i=1}^{k}\sum_{t=1}^{T}K_{h_s^*}(U_{sit}-u)\Delta_i^{s(1)}(-t)\\ \sum_{i=1}^{k}\sum_{t=1}^{T}\frac{U_{sit}-u}{h_s^*}K_{h_s^*}(U_{sit}-u)\Delta_i^{s(1)}(-t)\end{pmatrix}\\&\quad+(1,0)(\mathbf{D}_{su}^{*\tau}\mathbf{W}_{su}^{*}\mathbf{D}_{su}^{*})^{-1}\begin{pmatrix}\sum_{i=1}^{k}\sum_{t=1}^{T}K_{h_s^*}(U_{sit}-u)\Delta_i^{s(2)}(-t)\\ \sum_{i=1}^{k}\sum_{t=1}^{T}\frac{U_{sit}-u}{h_s^*}K_{h_s^*}(U_{sit}-u)\Delta_i^{s(2)}(-t)\end{pmatrix}\\&=J_{21}+J_{22},\quad\text{say.}\end{aligned}$$

By the root-k consistency of $\hat{\boldsymbol{\beta}}_s$ together with the proof for J_3 in the proof of Theorem 1, it is easy to see $J_{21}=O_p(k^{-1/2})$. Further,

$$\begin{aligned}&(1,0)(\mathbf{D}_{su}^{*\tau}\mathbf{W}_{su}^{*}\mathbf{D}_{su}^{*})^{-1}\sum_{i=1}^{k}\sum_{t=1}^{T}K_{h_s^*}(U_{sit}-u)\Delta_i^{s(2)}(-t)\\&\quad=(1,0)(\mathbf{D}_{su}^{*\tau}\mathbf{W}_{su}^{*}\mathbf{D}_{su}^{*})^{-1}\sum_{i=1}^{k}\sum_{t=1}^{T}\frac{K_{h_s^*}(U_{sit}-u)}{T-1}\\&\qquad\times\sum_{t_1\neq t}^{T}\left[g_s(U_{sit_1})-\mathbf{d}_{it_1}^{\tau}\left\{\mathbf{I}_{kT}-(\mathbf{I}_{kT}-\mathbf{S}_s\overline{\mathbf{P}}_{\mathbf{B}})^{-1}(\mathbf{I}_{kT}-\mathbf{S}_s)\right\}\mathbf{G}_s\right]\\&\qquad+(1,0)(\mathbf{D}_{su}^{*\tau}\mathbf{W}_{su}^{*}\mathbf{D}_{su}^{*})^{-1}\sum_{i=1}^{k}\sum_{t=1}^{T}K_{h_s^*}(U_{sit}-u)\\&\qquad\times\frac{1}{T-1}\sum_{t_1\neq t}^{T}\mathbf{d}_{it}^{\tau}\left\{\mathbf{I}_{kT}-(\mathbf{I}_{kT}-\mathbf{S}_s\overline{\mathbf{P}}_{\mathbf{B}})^{-1}(\mathbf{I}_{kT}-\mathbf{S}_s)\right\}\mathbf{X}_s(\boldsymbol{\beta}_s-\hat{\boldsymbol{\beta}}_s)\end{aligned}$$

https://doi.org/10.1017/S0266466613000352 Published online by Cambridge University Press

$$+(1,0)(\mathbf{D}_{su}^{*\tau}\mathbf{W}_{su}^{*}\mathbf{D}_{su}^{*})^{-1}\sum_{i=1}^{k}\sum_{t=1}^{T}\frac{K_{h_s^*}(U_{sit}-u)}{T-1}$$

$$\times\sum_{t_1\neq t}^{T}\mathbf{d}_{it}^{\tau}\left\{\mathbf{I}_{kT}-(\mathbf{I}_{kT}-\mathbf{S}_s\overline{\mathbf{P}}_{\mathbf{B}})^{-1}(\mathbf{I}_{kT}-\mathbf{S}_s)\right\}\boldsymbol{\nu}_s$$

$$=J_5+J_6+J_7,\quad \text{say.}$$

It is not difficult to show that $J_5=o_p\{h_s^{*2}+(kh_s^*)^{-1/2}\}$ and $J_7=o_p\{h_s^{*2}+(kh_s^*)^{-1/2}\}$ by the same argument as the proof of Theorem 1 in Horowitz and Mammen (2004), and $J_6=o_p(k^{-1/2})$ by Lemma 1 and the root-k consistency of $\hat{\boldsymbol{\beta}}_s$. It follows that $J_4=o_p\{h_s^{*2}+(kh_s^*)^{-1/2}\}$.

According to the usual nonparametric regression result,

$$(kh_s^*)^{1/2}\left[J_4-g_s(u)-\frac{h_s^{*2}}{2}\frac{\alpha_2^2-\alpha_1\alpha_3}{\alpha_2-\alpha_1^2}g_s''(u)\right]\to_p 0\quad\text{as } k\to\infty.$$

Thus it suffices to show that $(kh_s^*)^{1/2}J_1\to_D N(0,\zeta_s^{TS}(u))$ as $k\to\infty$. Let $c_0=\alpha_2/(\alpha_2-\alpha_1^2)$, $c_1=-\alpha_1/(\alpha_2-\alpha_1^2)$, and

$$Q=\frac{1}{k}\sum_{i=1}^{k}\sum_{t=1}^{T}\left[c_0+c_1\left(\frac{U_{sit}-u}{h_s^*}\right)\right]K_{h_s^*}(U_{sit}-u)\left(v_{sit}-\frac{1}{T-1}\sum_{t_1\neq t}^{T}v_{sit_1}\right).$$

Then

$$(kTh_s^*)^{1/2}\left[J_1+J_4-g_s(u)-\frac{h_s^{*2}}{2}\frac{\alpha^2-\alpha_1\alpha_3}{\alpha_2-\alpha_1^2}g_s''(u)\right]=f_s^{-1}(u)(kTh_s^*)^{1/2}Q+o_p(1).$$

The variance matrix of $(kh_s^*)^{1/2}Q$ is

$$\begin{aligned}\mathrm{Var}((kh_s^*)^{1/2}Q)&=\frac{h_s^*\sigma_{vss}^2c_0^2}{k}\sum_{i=1}^{k}\sum_{t=1}^{T}K_{h_s^*}^2(U_{sit}-u)\\&+\frac{h_s^*\sigma_{vss}^2c_1^2}{k}\sum_{i=1}^{k}\sum_{t=1}^{T}\left(\frac{U_{sit}-u}{h_s^*}\right)^2K_{h_s^*}^2(U_{sit}-u)\\&+\frac{h_s^*\sigma_{vss}^2c_0c_1}{k}\sum_{i=1}^{k}\sum_{t=1}^{T}\left(\frac{U_{sit}-u}{h_s^*}\right)K_{h_s^*}^2(U_{sit}-u)=J_8+J_9+J_{10},\quad\text{say.}\end{aligned}$$

It is easy to see that $J_8\to_p c_0^2\sigma_{vss}^2\delta_0$, $J_9\to_p c_1^2\sigma_{vss}^2\delta_2$, and $J_{10}\to_p c_1c_0\sigma_{vss}^2\delta_1$ as $k\to\infty$. Therefore $\mathrm{Var}((kh_s^*)^{1/2}Q)=p_s^{-1}(u)(c_0^2\delta_0+2c_0c_1\delta_1+c_1^2\delta_2)+o(1)$. Let $B_k^2=\sum_{i=1}^{k}Ea_i^2$ with

$$a_i=\sqrt{h_s^*}\sum_{t=1}^{T}\left[c_0+c_1\left(\frac{U_{sit}-u}{h_s^*}\right)\right]K_{h_s^*}(U_{sit}-u)\left(v_{sit}-\frac{1}{T-1}\sum_{t_1\neq t}^{T}v_{sit_1}\right).$$

Then $B_k^2=kTf_s^{-1}(u)(c_0^2\delta_0+2c_0c_1\delta_1+c_1^2\delta_2)\sigma_{vss}^2+o(k)$. Further, simple calculations show that

$$\sum_{i=1}^{k}\mathbb{E}|a_i|^3\le O(1)\cdot\sum_{i=1}^{k}\sum_{t=1}^{T}h_s^{*(3/2)}\left[|c_0|+|c_1|\cdot\left|\frac{U_{sit}-u}{h_s^*}\right|\right]K_{h_s^*}^3(U_{sit}-u)=O(h_s^{*-1/2}).$$

https://doi.org/10.1017/S0266466613000352 Published online by Cambridge University Press

Thus $\lim_{k\to\infty} B_k^{-3}\sum_{i=1}^{k}\mathbb{E}|a_i^3| = 0$. The theorem then follows from the central limit theorem. ■

Proof of Theorem 6. By Theorem 1 and the standard nonparametric regression results,

$$
\begin{aligned}
m\hat{g}_s^{WTS}(u) &= ((\hat{\sigma}_v^{2ss})^{-1}, 0)(\mathbf{D}_{su}^{*\tau}\mathbf{W}_{su}^{*}\mathbf{D}_{su}^{*})^{-1}\mathbf{D}_{su}^{*\tau}\mathbf{W}_{su}^{*} \\
&\quad\times\left\{\sum_{s_1=1}^{M}\sigma_v^{2ss_1}\left(v_{s_1}11-\frac{1}{T-1}\sum_{t_1\neq 1}^{T}v_{s_2}1t_1\right),\right. \\
&\qquad\ldots,\ \sum_{s_1=1}^{M}\sigma_v^{2ss_1}\left(v_{s_1}1T-\frac{1}{T-1}\sum_{t_1\neq T}^{T}v_{s_2}1t_1\right), \\
&\qquad\left.\ldots,\ \sum_{s_1=1}^{M}\sigma_v^{2ss_1}\left(v_{s_1}kT-\frac{1}{T-1}\sum_{t_1\neq T}^{T}v_{s_2}kt_1\right)\right\} \\
&\quad+(1,0)\left(\mathbf{D}_{su}^{*\tau}\mathbf{W}_{su}^{*}\mathbf{D}_{su}^{*}\right)^{-1}\mathbf{D}_{su}^{*\tau}\mathbf{W}_{su}^{*}\mathbf{G}_s+o_p\left\{h_s^{*2}+(kh_s^{*})^{-1/2}\right\} \\
&= J_1+J_2+o_p\left\{h_s^{*2}+(kh_s^{*})^{-1/2}\right\}, \quad \text{say.}
\end{aligned}
$$

It is easy to see that as $k\to\infty$, $(kh_s^{*})^{1/2}J_1\to_D N(0,\zeta_s^{WTS}(u))$ and

$$
(kh_s^{*})^{1/2}\left[J_2-g_s(u)-\frac{h_s^{*2}}{2}\frac{\alpha_2^2-\alpha_1\alpha_3}{\alpha_2-\alpha_1^2}g_s''(u)\right]\to_p 0.
$$

Thus the proof is complete. ■

https://doi.org/10.1017/S0266466613000352 Published online by Cambridge University Press