Efficient nonparametric three-stage estimation of fixed effects varying coefficient panel data models

Juan M. Rodriguez-Póo and Alexandra Soberón

Universidad de Cantabria

Supplementary Material

S1 Local linear estimators

We propose the criterion function,

$$\sum_{i=1}^{N}\sum_{t=2}^{T}\Big(\Delta Y_{it}^{*}-\Delta W_{X_{it}}^{\top}\gamma_1-\Big(W_{X_{it}}\otimes(Z_{it}-z)-W_{X_{i(t-1)}}\otimes(Z_{i(t-1)}-z)\Big)^{\top}\gamma_2\Big)^2 K_H(Z_{it}-z)K_H(Z_{i(t-1)}-z), \quad \text{(S1.1)}$$

where $\gamma_1 = m(z)$ and $\gamma_2 = D_m(z)$, with $D_m(z)$ being a $(dq \times 1)$ vector of partial derivatives of the $m(z)$ function respect to the elements of the $(q\times 1)$ vector z such that $D_m(z) = vec(\partial m(z)/\partial z^{\top})$. Replacing the unknown quantities by their corresponding estimators as before, we define $\Delta\widehat{W}_{XZ}(z)$

as a $n \times d(1+q)$ dimensional matrix of the form

$$\Delta\widehat{W}_{XZ}(z) = \begin{bmatrix} \Delta\widehat{W}_{X_{12}}^{\top} & \widehat{W}_{X_{12}}^{\top} \otimes (Z_{12}-z)^{\top} - \widehat{W}_{X_{11}}^{\top} \otimes (Z_{11}-z)^{\top} \\ \vdots & \vdots \\ \Delta\widehat{W}_{X_{NT}}^{\top} & \widehat{W}_{X_{NT}}^{\top} \otimes (Z_{NT}-z)^{\top} - \widehat{W}_{X_{N(T-1)}}^{\top} \otimes (Z_{N(T-1)}-z)^{\top} \end{bmatrix}.$$

Assuming that $\Delta\widehat{W}_{XZ}(z)^{\top}K(z;H_2)\Delta\widehat{W}_{XZ}$ is nonsingular, it is straightforward to show that the value of γ_1 that minimizes (S1.1) has the solution

$$\begin{aligned} \widehat{m}_{\widetilde{\beta},LL}(z;H_2) &= e_1^{\top}(\Delta\widehat{W}_{XZ}(z)^{\top}K(z;H_2)\Delta\widehat{W}_{XZ}(z))^{-1} \\ &\quad \times\Delta\widehat{W}_{XZ}(z)^{\top}K(z;H_2)(\Delta Y - \Delta\widehat{W}_U\widetilde{\beta}_{LL}), \end{aligned} \tag{S1.2}$$

where $e_1 = (I_d \vdots 0_{dq\times d})$ is a $(d(1+q)\times d)$ matrix, I_d is a $(d\times d)$ identity matrix, and $0_{dq\times d}$ is a $(dq\times d)$ matrix of zeros.

Also, $\widetilde{\beta}_{LL}$ is the corresponding profile least squares estimator for a local linear fitting given by

$$\begin{aligned} \widetilde{\beta}_{LL} &= [\Delta\widehat{W}_U^{\top}(I_n - \widehat{S}_{LL})^{\top}(I_n - \widetilde{S}_{LL})\Delta W_U]^{-1} \\ &\times \ \Delta U^{\top}(I_n - \widehat{S}_{LL})^{\top}(I_n - \widetilde{S}_{LL})\Delta Y, \end{aligned} \tag{S1.3}$$

where $\widetilde{S}_{LL} = (\widetilde{S}_{LL_{12}}^{\top}, \ldots, \widetilde{S}_{LL_{NT}}^{\top})^{\top}$ is a $n\times n$ matrix such as its it th element is of the form

$$\begin{aligned} \widetilde{S}_{LL_{it}} &= (X_{it}^{\top} \quad 0_{dq}^{\top})(\Delta\widehat{W}_{XZ}(Z_{it})^{\top}K(Z_{it};H_2)\Delta\widehat{W}_{XZ}(Z_{it}))^{-1}\Delta\widehat{W}_{XZ}(Z_{it})^{\top}K(Z_{it};H_2) \\ &- (X_{i(t-1)}^{\top} \quad 0_{dq}^{\top})(\Delta\widehat{W}_{XZ}(Z_{it})^{\top}K(Z_{it};H_2)\Delta\widehat{W}_{XZ}(Z_{it}))^{-1}\Delta\widehat{W}_{XZ}(Z_{it})^{\top}K(z_{i(t-1)};H_2), \end{aligned}$$

where 0_{dq} is a dq-dimensional vector of zeros. $\widehat{S}_{LL}$ is defined in a similar way as above, with $\widehat{W}_{X_{it}}$ and $\widehat{W}_{X_{i(t-1)}}$ instead of X_{it} and $X_{i(t-1)}$, respectively.

Under the previous assumptions, the asymptotic normality of $\widehat{m}_{\widetilde{\beta},LL}(z;H_2)$ and $\widehat{\beta}_{LL}$ is collected in the following Corollaries.

Corollary S1.1. *Suppose that Assumptions S2.1–S2.10 hold. When $Ntr(H_2)^2 \to 0$, because N tends to infinity and T is fixed, we have*

$$\sqrt{n}\left(\widetilde{\beta}_{LL}-\beta\right) \xrightarrow{d} \mathcal{N}\left(0, \Sigma^{-1}\Sigma^{*}\Sigma^{-1}\right).$$

Corollary S1.2. *Suppose that Assumptions S2.1–S2.10. Because N tends to infinity and T is fixed, we have*

$$\sqrt{n|H_2|}\left(\widehat{m}_{\widetilde{\beta},LL}(z;H_2)-m(z)-B_{LL}(z;H_2)(1+o_p(1))\right) \xrightarrow{d} \mathcal{N}\left(0, V_{LL}(z;H_2)\right),$$

where

$$\begin{aligned} B_{LL}(z;H_2) &= \frac{\mu_2(K)}{2} diag_d\left(tr\left(\mathcal{H}_{m_\kappa}(z)H_2\right)\right)\imath_d, \\ V_{LL}(z;H_2) &= 2\sigma_\epsilon^2 R^2(K)\mathcal{B}^{-1}_{\Delta W_X \Delta W_X}(z,z). \end{aligned}$$

The proofs of these corollaries follow a similar proof scheme as the corresponding for Theorems 3.1 and 3.2, respectively, and therefore they are omitted.

Assuming $Ntr(H_2)^{5/2} \to 0$, the higher-order bias can be ignored and the result of Corollary S1.2 can be rewritten as

$$\sqrt{n|H_2|}\left(\widehat{m}_{\widetilde{\beta},LL}(z;H_2)-m(z)\right) \xrightarrow{d} \mathcal{N}\left(0, \Sigma_{LL}(z;H_2)\right). \quad \text{(S1.4)}$$

We propose a consistent estimator for $\widehat{\Sigma}_{LL}(z,H_2)$ of the form

$$
\begin{aligned}
\widehat{\Sigma}_{LL}(z,H_2) &= R^2(K)e_1^\top\left(\frac{1}{n}\Delta\widehat{W}_{XZ}^\top K(z;H_2)\Delta\widehat{W}_{XZ}\right)^{-1} \\
&\times \left[\frac{1}{n}\Delta\widehat{W}_{XZ}(z)^\top K(z,H_2)\widehat{V}K(z,H_2)\widehat{W}_{XZ}(z)\right]\left(\frac{1}{n}\Delta\widehat{W}_{XZ}^\top K(z;H_2)\Delta\widehat{W}_{XZ}\right)^{-1}e_1.
\end{aligned}
$$

S2 Assumptions

Throughout this Supplemantary Material, we use the same notation as used in previous sections. Previously, we have used the following vectors,

$$
\begin{aligned}
\Delta W_{X_{it}} &= (E(\Delta X_{1it}^\top|\mathcal{L}_{it},\mathcal{L}_{i(t-1)}),\Delta X_{2it}^\top)^\top; & \Delta\widehat{W}_{X_{it}} &= (\widehat{E}(\Delta X_{1it}^\top|\mathcal{L}_{it},\mathcal{L}_{i(t-1)}),\Delta X_{2it}^\top)^\top, \\
\Delta W_{U_{it}} &= (E(\Delta U_{1it}^\top|\mathcal{L}_{it},\mathcal{L}_{i(t-1)}),\Delta U_{2it}^\top)^\top; & \Delta\widehat{W}_{U_{it}} &= (\widehat{E}(\Delta U_{1it}^\top|\mathcal{L}_{it},\mathcal{L}_{i(t-1)}),\Delta U_{2it}^\top)^\top, \\
\widehat{W}_{X_{it}}^\top &= (\widehat{E}(X_{1it}|\mathcal{L}_{it},\mathcal{L}_{i(t-1)})^\top,X_{2it}^\top)^\top; & W_{X_{it}}^\top &= (E(X_{1it}|\mathcal{L}_{it},\mathcal{L}_{i(t-1)})^\top,X_{2it}^\top)^\top
\end{aligned}
$$

Assumption S2.1. *Let* $(Y_{it},X_{1it},U_{1it},\mathcal{L}_{it})_{i=1,\ldots,N;t=1,\ldots,T}$ *be a set of independent and identically distributed* $\mathbb{R}^{1+d_1+k_1+\ell}$*-random variables in the subscript* i *for each fixed* t *and strictly stationary over* t *for fixed* i*, where* $\mathcal{L}_{it}$ *is a* $(\ell\times 1)$ *vector which contains all exogenous variables (i.e.* Z_{it}*,* X_{2it}*, and* U_{2it}*) and other IVs such that* $\ell = q + d_2 + k_2 + M$*.*

Assumption S2.2. *The idiosyncratic error terms,* ϵ_{it}*, are independent and identically distributed with constant variance,* σ_ϵ^2*. Furthermore,* $E\left(\epsilon_{it}|\,\mathcal{L}_{it}\right) = 0$ *and* $E\left(\epsilon_{it}^2|\,\mathcal{L}_{it}\right) = \sigma_\epsilon^2$*.*

Assumption S2.3. *Let* $f_{Z_{it}}(\cdot)$ *be the probability density function of* Z_{it}*. Moreover, let* $f_{Z_{it},Z_{i(t-1)}}(\cdot,\cdot)$ *be the probability density function of* $\left(Z_{it}, Z_{i(t-1)}\right)$*, respectively. All density functions are continuously differentiable in all their arguments and they are bounded from above and below in any point of their support.*

Assumption S2.4. *Let* z *be an interior point in the support of* $f_{Z_{it}}(\cdot)$*. All second-order derivatives of* $m(z)$*,* $E(\Delta X_{1it}|\mathcal{L}_{it}, \mathcal{L}_{i(t-1)})$*, and* $E(\Delta U_{1it}|\mathcal{L}_{it}, \mathcal{L}_{i(t-1)})$ *are bounded and uniformly continuous and they satisfy a Lipschitz condition.*

Assumption S2.5. *The kernel function* K *is the product of univariate kernels, symmetric around zero and compactly supported. Also, the kernel is bounded such that* $\int uu^\top K(u)du = \mu_2(K)I$ *and* $\int K^2(u)du = R(K)$*, where* $\mu_2(K)$ *and* $R(K)$ *are scalars and* I *the identity matrix. In addition, all odd-order moments of* K *vanish, that is* $\int u_1^{\iota_1}\ldots u_q^{\iota_q}K(u)du = 0$*, for all nonnegative integers* $\iota_1,\ldots,\iota_q$ *such that their sum is odd.*

Assumption S2.6. *The kernel function satisfies the property that* $|K(u)| \leq \overline{K} < \infty$ *and* $\int |K(u)|du < \kappa < \infty$*. Further, for some* $\Lambda_1 < \infty$ *and* $\Psi < \infty$*, either* $K(u) = 0$ *for* $\|u\| > \Psi$ *and for all* $u, u' \in I\!R^q$*,* $|K(u) - K(u')| \leq \Lambda_1\|u - u'\|$*, or* $K(u)$ *is differentiable,* $|(\partial/\partial u)K(u)| \leq \Lambda_1$*, and for some* $\varsigma > 1$*,* $|(\partial/\partial u)K(u)| \leq \Lambda_1\|u\|^{-\varsigma}$ *for* $\|u\| > \Psi$*.*

Assumption S2.7. *Let* $\|A\| = \sqrt{tr\left(A^\top A\right)}$. $E\left[\|W_{X_{it}}W_{X_{it}}^\top\|^2 | Z_{it} = z_1, Z_{i(t-1)} = z_2\right]$ *is bounded and uniformly continuous in its support. Furthermore, let* $\mathbb{W}_X = (W_{X_{it}}^\top, W_{X_{i(t-1)}}^\top)^\top$ *and* $\Delta\mathbb{W}_X = (\Delta W_{X_{it}}^\top, \Delta W_{X_{i(t-1)}}^\top)^\top$. *The matrix functions* $E[\mathbb{W}_{X_{it}}\mathbb{W}_{X_{it}}^\top | Z_{it} = z_1, Z_{i(t-1)} = z_2]$, $E[\Delta\mathbb{W}_{X_{it}}\Delta\mathbb{W}_{X_{it}}^\top | Z_{it} = z_1, Z_{i(t-1)} = z_2]$, $E[\mathbb{W}_{X_{it}}\Delta\mathbb{W}_{X_{it}}^\top | Z_{it} = z_1, Z_{i(t-1)} = z_2]$, $E[\mathbb{W}_{X_{it}}\mathbb{W}_{X_{it}}^\top | Z_{i(t+1)} = z_1, Z_{it} = z_2, Z_{i(t-1)} = z_3]$,$E[\mathbb{W}_{X_{it}}\Delta\mathbb{W}_{X_{it}}^\top | Z_{i(t+1)} = z_1, Z_{it} = z_2, Z_{i(t-1)} = z_3]$ *are bounded and uniformly continuous at any interior point,* (z_1, z_2) *or* (z_1, z_2, z_3), *in the support of* $f_{Z_{it},Z_{i(t-1)}}(z_1, z_2)$ *and* $f_{Z_{i(t+1)},Z_{it},Z_{i(t+1)}}(z_1, z_2, z_3)$, *respectively.*

Assumption S2.8. *The bandwidth matrices* H_1 *and* H_2 *are symmetric and strictly positive-definite. Also, let* h_1 *and* h_2 *be each entry of the matrices* H_1 *and* H_2, *respectively,* $h_1 \to 0$ *and* $h_2 \to 0$. *As* $N \to \infty$, $|H_2| log N (N|H_1|^2)^{-1} \to c \in [0, \infty)$, $tr(H_2)(N|H_2|)^{1/2} \to c \in [0, \infty)$, *and* $tr(H_1) = o_p\left(tr(H_2)\right)$.

Assumption S2.9. *Let*

$$
\begin{aligned}
\mathcal{B}_{\Delta W_X \Delta W_X}(z_1, z_2) &= E[\Delta W_{X_{it}} \Delta W_{X_{it}}^\top | Z_{it} = z_1, Z_{i(t-1)} = z_2] f_{Z_{it},Z_{i(t-1)}}(z_1, z_2), \\
\mathcal{B}_{\Delta W_{X_{-1}} \Delta W_{X_{-1}}}(z_1, z_2) &= E[\Delta W_{X_{i(t-1)}} \Delta W_{X_{i(t-1)}}^\top | Z_{it} = z, Z_{i(t-1)} = z] f_{Z_{it},Z_{i(t-1)}}(z_1, z_2).
\end{aligned}
$$

The matrices $\mathcal{B}_{\Delta W_X \Delta W_X}(z_1, z_2)$ *and* $\mathcal{B}_{\Delta W_{X_{-1}} \Delta W_{X_{-1}}}(z_1, z_2)$ *are positive-definite at any interior point,* (z_1, z_2), *in the support of* $f_{Z_{1t},Z_{1(t-1)}}(z_1, z_2)$.

Assumption S2.10. *For some* $\xi > 0$, *the function* $E\left[|\epsilon_{it}|^{2+\xi} | Z_{it} = z_1, Z_{i(t-1)} = z_2\right]$

is bounded and it is uniformly continuous at any point, (z_1, z_2)*, in the support of* $f_{Z_{it},Z_{i(t-1)}}(z_1, z_2)$.

Assumptions **S2.1** and **S2.2** are standard in the nonparametric panel data regression analysis and characterizes the data-generating process. Specifically, it states that individuals are independent and, for a fixed individual, correlation along time is allowed. Other time-series structures can also be considered; see, e.g., Cai and Li (2008) or Cai et al. (2009). Also, for the estimation of the fully nonlinear part in the one-step backfitting algorithm some further assumptions about the density functions than the usual Lipschitz continuity are needed. The smoothness and boundedness conditions established in Assumptions **S2.3-S2.7** for the kernel function, conditional moments, and densities are standard in the literature of local linear regression estimates; see Ruppert and Wand (1994). Also, they allow us to claim the uniform convergence results established in Hansen (2008, Theorems 8 and 10). Assumption **S2.8** contains bandwidth conditions. They are required to show consistency of the different estimators. For example, the condition $\mathrm{tr}\,(H_1) = o_p\,(\mathrm{tr}\,(H_2))$ is needed to ensure that $\widehat{m}_{\widetilde{\beta}}(z; H_2)$ is not sensitive to the choice of H_1. Assumption **S2.9** is the sufficient condition for the model identification. Assumption **S2.10** is required to show that the Lyapunov condition holds.

Assumption S2.11. *The bandwidth matrix H_3 is symmetric and strictly positive-definite. Also, each entry of the matrix tends to zero as N tends to infinity in such a way that $N|H_3| \to \infty$. As $N \to \infty$, $|H_3| logN(N|H_2|^2)^{-1} \to c \in [0, \infty)$, $tr(H_3)(N|H_3|)^{1/2} \to c \in [c, \infty)$, and $tr(H_2) = o_p(tr(H_3))$.*

Assumption S2.12. *Let*

$$\begin{aligned}
\mathcal{B}_{W_X W_X}(z) &= E[W_{X_{it}} W_{X_{it}}^\top | Z_{it} = z] f_{Z_{it}}(z), \\
\mathcal{B}_{W_{X_{-1}} W_{X_{-1}}}(z) &= E[W_{X_{i(t-1)}} W_{X_{i(t-1)}}^\top | Z_{i(t-1)} = z] f_{Z_{i(t-1)}}(z).
\end{aligned}$$

The matrices $\mathcal{B}_{W_X W_X}(z)$ and $\mathcal{B}_{W_{X_{-1}} W_{X_{-1}}}(z)$ are positive-definite at any interior point, z_1 and z_2, in the support of $f_{Z_{it}}(z_1)$ and $f_{Z_{i(t-1)}}(z_2)$, respectively.

To prove Theorems 3.1-4.2, the following lemmas are needed. For the sake of simplicity, let us denote

$$K_{it} = |H_2|^{-1/2} K(H_2^{-1/2}(Z_{it} - z)) \quad , \quad K_{i(t-1)} = |H_2|^{-1/2} K(H_2^{-1/2}(Z_{i(t-1)} - z)),$$

and $\widehat{\Gamma}_n = n^{-1} \sum_{it} K_{it} K_{i(t-1)} \Delta \widehat{W}_{X_{it}} \Delta \widehat{W}_{X_{it}}^\top$.

S3 Lemmas

Lemma S3.1. *Suppose that Assumptions S2.1–S2.10 hold. Because $N \to \infty$, we have*

$$\widehat{\Gamma}_n = \mathcal{B}_{\Delta W_X \Delta W_X}(z, z)(1 + O_p(a_{2N})),$$

holds uniformly for $z \in \mathcal{A}$, *where* $a_{2N} = (N|H_2|)^{-1/2} + tr(H_2)$ *and*

$$\mathcal{B}_{\Delta W_X \Delta W_X}(z,z) = E[\Delta W_{X_{it}} \Delta W_{X_{it}}^\top | Z_{it} = z, Z_{i(t-1)} = z] f_{Z_{it},Z_{i(t-1)}}(z,z).$$

Proof of Lemma S3.1: Following the same arguments as in the proof of Theorem 8 in (Hansen (2008))

$$\Delta\widehat{W}_{X_{it}} = \Delta W_{X_{it}} + O_p(\delta_N^{-1} a_{1N}^*), \tag{S3.5}$$

where $a_{1N}^* = \left(\frac{logN}{N|H_1|^\ell}\right)^{1/2} + tr(H_1)$ and $\delta_N = \inf_{|\mathcal{L}_1,\mathcal{L}_2| \leq c_N} f(\mathcal{L}_1, \mathcal{L}_2) > 0$ with $c_N = ((lnN)^{1/\ell} N^{1/2r})$, for some $r > 0$.

Then, it is possible to write

$$\widehat{\Gamma}_n = \Gamma_n + O_p(\delta_N^{-1} a_{1N}^*), \tag{S3.6}$$

where $\Gamma_n = n^{-1} \Delta W_X^\top K(z; H_2) \Delta W_X$.

To prove this lemma, we first show that, as $N \to \infty$,

$$\Gamma_n = \mathcal{B}_{\Delta W_X \Delta W_X}(z,z)(1 + O_p(a_{2N})). \tag{S3.7}$$

To this end, we follow the usual Taylor expansion, i.e.,

$$f(z + H_2^{1/2} u) = f(z) + D_f(z) H_2^{1/2} u + O_p(tr(H_2)), \quad \text{as} \quad H_2 \to 0,$$

where $D_f^\top(z) = \partial f(z)/\partial z$ is a q-dimensional vector which contains the first-order derivative vector of $f(\cdot)$.

By Assumption **S2.1** Z_{it} is *i.i.d.* across i and strict stationary in t.

Then, by the law of iterated expectations we get that

$$
\begin{aligned}
E(\Gamma_n) &= = \frac{1}{n}\sum_{it} \Delta W_{X_{it}} \Delta W_{X_{it}}^{\top} K_{it} K_{i(t-1)} \\
&= \int\int E[\Delta W_{X_{it}} \Delta W_{X_{it}}^{\top} | Z_{it} = z + H_2^{1/2}u, Z_{i(t-1)} = z + H_2^{1/2}v] \\
&\times \; f_{Z_{it},Z_{i(t-1)}}(Z_{it} = z + H_2^{1/2}u, Z_{i(t-1)} = z + H_2^{1/2}v) K(u)K(v)dudv \\
&= \mathcal{B}_{\Delta W_X \Delta W_X}(z,z) + O_p(tr(H_2)). \qquad \text{(S3.8)}
\end{aligned}
$$

Also, under Assumption **S2.1**,

$$
\begin{aligned}
Var(\Gamma_n) &= n^{-1} Var(K_{it} K_{i(t-1)} \Delta W_{X_{it}} \Delta W_{X_{it}}^{\top}) \\
&+ \; n^{-1} \sum_{\kappa=1}^{T-2} (T-\kappa) Cov(K_{i2} K_{i1} \Delta W_{X_{i2}} \Delta W_{X_{i2}}^{\top}, K_{i(2+\kappa)} K_{i(1+\kappa)} \Delta W_{X_{i(2+\kappa)}} \Delta W_{X_{i(2+\kappa)}}^{\top}),
\end{aligned}
$$

for $\kappa = |t-s|$, where $\kappa \in \{2, \ldots, (T-2)\}$. Under conditions **S2.8**–**S2.9**, it holds

$$
n^{-1} Var(K_{it} K_{i(t-1)} \Delta W_{X_{it}} \Delta W_{X_{it}}^{\top}) = O_p\left(\frac{1}{N|H_2|}\right)
$$

and

$$
n^{-1} Cov(K_{i2} K_{i1} \Delta W_{X_{i2}} \Delta W_{X_{i2}}^{\top}, K_{i(2+\kappa)} K_{i(1+\kappa)} \Delta W_{X_{i(2+\kappa)}} \Delta W_{X_{i(2+\kappa)}}^{\top}) = O_p\left(\frac{1}{N|H_2|}\right).
$$

Since $N|H_2| \to \infty$, this variance term tends to zero so (S3.7) is proved. Then, replacing (S3.7) in (S3.6) and using the bandwidth conditions of Assumption S2.8, it is easy to prove that

$$
\widehat{\Gamma}_n = \mathcal{B}_{\Delta W_X \Delta W_X}(z,z) + o_p(\delta_N^{-1} a_{2N}) + O_p(a_{2N}). \qquad \text{(S3.9)}
$$

∎

Lemma S3.2. *Suppose that Assumptions S2.1–S2.10 hold. Because $N \to \infty$, we have*

$$\frac{1}{n}\Delta\widehat{W}_U^\top (I_n - \widehat{S})^\top (I_n - \widetilde{S})\Delta U \quad \xrightarrow{p} \quad \Sigma,$$

where $\Sigma = E(\Upsilon_{it}\Upsilon_{it}^\top)$ *with* $\Upsilon_{it} = \Delta W_{U_{it}} - \mathcal{B}_{\Delta W_X \Delta W_U}(z,z)^\top \mathcal{B}_{\Delta W_X \Delta W_X}^{-1}(z,z)\Delta W_{X_{it}}$.

Proof of Lemma S3.2: Note that

$$n^{-1}\Delta\widehat{W}_U^\top (I_n - \widehat{S})^\top (I_n - \widetilde{S})\Delta U = n^{-1}(\Delta\widehat{W}_U - \widehat{S}\Delta\widehat{W}_U)^\top (\Delta U - \widetilde{S}\Delta U). \quad \text{(S3.10)}$$

Then, we analyze $\widehat{S}\Delta\widehat{W}_U$ and $\widetilde{S}\Delta U$ separately. To this end, let

$$\widehat{S}\Delta\widehat{W}_U = \begin{pmatrix} \widehat{W}_{X_{12}}^\top \widehat{\Gamma}_{12}^{-1}\Delta\widehat{W}_X^\top K(z_{12};H_2)\Delta\widehat{W}_U - \widehat{W}_{X_{11}}^\top \widehat{\Gamma}_{11}^{-1}\Delta\widehat{W}_X^\top K(z_{11};H_2)\Delta\widehat{W}_U \\ \vdots \\ \widehat{W}_{X_{NT}}^\top \widehat{\Gamma}_{NT}^{-1}\Delta\widehat{W}_X^\top K(z_{NT};H_2)\Delta\widehat{W}_U - \widehat{W}_{X_{N(T-1)}}^\top \widehat{\Gamma}_{N(T-1)}^{-1}\Delta\widehat{W}_X^\top K(z_{N(T-1)};H_2)\Delta\widehat{W}_U \end{pmatrix}. \quad \text{(S3.11)}$$

By Lemma S3.1 we know that, uniformly in $z \in \mathcal{A}$,

$$\Delta\widehat{W}_X^\top K(z;H_2)\Delta\widehat{W}_X = n\mathcal{B}_{\Delta W_X \Delta W_X}(z,z)(1 + O_p(a_{2N})). \quad \text{(S3.12)}$$

Following the same arguments as in the proof of Theorem 8 in (Hansen (2008)), as $N \to \infty$,

$$\Delta\widehat{W}_{U_{it}} = \Delta W_{U_{it}} + O_p(\delta_N^{-1} a_{1N}^*). \quad \text{(S3.13)}$$

Then, under the same reasoning as in (S3.12) it can be shown that, uniformly in $z \in \mathcal{A}$,

$$\begin{aligned} \Delta \widehat{W}_X^\top K(z; H_2) \Delta \widehat{W}_U &= \sum_{it} \Delta W_{X_{it}} \Delta W_{U_{it}}^\top K_{it} K_{i(t-1)} + O_p(\delta_N^{-1} a_{1N}^*) \\ &= n \mathcal{B}_{\Delta W_X \Delta W_U}(z, z)(1 + O_p(a_{2N})), \end{aligned} \quad \text{(S3.14)}$$

where

$$\mathcal{B}_{\Delta W_X \Delta W_U}(z, z) = E[\Delta W_{X_{it}} \Delta W_{U_{it}}^\top | Z_{it} = z, Z_{i(t-1)} = z] f_{Z_{it}, Z_{i(t-1)}}(z, z).$$

Replacing (S3.12)–(S3.14) into (S3.11) and rearranging terms, we get that, uniformly in $z \in \mathcal{A}$,

$$\widehat{S} \Delta \widehat{W}_U = \begin{pmatrix} \Delta W_{X_{12}}^\top \mathcal{B}_{\Delta W_X \Delta W_X}^{-1}(z, z) \mathcal{B}_{\Delta W_X \Delta W_U}(z, z) \\ \vdots \\ \Delta W_{X_{NT}}^\top \mathcal{B}_{\Delta W_X \Delta W_X}^{-1}(z, z) \mathcal{B}_{\Delta W_X \Delta W_U}(z, z) \end{pmatrix} (1 + O_p(a_{2N})). \quad \text{(S3.15)}$$

Similarly, it is straightforward to show that

$$\widetilde{S} \Delta U = \begin{pmatrix} \Delta X_{12}^\top \mathcal{B}_{\Delta W_X \Delta W_X}^{-1}(z, z) \mathcal{B}_{\Delta W_X \Delta W_U}(z, z) \\ \vdots \\ \Delta X_{NT}^\top \mathcal{B}_{\Delta W_X \Delta W_X}^{-1}(z, z) \mathcal{B}_{\Delta W_X \Delta W_U}(z, z) \end{pmatrix} (1 + O_p(a_{2N})). \quad \text{(S3.16)}$$

Using (S3.15) and (S3.16) into (S3.10) and some algebra, we obtain that, uniformly in $z \in \mathcal{A}$,

$$\begin{aligned} n^{-1}\Delta\widehat{W}_U^\top(I_n-\widehat{S})^\top(I_n-\widetilde{S})\Delta U &= \frac{1}{n}\sum_{it}[\Delta W_{U_{it}} - \Delta W_{X_{it}}^\top \mathcal{B}^{-1}_{\Delta W_X \Delta W_X}(z,z)\mathcal{B}_{\Delta W_X \Delta W_U}(z,z)] \\ &\times [\Delta U_{it} - \Delta U_{it}^\top \mathcal{B}^{-1}_{\Delta W_X \Delta W_X}(z,z)\mathcal{B}_{\Delta W_X \Delta W_U}(z,z)] \\ &+ o_p(\delta_N^{-1}a_{2N}) + O_p(a_{2N}) \\ &\to E(\Upsilon_{it}\Upsilon_{it}^\top), \end{aligned}$$

where the law of iterated expectations has been used and we denote

$$\Upsilon_{it} = \Delta W_{U_{it}} - \Delta W_{X_{it}}^\top \mathcal{B}^{-1}_{\Delta W_X \Delta W_X}(z,z)\mathcal{B}_{\Delta W_X \Delta W_U}(z,z).$$

Then, the proof is done.

∎

Lemma S3.3. *Suppose that Assumptions S2.1–S2.10 hold. Because $N \to \infty$, we have*

$$\frac{1}{n}\Delta\widehat{W}_U^\top(I_n-\widehat{S})^\top(I_n-\widetilde{S})M = O_p(a_{2N}^2),$$

where $M = [M_{12}^\top, \ldots, M_{NT}^\top]^\top$ is a n-dimensional vector whose it th element is $M_{it} = X_{it}^\top m(Z_{it}) - X_{i(t-1)}^\top m(Z_{i(t-1)})$.

Proof of Lemma S3.3: Note that

$$\frac{1}{n}\Delta\widehat{W}_U^\top(I_n-\widehat{S})^\top(I_n-\widetilde{S})M = \frac{1}{n}(\Delta\widehat{W}_U - \widehat{S}\Delta\widehat{W}_U)^\top(M-\widetilde{S}M), \quad \text{(S3.17)}$$

where

$$\widetilde{S}M = \begin{pmatrix} X_{12}^{\top}\widehat{\Gamma}_{12}^{\top}\Delta\widehat{W}_X^{\top}K(z_{12};H_2)M - X_{11}^{\top}\widehat{\Gamma}_{11}^{\top}\Delta\widehat{W}_X^{\top}K(z_{11};H_2)M \\ \vdots \\ X_{NT}^{\top}\widehat{\Gamma}_{NT}^{\top}\Delta\widehat{W}_X^{\top}K(z_{NT};H_2)M - X_{N(T-1)}^{\top}\widehat{\Gamma}_{N(T-1)}^{\top}\Delta\widehat{W}_X^{\top}K(z_{N(T-1)};H_2)M \end{pmatrix}$$

Following a similar procedure as in the proof of Lemma S3.1 and using (S3.13) and Assumption **S2.8**, we obtain that, uniformly in $z \in \mathcal{A}$,

$$\begin{aligned} \frac{1}{n}\Delta\widehat{W}_X^{\top}K(z;H_2)M &= \frac{1}{n}\sum_{it} K_{it}K_{i(t-1)}\Delta W_{X_{it}}(X_{it}^{\top}m(Z_{it}) - X_{i(t-1)}^{\top}m(Z_{i(t-1)})) \\ &+ O_p(\delta_N^{-1}a_{1N}^*) \\ &= \mathcal{B}_{\Delta W_X \Delta W_X}(z,z)m(z)(1+O_p(a_{2N})) \qquad \text{(S3.18)} \end{aligned}$$

Using Lemma S3.1, the equation (S3.18), and the usual Taylor expansion, we can show that, uniformly in $z \in \mathcal{A}$,

$$\widetilde{S}M = \left[(\Delta X_{12}^{\top}m(z))^{\top}, \ldots, (\Delta X_{NT}^{\top}m(z))^{\top}\right]^{\top}(1+O_p(a_{2N})). \qquad \text{(S3.19)}$$

By (S3.11) and (S3.12), we have

$$\begin{aligned} \frac{1}{n}(\Delta\widehat{W}_U - \widehat{S}\Delta\widehat{W}_U) &= (\Delta W_U - \Delta W_X \mathcal{B}_{\Delta W_X \Delta W_X}^{-1}(z,z)\mathcal{B}_{\Delta W_X \Delta W_U}(z,z)) \\ &\times (1+O_p(a_{2N})), \qquad \text{(S3.20)} \end{aligned}$$

where $\Delta W_U = (\Delta W_{U_{12}}^{\top}, \ldots, \Delta W_{U_{NT}}^{\top})^{\top}$ and $\Delta W_X = (\Delta W_{X_{12}}^{\top}, \ldots, \Delta W_{X_{NT}}^{\top})^{\top}$ are matrices of $(n \times (k_1+k_2))$ and $(n \times (d_1+d_2))$ dimension, respectively.

Replacing (S3.19)–(S3.20) into (S3.17) and using the bandwidth conditions in Assumption **S2.8**, we get

$$\begin{aligned}
&\frac{1}{n}\Delta\widehat{W}_U^\top(I_n-\widehat{S})^\top(I_n-\widetilde{S})M \\
&=\frac{1}{n}\sum_{it}(\Delta W_{U_{it}}-\mathcal{B}_{\Delta W_X\Delta W_U}(z,z)^\top\mathcal{B}^{-1}_{\Delta W_X\Delta W_X}(z,z)\Delta W_{X_{it}})\Delta X_{it}^\top m(z) \\
&\times(1+O_p(a_{2N}))O_p(a_{2N}) \\
&=O_p(a_{2N}^2). \qquad\qquad (S3.21)
\end{aligned}$$

∎

Lemma S3.4. *Under conditions S2.1–S2.6, and S2.8. For some* $r>0$, *we define* $c_N=O((lnN)^{1/q}N^{1/2r})$. *For all* z *such that* $\|z\|\le c_N$, *where* $\|z\|=\max(|z_1|,\dots,|z_q|)$, *as* $N\to\infty$ *and* T *is fixed,*

$$\sup_{\|z\|\le c_N}\left|\widehat{m}_{\widehat{\beta}}(z;H_2)-\widetilde{m}_\beta(z;H_2)\right|=o_p(\delta_N^{-1}a_{2N}).$$

Proof of Lemma S3.4. Note that

$$\widehat{m}_{\widetilde{\beta}}(z;H_2)-\widetilde{m}_\beta(z;H_2)=\widehat{\Gamma}_n^{-1}\widehat{T}_n-\Gamma_n^{-1}T_n, \qquad (S3.22)$$

where

$$\begin{aligned}
\widehat{T}_n &= n^{-1}\sum_{it}K_{it}K_{i(t-1)}\Delta\widehat{W}_{X_{it}}(\Delta Y_{it}-\Delta\widehat{W}_{U_{it}}\widetilde{\beta}), \\
T_n &= n^{-1}\sum_{it}K_{it}K_{i(t-1)}\Delta W_{X_{it}}(\Delta Y_{it}-\Delta W_{U_{it}}\beta).
\end{aligned}$$

Using Lemma S3.1 and equation (S3.7). By the Slutsky theorem, we get

$$\widehat{m}_{\widetilde{\beta}}(z;H_2)-\widetilde{m}_\beta(z;H_2) = \mathcal{B}^{-1}_{\Delta W_X\Delta W_X}(z,z)(\widehat{T}_n-T_n)(1+O_p(a_{2N})). \qquad (S3.23)$$

Let us consider now

$$\begin{aligned}\widehat{T}_n - T_n &= n^{-1}\sum_{it} K_{it}K_{i(t-1)}\Delta\widehat{W}_{X_{it}}(\Delta Y_{it} - \Delta\widehat{W}_{U_{it}}^{\top}\widetilde{\beta}) \\ &- n^{-1}\sum_{it} K_{it}K_{i(t-1)}\Delta W_{X_{it}}(\Delta Y_{it} - \Delta W_{U_{it}}^{\top}\beta) \\ &= n^{-1}\sum_{it} K_{it}K_{i(t-1)}(\Delta\widehat{W}_{X_{it}} - \Delta W_{X_{it}})\Delta Y_{it} \\ &- I\!I_n^{(1)} - I\!I_n^{(2)} - I\!I_n^{(3)} - I\!I_n^{(4)} - I\!I_n^{(5)} - I\!I_n^{(6)} - I\!I_n^{(7)},\end{aligned}$$

where

$$\begin{aligned}I\!I_n^{(1)} &= n^{-1}\sum_{it} K_{it}K_{i(t-1)}(\Delta\widehat{W}_{X_{it}} - \Delta W_{X_{it}})(\Delta\widehat{W}_{U_{it}} - \Delta W_{U_{it}})^{\top}(\widetilde{\beta} - \beta), \\ I\!I_n^{(2)} &= n^{-1}\sum_{it} K_{it}K_{i(t-1)}(\Delta\widehat{W}_{X_{it}} - \Delta W_{X_{it}})(\Delta\widehat{W}_{U_{it}} - \Delta W_{U_{it}})^{\top}\beta, \\ I\!I_n^{(3)} &= n^{-1}\sum_{it} K_{it}K_{i(t-1)}(\Delta\widehat{W}_{X_{it}} - \Delta W_{X_{it}})\Delta W_{U_{it}}^{\top}(\widetilde{\beta} - \beta), \\ I\!I_n^{(4)} &= n^{-1}\sum_{it} K_{it}K_{i(t-1)}(\Delta\widehat{W}_{X_{it}} - \Delta W_{X_{it}})\Delta W_{U_{it}}^{\top}\beta, \\ I\!I_n^{(5)} &= n^{-1}\sum_{it} K_{it}K_{i(t-1)}\Delta W_{X_{it}}(\Delta\widehat{X}_{U_{it}} - \Delta W_{U_{it}})^{\top}(\widetilde{\beta} - \beta), \\ I\!I_n^{(6)} &= n^{-1}\sum_{it} K_{it}K_{i(t-1)}\Delta W_{X_{it}}(\Delta\widehat{X}_{U_{it}} - \Delta W_{U_{it}})^{\top}\beta, \\ I\!I_n^{(7)} &= n^{-1}\sum_{it} K_{it}K_{i(t-1)}\Delta W_{X_{it}}\Delta W_{U_{it}}^{\top}(\widetilde{\beta} - \beta).\end{aligned}$$

Following the same lines as in Lemma S3.1 and the $\sqrt{n}$-consistency of $\widetilde{\beta}$, we claim that

$$\widehat{T}_n - T_n = o_p(\delta_N^{-1}a_{1N}^*). \tag{S3.24}$$

More precisely, using (S3.5) we get

$$n^{-1}\sum_{it} K_{it}K_{i(t-1)}(\Delta\widehat{W}_{X_{it}} - \Delta W_{X_{it}})\Delta Y_{it} = O_p(\delta_N^{-1}a_{1N}^*)$$

since it is straightforward to show that $n^{-1}\sum_{it} K_{it}K_{i(t-1)}\Delta Y_{it} = O_p(1)$. Similarly, it is straightforward to show that $I\!I_n^{(1)} = o_p(\delta_N^{-2}a_{1N}^* a_{1N}^*)$, $I\!I_n^{(2)} = O_p(\delta_N^{-2}a_{1N}^* a_{1N}^*)$, $I\!I_n^{(3)} = I\!I_n^{(5)} = o_p(\delta_N^{-1}a_{1N}^*)$, $I\!I_n^{(4)} = I\!I_n^{(6)} = O_p(\delta_N^{-1}a_{1N}^*)$, and $I\!I_n^{(7)} = o_p(1)$.

Then, replacing (S3.24) in (S3.23), following some straightforward calculations, and Assumption **(S2.8)**,

$$\widehat{m}_{\widetilde{\beta}}(z;H_2) - \widetilde{m}_{\beta}(z;H_2) = o_p(\delta_N^{-1}a_{2N}),$$

so Lemma S3.4 is proved.

∎

Lemma S3.5. *Assume conditions S2.1–S2.12 hold. For some $r > 0$, we define $c_N = O((lnN)^{1/q}N^{1/2r})$. For all z such that $\|z\| \leq c_N$, where $\|z\| = \max(|z_1|,\ldots,|z_q|)$, as $N \to \infty$ and T is fixed,*

$$\sup_{\|z\|\leq a_N} \left| \widehat{m}_{\widehat{\beta}}^{(1)}(z;H_3) - \widetilde{m}_{\beta}^{(1)}(z;H_3) \right| = o_p(\delta_N^{-1}a_{3N}),$$

where $a_{3N} = (N|H_3|^{1/2})^{-1/2} + tr(H_3)$.

To proof this lemma we follow the same line as in the proof of Lemma S3.4 and is therefore omitted.

∎

S4 Proof of Theorem 3.1

The estimator $\widetilde{\beta}$ can be written as

$$\sqrt{n}\left(\widetilde{\beta}-\beta\right) = \sqrt{n}\left(\Delta\widehat{W}_U^\top(I_n-\widehat{S})^\top(I_n-\widetilde{S})\Delta U\right)^{-1} \times \left[\Delta\widehat{W}_U^\top(I_n-\widehat{S})^\top(I_n-\widetilde{S})M+\Delta\widehat{W}_U^\top(I_n-\widehat{S})^\top(I_n-\widetilde{S})\Delta\epsilon\right], \quad \text{(S4.25)}$$

By Lemmas S3.2–S3.3, the bias term is

$$\sqrt{n}\left(\Delta\widehat{W}_U^\top(I_n-\widehat{S})^\top(I_n-\widetilde{S})\Delta U\right)^{-1}\Delta\widehat{W}_U^\top(I_n-\widehat{S})^\top(I_n-\widetilde{S})M = O_p(\sqrt{n}a_{2N}^*). \quad \text{(S4.26)}$$

Consider now the second term of the right-hand side of (S4.25), we write

$$\Delta\widehat{W}_U^\top(I_n-\widehat{S})^\top(I_n-\widetilde{S})\Delta\epsilon = \Delta\widehat{W}_U^\top(I_n-\widehat{S})^\top(\Delta\epsilon-\widetilde{S}\Delta\epsilon).$$

Following a similar reasoning as in (S3.18) we obtain that, uniformly in $z\in\mathcal{A}$, $\widetilde{S}\Delta\epsilon=[\Delta X_{12},\ldots,\Delta X_{NT}]^\top O_p(a_{2N})$. Using this result and (S3.20),

$$\frac{1}{n}(\Delta\widehat{W}_U-\widehat{S}\Delta\widehat{W}_U)^\top\widetilde{S}\Delta\epsilon = \frac{1}{n}\sum_{it}(\Delta W_{U_{it}}-\mathcal{B}_{\Delta W_X\Delta W_U}(z,z)^\top\mathcal{B}^{-1}_{\Delta W_X\Delta W_X}(z,z)\Delta W_{X_{it}}) \times\Delta X_{it}^\top(1+O_p(a_{2N}))O_p(a_{2N}). \quad \text{(S4.27)}$$

Using (S4.27) and given that $E(\Delta\epsilon_{it}|Z_{it},Z_{i(t-1)})=0$, it is straightforward to show

$$\frac{1}{\sqrt{n}}\Delta\widehat{W}_U^\top(I_n-\widehat{S})^\top(I_n-\widetilde{S})\Delta\epsilon \quad = \quad \frac{1}{\sqrt{n}}\sum_{it}\Upsilon_{it}\Delta\epsilon_{it}+o_p(\delta_N^{-1}a_{2N})+O_p(a_{2N})$$
$$\xrightarrow{d} \quad \mathcal{N}(0,\Sigma^*), \qquad \text{(S4.28)}$$

where $\Sigma^*=2\sigma_\epsilon^2E(\Upsilon_{it}\Upsilon_{it}^\top)-\sigma_\epsilon^2E(\Upsilon_{it}\Upsilon_{i(t+1)}^\top)$.

Finally, by the Slutsky theorem and the central limit theorem, we get

$$\sqrt{n}(\widetilde{\beta}-\beta) \quad \xrightarrow{d} \quad \mathcal{N}(0,\Sigma^{-1}\Sigma^*\Sigma^{-1}).$$

∎

S5 Proof of Theorem 3.2

In order to obtain the main asymptotic properties of the three-stage nonparametric estimator we can write

$$\sqrt{n|H_2|}\left(\widehat{m}_{\widetilde{\beta}}(z;H_2)-m(z)\right) \quad = \quad \sqrt{n|H_2|}\left(\widehat{m}_{\widetilde{\beta}}(z;H_2)-\widetilde{m}_\beta(z;H_2)\right)$$
$$+ \quad \sqrt{n|H_2|}\left(\widetilde{m}_\beta(z;H_2)-m(z)\right). \quad \text{(S5.29)}$$

In Lemma S3.4 it is shown the asymptotic equivalence between $\widehat{m}_{\widetilde{\beta}}(z;H_2)$ and $\widetilde{m}_\beta(z;H_2)$. Hence, the first element of the right-hand side of (S5.29) is asymptotically negligible. In the following, we consider the asymptotic distribution of the two-step feasible estimator, i.e., $\widehat{m}_{\widetilde{\beta}}(z,H_2)$.

With this aim, we first focus on the asymptotic bias of this estimator. The Taylor's approximation of the smooth functions implies

$$\begin{aligned}
\Delta Y_{it} &\simeq \Delta X_{it}^{\top} m(z) + \left[X_{it} \otimes (Z_{it} - z) - X_{i(t-1)} \otimes (Z_{i(t-1)} - z)\right]^{\top} D_m(z) \\
&+ \frac{1}{2}\left[X_{it}^{\top} \otimes (Z_{it} - z)^{\top} \mathcal{H}_m(z)(Z_{it} - z) - X_{i(t-1)}^{\top}\right. \\
&\left.\otimes (Z_{i(t-1)} - z)^{\top} \mathcal{H}_m(z)(Z_{i(t-1)} - z)\right] + \Delta U_{it}^{\top}\beta + \Delta\epsilon_{it} + R_m(z), \quad \text{(S5.30)}
\end{aligned}$$

where $D_m(z)$ and $\mathcal{H}_m(z)$ are the first-order derivatives vector and the Hessian matrix of $(dq \times 1)$ and $(dq \times q)$ dimension, respectively, for $D_m(z) = vec(\partial m(z)/\partial z^{\top})$ and $\mathcal{H}_m(z) = \partial^2 m(z)/\partial z z^{\top}$. Also, $R_m(z)$ is a vector of Taylor series remainder terms and we denote

$$\begin{aligned}
X_{it}^{\top} \otimes (Z_{it} - z)^{\top} D_m(z) &= \left[X_{1it}^{\top} \otimes (Z_{it} - z)^{\top} D_{m_1}(z), \quad X_{2it}^{\top} \otimes (Z_{it} - z)^{\top} D_{m_2}(z)\right], \\
X_{it}^{\top} \otimes (Z_{it} - z)^{\top} \mathcal{H}_m(z) (Z_{it} - z) &= \left[X_{1it}^{\top} \otimes (Z_{it} - z)^{\top} \mathcal{H}_{m_1}(z) (Z_{it} - z),\right. \\
&\quad \left. X_{2it}^{\top} \otimes (Z_{it} - z)^{\top} \mathcal{H}_{m_2}(z) (Z_{it} - z)\right].
\end{aligned}$$

Similar definitions for $X_{i(t-1)}^{\top} \otimes (Z_{i(t-1)} - z)^{\top} D_m(z)$ and $X_{i(t-1)}^{\top} \otimes (Z_{i(t-1)} - z)^{\top} \mathcal{H}_m(z)(Z_{i(t-1)} - z)$.

Using (S5.30) and subtracting $\Delta W_{X_{it}}^{\top} m(z)$ from both sides of $\widetilde{m}_{\beta}(z, H_2)$, this estimator can be rewritten as

$$\begin{aligned}
\widetilde{m}_{\beta}(z; H_2) - m(z) &= \left(\sum_{it} K_{it} K_{i(t-1)} \Delta W_{X_{it}} \Delta W_{X_{it}}^{\top}\right)^{-1} \\
&\times \sum_{it} K_{it} K_{i(t-1)} \Delta W_{X_{it}} \left(G_{it} + \Delta\epsilon_{it}\right), \quad \text{(S5.31)}
\end{aligned}$$

where

$$\begin{aligned}G_{it} &= (\Delta X_{it} - \Delta W_{X_{it}})^\top m(z) + \left[X_{it} \otimes (Z_{it} - z) - X_{i(t-1)} \otimes (Z_{i(t-1)} - z)\right]^\top D_m(z) \\ &+ \frac{1}{2}\left[X_{it}^\top \otimes (Z_{it} - z)^\top \mathcal{H}_m(z)(Z_{it} - z) - X_{i(t-1)}^\top \otimes (Z_{i(t-1)} - z)^\top \mathcal{H}_m(z)(Z_{i(t-1)} - z)\right] \\ &+ R_m(z).\end{aligned}$$

For the sake of simplicity, let us denote

$$\widetilde{m}_\beta(z; H_2) - m(z) = \Gamma_n^{-1}\left(B_n + \Psi_n\right), \tag{S5.32}$$

where

$$\begin{aligned}B_n &= n^{-1}\sum_{it} K_{it}K_{i(t-1)}\Delta W_{X_{it}} G_{it} \\ \Psi_n &= n^{-1}\sum_{it} K_{it}K_{i(t-1)}\Delta W_{X_{it}} \Delta\epsilon_{it}.\end{aligned}$$

Thus, to complete the proof of Theorem 3.2 it is enough to show

$$(\widetilde{m}_\beta(z; H_2) - m(z)) - \Gamma_n^{-1}B_n = \Gamma_n^{-1}\Psi_n, \tag{S5.33}$$

where we will demonstrate that $\Gamma_n^{-1}B_n$ contributes to the asymptotic bias, whereas the term of the right-hand side of (S5.33) is asymptotically normal.

Considering now the behavior of B_n, it can be decomposed into five different terms, each one has to be analyzed separately,

$$B_n = B_n^{(1)} + B_n^{(2)} + B_n^{(3)} + B_n^{(4)} + B_n^{(5)}, \tag{S5.34}$$

where

$$\begin{aligned}
B_n^{(1)} &= n^{-1}\sum_{it} K_{it}K_{i(t-1)}\Delta W_{X_{it}}(\Delta X_{it}-\Delta W_{X_{it}})^\top m(z),\\
B_n^{(2)} &= n^{-1}\sum_{it} K_{it}K_{i(t-1)}\Delta W_{X_{it}}(\Delta U_{it}-\Delta W_{U_{it}})^\top \beta,\\
B_n^{(3)} &= n^{-1}\sum_{it} K_{it}K_{i(t-1)}\Delta W_{X_{it}}\left[X_{it}\otimes(Z_{it}-z)-X_{i(t-1)}\otimes\left(Z_{i(t-1)}-z\right)\right]^\top D_m(z),\\
B_n^{(4)} &= (2n)^{-1}\sum_{it} K_{it}K_{i(t-1)}\Delta W_{X_{it}}\Big[X_{it}^\top\otimes(Z_{it}-z)^\top\mathcal{H}_m(z)\left(Z_{it}-z\right)\\
&\quad -X_{i(t-1)}^\top\otimes\left(Z_{i(t-1)}-z\right)^\top\mathcal{H}_m(z)\left(Z_{i(t-1)}-z\right)\Big],\\
B_n^{(5)} &= n^{-1}\sum_{it} K_{it}K_{i(t-1)}\Delta W_{X_{it}}R_m(z).
\end{aligned}$$

By the law of iterated expectations and the stationarity condition, it is easy to show that $E(B_n^{(1)})$ and $E(B_n^{(2)})$ are $o_p(1)$. Similarly,

$$\begin{aligned}
E(B_n^{(3)}) &= \int\Big(E(\Delta W_{X_{it}}X_{it}^\top|Z_{it}=z,Z_{i(t-1)}=z)D_f(z)(H_2^{1/2}u)\Big)\\
&\quad \otimes(H_2^{1/2}u)^\top D_m(z)K(u)K(v)dudv\\
&- \int\Big(E(\Delta W_{X_{it}}X_{i(t-1)}^\top|Z_{it}=z,Z_{i(t-1)}=z)D_f(z)(H_2^{1/2}v)\Big)\\
&\quad \otimes(H_2^{1/2}v)^\top D_m(z)K(u)K(v)dudv\\
&= \mu_2(K)\mathcal{B}_{\Delta W_X\Delta W_X}(z,z)diag_d(D_f(z)H_2D_{m_\kappa}(z))\imath_d f_{Z_{it},Z_{i(t-1)}}^{-1}(z,z)\\
&+ o_p(tr(H_2)). \qquad \text{(S5.35)}
\end{aligned}$$

Following a similar procedure as above,

$$\begin{aligned}
E(B_n^{(4)}) &= \frac{1}{2}E\Big[K_{it}K_{i(t-1)}\Big(E(\Delta W_{X_{it}}X_{it}^\top|Z_{it},Z_{i(t-1)})\otimes(Z_{it}-z)^\top\mathcal{H}_m(z)(Z_{it}-z)\\
&- E(\Delta W_{X_{it}}X_{i(t-1)}^\top|Z_{it},Z_{i(t-1)})\otimes(Z_{i(t-1)}-z)^\top\mathcal{H}_m(z)(Z_{it}-z)\Big)\Big]\\
&= \frac{1}{2}\mathcal{B}_{\Delta W_X\Delta W_X}(z,z)diag_d(tr(\mathcal{H}_{m_\kappa}(z)H_2))\imath_d\mu_2(K)+o_p(tr(H_2)). \quad \text{(S5.36)}
\end{aligned}$$

On its part, as it is shown in Rodriguez-Poo and Soberon (2014), $E(B_n^{(5)}) = o_p(tr(H_2))$.

Following a similar procedure as in the proof of (S3.7) and assuming $H_2 \to 0$ and $N|H_2| \to \infty$, it is easy to prove that any component of the variance of B_n converges to zero. Then, replacing (S5.35)-(S5.36) in B_n, using (S3.7), and applying the Slutsky theorem we have that

$$\begin{aligned}\Gamma_n^{-1} B_n &= \mu_2(K) \mathcal{B}_{\Delta W_X \Delta W_X}(z,z)^{-1} \mathcal{B}_{\Delta W_X \Delta W_X}(z,z) \left[diag_d(D_f(z) H_2 D_{m_\kappa}(z)) \right. \\ & \left. \times \imath_d f^{-1}_{Z_{it}, Z_{i(t-1)}}(z,z) + \frac{1}{2} diag_d(tr(\mathcal{H}_{m_\kappa}(z) H_2)) \imath_d \right] \\ &+ o_p(tr(H_2)), \end{aligned} \tag{S5.37}$$

where $\kappa = 1, \ldots, d$. Then, the first part of the proof is done.

To finish the proof of the theorem, all we have to do is to prove the convergence in distribution of $\sqrt{N|H_2|}\Gamma_n^{-1}\Psi_n$. In order to do so, we first calculate the asymptotic variance of $\Gamma_n^{-1}\Psi_n$ and then we will check the Lyapunov condition. With this aim, let us denote $\Delta\epsilon = (\Delta\epsilon_1, \ldots, \Delta\epsilon_N)$ as the $(n \times 1)$ vector with $\Delta\epsilon_i = (\Delta\epsilon_{i2}, \ldots, \Delta\epsilon_{iT})^\top$,

$$E(\Delta\epsilon_i \Delta\epsilon_i^\top | \mathcal{L}_{it}, \mathcal{L}_{i(t-1)}) = \begin{cases} 2\sigma_\epsilon^2, & for \quad i = i', \quad t = t', \\ -\sigma_\epsilon^2, & for \quad i = i', \quad |t - t'| < 2, \\ 0, & for \quad i = i', \quad |t - t'| \geq 2. \end{cases} \tag{S5.38}$$

When Ψ_n is analyzed we claim that by the law of iterated expectations, condition (2.2), and Assumptions **S2.1**, **S2.3** and **S2.5**–**S2.9**, we have that

$$\begin{aligned} n|H_2|Var(\Psi_n) &= |H_2|n^{-1}\sum_{ii'}\sum_{tt'}E\left[\Delta W_{X_{it}}\Delta\epsilon_{it}\Delta\epsilon_{i't'}\Delta W^{\top}_{X_{i't'}}K_{it}K_{i(t-1)}K_{i't'}K_{i'(t'-1)}\right] \\ &= 2\sigma^2_{\epsilon}R^2(K)\mathcal{B}_{\Delta W_X\Delta W_X}(z,z)(1+o_p(1)). \qquad \text{(S5.39)} \end{aligned}$$

To show this result note that the covariance between different individuals are clearly zero by the independence condition. Therefore, for $i = i'$ we consider two different cases: $t = t'$ and $t \neq t'$. For $t = t'$ and Assumptions **S2.1**, **S2.3** and **S2.5**–**S2.9**, by standard kernel methods we obtain

$$\begin{aligned} &|H_2|(T-1)^{-1}\sum_{t=2}^{T}E\left[\Delta W_{X_{it}}E(\Delta\epsilon^2_{it}|\mathcal{L}_{it},\mathcal{L}_{i(t-1)})\Delta W^{\top}_{X_{it}}K^2_{it}K^2_{i(t-1)}\right] \\ &= 2\sigma^2_{\epsilon}|H_2|E\left[E(\Delta W_{X_{it}}\Delta W^{\top}_{X_{it}}|Z_{it},Z_{i(t-1)})K^2_{it}K^2_{i(t-1)}\right] \\ &= 2\sigma^2_{\epsilon}R^2(K)\mathcal{B}_{\Delta W_X\Delta W_X}(z,z)(1+o_p(1)). \end{aligned}$$

Meanwhile, for $t \neq t'$, and proceeding in the same way as in the previous equation, if we consider again the stationary assumption

$$\begin{aligned} &2|H_2|E\left[\Delta W_{X_{i2}}E(\Delta\epsilon_{i2}\Delta\epsilon_{i3}|\mathcal{L}_{i1},\mathcal{L}_{i2},\mathcal{L}_{i3})\Delta W^{\top}_{X_{i3}}K^2_{i2}K_{i1}K_{i3}\right] \\ &\quad = -\sigma^2_{\epsilon}|H_2|^{1/2}R(K_u)E[\Delta W_{X_{it}}\Delta W^{\top}_{X_{it}}|Z_{i1}=z,Z_{i2}=z,Z_{i3}=z] \\ &\quad \times f_{Z_{i1},Z_{i2},Z_{i3}}(z,z,z)(1+o_p(1)). \end{aligned}$$

Note that only those terms of the variance-covariance matrix in which $|T-\kappa| < 2$ holds are nonzero. The remaining terms of this matrix are zero by the structure of the error term in first differences established in (S5.38).

Then, applying the Slutsky theorem and using (S3.8) and (S5.39), as $N|H_2| \to \infty$,

$$n|H_2|Var\left(\Gamma_n^{-1}\Psi_n\right) = 2\sigma_\epsilon^2 R^2(K)\mathcal{B}^{-1}_{\Delta W_X \Delta W_X}(z,z)(1+o_p(1))$$

Finally, in order to obtain the asymptotic distribution of $\widetilde{m}_\beta(z;H_2)$, it is suffices to check the Lyapunov condition. Let

$$\lambda^*_{n,i} = T^{-1/2}\sum_{it}\lambda_{it},$$

where

$$\lambda_{it} = K_{it}K_{i(t-1)}\Delta W_{X_{it}}\Delta\epsilon_{it}|H_2|^{1/2}, \qquad i=1,\ldots,N \quad ; \quad t=2,\ldots,T.$$

By Minkowski's inequality,

$$E|\lambda^*_{n,i}|^{2+\xi} \le CT^{(2+\xi)/2}E|\lambda_{it}|^{2+\xi},$$

where using similar derivations to those used in the proof of Theorem 3.2 it is obtained

$$\begin{aligned}
E|\lambda_{it}|^{2+\xi} &\le |H_2|^{(2+\xi)/2}E|K_{it}K_{i(t-1)}\Delta W_{X_{it}}\Delta\epsilon_{it}|^{2+\xi} \\
&= |H_2|^{-\xi/2}\int E\left(|\Delta W_{X_{it}}\Delta\epsilon_{it}|^{2+\xi}|Z_{it} = z + H_2^{1/2}u, Z_{i(t-1)} = z + H_2^{1/2}v\right) \\
&\times f_{Z_{it},Z_{i(t-1)}}(z+H_2^{1/2}u, z+H_2^{1/2}v)K^{2+\xi}(u)K^{2+\xi}(v)dudv \\
&= |H_2|^{-\xi/2}E\left(|\Delta W_{X_{it}}\Delta\epsilon_{it}|^{2+\xi}|Z_{it} = z, Z_{i(t-1)} = z\right) f_{Z_{it},Z_{i(t-1)}}(z,z) \\
&\times \int K^{2+\xi}(u)K^{2+\xi}(v)dudv + o_p(|H_2|^{-\xi/2}).
\end{aligned}$$

Then, it is proved that

$$E|\lambda^*_{n,i}|^{2+\xi} \leq CT^{(2+\xi)/2}|H_2|^{-\xi/2}.$$

Therefore, $N^{-(2+\xi)/2}\sum_{i=1}^{N} E|\lambda^*_{n,i}|^{(2+\xi)} \leq C(N|H_2|)^{-\xi/2} \to 0$ given that, as N tends to infinity, $N|H_2| \to \infty$. Thus, it is shown that the Lyapunov condition holds. Using this result and Lemma S3.4, we obtain

$$\sqrt{n|H_2|}\left(\widehat{m}_{\widehat{\beta}}(z;H_2) - m(z) - B(z,H_2)(1+o_p(1))\right)$$
$$\xrightarrow{d} \mathcal{N}\left(0, 2\sigma^2_\epsilon R^2(K)\mathcal{B}^{-1}_{\Delta W_X \Delta W_X}(z,z)\right)$$

so the proof of Theorem 3.2 is done.

∎

S6 Proof of Theorem 4.1

In order to prove Theorem 4.1, let us denote

$$\widetilde{m}^{(1)}_{\beta}(z;H_3) = \left(\sum_{i=1}^{N}\sum_{t=2}^{T} K_{it}W_{X_{it}}W^\top_{X_{it}}\right)^\top \sum_{i=1}^{N}\sum_{t=2}^{T} K_{it}W_{X_{it}}(\Delta\widehat{Y}_{1it} - \Delta W^\top_{U_{it}}\beta), \quad \text{(S6.40)}$$

where now

$$K_{it} = \frac{1}{|H_3|^{1/2}}K\left(H_3^{-1/2}(Z_{it}-z)\right).$$

Then, the one-step backfitting estimator can be written as

$$\begin{aligned}\widehat{m}^{(1)}_{\widehat{\beta}}(z;H_3) - m(z) &= \left(\widehat{m}^{(1)}_{\widehat{\beta}}(z;H_3) - \widetilde{m}^{(1)}_{\beta}(z;H_3)\right)\\ &+ \left(\widetilde{m}^{(1)}_{\beta}(z;H_3) - m(z)\right). \qquad \text{(S6.41)}\end{aligned}$$

Using Lemma S3.5, it can be shown that the first element of the right-hand side of (S6.41) is asymptotically negligible. Then, to show the results in Theorem 4.1, the asymptotic behavior of the latter term of (S6.41) needs to be proved. To do it, we first focus on the asymptotic bias of the one-step backfitting estimator and later on the corresponding variance.

Taylor's approximation of the smooth functions implies

$$\begin{aligned}\Delta\widehat{Y}_{1it} &= X_{it}^\top m(z) + X_{it}^\top \otimes (Z_{it}-z)^\top D_m(z) + \frac{1}{2}X_{it}^\top \otimes (Z_{it}-z)^\top \mathcal{H}_m(z)(Z_{it}-z) \\ &+ X_{i(t-1)}^\top \left[\widehat{m}_{\widehat{\beta}}(Z_{i(t-1)};H_2) - m(Z_{i(t-1)})\right] + \Delta U_{it}^\top \beta + \Delta\epsilon_{it} + R_m(z), \quad \text{(S6.42)}\end{aligned}$$

where $X_{it}^\top \otimes (Z_{it}-z)^\top D_m(z)$ and $X_{it}^\top \otimes (Z_{it}-z)^\top \mathcal{H}_m(z)(Z_{it}-z)$, $D_m(z)$, $\mathcal{H}_m(z)$, $R_m(z)$ are defined in (S5.30).

Replacing (S6.42) in (S6.40) and subtracting $W_{X_{it}}^\top m(z)$ from both sides of (S6.40), $\widetilde{m}_{\beta}^{(1)}(z;H_3)$ can be rewritten as

$$\begin{aligned}\widetilde{m}_{\beta}^{(1)}(z;H_3) - m(z) &= \left(\sum_{i=1}^{N}\sum_{t=2}^{T} K_{it}W_{X_{it}}W_{X_{it}}^\top\right)^{-1} \\ &\times \sum_{i=1}^{N}\sum_{t=2}^{T} K_{it}W_{X_{it}}\left(G_{it}^{(1)} + Q_{it} + \Delta\epsilon_{it}\right) \quad \text{(S6.43)}\end{aligned}$$

where

$$\begin{aligned}G_{it}^{(1)} &= (X_{it} - W_{X_{it}})^\top m(z) + X_{it}^\top \otimes (Z_{it}-z)^\top D_m(z) \\ &+ \frac{1}{2}X_{it}^\top \otimes (Z_{it}-z)^\top \mathcal{H}_m(z)(Z_{it}-z) + R_m(z), \\ Q_{it} &= X_{i(t-1)}^\top \left[\widehat{m}_{\widehat{\beta}}(Z_{i(t-1)};H_2) - m(Z_{i(t-1)})\right].\end{aligned}$$

For the sake of simplicity, let us denote

$$\widetilde{m}_{\beta}^{(1)}(z;H_3) - m(z) = \widetilde{\Gamma}_{n_1}^{-1}(\widetilde{B}_{n_1} + \widetilde{M}_{n_1} + \widetilde{U}_{n_1}), \tag{S6.44}$$

where

$$\begin{aligned} \widetilde{B}_{n_1} &= n^{-1}\textstyle\sum_{it} K_{it}W_{X_{it}}G_{it}^{(1)}, & \widetilde{M}_{n_1} &= n^{-1}\textstyle\sum_{it} K_{it}W_{X_{it}}Q_{it}, \\ \widetilde{U}_{n_1} &= n^{-1}\textstyle\sum_{it} K_{it}W_{X_{it}}\Delta v_{it}, & \widetilde{\Gamma}_{n_1} &= n^{-1}\textstyle\sum_{it} K_{it}W_{X_{it}}W_{X_{it}}^{\top}. \end{aligned}$$

Therefore, to complete the proof of this lemma it is necessary to show

$$\begin{aligned} & \sqrt{n|H_3|^{1/2}}(\widetilde{m}_{\beta}^{(1)}(z;H_3) - m(z)) - \sqrt{n|H_3|^{1/2}}\widetilde{\Gamma}_{n_1}^{-1}(\widetilde{B}_{n_1} + \widetilde{M}_{n_1}) \\ = & \sqrt{n|H_3|^{1/2}}\widetilde{\Gamma}_{n_1}^{-1}\widetilde{U}_{n_1}. \end{aligned} \tag{S6.45}$$

To obtain the bias term we first focus on the inverse term of (S6.45) and later analyze the behavior of $\widetilde{B}_{n_1}$ and $\widetilde{M}_{n_1}$. Then, following the same reasoning as in (S3.7), it can be shown that, as N tends to infinity,

$$\widetilde{\Gamma}_{n_1}^{-1} = \mathcal{B}_{W_XW_X}^{-1}(z) + o_p(1), \tag{S6.46}$$

given that

$$\widetilde{\Gamma}_{n_1} = n^{-1}\sum_{it} E[K_{it}W_{X_{it}}W_{X_{it}}^{\top}] = \mathcal{B}_{W_XW_X}(z) + o_p(1),$$

where

$$\mathcal{B}_{W_XW_X}(z) = E[W_{X_{it}}W_{X_{it}}^{\top}|Z_{it} = z]f_{Z_{it}}(z).$$

Focus now on the behavior of $\widetilde{B}_n$, it can be splitted up into five terms, i.e.

$$\widetilde{B}_{n_1} = n^{-1} \sum_{it} K_{it} W_{it} G_{it}^{(1)} = \widetilde{B}_{n_1}^{(1)} + \widetilde{B}_{n_1}^{(2)} + \widetilde{B}_{n_1}^{(3)} + \widetilde{B}_{n_1}^{(4)} + \widetilde{B}_{n_1}^{(5)}, \quad \text{(S6.47)}$$

where

$$\begin{aligned}
\widetilde{B}_{n_1}^{(1)} &= n^{-1} \sum_{it} K_{it} W_{X_{it}} (X_{it} - W_{X_{it}})^\top m(z), \\
\widetilde{B}_{n_1}^{(2)} &= n^{-1} \sum_{it} K_{it} W_{X_{it}} (\Delta U_{it} - \Delta W_{U_{it}})^\top \beta, \\
\widetilde{B}_{n_1}^{(3)} &= n^{-1} \sum_{it} K_{it} W_{X_{it}} (X_{it} \otimes (Z_{it} - z))^\top D_m(z), \\
\widetilde{B}_{n_1}^{(4)} &= (2n)^{-1} \sum_{it} K_{it} W_{X_{it}} (X_{it} \otimes (Z_{it} - z))^\top \mathcal{H}_m(z)(Z_{it} - z), \\
\widetilde{B}_{n_1}^{(5)} &= (2n)^{-1} \sum_{it} K_{it} W_{X_{it}} R_m(z).
\end{aligned}$$

Analyzing each of these terms separately and using the law of iterated expectations it can be proved that, as $N \to \infty$, $E(\widetilde{B}_{n_1}^{(1)})$ and $E(\widetilde{B}_{n_1}^{(2)})$ are $o_p(1)$. Further,

$$\begin{aligned}
E(\widetilde{B}_{n_1}^{(3)}) &= E\left[K_{it} E(W_{X_{it}} X_{it}^\top | Z_{it}) \otimes (Z_{it} - z)^\top D_m(z)\right] \\
&= \mu_2(K) \mathcal{B}_{W_X W_X}(z) diag_d\left(D_f(z) H_3 D_{m_\kappa}(z)\right) \imath_d f_{Z_{it}}^{-1}(z) + o_p(tr(H_3)),
\end{aligned} \quad \text{(S6.48)}$$

$$\begin{aligned}
E(\widetilde{B}_{n_1}^{(4)}) &= \frac{1}{2} E\left[K_{it} E(W_{X_{it}} X_{it}^\top | Z_{it}) \otimes (Z_{it} - z)^\top \mathcal{H}_{m_r}(z)(Z_{it} - z)\right] \\
&= \frac{1}{2} \mu_2(K) \mathcal{B}_{W_X W_X}(z) diag_d\left(tr(\mathcal{H}_{m_\kappa}(z) H_3)\right) \imath_d + o_p(tr(H_3)), \quad \text{(S6.49)}
\end{aligned}$$

and $E(\widetilde{B}_{n_1}^{(5)}) = o_p(tr(H_3))$. Also, assuming $H_3 \to 0$ and $N|H_3|^{1/2} \to \infty$, it is easy to prove that any component of the variance of $\widetilde{B}_{n_1}$ converges to zero.

To complete the proof of the asymptotic bias we have to analyze $\widetilde{M}_{n_1}$. Specifically, using the results in Lemma S3.4 it can be shown that, as N tends to infinity,

$$\begin{aligned} E(\widetilde{M}_{n_1}) &= E\left[K_{it}W_{X_{it}}X_{i(t-1)}^{\top}(\widehat{m}_{\widehat{\beta}}(z;H_2)-m(z;H_2))\right] \\ &= E\left[K_{it}W_{X_{it}}W_{X_{i(t-1)}}^{\top}(\widehat{m}_{\widehat{\beta}}(z;H_2)-m(z;H_2))\right] \\ &= o_p(\delta_N^{-1}a_{2N}), \end{aligned} \tag{S6.50}$$

since it can be proved that $n^{-1}\sum_{it}|K_{it}W_{X_{it}}W_{X_{i(t-1)}}^{\top}| = O_p(1)$.

Using the fact that $tr(H_2) \to 0$ and $tr(H_3) \to 0$ in the sense that $N|H_2| \to \infty$ and $N|H_3| \to \infty$, it is proved that $\widetilde{M}_{n_1}$ is asymptotically negligible. Then, if we substitute the asymptotic results of (S6.46) and (S6.47)–(S6.49) into (S6.44) by the Slutsky theorem, we obtain

$$\begin{aligned} \widetilde{\Gamma}_{n_1}^{-1}(\widetilde{B}_{n_1}+\widetilde{M}_{n_1}) &= \mu_2(K)\Big(diag_d\left(D_f(z)H_3D_{m_\kappa}(z)\right)\imath_d f_{Z_{it}}^{-1}(z) \\ &+ \frac{1}{2}diag_d\left(tr(\mathcal{H}_{m_\kappa}(z)H_3)\right)\imath_d\Big) + o_p(tr(H_3)). \end{aligned} \tag{S6.51}$$

Therefore, it is proved that the bias rate of the one-step backfitting estimator in (4.11) is the same as the corresponding of the three-stage nonparametric estimator in (2.10), as we expected.

Focus now on the asymptotic variance, under assumptions of Theorem

3.1 and by the law of iterated expectations, it is obtained

$$
\begin{aligned}
n|H_3|^{1/2}Var(\widetilde{U}_{n_1}) &= |H_3|^{1/2}n^{-1}\sum_{ii'}\sum_{tt'}E\left[W_{X_{it}}W_{X_{i't'}}^{\top}\Delta\epsilon_{it}\Delta\epsilon_{i't'}K_{it}K_{i't'}\right] \\
&= 2\sigma_{\epsilon}^2 R(K)\mathcal{B}_{W_X W_X}(z)(1+o_p(1)). \qquad \text{(S6.52)}
\end{aligned}
$$

Using (S6.45) and (S6.51) and by the Slutsky theorem,

$$
n|H_3|Var(\widetilde{\Gamma}_{n_1}^{-1}\widetilde{U}_{n_1}) = 2\sigma_{\epsilon}^2 R(K)\mathcal{B}_{W_X W_X}^{-1}(z)(1+o_p(1)). \qquad \text{(S6.53)}
$$

Finally, following similar derivations as in the proof of Theorem 3.2 it is straightforward to show that the Lyapunov condition holds. Therefore, under the assumptions of Theorem 4.1 we obtain

$$
\begin{aligned}
&\sqrt{n|H_3|^{1/2}}\left(\widehat{m}_{\widehat{\beta}}^{(1)}(z;H_3) - m(z) - B(z,H_3)(1+o_p(1))\right) \\
&\qquad \xrightarrow{d} \mathcal{N}\left(0, 2\sigma_{\epsilon}^2 R(K)\mathcal{B}_{W_X W_X}^{-1}(z)\right)
\end{aligned}
$$

and the proof is done.

∎

S7 Proof of Theorem 4.2

By subtracting in both terms of (4.18) the quantity $m(z)$ and noting that $\mathcal{G}_m^{-1}(z)\left(\mathcal{J}_{1m}(z)+\mathcal{J}_{2m}(z)\right) = I$ we obtain

$$
\begin{aligned}
\widehat{m}_{\widehat{\beta}}^{(mde)}(z;H_3) - m(z) &= \mathcal{G}_m^{-1}(z)\mathcal{J}_{1m}(z)\left(\widehat{m}_{\widehat{\beta}}^{(1)}(z;H_3) - m(z)\right) \\
&+ \mathcal{G}_m^{-1}(z)\mathcal{J}_{2m}(z)\left(\widehat{m}_{\widehat{\beta}}^{(2)}(z;H_3) - m(z)\right).
\end{aligned}
$$

Note that the asymptotic variances of the backfitting estimators have been already obtained in Theorem 4.1. Then, in order to proof this result first we need to obtain the asymptotic covariance between $\widehat{m}^{(1)}_{\widehat{\beta}}(z; H_3)$ and $\widehat{m}^{(2)}_{\widehat{\beta}}(z; H_3)$. Let

$$\text{Cov}\left(\widehat{m}^{(1)}_{\widehat{\beta}}(z; H_3), \widehat{m}^{(2)}_{\widehat{\beta}}(z; H_3)\right) = E\left(\left(\widehat{m}^{(1)}_{\widehat{\beta}}(z; H_3) - m(z)\right)\left(\widehat{m}^{(2)}_{\widehat{\beta}}(z; H_3) - m(z)\right)^\top\right).$$

In order to obtain the asymptotic covariance term, using (S6.44) for $\widehat{m}^{(1)}_{\widehat{\beta}}(z; H_3)$ and the corresponding expression for $\widehat{m}^{(2)}_{\widehat{\beta}}(z; H_3)$ we have

$$\text{Cov}\left(\widehat{m}^{(1)}_{\widehat{\beta}}(z; H_3), \widehat{m}^{(2)}_{\widehat{\beta}}(z; H_3)\right) = Cov\left(\widetilde{\Gamma}^{-1}_{n_1}\widetilde{U}_{n_1}, \widetilde{\Gamma}^{-1}_{n_2}\widetilde{U}_{n_2}\right), \qquad \text{(S7.54)}$$

where for

$$\begin{aligned} K_{it} &= |H_3|^{-1/2} K(H_3^{-1/2}(Z_{it} - z)), \\ K_{i(t-1)} &= |H_3|^{-1/2} K(H_3^{-1/2}(Z_{i(t-1)} - z)), \end{aligned}$$

we have

$$\begin{aligned} \widetilde{\Gamma}_{n_2} &= (N_2T)^{-1} \sum_{it} K_{i(t-1)} W_{X_{i(t-1)}} W^\top_{X_{i(t-1)}}, \\ \widetilde{U}_{n_2} &= -(N_2T)^{-1} \sum_{it} K_{i(t-1)} W_{X_{i(t-1)}} \Delta\epsilon_{it}. \end{aligned}$$

Focus on the middle term of (S7.54), and under assumptions of Theorem 4.1, it can be proved that by the law of iterated expectations, the strict stationarity in t and that $E\left[\Delta\epsilon_{it}\Delta\epsilon_{i't'}\right] = 0$, for $i \neq i'$.

$$\begin{aligned} n|H_3|^{1/2}\mathrm{Cov}(\widetilde{U}_{n_1}, \widetilde{U}_{n_2}) &= -\left|H_3\right|^{1/2} n^{-1} \sum_{ii'tt'} E\left[K_{it}K_{i(t'-1)}W_{X_{it}}\Delta\epsilon_{it}\Delta\epsilon_{i't'}W^{\top}_{X_{i(t'-1)}}\right] \\ &= 0. \end{aligned} \tag{S7.55}$$

In addition, under the same reasoning as in (S3.7)

$$\widetilde{\Gamma}^{-1}_{n_2} = \mathcal{B}^{-1}_{W_{X_{-1}}W_{X_{-1}}}(z)(1 + o_p(1)), \tag{S7.56}$$

where

$$\mathcal{B}_{W_{X_{-1}}W_{X_{-1}}}(z) = E[W_{X_{i(t-1)}}W^{\top}_{X_{i(t-1)}}|Z_{i(t-1)} = z]f_{Z_{i(t-1)}}(z).$$

If we substitute (S6.46) and (S7.55)–(S7.56) into (S7.54) and by the Crámer-Wold device, as $N \to \infty$,

$$n|H_3|^{1/2}\mathrm{Cov}\left(\widehat{m}^{(1)}_{\widehat{\beta}}(z; H_3), \widehat{m}^{(2)}_{\widehat{\beta}}(z; H_3)\right) = o_p(1). \tag{S7.57}$$

Now, apply Theorem A from (Serfling (1980), p. 122) and the proof is done.

∎

References

Cai, Z. and Q. Li (2008). "Nonparametric estimation of varying coefficient dynamic panel data models". *Econometric Theory 24*, 1321–1342.

Cai, Z., Q. Li, and J. Y. Park (2009). "Functional-coefficient models for nonstationary time series data". *Journal of Econometrics 148*, 101–113.

Hansen, B. (2008). "Uniform convergence rates for kernel estimation with dependent data". *Econometric Theory 24*, 726–748.

Rodriguez-Poo, J. M. and A. Soberon (2014). "Direct semi-parametric estimation of fixed effects panel data varying coefficient models". *Econometrics Journal 17*, 107–138.

Ruppert, D. and M. P. Wand (1994). "Multivariate locally weighted least squares regression". *The Annals of Statistics 22*(3), 1346–1370.

Serfling, R. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley Series in Probability and Mathematical Statistics. John Wiley and Sons, Inc.