Computational Statistics and Data Analysis 100 (2016) 401–423

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Semiparametric GMM estimation and variable selection in dynamic panel data models with fixed effects

Rui Li [a,b], Alan T.K. Wan [c,*], Jinhong You [b]

[a] School of Business Information, Shanghai University of International Business and Economics, Shanghai 201620, China

[b] School of Statistics and Management and Key Laboratory of Mathematical Economics, Shanghai University of Finance and Economics, Shanghai 200433, China

[c] Department of Management Sciences, City University of Hong Kong, Kowloon, Hong Kong

ARTICLE INFO

Article history:
Received 10 December 2014
Received in revised form 6 February 2016
Accepted 6 February 2016
Available online 16 February 2016

Keywords:
B-spline
GMM
Instrument matrix
SCAD penalty
Variable selection

ABSTRACT

Often fixed-effects dynamic panel data model assumes parametric structures and an AR(1) dynamic order. The latter assumption is mainly for convenience and not consistent with many sampling processes especially when longer panels are available. A fixed-effects dynamic partially linear additive model with a finite autoregressive lag order is considered. Based on this setup, semiparametric Generalized Method of Moments (GMM) estimators of the unknown coefficients and functions using the B(asis)-spline approximation are developed. The asymptotic properties of these estimators are established. A procedure to identify the dynamic lag order and significant exogenous variables by employing the smoothly clipped absolute deviation (SCAD) penalty is developed. It is proven that the SCAD-based GMM estimators achieve the oracle property and are selection consistent. The usefulness of proposed procedure is further illustrated in Monte Carlo studies and a real data example.

1. Introduction

Longitudinal or panel data models have been a key development in econometrics and statistics. These models have found widespread applications in economics, finance, biomedicine and other fields of research. One typical feature of panel data is that the sample of individuals is large but the number of time periods is short. It is well-known that due to the correlation between the lagged dependent variable and individual effect in the errors, when one estimates parameters of a dynamic panel data model, traditional techniques such as ordinary least squares (OLS) yield estimators that are biased and inconsistent. On the other hand, the generalized method of moments (Hansen, 1982) allows consistent estimation of parameters. Popular GMM estimators in dynamic panel data models include the first-differenced (DIF) GMM estimator of Arellano and Bond (1991), which transforms the model into first difference to remove the individual effect, and the system (SYS) GMM estimator of Arellano and Bover (1995) and Blundell and Bond (1998), which uses extra moment conditions that assume certain stationarity conditions of the initial observations. Of these two estimators, the SYS estimator has superior finite sample properties when stationarity of data is assumed. Other important contributions in this burgeoning literature on dynamic panel data models include but are not limited to the work of Blundell et al. (2001), Bun and Windmeijer (2010), Gouriéroux et al. (2010), Han and Phillips (2010), Everaert (2013), and Lee and Yu (2014).

* Corresponding author.
E-mail address: msawan@cityu.edu.hk (A.T.K. Wan).

http://dx.doi.org/10.1016/j.csda.2016.02.001
0167-9473/

It is noteworthy that studies on dynamic panels almost invariably assume an AR(1) dynamic order. This assumption, which is mainly for convenience, is not consistent with many sampling processes by which the data are generated, especially when a longer panel is available. Indeed, first-order models are most likely misspecified when the lag order is unknown. Lee (2012) showed that fitting a high order AR panel process by a lower order model can bias the fixed effect estimate substantially, and the established methods for correcting the bias under first-order dynamics can worsen the bias when the dynamics are of higher order. The question remains as to how the lag order in a dynamic panel data model can be selected accurately. One key objective of this paper is to address this question.

Recent years have also witnessed a stream of studies on the semiparametric partially linear models (e.g., Newey, 1994; Linton and Nielsen, 1995; Fan and Li, 1996; Li, 2000; Fan and Li, 2003; and Carroll et al., 2009), which can alleviate the familiar curse of dimensionality inherent to many nonparametric and semiparametric approaches. This is one major advantage of the partially linear models over conventional methods, in addition to their flexibility, ease of interpretation, and ease of use. Extensions of these models for static panel data have been considered by Fan et al. (2005), Su and Ullah (2006), Henderson et al. (2008), You et al. (2011) and Ai et al. (2014). There have been few extensions into the area of dynamic panels in the partially linear model literature with the most well-cited work of this nature being Baltagi and Li (2002), who developed consistent estimators of the unknowns in an AR(1) partially linear panel data model using the series method. Recently, Baglan (2010) proposed a GMM estimator of the linear component for the same model. To our knowledge, there has been no extension of the partially linear model to cases of higher order panel dynamics, and this paper attempts to make some progress in this direction.

The objective of this paper is twofold. First, we develop semiparametric GMM estimators for the parametric and nonparametric components of model (2.1) to be described ahead based on an approach that uses the B(asis)-spline (abbreviated as B-spline hereafter) series approximation. Spline approximation and kernel-based methods resulting in weighted least squares procedure are widely used in the nonparametric literature. Relative to kernel methods, spline approximation has an advantage in regard to the ease with which one can impose structure on the resulting estimate. In particular, B-spline approximation has the advantage of a stable numerical property provided that the knots sequence can be appropriately specified. We prove that the resultant estimator of the parametric component is $\sqrt{n}$-consistent and that of the unknown function achieves the optimal nonparametric convergence rate. In the same context, we also construct an estimator of the error variance. Second, we develop a procedure to select the lag order in the dynamic component and identify significant exogenous variables in the parametric and nonparametric components of the model using the smoothly clipped absolute deviation (SCAD) penalty. The SCAD penalty is a non-concave penalty proposed by Fan and Li (2001) in a general parametric framework for variable selection and efficient estimation, and has the advantages of being computationally feasible even for high dimensional data and more stable than subset selection. We also demonstrate that the SCAD-based estimators have an oracle property, the best possible theoretical performance of any variable selection procedure.

This remainder is organized as follows. We develop the semiparametric GMM estimators in Section 2 and study the asymptotic properties of these estimators in Section 3. In Section 4, we use the SCAD penalty to determine the lag order of dynamics and select variables. Section 5 reports results of a Monte Carlo study and a real data application of proposed method. Proofs of results are relegated to Appendix.

2. Semiparametric GMM estimation

Our framework of analysis is a fixed-effects dynamic panel partially linear model with a finite lag order. This framework extends that of Ai et al. (2014) to a dynamic set-up, and also includes several other models as special cases. Specifically, we consider the following panel process:

$$y_{it} = \mu_i + \sum_{s=1}^{d} \gamma_s y_{i(t-s)} + \sum_{k=1}^{p} x_{itk}\beta_k + \sum_{r=1}^{q} \phi_r(z_{itr}) + \varepsilon_{it}, \tag{2.1}$$

for $i = 1, \ldots, n,\ t = d+1, \ldots, T$, where y_{it} is an observation of the response Y on individual i at time t (with $y_{i(t-s)}$ being its lag value at time $t-s, s \geq 1$), $\mathbf{x}_{it} = (x_{it1}, \ldots, x_{itp})^\top \in \mathcal{R}^p$ and $\mathbf{z}_{it} = (z_{it1}, \ldots, z_{itq})^\top \in \mathcal{R}^q$ are realizations of the covariate vectors $\boldsymbol{X} = (X_1, \ldots, X_p)^\top$ and $\boldsymbol{Z} = (Z_1, \ldots, Z_q)^\top$ respectively, and are assumed to be strictly exogenous. μ_i's are unobserved individual fixed effects, possibly correlated with $\mathbf{x}_{it}$ and/or $\mathbf{z}_{it}$ with an unknown correlation structure, and satisfy $\sum_{i=1}^{n} \mu_i = 0$, $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_d)^\top$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ are unknown parameter vectors, $\{\phi_r(z_r)\}_{r=1}^{q}$ are smooth functions satisfying $\mathrm{E}\phi_r(z_r) = 0$ for identification purpose, and ε_{it} is an independent and identically distributed error term with mean 0 and variance σ^2, and satisfies $\mathrm{E}(y_{is}\varepsilon_{it}) = 0$ for $i = 1, \ldots, n, s = 1, \ldots, d, t = d+1, \ldots, T$ (see, for example, Ahn and Schmidt, 1995).

The panel process of (2.1) includes a number of other models as special cases. When $d = 0$ and $p = 0$, i.e., both the lagged dependent variables and parametric components are absent, (2.1) reduces to the conventional additive model (Hastie and Tibshirani, 1990). When $d = 1$ and $q = 1$, (2.1) becomes the AR(1) panel data partially linear model considered by Baltagi and Li (2002) and Baglan (2010). When $d = 0$ and $q = 1$, model (2.1) reduces to the fixed effects panel data partially linear

model that has been widely studied in the literature. See You et al. (2011), Ai et al. (2014) and the references therein. When $d = 1$ and $q = 0$, model (2.1) reduces to the fixed effects dynamic panel data linear model (Nickell, 1981).

Considering a knot sequence $\xi_0 < \xi_1 < \cdots < \xi_m < \xi_{m+1}$ that satisfies $\max(\xi_{j+1} - \xi_j)/\min(\xi_{j+1} - \xi_j) \leq c$ for $j = 0, \ldots, m$, with $m = m_n$ being a monotonic increasing function of n, the sample size. For simplicity, we assume that all knots are equally spaced. Let $\mathcal{H}_n^0$ be the space of polynomial splines with dimension $J_n = m_n + \varrho$ such that each function $h(\cdot)$ in $\mathcal{H}_n^0$ satisfies the following conditions: (i) $h(\cdot)$ is a polynomial of degree $\varrho > 1$ on each subinterval $I_j = [\xi_j, \xi_{j+1}), j = 0, \ldots, m_n - 1$, with $I_{m_n} = [\xi_{m_n}, \xi_{m_n+1}]$; (ii) for $\varrho \geq 2$, $h(\cdot)$ is $\varrho - 1$ times continuously differentiable on $[\xi_0, \xi_{m_n+1}]$. A group of splines spanning the whole space $\mathcal{H}_n^0$ is called a spline basis function. A commonly used basis function for fitting smoothing curves is the B-spline, which has the advantage of permitting every spline function of a given degree, smoothness and domain partition to be uniquely represented as a linear combination of B-splines of that same degree and smoothness and over that same partition. Denote the B-spline basis function as $\{N_j(z)\}_{j=1}^{J_n}$. The B-spline representation of $h(\cdot)$ is $h(\cdot) = \sum_{j=1}^{J_n} N_j(\cdot)\theta_j$, with $\{\theta_j\}_{j=1}^{J_n}$ being the unknown coefficients. Similarly, the spline basis $N_{rj}(\cdot)$ and θ_{rj} used in (2.2) are defined for the rth function $\phi_r(\cdot)$ for $r = 1, \ldots, q$.

The following approximation is legitimate when $\phi_r(z)$ is sufficiently smooth:

$$\phi_r(z_{itr}) \approx h_r(z_{itr}) = \sum_{j=1}^{J_n} N_{rj}(z_{itr})\theta_{rj} = \boldsymbol{\theta}_r^\top \mathbf{N}_r(z_{itr}), \tag{2.2}$$

where $\mathbf{N}_r(z) = (N_{r1}(z), \ldots, N_{rJ_n}(z))^\top$ and $\boldsymbol{\theta}_r = (\theta_{r1}, \ldots, \theta_{rJ_n})^\top$. Using (2.2), model (2.1) can be re-expressed as

$$y_{it} = \mu_i + \sum_{s=1}^{d} \gamma_s y_{it-s} + \sum_{k=1}^{p} x_{itk}\beta_k + \sum_{r=1}^{q}\sum_{j=1}^{J_n} N_{rj}(z_{itr})\theta_{rj} + \varepsilon_{it}^*, \tag{2.3}$$

with $\varepsilon_{it}^* = \varepsilon_{it} + \sum_{r=1}^{q}\{h_r(z_{itr}) - \phi_r(z_{itr})\}$. Let $y_{i0} = 0$. Taking first difference on t to remove the nuisance parameter μ_i yields

$$\begin{aligned} y_{it} - y_{i(t-1)} = {} & \sum_{s=1}^{d} \gamma_s (y_{i(t-s)} - y_{i(t-s-1)}) + \sum_{k=1}^{p} (x_{itk} - x_{i(t-1)k})\beta_k \\ & + \sum_{r=1}^{q}\sum_{j=1}^{J_n} \{N_{rj}(z_{itr}) - N_{rj}(z_{i(t-1)r})\}\theta_{rj} + \varepsilon_{it}^* - \varepsilon_{i(t-1)}^*. \end{aligned} \tag{2.4}$$

Now, write $\Delta\mathbf{y} = (\Delta\mathbf{y}_1^\top, \ldots, \Delta\mathbf{y}_n^\top)^\top$ with $\Delta\mathbf{y}_i = (y_{i(d+3)} - y_{i(d+2)}, \ldots, y_{iT} - y_{i(T-1)})^\top$, and $\Delta\mathbf{Y}^d = (\Delta\mathbf{y}_{1(d+3)}^d, \ldots, \Delta\mathbf{y}_{1T}^d, \ldots, \Delta\mathbf{y}_{nT}^d)^\top$ with $\Delta\mathbf{y}_{it}^d = (y_{i(t-1)} - y_{i(t-2)}, \ldots, y_{i(t-d)} - y_{i(t-d-1)})^\top$. In the same manner, write $\Delta\mathbf{X} = (\Delta\mathbf{x}_{1(d+3)}, \ldots, \Delta\mathbf{x}_{1T}, \ldots, \Delta\mathbf{x}_{nT})^\top$ with $\Delta\mathbf{x}_{it} = (x_{it1} - x_{1(t-1)1}, \ldots, x_{itp} - x_{i(t-1)p})^\top$ and $\Delta\boldsymbol{\varepsilon}^* = (\varepsilon_{1(d+3)}^* - \varepsilon_{1(d+2)}^*, \ldots, \varepsilon_{1T}^* - \varepsilon_{1(T-1)}^*, \ldots, \varepsilon_{nT}^* - \varepsilon_{n(T-1)}^*)^\top$. As well, denote $\tilde{\mathbf{N}}_r = (\tilde{\mathbf{N}}_r(z_{1(d+3)r}), \ldots, \tilde{\mathbf{N}}_r(z_{1Tr}), \ldots, \tilde{\mathbf{N}}_r(z_{nTr}))^\top$ with $\tilde{\mathbf{N}}_r(z_{itr}) = (N_{r1}(z_{itr}) - N_{r1}(z_{i(t-1)r}), \ldots, N_{rJ_n}(z_{itr}) - N_{rJ_n}(z_{i(t-1)r}))^\top$, and write $\tilde{\mathbf{N}} = (\tilde{\mathbf{N}}_1, \ldots, \tilde{\mathbf{N}}_q)$ and $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \ldots, \boldsymbol{\theta}_q^\top)^\top$. These notations permit the following matrix representation of (2.4):

$$\Delta\mathbf{y} = \Delta\mathbf{Y}^d\boldsymbol{\gamma} + \Delta\mathbf{X}\boldsymbol{\beta} + \tilde{\mathbf{N}}\boldsymbol{\theta} + \Delta\boldsymbol{\varepsilon}^*. \tag{2.5}$$

Any estimator that ignores the endogeneity due to the correlation between $\Delta\mathbf{Y}^d$ and $\Delta\boldsymbol{\varepsilon}^*$ is biased and inefficient. Here, we estimate the unknowns in (2.5) by GMM. Now, by integrating $\mathrm{E}(y_{it-s}\Delta\varepsilon_{it}) = 0$ for $2 \leq s < t - 1, t = d + 3, \ldots, T$, assuming exogeneity of $\mathbf{x}_{it}$ and $\mathbf{z}_{it}$, one can construct an instrument matrix $\mathbf{W} = (\mathbf{W}_1^\top, \mathbf{W}_2^\top, \ldots, \mathbf{W}_n^\top)^\top$, where

$$\mathbf{W}_i = \begin{pmatrix} \Delta y_{i(d+1)} & 0 & 0 & \cdots & 0 & \cdots & 0 & \Delta\mathbf{x}_{i(d+3)}^\top & \Delta\mathbf{z}_{i(d+3)}^\top \\ 0 & \Delta y_{i(d+1)} & \Delta y_{i(d+2)} & \cdots & 0 & \cdots & 0 & \Delta\mathbf{x}_{i(d+4)}^\top & \Delta\mathbf{z}_{i(d+4)}^\top \\ \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \Delta y_{i(d+1)} & \cdots & \Delta y_{iT-2} & \Delta\mathbf{x}_{iT}^\top & \Delta\mathbf{z}_{iT}^\top \end{pmatrix}_{T^* \times L}$$

with $\Delta y_{it} = y_{it} - y_{i(t-1)}$, $\Delta\mathbf{z}_{it} = (z_{it1} - z_{1(t-1)1}, \ldots, z_{itq} - z_{i(t-1)q})^\top$, $T^* = T - d - 2$ and $L = T^*(T^* + 1)/2 + p + q$. Other studies on system GMM, including Blundell and Bond (1998) and Bun and Windmeijer (2010), used both differenced and level instruments. We only use differenced instruments here as they are practically equivalent to level instruments when the number of time periods is large. The use of differenced instruments also poses no additional technical challenge compared to level instruments although differenced instruments do require more time periods than in the case of levels.

Writing $\mathbf{M} = \tilde{\mathbf{N}}(\tilde{\mathbf{N}}^\top\tilde{\mathbf{N}})^{-1}\tilde{\mathbf{N}}^\top$ and premultiplying (2.5) by $\mathbf{W}^{*\top} = \mathbf{W}^\top(\mathbf{I}_{nT^*} - \mathbf{M})$ yields

$$\mathbf{W}^{*\top}\Delta\mathbf{y} = \mathbf{W}^{*\top}\Delta\mathbf{Y}^d\boldsymbol{\gamma} + \mathbf{W}^{*\top}\Delta\mathbf{X}\boldsymbol{\beta} + \mathbf{W}^{*\top}\Delta\boldsymbol{\varepsilon}^*.$$

The semiparametric GMM estimator of $(\boldsymbol{\gamma}, \boldsymbol{\beta})$ can be constructed by minimizing

$$Q_n(\boldsymbol{\gamma}, \boldsymbol{\beta}) = \frac{1}{nT^*}(\mathbf{W}^{*\top}\Delta\boldsymbol{\varepsilon}^*)^\top \boldsymbol{\Omega}^{-1}(\mathbf{W}^{*\top}\Delta\boldsymbol{\varepsilon}^*), \tag{2.6}$$

where $\mathbf{W}^{*\top} = (\mathbf{W}_1^{*\top}, \ldots, \mathbf{W}_n^{*\top})$ and $\boldsymbol{\Omega} = \sum_{i=1}^n \mathbf{W}_i^{*\top}\mathbf{B}\mathbf{B}^\top\mathbf{W}_i^*/(nT^*)$ with

$$\mathbf{B} = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & -1 & 1 \end{pmatrix}_{T^*\times(T^*+1)}.$$

Applying the weighted least squares procedure, we obtain

$$\begin{aligned}\begin{pmatrix} \widehat{\boldsymbol{\gamma}}_n \\ \widehat{\boldsymbol{\beta}}_n \end{pmatrix} &= \left\{(\Delta\mathbf{Y}^d, \Delta\mathbf{X})^\top(\mathbf{I}_{nT^*} - \mathbf{M})\mathbf{W}\boldsymbol{\Omega}^{-1}\mathbf{W}^\top(\mathbf{I}_{nT^*} - \mathbf{M})(\Delta\mathbf{Y}^d, \Delta\mathbf{X})\right\}^{-1} \\ &\quad \cdot (\Delta\mathbf{Y}^d, \Delta\mathbf{X})^\top(\mathbf{I}_{nT^*} - \mathbf{M})\mathbf{W}\boldsymbol{\Omega}^{-1}\mathbf{W}^\top(\mathbf{I}_{nT^*} - \mathbf{M})\Delta\mathbf{y}, \end{aligned}\tag{2.7}$$

which in turn yields $\widehat{\boldsymbol{\theta}}_n = (\widehat{\boldsymbol{\theta}}_{1,n}, \ldots, \widehat{\boldsymbol{\theta}}_{q,n})^\top = (\tilde{\mathbf{N}}^\top\tilde{\mathbf{N}})^{-1}\tilde{\mathbf{N}}^\top(\Delta\mathbf{y} - \Delta\mathbf{Y}^d\widehat{\boldsymbol{\gamma}}_n - \Delta\mathbf{X}\widehat{\boldsymbol{\beta}}_n)$ and

$$\widehat{\phi}_{r,n}(z_{itr}) = \widehat{\boldsymbol{\theta}}_{r,n}^\top\mathbf{N}_r(z_{itr}), \quad r = 1, \ldots, q. \tag{2.8}$$

In addition, an estimator of the error variance $\sigma^2 = \mathrm{var}(\varepsilon_{it}^2)$ is

$$\widehat{\sigma}_n^2 = \frac{\|\Delta\mathbf{y} - \Delta\mathbf{Y}^d\widehat{\boldsymbol{\gamma}}_n - \Delta\mathbf{X}\widehat{\boldsymbol{\beta}}_n - \tilde{\mathbf{N}}\widehat{\boldsymbol{\theta}}_n\|_2^2}{2(nT^* - d - p - qJ_n)}. \tag{2.9}$$

Remark 1. Several approaches exist for choosing the optimal number of B-spline basis. A popular approach is the cross-validation approach considered by Huang and Shen (2004) and Huang et al. (2004). The approach we use, which follows from Wang and Yang (2007), does not require any iteration, and is thus more straightforward. See Section 5 for a description of the method.

3. Theoretical results

This section focuses on the asymptotic properties of the estimators constructed in Section 2. Let us consider the following definition taken from Li (2000) and Baltagi and Li (2002).

Definition 1. A function $\varphi(\mathbf{z})$, $\mathbf{z} = (z_1, \ldots, z_q)$, is said to belong to an additive class $\mathcal{A}$ if it satisfies (i) $\varphi(\mathbf{z}) = \sum_{r=1}^q \varphi_r(z_r)$, where $\varphi_r(\cdot)$ is twice differentiable in the interior of its support and square integrable; and (ii) $\mathrm{E}(\varphi_r(z_r)) = 0$ for $r = 1, \ldots, q$.

We use $\mathrm{E}_{\mathcal{A}}\psi(\mathbf{z})$ to denote the projection of a scalar or vector function $\psi(\mathbf{z})$ onto the additive class $\mathcal{A}$. Thus, $\mathrm{E}_{\mathcal{A}}\psi(\mathbf{z})$ is an element belonging in $\mathcal{A}$ and the closest function to $\psi(\mathbf{z})$ among all the functions in $\mathcal{A}$. More precisely,

$$\mathrm{E}\{(\psi(\mathbf{z}) - \mathrm{E}_{\mathcal{A}}\psi(\mathbf{z}))(\psi(\mathbf{z}) - \mathrm{E}_{\mathcal{A}}\psi(\mathbf{z}))^\top\} = \inf_{\varphi(\mathbf{z})\in\mathcal{A}} \mathrm{E}\{(\psi(\mathbf{z}) - \varphi(\mathbf{z}))(\psi(\mathbf{z}) - \varphi(\mathbf{z}))^\top\}. \tag{3.1}$$

Now, let $\|\phi_r\|_2 = \{\int_{\mathcal{R}^1}\phi_r^2(z)dz\}^{1/2}$ be the L_2-norm of a square integrable function $\phi_r(z)$ on $\mathcal{R}^1$. We say that an estimator $\widehat{\phi}_{r,n}$ of ϕ_r is consistent if $\|\widehat{\phi}_{r,n} - \phi_r\|_2 = o_p(1)$.

The following regularity conditions are needed for establishing the asymptotic properties of the resultant estimators.

Assumption (A1). $\{(Y_{it}, X_{itk}, Z_{itr}) : k = 1, \ldots, p; r = 1, \ldots, q\}$ are independent and identically distributed, and $\sup_t \mathrm{E}(\|\mathbf{x}_{it}\|_2^4 + \|\mathbf{z}_{it}\|_2^4 + |\varepsilon_{it}|^4) < \infty$.

Assumption (A2). The marginal densities of Z_r for $r = 1, \ldots, q$, defined as $f_r(z)$, are Lipschitz continuous and satisfy $0 < \min_r \inf_z f_r(z) \le \max_r \sup_z f_r(z) < \infty$.

Assumption (A3). The functions $\{\phi_r(z)\}_{r=1}^q$ and $g(z) = \mathrm{E}_{\mathcal{A}}(\mathrm{E}(y, x|z))$ have continuous and bounded ℓth derivative ($\ell \ge 3$). Moreover, $J_n^2/n \to 0$ and $\sqrt{n}J_n^{-\ell} \to 0$ as $n \to \infty$.

Assumption (A4). $\boldsymbol{\Gamma}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$ are positive definite matrices, and their eigenvalues are bounded away from 0 and infinity, where $\boldsymbol{\Sigma} = (T^*)^{-1}\mathrm{E}\{(\mathbf{W}_1 - \mathrm{E}_{\mathcal{A}}\mathbf{W}_1)^\top\mathbf{B}\mathbf{B}^\top(\mathbf{W}_1 - \mathrm{E}_{\mathcal{A}}\mathbf{W}_1)\}$, $\boldsymbol{\Gamma} = (T^*)^{-1}\mathrm{E}\{((\Delta\mathbf{Y}_1^d, \Delta\mathbf{X}_1) - \mathrm{E}_{\mathcal{A}}(\Delta\mathbf{Y}_1^d, \Delta\mathbf{X}_1))^\top(\mathbf{W}_1 - \mathrm{E}_{\mathcal{A}}\mathbf{W}_1)\}$, $\Delta\mathbf{X}_1 = (\Delta\mathbf{x}_{1(d+3)}, \ldots, \Delta\mathbf{x}_{1T})^\top$ and $\Delta\mathbf{Y}_1^d = (\Delta\mathbf{y}_{1(d+3)}^d, \ldots, \Delta\mathbf{y}_{1T}^d)^\top$.

Assumption (A5). For any given consistent estimator $\widehat{\boldsymbol{\delta}}_n$ of $\boldsymbol{\delta} = (\boldsymbol{\gamma}^\top, \boldsymbol{\beta}^\top, \boldsymbol{\theta}^\top)^\top$, $\mathbf{K}^\top\mathbf{W}\boldsymbol{\Omega}^{-1}\mathbf{W}^\top\mathbf{K} + nT^*\boldsymbol{\Xi}_{\lambda_n}(\widehat{\boldsymbol{\delta}}_n)/2$ is positive definite in the sense that its eigenvalues are bounded away from 0 and infinity, where $\mathbf{K} = (\Delta\mathbf{Y}^d, \Delta\mathbf{X}, \tilde{\mathbf{N}})$, and $\boldsymbol{\Xi}_{\lambda_{1,n},\lambda_{2,n}}(\widehat{\boldsymbol{\delta}}_n)$ is a block diagonal matrix defined in (4.5).

Remark 2. Assumptions (A1) and (A2) are standard conditions for nonparametric and semiparametric models (e.g., Li, 2000; Baltagi and Li, 2002 and Huang and Shen, 2004). Assumption (A3) determines the rate of convergence of nonparametric estimators, which is similar to Assumption 3 in Li (2000), Assumption 2.3 in Baltagi and Li (2002) and Assumption (d) of Huang and Yang (2004). Assumptions (A4) and (A5) are needed to circumvent the singularity of matrices when constructing estimators and establishing their asymptotic properties.

Theorem 1. (i) *Suppose that Assumptions (A1)–(A4) hold. Then as* $n \to \infty$,

$$\sqrt{nT^*}\begin{pmatrix}\widehat{\boldsymbol{\gamma}}_n - \boldsymbol{\gamma}\\ \widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\end{pmatrix} \to_D N(\mathbf{0}, \sigma^2(\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Gamma}^\top)^{-1}) \tag{3.2}$$

where

$$\boldsymbol{\Sigma} = \frac{1}{T^*}\mathrm{E}\{(\mathbf{W}_1 - \mathrm{E}_{\mathcal{A}}\mathbf{W}_1)^\top\mathbf{B}\mathbf{B}^\top(\mathbf{W}_1 - \mathrm{E}_{\mathcal{A}}\mathbf{W}_1)\}$$

and

$$\boldsymbol{\Gamma} = \frac{1}{T^*}\mathrm{E}\{((\Delta\mathbf{Y}_1^d, \Delta\mathbf{X}_1) - \mathrm{E}_{\mathcal{A}}(\Delta\mathbf{Y}_1^d, \Delta\mathbf{X}_1))^\top(\mathbf{W}_1 - \mathrm{E}_{\mathcal{A}}\mathbf{W}_1)\}.$$

(ii) *The random variables*

$$\widehat{\boldsymbol{\Sigma}}_n = \frac{1}{nT^*}\mathbf{W}^\top(\mathbf{I}_{nT^*} - \mathbf{M})(\mathbf{I}_n \otimes \mathbf{B}\mathbf{B}^\top)(\mathbf{I}_{nT^*} - \mathbf{M})\mathbf{W}$$

and

$$\widehat{\boldsymbol{\Gamma}}_n = \frac{1}{nT^*}(\Delta\mathbf{Y}^d, \Delta\mathbf{X})^\top(\mathbf{I}_{nT^*} - \mathbf{M})\mathbf{W}$$

are consistent estimators of $\boldsymbol{\Gamma}$ *and* $\boldsymbol{\Sigma}$ *respectively.*

Theorem 2. *Assume that (A1)–(A4) hold, and* ϱ*, the degree of the B-spline, is no smaller than* ℓ *defined in Assumption (A3). Then as* $n \to \infty$*, we have*

$$\max_{1\le r\le q}\|\widehat{\phi}_{r,n} - \phi_r\|_2 = O_p(\sqrt{J_n/n} + J_n^{-\ell}). \tag{3.3}$$

Theorem 2 provides a consistent estimator of $\phi_r(\cdot)$ when $J_n \asymp n^\tau$ with $\tau \ge 1/(2\ell+1)$. In particular, if $J_n \asymp n^{1/(2\ell+1)}$, then $\|\widehat{\phi}_{r,n} - \phi_r\|_2 = O_p(n^{-\ell/(2\ell+1)})$.

Theorem 3. (i) *Under the same conditions as in Theorem 1. As* $n \to \infty$,

$$\sqrt{nT^*}(\widehat{\sigma}_n^2 - \sigma^2) \to_D N(0, \kappa), \tag{3.4}$$

where $\kappa = \{(2T^* - 1)\mathrm{E}(\varepsilon_{11}^4) + \sigma^4\}/2T^*$.

(ii) *A consistent estimator of* κ *is*

$$\widehat{\kappa}_n = \frac{2T^* - 1}{4T^*}\left\{\frac{1}{nT^*}\sum_{i=1}^{n}\sum_{t=d+3}^{T}\widehat{\Delta\varepsilon_{it}}^4 - 6\widehat{\sigma}_n^4\right\} + \frac{1}{2T^*}\widehat{\sigma}_n^4, \tag{3.5}$$

where

$$\widehat{\Delta\varepsilon_{it}} = \Delta y_{it} - (\Delta\mathbf{y}_{it}^d)^\top\widehat{\boldsymbol{\gamma}}_n - \Delta\mathbf{x}_{it}^\top\widehat{\boldsymbol{\beta}}_n - \tilde{\mathbf{N}}_{(i-1)T^*+t,*}\widehat{\boldsymbol{\theta}}_n,$$

with $\tilde{\mathbf{N}}_{(i-1)T^*+t,*}$ *being the* $((i-1)T^* + t)$*th row of* $\tilde{\mathbf{N}}$.

Theorem 3 shows the property of $\sqrt{n}$-consistency of the variance estimator that would be required for inference purposes but further investigation of its properties is beyond the scope of this paper.

4. Identifying the lag order and significant covariates

This section is devoted to a discussion on the selection of the lag order for the dynamic component and exogenous variables of the parametric and nonparametric components of (2.1). As mentioned in Section 1, we adopt the semiparametric

SCAD method to achieve this task. The objective function for estimating the unknowns with the SCAD penalty is

$$Q_n^*(\boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\theta}) = \frac{1}{nT^*}(\Delta\mathbf{y} - \Delta\mathbf{Y}^d\boldsymbol{\gamma} - \Delta\mathbf{X}\boldsymbol{\beta} - \tilde{\mathbf{N}}\boldsymbol{\theta})^\top \mathbf{W}\boldsymbol{\Omega}^{-1}\mathbf{W}^\top(\Delta\mathbf{y} - \Delta\mathbf{Y}^d\boldsymbol{\gamma} - \Delta\mathbf{X}\boldsymbol{\beta} - \tilde{\mathbf{N}}\boldsymbol{\theta}) + \sum_{s=1}^{d} pe_{\lambda_{1n}}(|\gamma_s|) + \sum_{k=1}^{p} pe_{\lambda_{1n}}(|\beta_k|) + \sum_{r=1}^{q} pe_{\lambda_{2n}}(\|\boldsymbol{\theta}_r\|_{\mathrm{N}}), \tag{4.1}$$

where λ_{1n} and λ_{2n} are tuning parameters controlling the model complexity and can be selected by data driven methods, and $\|\boldsymbol{\theta}_r\|_{\mathrm{N}} = (\boldsymbol{\theta}_r^\top \mathbf{N}_r^\top \mathbf{N}_r \boldsymbol{\theta}_r)^{1/2}$. The penalized semiparametric GMM estimator would satisfy

$$(\breve{\boldsymbol{\gamma}}_n^\top, \breve{\boldsymbol{\beta}}_n^\top, \breve{\boldsymbol{\theta}}_n^\top)^\top = \arg\min Q_n^*(\boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\theta}). \tag{4.2}$$

We use the local quadratic approximation-based iterative algorithm (Fan and Li, 2001) to find solutions to (4.2). Now, given an initial $\boldsymbol{\gamma}_0 = (\gamma_{0,1}, \ldots, \gamma_{0,d})^\top$, for some $\gamma_{0,s}$, $s = 1, \ldots, d$, and γ_s in a close neighbourhood of $\gamma_{0,s}$, we have

$$pe_{\lambda_{1n}}(|\gamma_s|) \approx pe_{\lambda_{1n}}(|\gamma_{0,s}|) + \frac{pe'_{\lambda_{1n}}(|\gamma_{0,s}|)}{2|\gamma_{0,s}|}(\gamma_s^2 - \gamma_{0,s}^2). \tag{4.3}$$

In the same manner we can obtain the approximation for $pe_{\lambda_{1n}}(|\beta_k|)$ and $pe_{\lambda_{2n}}(\|\boldsymbol{\theta}_r\|_{\mathrm{N}})$. Substituting (4.3) into (4.1) and removing the irrelevant constant terms, the desired estimator can be written as

$$\breve{\boldsymbol{\delta}}_n = \left\{\mathbf{K}^\top\mathbf{W}\boldsymbol{\Omega}^{-1}\mathbf{W}^\top\mathbf{K} + \frac{nT^*}{2}\boldsymbol{\Xi}_{\lambda_n}(\widehat{\boldsymbol{\delta}}_n)\right\}^{-1}\mathbf{K}^\top\mathbf{W}\boldsymbol{\Omega}^{-1}\mathbf{W}^\top\Delta\mathbf{y}, \tag{4.4}$$

which resembles a ridge regression estimator, where $\widehat{\boldsymbol{\delta}}_n$ is an initial estimator of $\boldsymbol{\delta} = (\boldsymbol{\gamma}^\top, \boldsymbol{\beta}^\top, \boldsymbol{\theta}^\top)^\top$ and

$$\boldsymbol{\Xi}_{\lambda_{1,n},\lambda_{2,n}}(\widehat{\boldsymbol{\delta}}_n) = \mathrm{blkdiag}\left\{\frac{pe'_{\lambda_{1n}}(|\widehat{\gamma}_{1,n}|)}{|\widehat{\gamma}_{1,n}|}, \ldots, \frac{pe'_{\lambda_{1n}}(|\widehat{\gamma}_{d,n}|)}{|\widehat{\gamma}_{d,n}|}, \frac{pe'_{\lambda_{1n}}(|\widehat{\beta}_{1,n}|)}{|\widehat{\beta}_{1,n}|}, \ldots, \frac{pe'_{\lambda_{1n}}(|\widehat{\beta}_{p,n}|)}{|\widehat{\beta}_{p,n}|}, \frac{pe'_{\lambda_{2n}}(\|\widehat{\boldsymbol{\theta}}_{1,n}\|_{\mathrm{N}})}{\|\widehat{\boldsymbol{\theta}}_{1,n}\|_{\mathrm{N}}}\mathbf{N}_1^\top\mathbf{N}_1, \ldots, \frac{pe'_{\lambda_{2n}}(\|\widehat{\boldsymbol{\theta}}_{q,n}\|_{\mathrm{N}})}{\|\widehat{\boldsymbol{\theta}}_{q,n}\|_{\mathrm{N}}}\mathbf{N}_q^\top\mathbf{N}_q\right\}. \tag{4.5}$$

It is well-known that the choice of tuning parameter plays an important role for shrinkage estimators. Wang et al. (2007) proposed a BIC-based selector and demonstrated its selection consistency in the penalized least squares context. Here, we choose λ_{1n} and λ_{2n} such that

$$\mathrm{BIC}_{\lambda_{1,n},\lambda_{2,n}} = \log(\breve{\sigma}_n^2) + \mathrm{DF}_{\lambda_{1,n},\lambda_{2,n}}\frac{\log(nT^*)}{nT^*} \tag{4.6}$$

attains a minimum, where $\breve{\sigma}_n^2 = \|\Delta\mathbf{y} - \Delta\mathbf{Y}^d\breve{\boldsymbol{\gamma}}_n - \Delta\mathbf{X}\breve{\boldsymbol{\beta}}_n - \tilde{\mathbf{N}}\breve{\boldsymbol{\theta}}_n\|_2^2/2(nT^* - d - p - qJ_n)$ and $\mathrm{DF}_{\lambda_{1,n},\lambda_{2,n}}$ is the generalized degrees of freedom (Fan and Li, 2001), defined as

$$\mathrm{DF}_{\lambda_{1,n},\lambda_{2,n}} = \mathrm{tr}\left\{\mathbf{K}\left(\mathbf{K}^\top\mathbf{W}\boldsymbol{\Omega}^{-1}\mathbf{W}^\top\mathbf{K} + nT^*\boldsymbol{\Xi}_{\lambda_{1,n},\lambda_{2,n}}(\widehat{\boldsymbol{\delta}}_n)/2\right)^{-1}\mathbf{K}^\top\mathbf{W}\boldsymbol{\Omega}^{-1}\mathbf{W}^\top\right\}. \tag{4.7}$$

For notational convenience, let $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_{(1)}^\top, \boldsymbol{\gamma}_{(0)}^\top)^\top$ with $\boldsymbol{\gamma}_{(1)} = (\gamma_1, \ldots, \gamma_{d_0})^\top$ denoting the nonzero component and $\boldsymbol{\gamma}_{(0)} = (\gamma_{d_0+1}, \ldots, \gamma_d)^\top = \mathbf{0}_{d-d_0}$. Similarly, write $\boldsymbol{\beta} = (\boldsymbol{\beta}_{(1)}^\top, \boldsymbol{\beta}_{(0)}^\top)^\top$ with $\boldsymbol{\beta}_{(1)} = (\beta_1, \ldots, \beta_{p_0})^\top$, and $\boldsymbol{\phi} = (\boldsymbol{\phi}_{(1)}^\top, \boldsymbol{\phi}_{(0)}^\top)^\top$ with $\boldsymbol{\phi}_{(1)}^\top = (\phi_1, \ldots, \phi_{q_0})^\top$. Let $\Lambda_1 = \{(s,k,r) : \gamma_s \neq 0, \beta_k \neq 0, \phi_r(z) \not\equiv 0\}$, $\Lambda_2 = \{(s,k,r) : \gamma_s = 0, \beta_k = 0, \phi_r \equiv 0\}$, and $\breve{\Lambda}_{1,n}$ and $\breve{\Lambda}_{2,n}$ be their estimated sets. Moreover, let $\Delta\mathbf{Y}_{(1)}^d$ and $\Delta\mathbf{X}_{(1)}$ be the columns of $\Delta\mathbf{Y}^d$ and $\Delta\mathbf{X}$ corresponding to the selected nonzero coefficients respectively. Now, define

$$a_n = \max_{s,k,r}\{pe'_{\lambda_{1n}}(|\gamma_s|), pe'_{\lambda_{1n}}(|\beta_k|), pe'_{\lambda_{2n}}(\|\boldsymbol{\theta}_r\|_{\mathrm{N}}) : s,k,r \in \Lambda_1\}$$

and

$$b_n = \max_{s,k,r}\{pe''_{\lambda_{1n}}(|\gamma_s|), pe''_{\lambda_{1n}}(|\beta_k|), pe''_{\lambda_{2n}}(\|\boldsymbol{\theta}_r\|_{\mathrm{N}}) : s,k,r \in \Lambda_1\}.$$

Also, assume, for $s, k, r \in \Lambda_2$, that

$$\min\left\{\liminf_{n\to\infty}\liminf_{\gamma_s\to 0+}\frac{pe'_{\lambda_{1n}}(|\gamma_s|)}{\lambda_{1n}}, \liminf_{n\to\infty}\liminf_{\beta_k\to 0+}\frac{pe'_{\lambda_{1n}}(|\beta_k|)}{\lambda_{1n}}, \liminf_{n\to\infty}\liminf_{(\mathbf{N}_r\boldsymbol{\theta}_r)\to\mathbf{0}+}\frac{pe'_{\lambda_{2n}}(\|\boldsymbol{\theta}_r\|_{\mathrm{N}})}{\lambda_{2n}}\right\} > 0. \tag{4.8}$$

Theorem 4. *Let Assumptions (A1)–(A5) hold, and $b_n \to 0$ as $n \to \infty$. Then*

(i) $\|\breve{\boldsymbol{\gamma}}_n - \boldsymbol{\gamma}\|_2 = O_p((nT^*)^{-1/2} + a_n)$;
(ii) $\|\breve{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\|_2 = O_p((nT^*)^{-1/2} + a_n)$; *and*
(iii) $\|\breve{\phi}_{r,n} - \phi_r\|_2 = O_p((nT^*)^{-2/5} + a_n)$, *for* $r = 1, \ldots, q$ *when* $J_n = (nT^*)^{1/5}$.

Theorem 5. *Suppose that* $\max(\lambda_{1n}, \lambda_{2n}) \to 0$ *and* $\min(\lambda_{1n}, \lambda_{2n})/\max(J_n^{-\ell}, \sqrt{J_n/n}) \to \infty$ *as* $n \to \infty$. *Further, assume that* $\|\breve{\boldsymbol{\gamma}}_n - \boldsymbol{\gamma}\|_2 \asymp (nT^*)^{-1/2}$, $\|\breve{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\|_2 \asymp (nT^*)^{-1/2}$ *and* $\|\breve{\phi}_{r,n}(z) - \phi_r(z)\|_2 \asymp (nT^*)^{-\ell/(2\ell+1)}$. *Combined with Assumptions (A1)–(A5) and (4.8), these conditions lead to the following so-called variable selection consistency property:*

$$\mathrm{P}(\breve{\Lambda}_{2,n} = \Lambda_2) \to 1.$$

By Theorem 5, the SCAD-based semiparametric GMM estimator is "selection consistent" in the sense that it can identify the significant regressors correctly with probability approaching 1 as the sample size grows to infinity.

Theorem 6. *Assume that (A1)–(A5) and (4.8) hold. If $b_n \to 0$ as $n \to \infty$, then*

$$\sqrt{nT^*}\begin{pmatrix}\breve{\boldsymbol{\gamma}}_n - \boldsymbol{\gamma}_{(1)} \\ \breve{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_{(1)}\end{pmatrix} \to_D N(\mathbf{0}, \sigma^2(\boldsymbol{\Gamma}_{(1)}\boldsymbol{\Sigma}_{(1)}^{-1}\boldsymbol{\Gamma}_{(1)}^\top)^{-1}), \tag{4.9}$$

with

$$\boldsymbol{\Sigma}_{(1)} = \frac{1}{T^*}\mathrm{E}\{(\mathbf{W}_{(1),1} - \mathrm{E}_{\mathcal{A}}\mathbf{W}_{(1),1})^\top \mathbf{B}\mathbf{B}^\top(\mathbf{W}_{(1),1} - \mathrm{E}_{\mathcal{A}}\mathbf{W}_{(1),1})\}$$

and

$$\boldsymbol{\Gamma}_{(1)} = \frac{1}{T^*}\mathrm{E}\{((\Delta\mathbf{Y}^d_{(1),1}, \Delta\mathbf{X}_{(1),1}) - \mathrm{E}_{\mathcal{A}}(\Delta\mathbf{Y}^d_{(1),1}, \Delta\mathbf{X}_{(1),1}))^\top(\mathbf{W}_{(1),1} - \mathrm{E}_{\mathcal{A}}\mathbf{W}_{(1),1})\},$$

where $\mathbf{W}_{(1),1}$, $\Delta\mathbf{Y}^d_{(1),1}$ and $\Delta\mathbf{X}_{(1),1}$ are defined analogously to $\mathbf{W}$, $\Delta\mathbf{Y}^d$ and $\Delta\mathbf{X}$ respectively.

Remark 3. Theorem 6 shows that the identified significant estimators have an oracle property in the sense that they have the same joint asymptotic distribution as when the zero coefficients were known in advance. That being said, we do not know to which estimated parameters this theorem can be applied because in practice, it is unknown as to which coefficients are zero and which are non-zero.

5. Numerical experiments

5.1. Monte Carlo simulations

This section reports the results of Monte Carlo studies undertaken to investigate the empirical performance of the proposed procedure. Designs 1 and 2 are for the purpose of investigating the asymptotic properties of the GMM estimators assuming that the lag order and significant exogenous variables are known in advance. The effectiveness of the SCAD-GMM procedure is examined under Design 3.

Design 1. The data were generated by the panel process with an AR(1) dynamic order:

$$y_{it} = \mu_i + \gamma y_{i(t-1)} + \beta x_{it} + \phi_1(z_{it1}) + \phi_2(z_{it2}) + \varepsilon_{it}, \quad i = 1, \ldots, n;\ t = 2, \ldots, T,$$

where $\mu_i = (T-1)^{-1}\sum_{t=2}^{T} x_{it}$, $x_{it} \sim$ I.I.D. Bernoulli(1, 0.5), $z_{it1} = -3 + 5u_{it1}$, $z_{it2} = 2 - 4u_{it2}$, $u_{it} \sim$ I.I.D.$U(0, 1)$, $\phi_1(z_1) = (z_1 - 0.5)^2$ and $\phi_2(z_2) = z_2(e^{z_2} - 1)$. Let $\beta = 1.5$ and $\gamma = 0.1, -0.6, 0.9$. We consider the following two error scenarios:

Scenario I (homoscedastic errors): $\varepsilon_{it} \sim$ I.I.D.$N(0, 1)$;
Scenario II (heteroscedastic errors): $\varepsilon_{it} = \sqrt{h_{it}}\epsilon_{it}$, where $\epsilon_{it} \sim$ I.I.D.$N(0, 1)$, and h_{it} is described by: (1) $h_{it} = 0.1 + 0.3y^2_{it-1}$ or (2) $h_{it} = 0.1 + 0.1y^2_{it-1} + 0.1z^2_{it1} + 0.1z^2_{it2}$.

We consider sample sizes of $n = 50, 100, 200$ and $T = 5, 10, 15$. The experiment is replicated for $R = 1000$ times for $\gamma = 0.1$ and -0.6, and $R = 100\,000$ times for $\gamma = 0.9$; it is found that when γ is close to the unit-root, a larger number of replications are required for the experiment to yield stable results. Our evaluations of the estimators $\widehat{\gamma}_n$, $\widehat{\beta}_n$ and $\widehat{\sigma}^2_n$ are in terms the average of estimates (mean), standard deviation (std) of the estimates, standard errors (se), which are the square roots of the diagonal elements in the asymptotic variance matrix $\widehat{\sigma}^2_n(\widehat{\boldsymbol{\Gamma}}_n\widehat{\boldsymbol{\Sigma}}_n^{-1}\widehat{\boldsymbol{\Gamma}}_n^\top)$, and proximity of actual confidence interval

coverage to the targeted coverage of 0.95 (cp). For the estimators of $\phi_1(z_1)$ and $\phi_2(z_2)$, evaluation of efficiency is based on the mean of average squared error (MASE) measure defined as

$$\text{MASE}(\widehat{\phi}_{r,n}(z_r)) = \frac{1}{RnT^*}\sum_{\hbar=1}^{R}\sum_{i=1}^{n}\sum_{t=d+3}^{T}\left\{\widehat{\phi}_{r,n}^{(\hbar)}(z_{itr}) - \phi_r(z_{itr})\right\}^2,$$

where T^*, as defined previously, is $T-d-2$, and $\widehat{\phi}_{r,n}^{(\hbar)}(\cdot)$ is a GMM estimator of $\phi_r(\cdot)$ in the $\hbar$th simulation. We denote the standard deviation of the average squared error as SASE.

In our implementation we use the univariate quadratic B-splines with uniform knots. We determine the number of interior knots m_n by the method prescribed by Wang and Yang (2007) via the formula

$$m_n = \min(\lfloor c(n(T-d))^{1/5}\rfloor + 1, \lfloor (n(T-d)/2 - (p+d+q))/2\rfloor),$$

where $n(T-d)$ is the total sample size, c is a tuning constant, and $\lfloor a\rfloor$ denotes the largest integer of magnitude not greater than a. For simplicity, we set $c=1$ as varying the value of c generally makes no difference to the results. The constraint of $m_n \le (n(T-d)/2 - (p+d+q))/2$ ensures that the number of terms in (2.6) is not larger than $n(T-d)/2$. This condition is needed when the sample size is not large.

Tables 1(a) and 1(b) report the Monte Carlo results under the scenarios of homoscedastic and heteroscedastic errors respectively. Turning first to the efficiency of the parametric estimates, the following observations are apparent. First, other things being equal, the larger the value of n or T, the closer are the means of $\widehat{\gamma}_n$, $\widehat{\beta}_n$ and $\widehat{\sigma}_n^2$ to their true respective parameter values. We take this as an indication that $\widehat{\gamma}_n$, $\widehat{\beta}_n$ and $\widehat{\sigma}_n^2$ are asymptotically unbiased. Second, as n or T increases, the standard deviations of the estimators decrease, and std's are invariably very close to the corresponding se's. Third, in most cases, the actual confidence interval coverages obtained based on the proposed estimators differ only marginally from the targeted nominal level of 0.95, and as expected, their deviations from 0.95 decrease as n or T increases. The above comments also apply to the estimators for the nonparametric component. It is observed that the MASE's of $\widehat{\phi}_{1,n}(z_{it1})$ and $\widehat{\phi}_{2,n}(z_{it2})$ decrease as n or T increases, and their corresponding SASE's decrease as the number of observations increases. A comparison of results in Table 1(a) with the corresponding results in Table 1(b) shows that other things being equal, heteroscedasticity in the errors generally has the effect of inflating the standard deviations and standard errors of the estimator although the estimator remains asymptotically unbiased when the errors are heteroscedastic.

Additionally, we draw comparisons with results based on lagged level instruments as in Arellano and Bond (1991) under the homoscedastic error scenario. The results are displayed in Table 1(c), where it is shown that when the panel is relatively short (e.g., $T=5$), the first difference-based instruments generally lead to an estimator (labelled as BB) with smaller standard deviations than the estimator arising from lagged level-based instruments (labelled as AB). For longer panels, results based on the two procedures are comparable. This finding concurs with that of Blundell and Bond (1998) under the parametric dynamic model.

One reviewer has pointed that simulations with different ratios of the variances of idiosyncratic errors and individual effects are also useful for evaluating the performance of our approach. For this purpose we consider cases of σ^2/μ taking on 0.5, 1 and 5. The results, displayed in Table 1(d), indicate that the estimators of the parametric and nonparametric components commonly have smaller standard deviations when σ^2/μ is small than when it is large.

Design 2. This experiment extends the last experiment to a more complex model setup where the lag order is increased to 2, and the model contains a second exogenous variable. We consider the model

$$y_{it} = \mu_i + \gamma_1 y_{i(t-1)} + \gamma_2 y_{i(t-2)} + \beta_1 x_{it1} + \beta_2 x_{it2} + \phi_1(z_{it1}) + \phi_2(z_{it2}) + \varepsilon_{it},$$

where $\mu_i = (T-2)^{-1}\sum_{t=3}^{T}(x_{it1}+x_{it2})$ represents the fixed effect, x_{it1} and x_{it2} are each generated by a no-drift AR(1) process with I.I.D.$N(0,1)$ errors, and an autocorrelation coefficient of 0.3 (for x_{it1}) and 0.5 (for x_{it2}), $\phi_1(z) = \sin(\pi z) + \cos^2(2z)$ and $\phi_2(z) = \sin^2(2\pi z) + \cos(\pi z)$, and z is defined as under Design 1. We let $\boldsymbol{\beta}^\top = (1.2, -0.5)$ and $\boldsymbol{\gamma}^\top = (0.7, -0.4)$ or $(-0.9, 0.1)$. As in Design 1, we consider the scenarios of homoscedastic errors for which $\varepsilon_{it} \sim$ I.I.D.$N(0,1)$, and heteroscedastic errors for which we assume $\varepsilon_{it} = \sqrt{h_{it}}\epsilon_{it}$ with $h_{it} = 0.1 + 0.3y_{it-1}^2$. We set $n = 50, 100$ and $T = 10, 15$, and apply the same number of replications as in Design 1.

Table 2 reports the simulation results. As far as the performance of the various estimators is concerned, all of the general comments reported under Design 1 above apply in broad terms. Specifically, as n or T increases, the biases of all estimates decrease, as do the standard deviations of the estimates, which are never very different from the corresponding standard errors, and in all cases, the actual confidence interval coverages exhibit close proximity to the targeted 0.95 level. Other things being equal, heteroscedasticity has the effect of inflating the standard deviations of the estimates.

Design 3. The purpose of this experiment is to evaluate the performance of the SCAD-GMM procedure described in Section 4. The panel process being considered is

$$y_{it} = \mu_i + \sum_{s=1}^{5}\gamma_s y_{it-s} + \sum_{k=1}^{10}x_{itk}\beta_k + \sum_{r=1}^{9}\phi_r(z_{itr}) + \varepsilon_{it}, \tag{5.1}$$

Table 1(a)
Monte Carlo results under Design 1 with homoscedastic errors.

	T	5			10			15		
	n	50	100	200	50	100	200	50	100	200
		$\gamma = 0.1$								
$\widehat{\gamma}_n$	mean	0.1006	0.1006	0.0993	0.0996	0.0996	0.0999	0.0994	0.0999	0.0999
	std	0.0328	0.0233	0.0152	0.0139	0.0097	0.0070	0.0100	0.0066	0.0050
	se	0.0301	0.0237	0.0147	0.0139	0.0098	0.0069	0.0100	0.0068	0.0052
	cp	0.9290	0.9480	0.9400	0.9460	0.9430	0.9500	0.9500	0.9610	0.9560
$\widehat{\beta}_n$	mean	1.5074	1.4959	1.4994	1.5036	1.4990	1.5012	1.4977	1.4992	1.5006
	std	0.1852	0.1306	0.0915	0.1225	0.0857	0.0596	0.0931	0.0627	0.0458
	se	0.1773	0.1372	0.0917	0.1193	0.0852	0.0588	0.0906	0.0649	0.0463
	cp	0.9390	0.9570	0.9490	0.9440	0.9420	0.9470	0.9400	0.9550	0.9510
$\widehat{\sigma}_n^2$	mean	0.9313	1.0148	0.9966	0.9521	0.9728	0.9886	0.9666	0.9824	0.9908
	std	0.1298	0.0975	0.0649	0.0796	0.0590	0.0406	0.0646	0.0468	0.0333
	se	0.1443	0.1121	0.0785	0.0854	0.0617	0.0445	0.0670	0.0483	0.0345
	cp	0.9050	0.9690	0.9380	0.9010	0.9210	0.9520	0.9110	0.9170	0.9560
$\widehat{\phi}_{1,n}(z)$	MASE	0.0670	0.0321	0.0163	0.0254	0.0132	0.0062	0.0155	0.0080	0.0039
	SASE	0.0369	0.0177	0.0085	0.0141	0.0070	0.0034	0.0085	0.0043	0.0021
$\widehat{\phi}_{2,n}(z)$	MASE	0.0667	0.0318	0.0160	0.0262	0.0120	0.0062	0.0159	0.0077	0.0040
	SASE	0.0347	0.0166	0.0086	0.0144	0.0065	0.0033	0.0088	0.0040	0.0022
		$\gamma = -0.6$								
$\widehat{\gamma}_n$	mean	−0.5989	−0.5994	−0.5996	−0.5992	−0.5996	−0.6000	−0.6000	−0.5999	−0.6001
	std	0.0209	0.0135	0.0096	0.0100	0.0063	0.0049	0.0074	0.0049	0.0037
	se	0.0209	0.0128	0.0095	0.0094	0.0063	0.0048	0.0072	0.0050	0.0037
	cp	0.9470	0.9390	0.9450	0.9330	0.9520	0.9460	0.9460	0.9590	0.9460
$\widehat{\beta}_n$	mean	1.5016	1.5044	1.5003	1.5036	1.5006	1.4973	1.5025	1.5009	1.5004
	std	0.1930	0.1490	0.0994	0.1169	0.0869	0.0588	0.0934	0.0664	0.0479
	se	0.1909	0.1437	0.0952	0.1158	0.0857	0.0595	0.0946	0.0668	0.0469
	cp	0.9460	0.9370	0.9340	0.9470	0.9560	0.9550	0.9600	0.9500	0.9430
$\widehat{\sigma}_n^2$	mean	0.9853	0.9823	0.9931	0.9721	0.9789	0.9864	0.9681	0.9818	0.9904
	std	0.1246	0.0939	0.0659	0.0842	0.0580	0.0413	0.0665	0.0472	0.0339
	se	0.1518	0.1084	0.0778	0.0872	0.0622	0.0444	0.0670	0.0482	0.0345
	cp	0.9570	0.9620	0.9770	0.9330	0.9360	0.9470	0.9090	0.9290	0.9300
$\widehat{\phi}_{1,n}(z)$	MASE	0.0644	0.0303	0.0158	0.0257	0.0126	0.0064	0.0165	0.0080	0.0039
	SASE	0.0354	0.0156	0.0081	0.0134	0.0068	0.0034	0.0090	0.0041	0.0020
$\widehat{\phi}_{2,n}(z)$	MASE	0.0679	0.0324	0.0160	0.0270	0.0125	0.0062	0.0155	0.0078	0.0041
	SASE	0.0349	0.0167	0.0088	0.0141	0.0067	0.0032	0.0083	0.0043	0.0021
		$\gamma = 0.9$								
$\widehat{\gamma}_n$	mean	0.8133	0.8522	0.8814	0.8828	0.8869	0.8890	0.8911	0.8946	0.8960
	std	0.2265	0.1563	0.0906	0.0353	0.0306	0.0285	0.0164	0.0118	0.0111
	se	0.2142	0.1553	0.0913	0.0349	0.0298	0.0281	0.0167	0.0123	0.0113
	cp	0.9430	0.9660	0.9440	0.9090	0.9230	0.9210	0.9080	0.9370	0.9420
$\widehat{\beta}_n$	mean	1.4117	1.4570	1.4874	1.4921	1.4858	1.4955	1.4915	1.4952	1.5000
	std	0.2842	0.1857	0.1077	0.1276	0.0852	0.0642	0.0911	0.0666	0.0472
	se	0.2894	0.1922	0.1110	0.1205	0.0877	0.0614	0.0907	0.0670	0.0479
	cp	0.9460	0.9490	0.9560	0.9340	0.9570	0.9400	0.9330	0.9510	0.9530
$\widehat{\sigma}_n^2$	mean	1.0428	1.0543	0.9942	0.9335	0.9712	0.9804	0.9646	0.9776	0.9892
	std	0.4936	0.3028	0.1159	0.0836	0.0597	0.0490	0.0646	0.0466	0.0332
	se	0.1588	0.1180	0.0784	0.0839	0.0618	0.0441	0.0664	0.0482	0.0344
	cp	0.7560	0.8190	0.8530	0.8430	0.9120	0.8810	0.9010	0.9160	0.9350
$\widehat{\phi}_{1,n}(z)$	MASE	0.2018	0.1237	0.0413	0.0294	0.0152	0.0091	0.0167	0.0081	0.0042
	SASE	0.2953	0.1724	0.0414	0.0163	0.0086	0.0061	0.0087	0.0043	0.0022
$\widehat{\phi}_{2,n}(z)$	MASE	0.1544	0.0823	0.0322	0.0290	0.0144	0.0075	0.0164	0.0081	0.0041
	SASE	0.2117	0.1033	0.0289	0.0159	0.0082	0.0044	0.0087	0.0043	0.0023

where $\boldsymbol{\gamma} = (0, 0.7, 0, -0.4, \gamma_5)^\top$ or $(0, 0.9, 0, 0.1, \gamma_5)^\top$, $\boldsymbol{\beta} = (0, 0, 1.2, 0, \beta_5, 0, -0.5, 0, 0, 0)^\top$, with γ_5 and β_5 both being $1/\sqrt{nT^* \log(nT^*)}$, $\phi_1(z_1) = 3\cos(z_1 - 0.5) + z_1^2$, $\phi_4(z_2) = 2z_2/(1.2 - \sin(z_2))$, $\phi_2(z_2) = \phi_3(z_3) = \phi_5(z_5) = \cdots = \phi_9(z_9) = 0$, $z_r \sim$ I.I.D.$U(0, 0.1)$ for $r = 1, \ldots, 9$, $\mu_i = (T-5)^{-1} \sum_{t=6}^{T} \sum_{k=1}^{9} x_{itk}$, with x_{itk}, $k = 6, \ldots, 10$, being I.I.D.$N(0, 0.5)$ distributed, and x_{itk}, $k = 1, \ldots, 5$, following an AR(1) process with I.I.D.$N(0, 1)$ disturbances, no drift and an autocorrelation coefficient of 0.5, and $\varepsilon_{it} \sim$ I.I.D.$N(0, 1)$. We consider the same values of $n = 50, 100, 200$, $T = 10, 15$ and the same number of replications as in the previous two designs. When fitting the model, we allow a maximum of 5 lag dependent variables, 10 exogenous covariates and 9 nonparametric functions. We also compare the selection performance of SCAD-GMM with a

Table 1(b)
Monte Carlo results under Design 1 with heteroscedastic errors.

		Case (1)				Case (2)			
	T	10		15		10		15	
	n	50	100	50	100	50	100	50	100
		$\gamma = 0.1$							
$\widehat{\gamma}_n$	mean	0.0935	0.0941	0.0921	0.0962	0.0928	0.0946	0.0945	0.0966
	std	0.0533	0.0408	0.0397	0.0293	0.0557	0.0394	0.0407	0.0290
	se	0.0535	0.0406	0.0381	0.0298	0.0545	0.0396	0.0392	0.0297
	cp	0.9440	0.9430	0.9330	0.9420	0.9390	0.9540	0.9350	0.9520
$\widehat{\beta}_n$	mean	1.4993	1.4976	1.5016	1.4893	1.4985	1.4990	1.5113	1.5031
	std	0.3152	0.2328	0.2153	0.1793	0.2991	0.2439	0.2360	0.1786
	se	0.3001	0.2221	0.2116	0.1791	0.2764	0.2311	0.2398	0.1746
	cp	0.9360	0.9410	0.9430	0.9510	0.9210	0.9420	0.9530	0.9430
$\widehat{\sigma}_n^2$	mean	0.9278	0.9588	0.9505	0.9761	0.9178	0.9556	0.9457	0.9762
	std	0.1533	0.1254	0.1316	0.0975	0.1547	0.1122	0.1282	0.0862
$\widehat{\phi}_{1,n}(z)$	MASE	0.2375	0.1096	0.1304	0.0757	0.2111	0.1147	0.1489	0.0728
	SASE	0.1406	0.0621	0.0736	0.0426	0.1239	0.0633	0.0842	0.0410
$\widehat{\phi}_{2,n}(z)$	MASE	0.2166	0.1069	0.1365	0.0724	0.2349	0.1157	0.1704	0.0695
	SASE	0.1270	0.0589	0.0759	0.0382	0.1455	0.0619	0.1019	0.0373
		$\gamma = -0.6$							
$\widehat{\gamma}_n$	mean	−0.6039	−0.6002	−0.6054	−0.6013	−0.6002	−0.5998	−0.6022	−0.6009
	std	0.0541	0.0402	0.0412	0.0293	0.0521	0.0392	0.0410	0.0294
	se	0.0496	0.0387	0.0381	0.0285	0.0503	0.0382	0.0384	0.0297
	cp	0.9250	0.9380	0.9310	0.9480	0.9410	0.9480	0.9380	0.9510
$\widehat{\beta}_n$	mean	1.5062	1.5081	1.5100	1.4957	1.5013	1.5216	1.4771	1.5155
	std	0.3524	0.2703	0.2845	0.2142	0.3677	0.2883	0.2895	0.2259
	se	0.3402	0.2625	0.2709	0.2093	0.3611	0.2803	0.2764	0.2212
	cp	0.9470	0.9440	0.9450	0.9450	0.9530	0.9440	0.9280	0.9480
$\widehat{\sigma}_n^2$	mean	0.9179	0.9526	0.9416	0.9669	0.9195	0.9576	0.9453	0.9644
	std	0.1657	0.1335	0.1473	0.1248	0.1604	0.1300	0.1291	0.0989
$\widehat{\phi}_{1,n}(z)$	MASE	0.3260	0.1877	0.2374	0.1106	0.3048	0.1912	0.2347	0.1194
	SASE	0.2187	0.1220	0.2133	0.0673	0.2266	0.1157	0.1495	0.0735
$\widehat{\phi}_{2,n}(z)$	MASE	0.3335	0.1535	0.2417	0.1103	0.3264	0.1665	0.2297	0.1207
	SASE	0.2097	0.0917	0.2048	0.0729	0.2183	0.0999	0.1654	0.0713
		$\gamma = 0.9$							
$\widehat{\gamma}_n$	mean	0.7430	0.7464	0.7714	0.7808	0.7513	0.7528	0.8171	0.8357
	std	0.1360	0.1175	0.0597	0.0556	0.1267	0.1213	0.0665	0.0619
	se	0.1114	0.1049	0.0585	0.0525	0.1156	0.1029	0.0618	0.0540
	cp	0.7010	0.7030	0.7170	0.7480	0.7230	0.7380	0.7220	0.6990
$\widehat{\beta}_n$	mean	1.4096	1.3780	1.4629	1.4659	1.4269	1.3986	1.4473	1.4363
	std	0.4559	0.4171	0.4343	0.3398	0.5173	0.3722	0.4410	0.3577
	se	0.4349	0.3924	0.4209	0.3359	0.4970	0.3597	0.4115	0.3386
	cp	0.9240	0.9250	0.9390	0.9370	0.9300	0.9250	0.9390	0.9270
$\widehat{\sigma}_n^2$	mean	0.8453	0.8699	0.8952	0.9292	0.8483	0.8675	0.9205	0.9261
	std	0.2769	0.2188	0.2533	0.1842	0.2449	0.2136	0.2390	0.1957
$\widehat{\phi}_{1,n}(z)$	MASE	0.7453	0.5034	0.5868	0.3981	0.8166	0.5893	0.7742	0.4722
	SASE	0.5177	0.3604	0.4824	0.2969	0.5360	0.3359	0.5288	0.3147
$\widehat{\phi}_{2,n}(z)$	MASE	0.6454	0.4332	0.5660	0.3910	0.8836	0.5896	0.7768	0.4791
	SASE	0.4487	0.2859	0.3769	0.2001	0.6418	0.4576	0.5919	0.3681

Lasso-based (Tibshirani, 1996) GMM. In all our simulations, we use the BIC criterion described in Section 4 to determine the tuning parameter.

Table 3 reports the number of times in 1000 replications where the SCAD-GMM and Lasso-GMM procedures (labelled as SD and LA respectively) can identify the non-zero coefficients and/or functions as non-zero (labelled as NZ) and zero coefficients and/or functions as zero (labelled as Z) correctly. For example, when $\boldsymbol{\gamma} = (0, 0.7, 0, -0.4, \gamma_5)^\top$, $T = 10$ and $n = 50$, the SCAD (Lasso) procedure can correctly identify both of the two non-zero $\phi(\cdot)$ functions as non-zero 967 (502) times, all of the seven zero $\phi(\cdot)$ functions as zero 575 (497) times, all of non-zero coefficients and functions as non-zero 486 (392) times, and all of the zero coefficients and functions as zero 503 (364) times. The table reveals that by our criterion, SCAD-GMM is clearly a superior procedure to Lasso-GMM. Other things being equal, the effectiveness of both procedures improves as n or T increases. There is no substantial difference in results under the two specifications of $\boldsymbol{\gamma}$.

Table 1(c)
Comparison of results based on lagged levels instruments and difference instruments under Design 1 with homoscedastic errors.

	n	50				100			
	T	5		10		5		10	
		AB	BB	AB	BB	AB	BB	AB	BB
		$\gamma = 0.5$							
$\widehat{\gamma}_n$	mean	0.5007	0.4980	0.4988	0.4987	0.4989	0.5002	0.4992	0.5000
	std	0.0651	0.0531	0.0174	0.0166	0.0363	0.0277	0.0123	0.0108
	se	0.0622	0.0497	0.0168	0.0157	0.0347	0.0303	0.0121	0.0107
	cp	0.9410	0.9320	0.9450	0.9410	0.9480	0.9650	0.9490	0.9450
$\widehat{\beta}_n$	mean	1.5126	1.5103	1.4995	1.5016	1.5059	1.5005	1.4983	1.5073
	std	0.2220	0.1835	0.1178	0.1175	0.1457	0.1304	0.0844	0.0837
	se	0.2089	0.1772	0.1200	0.1170	0.1371	0.1323	0.0837	0.0840
	cp	0.9340	0.9400	0.9560	0.9480	0.9310	0.9500	0.9470	0.9490
$\widehat{\sigma}_n^2$	mean	0.9073	0.8778	0.9558	0.9501	0.9482	0.9426	0.9797	0.9714
	std	0.1301	0.1266	0.0845	0.0817	0.0970	0.0947	0.0599	0.0555
	se	0.1406	0.1358	0.0856	0.0810	0.1048	0.1045	0.0625	0.0616
	cp	0.8570	0.8070	0.8880	0.8890	0.8990	0.9130	0.9320	0.9380
$\widehat{\phi}_{1,n}(z)$	MASE	0.0734	0.0665	0.0271	0.0255	0.0348	0.0343	0.0126	0.0123
	SASE	0.0403	0.0352	0.0148	0.0137	0.0190	0.0184	0.0067	0.0065
$\widehat{\phi}_{2,n}(z)$	MASE	0.0670	0.0666	0.0255	0.0260	0.0357	0.0326	0.0128	0.0129
	SASE	0.0375	0.0369	0.0138	0.0136	0.0192	0.0175	0.0070	0.0068
		$\gamma = 0.9$							
$\widehat{\gamma}_n$	mean	0.8698	0.8761	0.8828	0.8850	0.8868	0.8914	0.8882	0.8931
	std	0.1313	0.1061	0.0378	0.0328	0.0975	0.0898	0.0307	0.0223
	se	0.1249	0.1032	0.0359	0.0317	0.0963	0.0875	0.0290	0.0226
	cp	0.9520	0.9450	0.9160	0.9130	0.9480	0.9460	0.9380	0.9460
$\widehat{\beta}_n$	mean	1.4627	1.4795	1.4844	1.4866	1.4827	1.4931	1.4907	1.4933
	std	0.2410	0.1948	0.1282	0.1218	0.1723	0.1459	0.0861	0.0834
	se	0.2449	0.1862	0.1218	0.1166	0.1711	0.1483	0.0872	0.0845
	cp	0.9510	0.9350	0.9330	0.9360	0.9460	0.9580	0.9410	0.9540
$\widehat{\sigma}_n^2$	mean	0.9578	0.9112	0.9494	0.9358	1.0026	0.9660	0.9693	0.9688
	std	0.1839	0.1427	0.0853	0.0825	0.1328	0.1107	0.0646	0.0608
	se	0.1769	0.1416	0.0851	0.0838	0.1339	0.1071	0.0616	0.0615
	cp	0.8500	0.8590	0.8790	0.8500	0.9170	0.8830	0.8810	0.8960
$\widehat{\phi}_{1,n}(z)$	MASE	0.1077	0.0946	0.0313	0.0288	0.0561	0.0507	0.0153	0.0141
	SASE	0.0773	0.0664	0.0173	0.0151	0.0504	0.0452	0.0087	0.0079
$\widehat{\phi}_{2,n}(z)$	MASE	0.1002	0.0967	0.0283	0.0275	0.0500	0.0450	0.0149	0.0130
	SASE	0.0713	0.0662	0.0152	0.0142	0.0377	0.0325	0.0083	0.0069

5.2. A real data application

This section presents a real data application of the proposed methodology. The data, extracted from Cornwell and Trumbull (1994), contains panel information on crime rate, the ratio of FBI index crimes to population, and its attributes in 90 counties in the state of North Carolina over the period 1981–1987. Our benchmark model is

$$CR_{it} = \mu_i + \sum_{s=1}^{3} \gamma_s CR_{i(t-s)} + \beta_1 AS_{it} + \beta_2 POL_{it} + \beta_3 WF_{it} + \beta_4 WS_{it} + \beta_5 WL_{it} + \beta_6 PY_{it} + \phi_1(PA_{it}) + \phi_2(PC_{it}) + \phi_3(PP_{it}) + \phi_4(DEN_{it}) + \phi_5(TAX_{it}) + \phi_6(WM_{it}) + \varepsilon_{it}, \quad i = 1, \ldots, 90, \ t = 1, \ldots, 7,$$

where the response variable CR_{it} is the crime rate of county i in the tth year, AS is the average prison sentence in days, POL is the number of police per capita, WF, WS, WL and WM represent, respectively, the average weekly wages of employees of federal government, state government, local government and manufacturing industries, and PY is the percentage of male population between the ages of 15 and 24 in the county. In addition, PA is the probability of being arrested, PC is the probability of conviction after arrest, PP is the probability of imprisonment after conviction, DEN is the population density, being the county's population divided by the county's land area and TAX represents per capita tax revenue.

For ease of interpretation, we scale all values of the covariates to between 0 and 1. The results are reported in Table 4, where EST, SE and CI denote the coefficient estimate associated with the corresponding variable shown on the first column of the table, its standard error, and the associated 95% confidence interval constructed using the wild bootstrap procedure proposed by Härdle et al. (2004). Three sets of GMM results are presented: those without the implementation of SCAD or other diagnostic tests are shown on the far left panel of the table; the middle panel presents the re-estimated results after removing the insignificant variables; the far right panel are results based on the SCAD-GMM procedure. The GMM results without variable selection by SCAD shows that, POL, the number of police per resident, and WS, the state government salary level, exert no significant effect on crime rate, while WF, the federal government salary level exerts a positive effect on crime

Table 1(d)
Monte Carlo results under Design 1 with different ratio of variance of idiosyncratic errors and individual effects.

	$\sigma^2/\lvert\mu\rvert$	$T=5, n=50$			$T=5, n=100$			$T=10, n=50$		
		0.5	1	5	0.5	1	5	0.5	1	5
		$\gamma=0.1$								
$\widehat{\gamma}_n$	mean	0.1000	0.0999	0.1025	0.1001	0.1003	0.1010	0.1000	0.0999	0.0996
	std	0.0106	0.0135	0.0404	0.0073	0.0096	0.0205	0.0036	0.0041	0.0095
	se	0.0109	0.0130	0.0390	0.0070	0.0093	0.0211	0.0037	0.0040	0.0094
	cp	0.9600	0.9460	0.9500	0.9440	0.9350	0.9640	0.9570	0.9430	0.9340
$\widehat{\beta}_n$	mean	1.5000	1.5003	1.5018	1.4997	1.4978	1.5019	1.4989	1.4990	1.4988
	std	0.0604	0.0769	0.1837	0.0388	0.0574	0.1112	0.0265	0.0352	0.0772
	se	0.0607	0.0757	0.1733	0.0384	0.0592	0.1121	0.0270	0.0335	0.0757
	cp	0.9480	0.9450	0.9360	0.9430	0.9500	0.9500	0.9560	0.9400	0.9400
$\widehat{\sigma}_n^2$	mean	1.0337	0.9628	0.9220	0.9331	0.9921	1.0042	0.9963	0.9532	0.9637
	std	0.1450	0.1461	0.1533	0.1004	0.1036	0.1184	0.0991	0.1087	0.1099
$\widehat{\phi}_{1,n}(z)$	MASE	0.0072	0.0124	0.0680	0.0033	0.0061	0.0291	0.0017	0.0027	0.0117
	SASE	0.0042	0.0070	0.0385	0.0017	0.0032	0.0150	0.0010	0.0015	0.0060
$\widehat{\phi}_{2,n}(z)$	MASE	0.0071	0.0107	0.0593	0.0030	0.0067	0.0284	0.0017	0.0028	0.0112
	SASE	0.0041	0.0059	0.0330	0.0017	0.0035	0.0162	0.0009	0.0015	0.0061
		$\gamma=0.9$								
$\widehat{\gamma}_n$	mean	0.8917	0.8724	0.8755	0.8992	0.8970	0.8464	0.8994	0.8972	0.8883
	std	0.0607	0.1093	0.1437	0.0327	0.0491	0.1660	0.0086	0.0150	0.0285
	se	0.0767	0.1187	0.1557	0.0390	0.0506	0.1665	0.0083	0.0146	0.0285
	cp	0.9960	0.9810	0.9870	0.9930	0.9670	0.9620	0.9370	0.9480	0.9250
$\widehat{\beta}_n$	mean	1.4982	1.4674	1.4957	1.4992	1.4972	1.4549	1.4998	1.4990	1.4921
	std	0.0580	0.1588	0.1567	0.0405	0.0725	0.1815	0.0252	0.0405	0.0796
	se	0.0750	0.1773	0.1758	0.0479	0.0750	0.1934	0.0247	0.0395	0.0802
	cp	0.9890	0.9720	0.9690	0.9810	0.9590	0.9680	0.9440	0.9520	0.9530
$\widehat{\sigma}_n^2$	mean	1.3233	1.2456	1.2503	1.1479	1.0842	1.1062	0.9583	1.0240	0.9471
	std	0.2659	0.2715	0.2852	0.1525	0.1563	0.1832	0.1018	0.1037	0.1055
$\widehat{\phi}_{1,n}(z)$	MASE	0.0166	0.0504	0.1203	0.0057	0.0132	0.1022	0.0015	0.0037	0.0169
	SASE	0.0178	0.0742	0.1292	0.0052	0.0116	0.1414	0.0009	0.0021	0.0096
$\widehat{\phi}_{2,n}(z)$	MASE	0.0124	0.0375	0.0797	0.0044	0.0108	0.0882	0.0014	0.0035	0.0163
	SASE	0.0114	0.0505	0.0602	0.0032	0.0084	0.1193	0.0008	0.0019	0.0091

rate. On the other hand, *AS*, the duration of prison sentence, *WL*, the local government salary level, and *PY*, the percentage of young male population, all have the effect of reducing crime rate. We find the last result somewhat surprising as we would expect crime rate to be positively associated with the percentage of young male population, *ceteris paribus*. When SCAD-GMM is used, the SCAD procedure selects only *WS* and *PY*. Interestingly, the coefficient estimate of *PY* by SCAD-GMM is positive, which is the opposite to the result obtained without SCAD selection. Both the GMM without SCAD and SCAD-GMM methods find $d=1$ to be the most appropriate lag order.

Fig. 1 exhibits the estimated nonparametric functions based on the SCAD-GMM method. The procedure selects all but the variable *PP* specified for the nonparametric component of the model. Fig. 1 shows that when *PA*, the probability of being arrested, is greater than 0.4, an increase in *PA* has the effect of reducing crime rate, and the rate of reduction increases as *PA* increases. The variable *PC*, which represents the probability of conviction, does not appear to have any significant effect on crime rate, except when $PC > 0.8$, a noticeable decline in *CR* is observed as *PC* increases. On the other hand, *DEN*, the population density, has the effect of increasing crime rate at an increasing rate. As well, a very high value of *TAX*, the tax rate, can escalate crime rate, but an increase in *WM*, the salary level of manufacturing employees, generally reduces crime rate. The estimated nonparametric functions based on GMM without SCAD variable selection are shown in Fig. 2. There are similarities as well as differences between the results shown in Figs. 1 and 2; for example, the estimated function of ϕ_1 based on the GMM method without SCAD selection shows that for very large values of *PA*, crime rate decreases at a decreasing rate in absolute terms when *PA* increases; also, without using SCAD selection, the GMM results show that an increase in manufacturing salary level from a moderately high level can increase the crime rate. These differ from those observed under the SCAD-GMM estimation procedure.

Acknowledgements

Li's work was supported by the SHUFE Excellent Doctoral Dissertation Cultivation Funds (No. 14110045) and Innovation Program of Shanghai Municipal Education Commission (No. 13YZ125). Wan's work was supported by a strategic grant from the City University of Hong Kong (No. 7008134). You's work was supported by grants from the National Natural Science Foundation of China (NSFC) (No. 11471203), Program for New Century Excellent Talents in University (NCET), the Leading Academic Discipline Program, 211 Project for Shanghai University of Finance and Economics (the 4th phase), and

Table 2
Monte Carlo results under Design 2.

		Homoscedastic errors				Heteroscedastic errors			
	T	10		15		10		15	
	n	50	100	50	100	50	100	50	100
		$(\gamma_1, \gamma_2) = (0.7, -0.4)$							
$\widehat{\gamma}_{1,n}$	mean	0.7014	0.6994	0.6951	0.7007	0.6936	0.6916	0.6932	0.6981
	std	0.0339	0.0254	0.0217	0.0166	0.0635	0.0517	0.0464	0.0334
	se	0.0328	0.0260	0.0223	0.0165	0.0612	0.0486	0.0455	0.0330
	cp	0.9460	0.9570	0.9430	0.9530	0.9370	0.9480	0.9330	0.9520
$\widehat{\gamma}_{2,n}$	mean	−0.3948	−0.4092	−0.4088	−0.3958	−0.4038	−0.4051	−0.4028	−0.4013
	std	0.0303	0.0230	0.0211	0.0152	0.0449	0.0324	0.0319	0.0219
	se	0.0296	0.0233	0.0212	0.0157	0.0426	0.0323	0.0307	0.0214
	cp	0.9440	0.9410	0.9270	0.9550	0.9270	0.9390	0.9410	0.9420
$\widehat{\beta}_{1,n}$	mean	1.2011	1.1985	1.1974	1.2076	1.2033	1.1977	1.2068	1.1931
	std	0.0676	0.0522	0.0508	0.0355	0.1087	0.0803	0.0821	0.0622
	se	0.0681	0.0515	0.0516	0.0373	0.1050	0.0801	0.0816	0.0618
	cp	0.9470	0.9450	0.9570	0.9550	0.9390	0.9490	0.9460	0.9480
$\widehat{\beta}_{2,n}$	mean	−0.5183	−0.5074	−0.4994	−0.5119	−0.5089	−0.5076	−0.4963	−0.4985
	std	0.0760	0.0508	0.0558	0.0386	0.1091	0.0876	0.0869	0.0659
	se	0.0733	0.0517	0.0542	0.0390	0.1055	0.0866	0.0826	0.0621
	cp	0.9330	0.9490	0.9310	0.9470	0.9440	0.9480	0.9330	0.9300
$\widehat{\sigma}_n^2$	mean	0.9979	1.0337	1.0216	0.9993	0.9348	0.9807	0.9722	0.9917
	std	0.0936	0.0659	0.0737	0.0505	0.1528	0.1442	0.1618	0.1180
	se	0.0960	0.0706	0.0737	0.0536	–	–	–	–
	cp	0.9380	0.9530	0.9500	0.9460	–	–	–	–
$\widehat{\phi}_{1,n}(z)$	MASE	0.0564	0.0481	0.0475	0.0399	0.1226	0.0778	0.0900	0.0607
	SASE	0.0157	0.0081	0.0098	0.0045	0.0616	0.0269	0.0369	0.0187
$\widehat{\phi}_{2,n}(z)$	MASE	0.0668	0.0490	0.0542	0.0447	0.1304	0.0826	0.0845	0.0694
	SASE	0.0158	0.0078	0.0095	0.0046	0.0600	0.0256	0.0319	0.0181
		$(\gamma_1, \gamma_2) = (-0.9, 0.1)$							
$\widehat{\gamma}_{1,n}$	mean	−0.9112	−0.9091	−0.9063	−0.8991	−0.9059	−0.9051	−0.9055	−0.9032
	std	0.0475	0.0327	0.0306	0.0201	0.0844	0.0696	0.0699	0.0491
	se	0.0456	0.0321	0.0299	0.0209	0.0789	0.0654	0.0581	0.0457
	cp	0.9250	0.9410	0.9380	0.9620	0.9410	0.9330	0.9160	0.9320
$\widehat{\gamma}_{2,n}$	mean	0.0890	0.0909	0.0918	0.0986	0.0903	0.0920	0.0873	0.0903
	std	0.0522	0.0365	0.0323	0.0211	0.0823	0.0646	0.0508	0.0388
	se	0.0506	0.0355	0.0319	0.0221	0.0733	0.0625	0.0469	0.0377
	cp	0.9420	0.9400	0.9480	0.9500	0.9180	0.9450	0.9330	0.9320
$\widehat{\beta}_{1,n}$	mean	1.1878	1.2342	1.1955	1.1925	1.1970	1.1972	1.1885	1.2184
	std	0.0717	0.0497	0.0515	0.0365	0.1203	0.1009	0.1049	0.0817
	se	0.0692	0.0519	0.0528	0.0367	0.1132	0.0990	0.1017	0.0816
	cp	0.9440	0.9020	0.9590	0.9470	0.9260	0.9450	0.9250	0.9340
$\widehat{\beta}_{2,n}$	mean	−0.4880	−0.4965	−0.4923	−0.4918	−0.4930	−0.5004	−0.5000	−0.4954
	std	0.0634	0.0521	0.0518	0.0384	0.1270	0.0953	0.0987	0.0857
	se	0.0650	0.0531	0.0535	0.0389	0.1154	0.0925	0.0937	0.0831
	cp	0.9580	0.9560	0.9500	0.9500	0.9240	0.9400	0.9290	0.9310
$\widehat{\sigma}_n^2$	mean	0.9785	1.0115	1.0030	0.9974	0.9022	0.9506	0.9133	0.9539
	std	0.0900	0.0648	0.0721	0.0549	0.2056	0.2001	0.2689	0.2255
	se	0.0941	0.0692	0.0726	0.0511	–	–	–	–
	cp	0.9380	0.9640	0.9490	0.9280	–	–	–	–
$\widehat{\phi}_{1,n}(z)$	MASE	0.0604	0.0452	0.0482	0.0391	0.1528	0.1049	0.1863	0.1274
	SASE	0.0169	0.0079	0.0092	0.0050	0.0818	0.0528	0.2009	0.0842
$\widehat{\phi}_{2,n}(z)$	MASE	0.0649	0.0512	0.0532	0.0466	0.1650	0.1099	0.2073	0.1294
	SASE	0.0158	0.0075	0.0090	0.0044	0.1070	0.0457	0.1831	0.0908

the Program for Changjiang Scholars and Innovative Research Team (IRT13077). The authors thank the Associate Editor and two anonymous referees for helpful comments and suggestions. The usual disclaimer applies.

Appendix

Our proofs of results require the following lemmas.

Table 3
Monte Carlo results under Design 3.

T	n	γ				β				$\phi(z)$				Combined			
		NZ		Z		NZ		Z		NZ		Z		NZ		Z	
		SD	LA	SD	LA	SD	LA	SD	LA	SD	LA	SD	LA	SD	LA	SD	LA
		$\gamma^\top = (0, 0.7, 0, -0.4, \gamma_5)$															
10	50	984	571	841	397	978	651	648	416	967	502	575	497	486	392	503	364
	100	1000	769	860	581	984	730	745	713	979	890	742	588	906	571	681	491
	200	1000	887	903	793	996	901	816	841	989	914	819	832	925	712	724	665
15	50	1000	613	922	436	1000	567	653	536	965	544	646	423	965	436	534	389
	100	1000	690	939	713	1000	802	807	845	983	903	784	591	983	548	709	507
	200	1000	901	966	835	1000	879	863	904	997	950	843	790	997	771	782	689
		$\gamma^\top = (0, 0.9, 0, 0.1, \gamma_5)$															
10	50	526	439	968	455	978	599	682	451	986	674	789	543	515	408	603	372
	100	859	651	993	681	989	657	692	706	960	681	919	631	827	553	539	525
	200	983	793	934	856	993	803	850	817	979	785	821	895	962	639	737	634
15	50	800	563	981	728	1000	627	623	499	965	498	781	612	932	405	491	312
	100	957	711	965	819	1000	802	785	613	947	722	693	508	947	519	560	398
	200	968	846	977	905	1000	879	829	768	981	894	775	605	944	702	719	577

Table 4
Estimates for the parametric component for the real data example.

	GMM estimates without SCAD			GMM estimates without SCAD (after removing insignificant variables)			SCAD-GMM estimates		
	EST	SE	CI	EST	SE	CI	EST	SE	CI
CR_{-1}	0.4856	0.1603	[0.3629, 0.8038]	0.8025	0.0997	[0.7341, 0.9914]	0.4574	0.1335	[0.3796, 0.6995]
CR_{-2}	−0.0279	0.0368	[−0.0522, 0.0408]	–	–	–	–	–	–
CR_{-3}	0.0348	0.0593	[−0.0025, 0.1478]	–	–	–	–	– .	–
AS	−0.0346	0.0157	[−0.0446, −0.0042]	−0.0152	0.0097	[−0.0214, −0.0028]	–	–	–
POL	−0.0303	0.0626	[−0.0802, 0.1018]	–	–	–	–	–	–
WF	0.0496	0.0358	[0.0194, 0.1146]	0.0280	0.0328	[0.0070, 0.0923]	–	–	–
WS	−0.0615	0.0472	[−0.1028, 0.0078]	–	–	–	−0.1018	0.0342	[−0.1244, −0.0356]
WL	−0.0534	0.0227	[−0.0756, −0.0133]	−0.0282	0.0169	[−0.0390, −0.0064]	–	–	–
PY	−1.3523	0.4584	[−1.4567, −0.2382]	−0.7793	0.2909	[−0.9430, −0.2119]	0.0687	0.4060	[−0.1834, 0.8613]

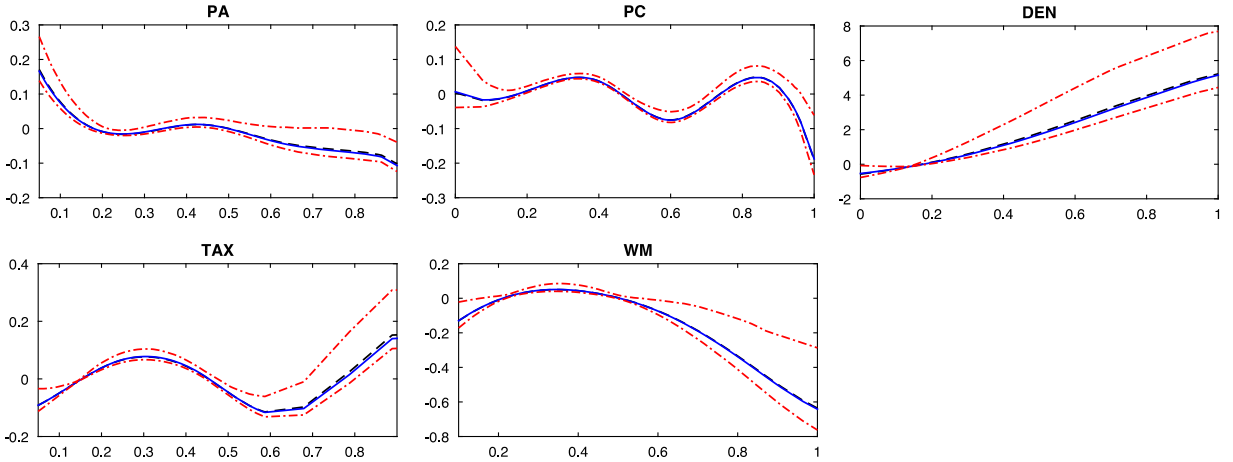

Fig. 1. The estimated nonparametric function (solid curve), its 95% point-wise confidence bands (dash-dotted lines) and the bootstrap-based nonparametric function (dashed curves) with SCAD selection.

Lemma 1. *Let Assumptions (A1)–(A3) hold. Then*

$$\sup_{\phi_r \in \mathcal{H}^0_{r,n}, r=1,\ldots,q} \left| \frac{n^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} (\sum_{r=1}^{q} \phi_r(z_{itr}))^2}{\mathrm{E} \sum_{t=1}^{T} (\sum_{r=1}^{q} \phi_r(z_{itr}))^2} - 1 \right| = o_p(1),$$

where $\mathcal{H}^0_{r,n}$, $r = 1, \ldots, q$ are the polynomial spline spaces. See Section 2 for similar definitions.

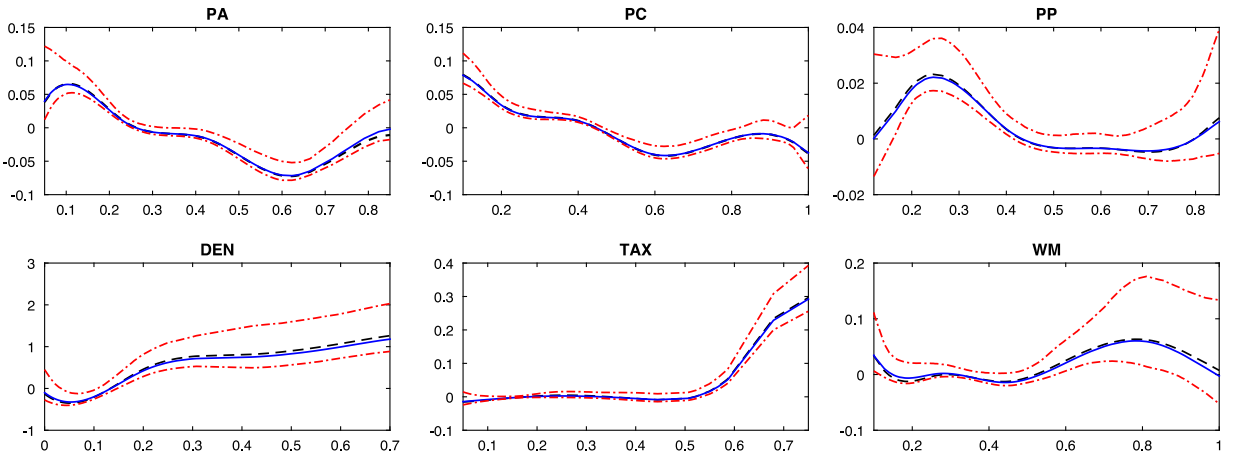

Fig. 2. The estimated nonparametric function (solid curve), its 95% point-wise confidence bands (dash-dotted lines) and the bootstrap-based nonparametric function (dashed curves) without SCAD selection.

Proof. This lemma can be proved by applying Lemma 1 of Huang and Shen (2004). We omit the details for brevity. ■

Lemma 2. *For each function $g(z)$ satisfying Assumptions (A1)–(A3), e.g., $g(z) = \sum_{r=1}^{q} \phi_r(z)$ and $g(z) = \mathrm{E}_{\mathcal{A}}(\mathrm{E}(y, x|z))$, there exist a quantity $C > 0$ and a polynomial spline function $h(z) \in \mathcal{H}_n^0$ such that*

$$\|g(z) - h(z)\|_\infty \leq CJ_n^{-\ell}.$$

This implies that there is $\boldsymbol{\theta} \in R^{J_n}$ such that

$$\sup_{z \in \mathcal{R}^1} |g(z) - \mathbf{N}^\top(z)\boldsymbol{\theta}| = O(J_n^{-\ell}).$$

Proof. Lemma 2 is a direct result of Theorem XII.1 of de Boor (2001). ■

Lemma 3. *Let Assumptions (A1)–(A4) hold. Then as $n \to \infty$, we have*

$$\frac{1}{nT^*}(\Delta\mathbf{Y}^d, \Delta\mathbf{X})^\top(\mathbf{I}_{nT^*} - \mathbf{M})\mathbf{W} \to_P \frac{1}{T^*}\mathrm{E}[\{(\Delta\mathbf{Y}_1^d, \Delta\mathbf{X}_1) - \mathrm{E}_{\mathcal{A}}(\Delta\mathbf{Y}_1^d, \Delta\mathbf{X}_1)\}^\top(\mathbf{W}_1 - \mathrm{E}_{\mathcal{A}}\mathbf{W}_1)].$$

Proof. Write $\mathbf{V} = (\Delta\mathbf{Y}^d, \Delta\mathbf{X})$ and $\mathbf{V}_i = (\Delta\mathbf{Y}_i^d, \Delta\mathbf{X}_i)$ with $\Delta\mathbf{Y}_i^d = (\Delta\mathbf{y}_{i(d+3)}^d, \ldots, \Delta\mathbf{y}_{iT}^d)^\top$ and $\Delta\mathbf{X}_i = (\Delta\mathbf{x}_{i(d+3)}, \ldots, \Delta\mathbf{x}_{iT})^\top$ for $i = 1, \ldots, n$. Note that $(nT^*)^{-1}(\Delta\mathbf{Y}^d, \Delta\mathbf{X})^\top(\mathbf{I}_{nT^*} - \mathbf{M})\mathbf{W}$ can be expressed as the summation of the following terms:

$$\mathbf{D}_1 = \frac{1}{nT^*}(\mathbf{V} - \mathrm{E}_{\mathcal{A}}\mathbf{V})^\top(\mathbf{I}_{nT^*} - \mathbf{M})(\mathbf{W} - \mathrm{E}_{\mathcal{A}}\mathbf{W}), \qquad \mathbf{D}_2 = \frac{1}{nT^*}(\mathbf{V} - \mathrm{E}_{\mathcal{A}}\mathbf{V})^\top(\mathbf{I}_{nT^*} - \mathbf{M})\mathrm{E}_{\mathcal{A}}\mathbf{W},$$
$$\mathbf{D}_3 = \frac{1}{nT^*}(\mathrm{E}_{\mathcal{A}}\mathbf{V})^\top(\mathbf{I}_{nT^*} - \mathbf{M})(\mathbf{W} - \mathrm{E}_{\mathcal{A}}\mathbf{W}), \quad \text{and} \quad \mathbf{D}_4 = \frac{1}{nT^*}(\mathrm{E}_{\mathcal{A}}\mathbf{V})^\top(\mathbf{I}_{nT^*} - \mathbf{M})\mathrm{E}_{\mathcal{A}}\mathbf{W}.$$

Using Assumption (A1) and the Law of Large Numbers, we can show that

$$\begin{aligned}\frac{1}{nT^*}(\mathbf{V} - \mathrm{E}_{\mathcal{A}}\mathbf{V})^\top(\mathbf{W} - \mathrm{E}_{\mathcal{A}}\mathbf{W}) &= \frac{1}{nT^*}\sum_{i=1}^{n}(\mathbf{V}_i - \mathrm{E}_{\mathcal{A}}\mathbf{V}_i)^\top(\mathbf{W}_i - \mathrm{E}_{\mathcal{A}}\mathbf{W}_i) \\ &\to_P \frac{1}{T^*}\mathrm{E}\{(\mathbf{V}_1 - \mathrm{E}_{\mathcal{A}}\mathbf{V}_1)^\top(\mathbf{W}_1 - \mathrm{E}_{\mathcal{A}}\mathbf{W}_1)\}, \end{aligned} \tag{A.1}$$

and

$$\begin{aligned}\frac{1}{nT^*}(\mathbf{V} - \mathrm{E}_{\mathcal{A}}\mathbf{V})^\top\mathbf{M}(\mathbf{V} - \mathrm{E}_{\mathcal{A}}\mathbf{V}) &= \frac{1}{nT^*}(\mathbf{V} - \mathrm{E}_{\mathcal{A}}\mathbf{V})^\top\tilde{\mathbf{N}}(\tilde{\mathbf{N}}^\top\tilde{\mathbf{N}})^{-1}\tilde{\mathbf{N}}^\top(\mathbf{V} - \mathrm{E}_{\mathcal{A}}\mathbf{V}) \\ &= \frac{1}{(nT^*)^2qJ_n}(\mathbf{V} - \mathrm{E}_{\mathcal{A}}\mathbf{V})^\top\tilde{\mathbf{N}}\tilde{\mathbf{N}}^\top(\mathbf{V} - \mathrm{E}_{\mathcal{A}}\mathbf{V}) \\ &= \frac{1}{(nT^*)^2qJ_n}(\sqrt{nT^*qJ_n})^2 = qJ_n/nT^* \to 0. \end{aligned} \tag{A.2}$$

Applying arguments similar to those in (A.2) implies that $(\mathbf{W}-\mathrm{E}_{\mathcal{A}}\mathbf{W})^{\top}\mathbf{M}(\mathbf{W}-\mathrm{E}_{\mathcal{A}}\mathbf{W})/(nT^*) = o_p(1)$, which, when combined with (A.2) and the Cauchy–Schwarz inequality, leads to

$$\frac{1}{nT^*}(\mathbf{V}-\mathrm{E}_{\mathcal{A}}\mathbf{V})^{\top}\mathbf{M}(\mathbf{W}-\mathrm{E}_{\mathcal{A}}\mathbf{W}) \to_P 0. \tag{A.3}$$

Now, by (A.1) and (A.3),

$$\mathbf{D}_1 \to_P \frac{1}{T^*}\mathrm{E}[\{(\Delta\mathbf{Y}_1^d, \Delta\mathbf{X}_1) - \mathrm{E}_{\mathcal{A}}(\Delta\mathbf{Y}_1^d, \Delta\mathbf{X}_1)\}^{\top}(\mathbf{W}_1 - \mathrm{E}_{\mathcal{A}}\mathbf{W}_1)].$$

The lemma holds if each of $\mathbf{D}_2$, $\mathbf{D}_3$ and $\mathbf{D}_4$ is of order $o_p(1)$. Consider the (u, v)th element of $\mathbf{D}_2$ first,

$$\begin{aligned}
(\mathbf{D}_2)_{u,v} &= \frac{1}{nT^*}(\mathbf{v}_{*,u} - \mathrm{E}_{\mathcal{A}}\mathbf{v}_{*,u})^{\top}(\mathbf{I}_{nT^*} - \mathbf{M})\mathrm{E}_{\mathcal{A}}\mathbf{w}_{*,v} \\
&\le \left\{\frac{1}{nT^*}(\mathbf{v}_{*,u} - \mathrm{E}_{\mathcal{A}}\mathbf{v}_{*,u})^{\top}(\mathbf{I}_{nT^*} - \mathbf{M})(\mathbf{v}_{*,u} - \mathrm{E}_{\mathcal{A}}\mathbf{v}_{*,u})\right\}^{1/2} \\
&\quad \cdot \left\{\frac{1}{nT^*}(\mathrm{E}_{\mathcal{A}}\mathbf{w}_{*,v} - \tilde{\mathbf{N}}\boldsymbol{\theta})^{\top}(\mathbf{I}_{nT^*} - \mathbf{M})(\mathrm{E}_{\mathcal{A}}\mathbf{w}_{*,v} - \tilde{\mathbf{N}}\boldsymbol{\theta})\right\}^{1/2} \\
&\le \left\{\frac{1}{nT^*}(\mathbf{v}_{*,u} - \mathrm{E}_{\mathcal{A}}\mathbf{v}_{*,u})^{\top}(\mathbf{v}_{*,u} - \mathrm{E}_{\mathcal{A}}\mathbf{v}_{*,u})\right\}^{1/2}\left\{\frac{1}{nT^*}(\mathrm{E}_{\mathcal{A}}\mathbf{w}_{*,v} - \tilde{\mathbf{N}}\boldsymbol{\theta})^{\top}(\mathrm{E}_{\mathcal{A}}\mathbf{w}_{*,v} - \tilde{\mathbf{N}}\boldsymbol{\theta})\right\}^{1/2} \\
&= \left\{\frac{1}{nT^*}\sum_{i=1}^{n}\|(\mathbf{V}_i)_{*,u} - \mathrm{E}_{\mathcal{A}}(\mathbf{V}_i)_{*,u}\|_2^2\right\}^{1/2}\left\{\frac{1}{nT^*}\sum_{i=1}^{n}\|\mathrm{E}_{\mathcal{A}}(\mathbf{W}_i)_{*,v} - \tilde{\mathbf{N}}_{[i]}\boldsymbol{\theta}\|_2^2\right\}^{1/2} \\
&= \frac{1}{\sqrt{T^*}}\mathrm{E}\|(\mathbf{V}_i)_{*,u} - \mathrm{E}_{\mathcal{A}}(\mathbf{V}_i)_{*,u}\|_2(1 + o_p(1))O_p(qJ_n^{-\ell}) = o_p(1),
\end{aligned}$$

where $\mathbf{v}_{*,u}$ and $\mathbf{w}_{*,v}$ are the uth column of $\mathbf{V}$ and the vth column of $\mathbf{W}$ respectively. Similarly, $(\mathbf{V}_i)_{*,u}$ and $(\mathbf{W}_i)_{*,v}$ are the uth and vth column of $\mathbf{V}_i$ and $\mathbf{W}_i$ respectively, and $\tilde{\mathbf{N}}_{[i]}$ is the same as $\mathbf{V}_i$. This implies that $\mathbf{D}_2 = o_p(1)$. In the same manner we can show that $\mathbf{D}_3 = \mathbf{D}_4 = o_p(1)$. The proof then follows. ■

Lemma 4. *Under the same conditionsas in Lemma 1, we have*

$$\frac{1}{nT^*}\mathbf{W}^{*\top}(\mathbf{I}_n \otimes \mathbf{H})\mathbf{W}^* \to_P \frac{1}{T^*}\mathrm{E}\left\{(\mathbf{W}_1 - \mathrm{E}_{\mathcal{A}}\mathbf{W}_1)^{\top}\mathbf{H}(\mathbf{W}_1 - \mathrm{E}_{\mathcal{A}}\mathbf{W}_1)\right\},$$

where $\mathbf{H} = \mathbf{B}\mathbf{B}^{\top}$.

Proof. Note that $(nT^*)^{-1}\mathbf{W}^{*\top}(\mathbf{I}_n \otimes \mathbf{H})\mathbf{W}^*$ can be decomposed into the following terms:

$$\begin{aligned}
\mathbf{E}_1 &= \frac{1}{nT^*}(\mathbf{W} - \mathrm{E}_{\mathcal{A}}\mathbf{W})^{\top}(\mathbf{I}_{nT^*} - \mathbf{M})(\mathbf{I}_n \otimes \mathbf{H})(\mathbf{I}_{nT^*} - \mathbf{M})(\mathbf{W} - \mathrm{E}_{\mathcal{A}}\mathbf{W}), \\
\mathbf{E}_2 &= \frac{1}{nT^*}(\mathrm{E}_{\mathcal{A}}\mathbf{W} - \tilde{\mathbf{N}}\boldsymbol{\theta})^{\top}(\mathbf{I}_{nT^*} - \mathbf{M})(\mathbf{I}_n \otimes \mathbf{H})(\mathbf{I}_{nT^*} - \mathbf{M})(\mathbf{W} - \mathrm{E}_{\mathcal{A}}\mathbf{W}), \\
\mathbf{E}_3 &= \frac{1}{nT^*}(\mathbf{W} - \mathrm{E}_{\mathcal{A}}\mathbf{W})^{\top}(\mathbf{I}_{nT^*} - \mathbf{M})(\mathbf{I}_n \otimes \mathbf{H})(\mathbf{I}_{nT^*} - \mathbf{M})(\mathrm{E}_{\mathcal{A}}\mathbf{W} - \tilde{\mathbf{N}}\boldsymbol{\theta}),
\end{aligned}$$

and

$$\mathbf{E}_4 = \frac{1}{nT^*}(\mathrm{E}_{\mathcal{A}}\mathbf{W} - \tilde{\mathbf{N}}\boldsymbol{\theta})^{\top}(\mathbf{I}_{nT^*} - \mathbf{M})(\mathbf{I}_n \otimes \mathbf{H})(\mathbf{I}_{nT^*} - \mathbf{M})(\mathrm{E}_{\mathcal{A}}\mathbf{W} - \tilde{\mathbf{N}}\boldsymbol{\theta}).$$

We can write

$$\begin{aligned}
\mathbf{E}_1 &= \frac{1}{nT^*}\{(\mathbf{W} - \mathrm{E}_{\mathcal{A}}\mathbf{W})^{\top}(\mathbf{I}_n \otimes \mathbf{H})(\mathbf{W} - \mathrm{E}_{\mathcal{A}}\mathbf{W}) - (\mathbf{W} - \mathrm{E}_{\mathcal{A}}\mathbf{W})^{\top}\mathbf{M}(\mathbf{I}_n \otimes \mathbf{H})(\mathbf{W} - \mathrm{E}_{\mathcal{A}}\mathbf{W}) \\
&\quad - (\mathbf{W} - \mathrm{E}_{\mathcal{A}}\mathbf{W})^{\top}(\mathbf{I}_n \otimes \mathbf{H})\mathbf{M}(\mathbf{W} - \mathrm{E}_{\mathcal{A}}\mathbf{W}) + (\mathbf{W} - \mathrm{E}_{\mathcal{A}}\mathbf{W})^{\top}\mathbf{M}(\mathbf{I}_n \otimes \mathbf{H})\mathbf{M}(\mathbf{W} - \mathrm{E}_{\mathcal{A}}\mathbf{W})\} \\
&\stackrel{def}{=} \mathbf{E}_{11} + \mathbf{E}_{12} + \mathbf{E}_{13} + \mathbf{E}_{14}.
\end{aligned}$$

Using Assumption (A1) and Lemma 1,

$$\mathbf{E}_{11} \to_P \frac{1}{T^*}\mathrm{E}\{(\mathbf{W}_1 - \mathrm{E}_{\mathcal{A}}\mathbf{W}_1)^{\top}\mathbf{H}(\mathbf{W}_1 - \mathrm{E}_{\mathcal{A}}\mathbf{W}_1)\}.$$

Moreover, the orthogonality of $\mathbf{W} - \mathrm{E}_{\mathcal{A}}\mathbf{W}$ and $\tilde{\mathbf{N}}$ implies that $\mathbf{E}_{12} = \mathbf{E}_{13} = \mathbf{E}_{14} = O_p(1/(nT^*)) = o_p(1)$. Additionally,

$$\begin{aligned}\mathbf{E}_4 = \frac{1}{nT^*}&\{(\mathrm{E}_{\mathcal{A}}\mathbf{W} - \tilde{\mathbf{N}}\boldsymbol{\theta})^\top(\mathbf{I}_n \otimes \mathbf{H})(\mathrm{E}_{\mathcal{A}}\mathbf{W} - \tilde{\mathbf{N}}\boldsymbol{\theta}) - (\mathrm{E}_{\mathcal{A}}\mathbf{W} - \tilde{\mathbf{N}}\boldsymbol{\theta})^\top\mathbf{M}(\mathbf{I}_n \otimes \mathbf{H})(\mathrm{E}_{\mathcal{A}}\mathbf{W} - \tilde{\mathbf{N}}\boldsymbol{\theta}) \\ &- (\mathrm{E}_{\mathcal{A}}\mathbf{W} - \tilde{\mathbf{N}}\boldsymbol{\theta})^\top(\mathbf{I}_n \otimes \mathbf{H})\mathbf{M}(\mathrm{E}_{\mathcal{A}}\mathbf{W} - \tilde{\mathbf{N}}\boldsymbol{\theta}) + (\mathrm{E}_{\mathcal{A}}\mathbf{W} - \tilde{\mathbf{N}}\boldsymbol{\theta})^\top\mathbf{M}(\mathbf{I}_n \otimes \mathbf{H})\mathbf{M}(\mathrm{E}_{\mathcal{A}}\mathbf{W} - \tilde{\mathbf{N}}\boldsymbol{\theta})\} \\ &\stackrel{def}{=} \mathbf{E}_{41} + \mathbf{E}_{42} + \mathbf{E}_{43} + \mathbf{E}_{44},\end{aligned}$$

where $\mathbf{E}_{41} = J_n^{-2\ell} \to 0$, and $\mathbf{E}_{42} = \mathbf{E}_{43} = \mathbf{E}_{44} = qJ_n^{1-2\ell}/(nT^*) = o_p(1)$. Because $\mathbf{E}_2$ and $\mathbf{E}_3$ are of the same order, it suffices to consider $\mathbf{E}_3$ only. Note that

$$\begin{aligned}\mathbf{E}_3 = \frac{1}{nT^*}&\{(\mathbf{W} - \mathrm{E}_{\mathcal{A}}\mathbf{W})^\top(\mathbf{I}_n \otimes \mathbf{H})(\mathrm{E}_{\mathcal{A}}\mathbf{W} - \tilde{\mathbf{N}}\boldsymbol{\theta}) - (\mathbf{W} - \mathrm{E}_{\mathcal{A}}\mathbf{W})^\top\mathbf{M}(\mathbf{I}_n \otimes \mathbf{H})(\mathrm{E}_{\mathcal{A}}\mathbf{W} - \tilde{\mathbf{N}}\boldsymbol{\theta}) \\ &- (\mathbf{W} - \mathrm{E}_{\mathcal{A}}\mathbf{W})^\top(\mathbf{I}_n \otimes \mathbf{H})\mathbf{M}(\mathrm{E}_{\mathcal{A}}\mathbf{W} - \tilde{\mathbf{N}}\boldsymbol{\theta}) + (\mathbf{W} - \mathrm{E}_{\mathcal{A}}\mathbf{W})^\top\mathbf{M}(\mathbf{I}_n \otimes \mathbf{H})\mathbf{M}(\mathrm{E}_{\mathcal{A}}\mathbf{W} - \tilde{\mathbf{N}}\boldsymbol{\theta})\} \\ &\stackrel{def}{=} \mathbf{E}_{31} + \mathbf{E}_{32} + \mathbf{E}_{33} + \mathbf{E}_{34},\end{aligned}$$

which, when combined with the fact that $(\mathbf{W} - \mathrm{E}_{\mathcal{A}}\mathbf{W}) \perp \mathrm{E}_{\mathcal{A}}\mathbf{W}$, leads to $\mathbf{E}_{31} = J_n^{-\ell}/(nT^*)$, and $\mathbf{E}_{32} = \mathbf{E}_{33} = \mathbf{E}_{34} = qJ_n^{1-\ell}/\sqrt{nT^*} = o_p(1)$. The proof is thus completed. ■

Write $\tilde{\boldsymbol{\phi}}_r(z_{itr}) = (\phi_{r1}(z_{itr}) - \phi_{r1}(z_{i(t-1)r}), \ldots, \phi_{rJ_n}(z_{itr}) - \phi_{rJ_n}(z_{i(t-1)r}))^\top$, and $\tilde{\boldsymbol{\phi}} = (\tilde{\boldsymbol{\phi}}_1, \ldots, \tilde{\boldsymbol{\phi}}_q)_{nT^* \times qJ_n}$ with $\tilde{\boldsymbol{\phi}}_r = (\tilde{\phi}_r(z_{1(d+3)r}), \ldots, \tilde{\phi}_r(z_{1Tr}), \ldots, \tilde{\phi}_r(z_{nTr}))^\top$.

Lemma 5. *Let Assumptions (A1)–(A4) hold. Then*

$$\frac{1}{\sqrt{nT^*}}\mathbf{W}^\top(\mathbf{I}_{nT^*} - \mathbf{M})\tilde{\boldsymbol{\phi}} = O_p(J_n^{-\ell}).$$

Proof. By Assumption (A3), there exists a vector $\boldsymbol{\theta} \in \mathcal{R}^{qJ_n}$ such that

$$\begin{aligned}\frac{1}{\sqrt{nT^*}}\mathbf{W}^\top(\mathbf{I}_{nT^*} - \mathbf{M})\tilde{\boldsymbol{\phi}} = &\frac{1}{\sqrt{nT^*}}(\mathbf{W} - \mathrm{E}_{\mathcal{A}}\mathbf{W})^\top(\mathbf{I}_{nT^*} - \mathbf{M})(\tilde{\boldsymbol{\phi}} - \tilde{\mathbf{N}}\boldsymbol{\theta}) \\ &+ \frac{1}{\sqrt{nT^*}}(\mathrm{E}_{\mathcal{A}}\mathbf{W})^\top(\mathbf{I}_{nT^*} - \mathbf{M})(\tilde{\boldsymbol{\phi}} - \tilde{\mathbf{N}}\boldsymbol{\theta}) \stackrel{def}{=} \mathbf{G}_1 + \mathbf{G}_2,\end{aligned}$$

where

$$\begin{aligned}\mathbf{G}_1 &\le \frac{1}{\sqrt{nT^*}}\{(\mathbf{W} - \mathrm{E}_{\mathcal{A}}\mathbf{W})^\top(\mathbf{I}_{nT^*} - \mathbf{M})(\mathbf{W} - \mathrm{E}_{\mathcal{A}}\mathbf{W})\}^{1/2}\left\{(\tilde{\boldsymbol{\phi}} - \tilde{\mathbf{N}}\boldsymbol{\theta})^\top(\mathbf{I}_{nT^*} - \mathbf{M})(\tilde{\boldsymbol{\phi}} - \tilde{\mathbf{N}}\boldsymbol{\theta})\right\}^{1/2} \\ &\le \frac{1}{\sqrt{nT^*}}\left\{(\mathbf{W} - \mathrm{E}_{\mathcal{A}}\mathbf{W})^\top(\mathbf{W} - \mathrm{E}_{\mathcal{A}}\mathbf{W})\right\}^{1/2}\left\{(\tilde{\boldsymbol{\phi}} - \tilde{\mathbf{N}}\boldsymbol{\theta})^\top(\tilde{\boldsymbol{\phi}} - \tilde{\mathbf{N}}\boldsymbol{\theta})\right\}^{1/2} \\ &\le \frac{C}{\sqrt{nT^*}}O_p(\sqrt{n}qJ_n^{-\ell}) = O_p(J_n^{-\ell}),\end{aligned}$$

and $\mathbf{G}_2$ has the property that

$$\begin{aligned}\mathbf{G}_2 &\le \frac{1}{\sqrt{nT^*}}\{(\mathrm{E}_{\mathcal{A}}\mathbf{W})^\top(\mathrm{E}_{\mathcal{A}}\mathbf{W})\}^{1/2}\{(\tilde{\boldsymbol{\phi}} - \tilde{\mathbf{N}}\boldsymbol{\theta})^\top(\tilde{\boldsymbol{\phi}} - \tilde{\mathbf{N}}\boldsymbol{\theta})\}^{1/2} \\ &\le \frac{1}{\sqrt{nT^*}}O_p(nqJ_n^{-\ell}) = O_p(\sqrt{n}J_n^{-\ell}).\end{aligned}$$

These yield the desired result. ■

Proof of Theorem 1. (i) Simple calculations show that

$$\begin{aligned}\sqrt{nT^*}\begin{pmatrix}\widehat{\boldsymbol{\gamma}}_n - \boldsymbol{\gamma} \\ \widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\end{pmatrix} = &\left\{(\Delta\mathbf{Y}^d, \Delta\mathbf{X})^\top(\mathbf{I}_{nT^*} - \mathbf{M})\mathbf{W}\boldsymbol{\Omega}^{-1}\mathbf{W}^\top(\mathbf{I}_{nT^*} - \mathbf{M})(\Delta\mathbf{Y}^d, \Delta\mathbf{X})\right\}^{-1} \\ &\cdot \frac{1}{\sqrt{nT^*}}(\Delta\mathbf{Y}^d, \Delta\mathbf{X})^\top(\mathbf{I}_{nT^*} - \mathbf{M})\mathbf{W}\boldsymbol{\Omega}^{-1}\left\{\mathbf{W}^\top(\mathbf{I}_{nT^*} - \mathbf{M})\tilde{\boldsymbol{\phi}} + \mathbf{W}^\top(\mathbf{I}_{nT^*} - \mathbf{M})\Delta\boldsymbol{\varepsilon}\right\}.\end{aligned}$$

Using results of Lemmas 1, 2, 3, and 4, Theorem 1 follows if we can prove that

$$\frac{1}{\sqrt{nT^*}}\mathbf{W}^\top(\mathbf{I}_{nT^*} - \mathbf{M})\Delta\boldsymbol{\varepsilon} \to_D N(\mathbf{0}, \sigma^2\boldsymbol{\Sigma}) \quad \text{as } n \to \infty. \tag{A.4}$$

Note that

$$\begin{aligned}
&\frac{1}{\sqrt{nT^*}}\mathbf{W}^\top(\mathbf{I}_{nT^*}-\mathbf{M})\Delta\boldsymbol{\varepsilon}\\
&=\frac{1}{\sqrt{nT^*}}(\mathbf{W}-\mathrm{E}_{\mathcal{A}}\mathbf{W})^\top(\mathbf{I}_{nT^*}-\mathbf{M})\Delta\boldsymbol{\varepsilon}+\frac{1}{\sqrt{nT^*}}(\mathrm{E}_{\mathcal{A}}\mathbf{W}-\tilde{\mathbf{N}}\boldsymbol{\theta})^\top(\mathbf{I}_{nT^*}-\mathbf{M})\Delta\boldsymbol{\varepsilon}\\
&=\frac{1}{\sqrt{nT^*}}(\mathbf{W}-\mathrm{E}_{\mathcal{A}}\mathbf{W})^\top\Delta\boldsymbol{\varepsilon}-\frac{1}{\sqrt{nT^*}}(\mathbf{W}-\mathrm{E}_{\mathcal{A}}\mathbf{W})^\top\mathbf{M}\Delta\boldsymbol{\varepsilon}+\frac{1}{\sqrt{nT^*}}(\mathrm{E}_{\mathcal{A}}\mathbf{W}-\tilde{\mathbf{N}}\boldsymbol{\theta})^\top(\mathbf{I}_{nT^*}-\mathbf{M})\Delta\boldsymbol{\varepsilon},
\end{aligned}$$

for which the leading term is the first term on the r.h.s., while the second and third terms can be readily shown to have orders $O_p(J_n/\sqrt{n})$ and $O_p(J_n^{-\ell})$ respectively.

It is easily seen that $\mathrm{E}\{(\mathbf{W}-\mathrm{E}_{\mathcal{A}}\mathbf{W})^\top\Delta\boldsymbol{\varepsilon}|\mathbf{X},\mathbf{Z}\}=\mathbf{0}$ by the exogeneity of $\mathbf{W}$. Moreover,

$$\begin{aligned}
\mathrm{var}\{(\mathbf{W}-\mathrm{E}_{\mathcal{A}}\mathbf{W})^\top\Delta\boldsymbol{\varepsilon}|\mathbf{X},\mathbf{Z}\}&=\sigma^2(\mathbf{W}-\mathrm{E}_{\mathcal{A}}\mathbf{W})^\top(\mathbf{I}\otimes\mathbf{H})(\mathbf{W}-\mathrm{E}_{\mathcal{A}}\mathbf{W})\\
&\to_P\frac{\sigma^2}{T^*}\mathrm{E}\{(\mathbf{W}_1-\mathrm{E}_{\mathcal{A}}\mathbf{W}_1)^\top\mathbf{B}\mathbf{B}^\top(\mathbf{W}_1-\mathrm{E}_{\mathcal{A}}\mathbf{W}_1)\}.
\end{aligned}$$

Hence (A.4) holds by utilizing Assumptions (A1) and (A4) and the Lindeberg–Levy Central Limit theorem. This completes the proof.

(ii) This is an obvious result and we omit the proof for brevity. ■

Proof of Theorem 2. The definition of $\mathcal{H}_n^0$ and (2.2) imply that

$$\|\widehat{\phi}_{r,n}-\phi_r\|_2\le\|\widehat{\phi}_{r,n}-\mathbf{N}_r\boldsymbol{\theta}_r\|_2+\|\mathbf{N}_r\boldsymbol{\theta}_r-\phi_r\|_2\le\|\widehat{\boldsymbol{\theta}}_{r,n}-\boldsymbol{\theta}_r\|+J_n^{-\ell}. \tag{A.5}$$

Thus, it suffices to consider only the term $\|\widehat{\boldsymbol{\theta}}_{r,n}-\boldsymbol{\theta}_r\|$. Note that

$$\begin{aligned}
\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}&=(\tilde{\mathbf{N}}^\top\tilde{\mathbf{N}})^{-1}\tilde{\mathbf{N}}^\top\left\{\tilde{\boldsymbol{\phi}}+\Delta\boldsymbol{\varepsilon}+\Delta\mathbf{Y}^d(\boldsymbol{\gamma}-\widehat{\boldsymbol{\gamma}}_n)-\Delta\mathbf{X}(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}_n)\right\}-\boldsymbol{\theta}\\
&=(\tilde{\mathbf{N}}^\top\tilde{\mathbf{N}})^{-1}\tilde{\mathbf{N}}^\top(\tilde{\boldsymbol{\phi}}-\tilde{\mathbf{N}}\boldsymbol{\theta})+(\tilde{\mathbf{N}}^\top\tilde{\mathbf{N}})^{-1}\tilde{\mathbf{N}}^\top\Delta\boldsymbol{\varepsilon}\\
&\quad+(\tilde{\mathbf{N}}^\top\tilde{\mathbf{N}})^{-1}\tilde{\mathbf{N}}^\top(\Delta\mathbf{Y}^d,\Delta\mathbf{X})\{(\boldsymbol{\gamma}^\top,\boldsymbol{\beta}^\top)^\top-(\widehat{\boldsymbol{\gamma}}_n^\top,\widehat{\boldsymbol{\beta}}_n^\top)^\top\}\\
&\overset{def}{=}\mathbf{e}_1+\mathbf{e}_2+\mathbf{e}_3.
\end{aligned}$$

Let $\mathbf{A}=(\mathbf{0}_{J_n,(r-1)J_n},\mathbf{I}_{J_n},\mathbf{0}_{J_n,(q-r)J_n})$, we have $\widehat{\boldsymbol{\theta}}_{r,n}-\boldsymbol{\theta}_r=\mathbf{A}(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta})$. Hence,

$$\begin{aligned}
\|\mathbf{A}\mathbf{e}_1\|_2^2&=\left(\tilde{\boldsymbol{\phi}}-\tilde{\mathbf{N}}\boldsymbol{\theta}\right)^\top\tilde{\mathbf{N}}(\tilde{\mathbf{N}}^\top\tilde{\mathbf{N}})^{-1}\mathbf{A}^\top\mathbf{A}(\tilde{\mathbf{N}}^\top\tilde{\mathbf{N}})^{-1}\tilde{\mathbf{N}}^\top\left(\tilde{\boldsymbol{\phi}}-\tilde{\mathbf{N}}\boldsymbol{\theta}\right)\\
&=O_p((nT^*)^{-2})\mathrm{tr}\left\{\left(\tilde{\boldsymbol{\phi}}-\tilde{\mathbf{N}}\boldsymbol{\theta}\right)\tilde{\mathbf{N}}\tilde{\mathbf{N}}^\top\left(\tilde{\boldsymbol{\phi}}-\tilde{\mathbf{N}}\boldsymbol{\theta}\right)^\top\right\}\\
&\le O_p((nT^*)^{-2})O_p(nT^*J_n^{-2\ell})O_p(nT^*J_n)\le J_nO_p(J_n^{-2\ell})=o_p(J_n^{-\ell}),
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}(\|\mathbf{A}\mathbf{e}_2\|_2^2|\mathbf{Z})&=O((nT^*)^{-2})\mathrm{E}\{\mathrm{tr}(\Delta\boldsymbol{\varepsilon}^\top\tilde{\mathbf{N}}\tilde{\mathbf{N}}^\top\Delta\boldsymbol{\varepsilon})|\mathbf{Z}\}=O((nT^*)^{-2})\mathrm{tr}\{\tilde{\mathbf{N}}\tilde{\mathbf{N}}^\top\mathrm{E}(\Delta\boldsymbol{\varepsilon}\Delta\boldsymbol{\varepsilon}^\top|\mathbf{Z})\}\\
&=O((nT^*)^{-2})\cdot O(nT^*J_n)=O(J_n/n)
\end{aligned}$$

and

$$\begin{aligned}
\|\mathbf{A}\mathbf{e}_3\|_2^2&=O_p((nT^*)^{-2})O_p((nT^*)^{-1})\mathrm{tr}\{(\Delta\mathbf{Y}^d,\Delta\mathbf{X})^\top\tilde{\mathbf{N}}\tilde{\mathbf{N}}^\top(\Delta\mathbf{Y}^d,\Delta\mathbf{X})\}\\
&=O_p((nT^*)^{-3})O((nT^*)^2J_n)=O(J_n/n).
\end{aligned}$$

Combining the equations above, we can readily show that

$$\|\widehat{\boldsymbol{\theta}}_{r,n}-\boldsymbol{\theta}_r\|_2^2\le O_p(J_n/n)+O_p(J_n^{-\ell}).$$ ■

Proof of Theorem 3. (i) Let $K_n=2(nT^*-d-p-qJ_n)$. From (2.9),

$$\begin{aligned}
\widehat{\sigma}_n^2&=K_n^{-1}\|\Delta\boldsymbol{\varepsilon}\|_2^2+K_n^{-1}\|\Delta\mathbf{Y}^d(\boldsymbol{\gamma}-\widehat{\boldsymbol{\gamma}}_n)+\Delta\mathbf{X}(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}_n)+\tilde{\mathbf{N}}(\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}}_n)\|_2^2\\
&\quad-2K_n^{-1}\{\Delta\mathbf{Y}^d(\boldsymbol{\gamma}-\widehat{\boldsymbol{\gamma}}_n)+\Delta\mathbf{X}(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}_n)+\tilde{\mathbf{N}}(\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}}_n)\}^\top\Delta\boldsymbol{\varepsilon}\\
&\overset{def}{=}F_1+F_2+F_3.
\end{aligned}$$

Using the results in Theorems 1 and 2, we have $F_2 \leq O_p(J_n/nT^*) + J_n^{-2\ell}$. Coupled with the Cauchy–Schwarz inequality, this implies that $|F_3| < |F_1|$ with probability approaching 1. Hence it suffices to prove the asymptotic normality of F_1. Clearly,

$$F_1 = K_n^{-1}\Delta\boldsymbol{\varepsilon}^\top\Delta\boldsymbol{\varepsilon} = K_n^{-1}\sum_{i=1}^{n}\sum_{t=d+3}^{T}(\varepsilon_{it}-\varepsilon_{i(t-1)})^2.$$

Write $\vartheta_i = \sum_{t=d+3}^{T}(\varepsilon_{it}-\varepsilon_{i(t-1)})^2$. Clearly, ϑ_i is an i.i.d. random variable with mean

$$\mathrm{E}(\vartheta_i) = \sum_{t=d+3}^{T}\mathrm{E}(\varepsilon_{it}-\varepsilon_{i(t-1)})^2 = 2T^*\sigma^2$$

and variance

$$\begin{aligned}
\mathrm{var}(\vartheta_i) &= \mathrm{E}\left\{\sum_{t=d+3}^{T}(\varepsilon_{it}-\varepsilon_{i(t-1)})^4 + 2\sum_{d+3\leq t'<t\leq T}(\varepsilon_{it}-\varepsilon_{i(t-1)})^2(\varepsilon_{it'}-\varepsilon_{it'-1})^2\right\} - 4T^{*2}\sigma^4\\
&= 2T^*\mathrm{E}(\varepsilon_{11}^4) + 6T^*\sigma^4 + 3T^*(T^*-1)\sigma^4 + 2\sum_{d+3\leq t'<t\leq T}\mathrm{E}(\varepsilon_{i(t-1)}^2\varepsilon_{it'}^2) - 4T^{*2}\sigma^4\\
&= 2T^*\mathrm{E}(\varepsilon_{11}^4) - (T^{*2}-3T^*)\sigma^4 + 2\sum_{d+4\leq t'+1<t\leq T}\mathrm{E}(\varepsilon_{i(t-1)}^2\varepsilon_{it'}^2) + 2\sum_{d+4\leq t'+1=t\leq T}\mathrm{E}(\varepsilon_{i(t-1)}^2\varepsilon_{it'}^2)\\
&= 2T^*\mathrm{E}(\varepsilon_{11}^4) - (T^{*2}-3T^*)\sigma^4 + (T^*-1)(T^*-2)\sigma^4 + 2(T^*-1)\mathrm{E}(\varepsilon_{11}^4)\\
&= (4T^*-2)\mathrm{E}(\varepsilon_{11}^4) + 2\sigma^4.
\end{aligned}$$

Further,

$$\mathrm{var}\left\{\frac{\sqrt{nT^*}}{K}\sum_{i=1}^{n}\vartheta_i\right\} = \frac{1}{2T^*}\{(2T^*-1)\mathrm{E}\varepsilon_{11}^4 + \sigma^4\} + o(1).$$

Combining $\mathrm{E}|\varepsilon_{11}|^4 < \infty$ in Assumption (A1) with the Lindeberg–Levy Central Limit Theorem, as $n \to \infty$, we obtain

$$\begin{aligned}
\sqrt{nT^*}(\widehat{\sigma}_n^2 - \sigma^2) &= \frac{\sqrt{nT^*}}{2(nT^*-d-p-qJ_n)} + o_p(1)\sum_{i=1}^{n}(\vartheta_i - 2T^*\sigma^2)\\
&\to_D N\left(0, \frac{1}{2T^*}\{(2T^*-1)\mathrm{E}(\varepsilon_{11}^4) + \sigma^4\}\right).
\end{aligned}$$

This completes the proof.

(ii) Applying the results in Theorems 1 and 2, we have

$$|(\Delta\mathbf{y}_{it}^d)^\top(\widehat{\boldsymbol{\gamma}}_n - \boldsymbol{\gamma}) + \Delta\mathbf{x}_{it}^\top(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) + \tilde{\mathbf{N}}_{(i-1)T^*+t,*}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})| \leq O_p(\sqrt{J_n/nT^*}) + J_n^{-\ell},$$

which leads to

$$\begin{aligned}
\frac{1}{nT^*}\sum_{i=1}^{n}\sum_{t=d+3}^{T}\widehat{\Delta\varepsilon_{it}}^4 &= \frac{1}{nT^*}\sum_{i=1}^{n}\sum_{t=d+3}^{T}\left(\Delta y_{it} - (\Delta\mathbf{y}_{it}^d)^\top\widehat{\boldsymbol{\gamma}}_n - \Delta\mathbf{x}_{it}^\top\widehat{\boldsymbol{\beta}}_n - \tilde{\mathbf{N}}_{(i-1)T^*+t,*}\widehat{\boldsymbol{\theta}}_n\right)^4\\
&= \frac{1}{nT^*}\sum_{i=1}^{n}\sum_{t=d+3}^{T}\{\Delta\varepsilon_{it} - (\Delta\mathbf{y}_{it}^d)^\top(\widehat{\boldsymbol{\gamma}}_n - \boldsymbol{\gamma}) - \Delta\mathbf{x}_{it}^\top(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) - \tilde{\mathbf{N}}_{(i-1)T^*+t,*}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})\}^4\\
&= \frac{1}{nT^*}\sum_{i=1}^{n}\sum_{t=d+3}^{T}\Delta\varepsilon_{it}^4 + o_p(1) = 2\mathrm{E}(\varepsilon_{11}^4) + 6\sigma^4 + o_p(1).
\end{aligned}$$

Hence,

$$\widehat{\mathrm{E}}(\varepsilon_{11}^4) = \frac{1}{2nT^*}\sum_{i=1}^{n}\sum_{t=d+3}^{T}\widehat{\Delta\varepsilon_{it}}^4 - 3\widehat{\sigma}_n^2,$$

yielding the desired result. ■

Proof of Theorem 4. Let $\eta_1 = (nT^*)^{-1/2} + a_n$, $\eta_2 = (nT^*)^{-2/5} + a_n$, $\boldsymbol{\gamma}^* = \boldsymbol{\gamma} + \eta_1\boldsymbol{\omega}_1$, $\boldsymbol{\beta}^* = \boldsymbol{\beta} + \eta_1\boldsymbol{\omega}_2$, $\boldsymbol{\theta}_r^* = \boldsymbol{\theta}_r + \eta_2\boldsymbol{\omega}_{3r}$, $r = 1, \ldots, q$ and $\boldsymbol{\omega}_3 = (\boldsymbol{\omega}_{31}^\top, \ldots, \boldsymbol{\omega}_{3q}^\top)^\top \in \mathcal{R}^{qJ_n}$. If we can show that for any $\epsilon > 0$, there is always a sufficiently large

constant C such that

$$P\left\{\inf_{\|\boldsymbol{\omega}_1\|+\|\boldsymbol{\omega}_2\|+\|\boldsymbol{\omega}_3\|=C} Q_n^*(\boldsymbol{\gamma}^*,\boldsymbol{\beta}^*,\boldsymbol{\theta}^*) \geq Q_n^*(\boldsymbol{\gamma},\boldsymbol{\beta},\boldsymbol{\theta})\right\} \geq 1-\epsilon,$$

then there must exist a local minimizer of $Q_n^*(\boldsymbol{\gamma}^*,\boldsymbol{\beta}^*,\boldsymbol{\theta}^*)$ for this same constant C. Specifically, recognizing the fact that $pe(0)=0$ and a non-decreasing penalty function, it holds that

$$\begin{aligned}
&Q_n^*(\boldsymbol{\gamma}^*,\boldsymbol{\beta}^*,\boldsymbol{\theta}^*) - Q_n^*(\boldsymbol{\gamma},\boldsymbol{\beta},\boldsymbol{\theta}) \\
&\geq \frac{1}{nT^*}\{(\Delta\mathbf{y}-\Delta\mathbf{Y}^d\boldsymbol{\gamma}^*-\Delta\mathbf{X}\boldsymbol{\beta}^*-\tilde{\mathbf{N}}\boldsymbol{\theta}^*)^\top\mathbf{W}\boldsymbol{\Omega}^{-1}\mathbf{W}^\top(\Delta\mathbf{y}-\Delta\mathbf{Y}^d\boldsymbol{\gamma}^*-\Delta\mathbf{X}\boldsymbol{\beta}^*-\tilde{\mathbf{N}}\boldsymbol{\theta}^*) \\
&\quad -(\Delta\mathbf{y}-\Delta\mathbf{Y}^d\boldsymbol{\gamma}-\Delta\mathbf{X}\boldsymbol{\beta}-\tilde{\mathbf{N}}\boldsymbol{\theta})^\top\mathbf{W}\boldsymbol{\Omega}^{-1}\mathbf{W}^\top(\Delta\mathbf{y}-\Delta\mathbf{Y}^d\boldsymbol{\gamma}-\Delta\mathbf{X}\boldsymbol{\beta}-\tilde{\mathbf{N}}\boldsymbol{\theta})\} + \sum_{s=1}^{d_0}\{pe_{\lambda_{1n}}(|\gamma_s^*|) \\
&\quad - pe_{\lambda_{1n}}(|\gamma_s|)\} + \sum_{k=1}^{p_0}\{pe_{\lambda_{1n}}(|\beta_k^*|) - pe_{\lambda_{1n}}(|\beta_k|)\} + \sum_{r=1}^{q_0}\{pe_{\lambda_{2n}}(\|\boldsymbol{\theta}_r^*\|_{\mathrm{N}}) - pe_{\lambda_{2n}}(\|\boldsymbol{\theta}_r\|_{\mathrm{N}})\} \\
&\stackrel{def}{=} R_1+R_2+R_3+R_4.
\end{aligned}$$

Some calculations show that

$$\begin{aligned}
R_1 &= -\frac{2}{nT^*}(\eta_1\Delta\mathbf{Y}^d\boldsymbol{\omega}_1+\eta_1\Delta\mathbf{X}\boldsymbol{\omega}_2+\eta_2\tilde{\mathbf{N}}\boldsymbol{\omega}_3)^\top\mathbf{W}\boldsymbol{\Omega}^{-1}\mathbf{W}^\top\left\{\Delta\boldsymbol{\varepsilon}+(\tilde{\boldsymbol{\phi}}-\tilde{\mathbf{N}}\boldsymbol{\theta})\right\} \\
&\quad + \frac{1}{nT^*}(\eta_1\Delta\mathbf{Y}^d\boldsymbol{\omega}_1-\eta_1\Delta\mathbf{X}\boldsymbol{\omega}_2-\eta_2\tilde{\mathbf{N}}\boldsymbol{\omega}_3)^\top\mathbf{W}\boldsymbol{\Omega}^{-1}\mathbf{W}^\top(\eta_1\Delta\mathbf{Y}^d\boldsymbol{\omega}_1-\eta_1\Delta\mathbf{X}\boldsymbol{\omega}_2-\eta_2\tilde{\mathbf{N}}\boldsymbol{\omega}_3) \\
&\stackrel{def}{=} R_{11}+R_{12}.
\end{aligned}$$

Using similar steps to those used for proving Lemmas 2 and 3 and Theorem 1, we can show that

$$\begin{aligned}
R_{11} &= \frac{-2}{nT^*}\eta_1(\boldsymbol{\omega}_1^\top,\boldsymbol{\omega}_2^\top)(\Delta\mathbf{Y}^d,\Delta\mathbf{X})^\top\mathbf{W}\boldsymbol{\Omega}^{-1}\mathbf{W}^\top\Delta\boldsymbol{\varepsilon} - \frac{2}{nT^*}\eta_1(\boldsymbol{\omega}_1^\top,\boldsymbol{\omega}_2^\top)(\Delta\mathbf{Y}^d,\Delta\mathbf{X})^\top\mathbf{W}\boldsymbol{\Omega}^{-1}\mathbf{W}^\top \\
&\quad\times(\tilde{\boldsymbol{\phi}}-\tilde{\mathbf{N}}\boldsymbol{\theta}) - \frac{2}{nT^*}\eta_2\boldsymbol{\omega}_3^\top\tilde{\mathbf{N}}^\top\mathbf{W}\boldsymbol{\Omega}^{-1}\mathbf{W}^\top\Delta\boldsymbol{\varepsilon} - \frac{2}{nT^*}\eta_2\boldsymbol{\omega}_3^\top\tilde{\mathbf{N}}^\top\mathbf{W}\boldsymbol{\Omega}^{-1}\mathbf{W}^\top(\tilde{\boldsymbol{\phi}}-\tilde{\mathbf{N}}\boldsymbol{\theta}) \\
&= -O_p\{\sqrt{nT^*}+J_n^\ell\}(\|\boldsymbol{\omega}_1\|+\|\boldsymbol{\omega}_2\|)\eta_1 - O_p\{\sqrt{nT^*}+J_n^\ell\}\|\boldsymbol{\omega}_3\|\eta_2 \\
&= -O_p\{\sqrt{nT^*}+J_n^\ell\}(\|\boldsymbol{\omega}_1\|+\|\boldsymbol{\omega}_2\|+\|\boldsymbol{\omega}_3\|)\max(\eta_1,\eta_2)
\end{aligned}$$

and

$$\begin{aligned}
R_{12} &= \frac{1}{nT^*}\{\eta_1^2(\boldsymbol{\omega}_1^\top,\boldsymbol{\omega}_2^\top)(\Delta\mathbf{Y}^d,\Delta\mathbf{X})^\top\mathbf{W}\boldsymbol{\Omega}^{-1}\mathbf{W}^\top(\Delta\mathbf{Y}^d,\Delta\mathbf{X})(\boldsymbol{\omega}_1^\top,\boldsymbol{\omega}_2^\top)^\top+\eta_2^2\boldsymbol{\omega}_3^\top\tilde{\mathbf{N}}^\top\mathbf{W}\boldsymbol{\Omega}^{-1}\mathbf{W}\tilde{\mathbf{N}}\boldsymbol{\omega}_3\} \\
&\quad + \frac{2}{nT^*}\eta_1\eta_2(\boldsymbol{\omega}_1^\top,\boldsymbol{\omega}_2^\top)(\Delta\mathbf{Y}^d,\Delta\mathbf{X})^\top\mathbf{W}\boldsymbol{\Omega}^{-1}\mathbf{W}^\top\tilde{\mathbf{N}}\boldsymbol{\omega}_3 \\
&= O_p(nT^*)(\|\boldsymbol{\omega}_1\|^2+\|\boldsymbol{\omega}_2\|^2)\eta_1^2 + O_p(nT^*)\|\boldsymbol{\omega}_3\|^2\eta_2^2 + O_p(nT^*)(\|\boldsymbol{\omega}_1\|\,\|\boldsymbol{\omega}_3\|+\|\boldsymbol{\omega}_2\|\,\|\boldsymbol{\omega}_3\|)\eta_1\eta_2 \\
&= O_p(nT^*)(\|\boldsymbol{\omega}_1\|^2+\|\boldsymbol{\omega}_2\|^2+\|\boldsymbol{\omega}_3\|^2)\max(\eta_1^2,\eta_2^2).
\end{aligned}$$

Furthermore,

$$\begin{aligned}
R_2 &= \sum_{s=1}^{d_0}\{pe_{\lambda_{1n}}(|\gamma_s+\eta_1\omega_{1s}|) - pe_{\lambda_{1n}}(|\gamma_s|)\} \\
&= \sum_{s=1}^{d_0}\{\eta_1 pe'_{\lambda_{1n}}(|\gamma_s|)\mathrm{sgn}(\gamma_s)\omega_{1s} + \eta_1^2 pe''_{\lambda_{1n}}(|\gamma_s|)\omega_{1s}^2\}(1+o_p(1)) \leq \sqrt{d_0}\eta_1 a_n\|\boldsymbol{\omega}_1\| + \eta_1^2 b_n\|\boldsymbol{\omega}_1\|^2.
\end{aligned}$$

Using similar arguments, we have $R_3 \leq \sqrt{d_0}\eta_1 a_n\|\boldsymbol{\omega}_2\| + \eta_1^2 b_n\|\boldsymbol{\omega}_2\|^2$. Moreover,

$$\begin{aligned}
R_4 &= \sum_{r=1}^{q_0}\{pe_{\lambda_{2n}}(\|\boldsymbol{\theta}_r+\eta_2\boldsymbol{\omega}_{3r}\|_{\mathrm{N}}) - pe_{\lambda_{2n}}(\|\boldsymbol{\theta}_r\|_{\mathrm{N}})\} \\
&\leq \sum_{r=1}^{q_0}\left\{\frac{\eta_2 pe'_{\lambda_{2n}}(\|\boldsymbol{\theta}_r\|_{\mathrm{N}})\boldsymbol{\omega}_{3r}^\top\mathbf{N}_r{}^\top\mathbf{N}_r\boldsymbol{\theta}_r}{\|\boldsymbol{\theta}_r\|_{\mathrm{N}}} + \frac{\eta_2^2 pe''_{\lambda_{2n}}(\|\boldsymbol{\theta}_r\|_{\mathrm{N}})\boldsymbol{\theta}_r^\top\mathbf{N}_r{}^\top\mathbf{N}_r\mathbf{N}_r{}^\top\mathbf{N}_r\boldsymbol{\theta}_r\|\boldsymbol{\omega}_{3r}\|^2}{\|\boldsymbol{\theta}_r\|_{\mathrm{N}}^2}\right.
\end{aligned}$$

$$+\frac{\eta_2^2 pe'_{\lambda_{2n}}(\|\boldsymbol{\theta}_r\|_{\mathrm{N}})\boldsymbol{\omega}_{3r}^{\top}\mathbf{N}_r{}^{\top}\mathbf{N}_r\boldsymbol{\omega}_{3r}}{\|\boldsymbol{\theta}_r\|_{\mathrm{N}}^2}+\frac{\eta_2^2 pe'_{\lambda_{2n}}(\|\boldsymbol{\theta}_r\|_{\mathrm{N}})\boldsymbol{\theta}_r^{\top}\mathbf{N}_r{}^{\top}\mathbf{N}_r\mathbf{N}_r{}^{\top}\mathbf{N}_r\boldsymbol{\theta}_r\|\boldsymbol{\omega}_{3r}\|^2}{\|\boldsymbol{\theta}_r\|_{\mathrm{N}}^2}\Bigg\}(1+o_p(1))$$

$$\le\sqrt{nT^*q_0}\eta_2 a_n\|\boldsymbol{\omega}_3\|+nT^*\sqrt{q_0}\eta_2^2(b_n+a_n)\|\boldsymbol{\omega}_3\|^2.$$

Hence, for a sufficiently large constant C, R_{12} has the order $\|\boldsymbol{\omega}_1\|^2+\|\boldsymbol{\omega}_2\|^2+\|\boldsymbol{\omega}_3\|^2$ that dominates the other terms uniformly. The required conclusion thus follows. ■

Proof of Theorem 5. It suffices to prove that for $(s,k,r)\in\Lambda_2$, the following results hold:

$$\begin{aligned}
&\mathrm{sgn}\{\partial Q_n^*(\breve{\boldsymbol{\gamma}}_n,\breve{\boldsymbol{\beta}}_n,\breve{\boldsymbol{\theta}}_n)/\partial\gamma_s\}=\mathrm{sgn}(\gamma_s),\quad \gamma_s\in(-c(nT^*)^{-1/2},c(nT^*)^{-1/2}),\\
&\mathrm{sgn}\{\partial Q_n^*(\breve{\boldsymbol{\gamma}}_n,\breve{\boldsymbol{\beta}}_n,\breve{\boldsymbol{\theta}}_n)/\partial\beta_k\}=\mathrm{sgn}(\beta_k),\quad \beta_k\in(-c(nT^*)^{-1/2},c(nT^*)^{-1/2}),\\
&\mathrm{sgn}\{\partial Q_n^*(\breve{\boldsymbol{\gamma}}_n,\breve{\boldsymbol{\beta}}_n,\breve{\boldsymbol{\theta}}_n)/\partial\boldsymbol{\theta}_r\}=\mathrm{sgn}(\boldsymbol{\theta}_r),\quad \text{and}\quad \boldsymbol{\theta}_r\in(-c(nT^*)^{-\ell/(2\ell+1)}\mathbf{1},c(nT^*)^{-\ell/(2\ell+1)}\mathbf{1}).
\end{aligned}$$

We only consider the first derivative of γ_s. Results relating to the derivatives of β_k and $\boldsymbol{\theta}_r$ can be similarly obtained. Following the proof in Theorem 1, $\Delta\mathbf{W}^{\top}\Delta\boldsymbol{\varepsilon}=O_p(\sqrt{nT^*})$, which, when combined with the consistency assumptions of $(\breve{\boldsymbol{\gamma}}_n,\breve{\boldsymbol{\beta}}_n,\breve{\boldsymbol{\theta}}_n)$ in Theorem 5, implies that, when $\gamma_s\neq 0$,

$$\begin{aligned}
\frac{\partial Q_n^*(\breve{\boldsymbol{\gamma}}_n,\breve{\boldsymbol{\beta}}_n,\breve{\boldsymbol{\theta}}_n)}{\partial\gamma_s}&=\frac{1}{nT^*}(\Delta\mathbf{Y}_{*,s}^d)^{\top}\mathbf{W}\boldsymbol{\Omega}^{-1}\mathbf{W}^{\top}(\Delta\mathbf{y}-\Delta\mathbf{Y}^d\breve{\boldsymbol{\gamma}}_n-\Delta\mathbf{X}\breve{\boldsymbol{\beta}}_n-\tilde{\mathbf{N}}\breve{\boldsymbol{\theta}}_n)+pe'_{\lambda_{1n}}(|\breve{\gamma}_{s,n}|)\mathrm{sgn}(\breve{\gamma}_{s,n})\\
&=\frac{1}{nT^*}(\Delta\mathbf{Y}_{*,s}^d)^{\top}\mathbf{W}\boldsymbol{\Omega}^{-1}\mathbf{W}^{\top}\{\Delta\boldsymbol{\varepsilon}+\Delta\mathbf{Y}^d(\boldsymbol{\gamma}-\breve{\boldsymbol{\gamma}}_n)-\Delta\mathbf{X}(\boldsymbol{\beta}-\breve{\boldsymbol{\beta}}_n)-(\tilde{\boldsymbol{\phi}}-\tilde{\mathbf{N}}\breve{\boldsymbol{\theta}}_n)\}\\
&\quad+pe'_{\lambda_{1n}}(|\breve{\gamma}_{s,n}|)\mathrm{sgn}(\breve{\gamma}_{s,n})\\
&=O_p(\sqrt{1/n})+O_p(\sqrt{J_n/n}+J_n^{-\ell})+pe'_{\lambda_{1n}}(|\breve{\gamma}_{s,n}|)\mathrm{sgn}(\breve{\gamma}_{s,n})\\
&=\lambda_{1n}\{\lambda_{1n}^{-1}pe'_{\lambda_{1n}}(|\breve{\gamma}_{s,n}|)\mathrm{sgn}(\breve{\gamma}_{s,n})+O_p(\lambda_{1n}^{-1}(\sqrt{J_n/n}+J_n^{-\ell}))\}.
\end{aligned}\tag{A.6}$$

Under the conditions in Theorem 5, we can show that $pe'_{\lambda_{1n}}/\lambda_{1n}>0$ and $\lambda_{1n}^{-1}(\sqrt{J_n/n}+J_n^{-\ell})\to 0$. Thus, the sign of (A.6) is determined by $\mathrm{sgn}(\breve{\gamma}_{s,n})$, meaning that for given values of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, $Q_n^*(\breve{\boldsymbol{\gamma}}_n,\boldsymbol{\beta},\boldsymbol{\theta})$ is minimized at $\breve{\boldsymbol{\gamma}}_n=(\breve{\boldsymbol{\gamma}}_{n,(1)}^{\top},\mathbf{0}^{\top})^{\top}$. ■

Proof of Theorem 6. Denote $\boldsymbol{\theta}_{(1)}=(\boldsymbol{\theta}_1^{\top},\ldots,\boldsymbol{\theta}_{q_0}^{\top})^{\top}$ as the significant coefficient vector and let $\breve{\boldsymbol{\theta}}_{n,(1)}$ be the shrinkage estimator of $\boldsymbol{\theta}_{(1)}$. By Theorems 4 and 5, we have

$$\left((\breve{\boldsymbol{\gamma}}_{n,(1)}^{\top},\mathbf{0}^{\top}),(\breve{\boldsymbol{\beta}}_{n,(1)}^{\top},\mathbf{0}^{\top}),(\breve{\boldsymbol{\theta}}_{n,(1)}^{\top},\mathbf{0}^{\top})\right)^{\top}=\arg\min Q_n^*(\boldsymbol{\gamma},\boldsymbol{\beta},\boldsymbol{\theta})$$

with probability tending to 1. Hence,

$$\begin{aligned}
&\frac{\partial Q_n^*((\breve{\boldsymbol{\gamma}}_{n,(1)}^{\top},\mathbf{0}^{\top})^{\top},(\breve{\boldsymbol{\beta}}_{n,(1)}^{\top},\mathbf{0}^{\top})^{\top},(\breve{\boldsymbol{\theta}}_{n,(1)}^{\top},\mathbf{0}^{\top})^{\top})}{\partial\boldsymbol{\gamma}_{(1)}}\\
&=(nT^*)^{-1}(\Delta\mathbf{Y}_{(1)}^d)^{\top}\mathbf{W}\boldsymbol{\Omega}^{-1}\mathbf{W}^{\top}\{\Delta\boldsymbol{\varepsilon}+\Delta\mathbf{Y}_{(1)}^d(\boldsymbol{\gamma}_{(1)}-\breve{\boldsymbol{\gamma}}_{n,(1)})-\Delta\mathbf{X}_{(1)}(\boldsymbol{\beta}_{(1)}-\breve{\boldsymbol{\beta}}_{n,(1)})\\
&\quad-(\tilde{\boldsymbol{\phi}}_{(1)}-\tilde{\mathbf{N}}_{(1)}\breve{\boldsymbol{\theta}}_{n,(1)})\}+(pe'_{\lambda_{1n}}(|\breve{\gamma}_{1,n}|)\mathrm{sgn}(\breve{\gamma}_{1,n}),\ldots,pe'_{\lambda_{1n}}(|\breve{\gamma}_{d_0,n}|)\mathrm{sgn}(\breve{\gamma}_{d_0,n}))^{\top}=\mathbf{0},\\
&\frac{\partial Q_n^*((\breve{\boldsymbol{\gamma}}_{n,(1)}^{\top},\mathbf{0}^{\top})^{\top},(\breve{\boldsymbol{\beta}}_{n,(1)}^{\top},\mathbf{0}^{\top})^{\top},(\breve{\boldsymbol{\theta}}_{n,(1)}^{\top},\mathbf{0}^{\top})^{\top})}{\partial\boldsymbol{\beta}_{(1)}}\\
&=(nT^*)^{-1}\Delta\mathbf{X}_{(1)}^{\top}\mathbf{W}\boldsymbol{\Omega}^{-1}\mathbf{W}^{\top}\{\Delta\boldsymbol{\varepsilon}+\Delta\mathbf{Y}_{(1)}^d(\boldsymbol{\gamma}_{(1)}-\breve{\boldsymbol{\gamma}}_{n,(1)})-\Delta\mathbf{X}_{(1)}(\boldsymbol{\beta}_{(1)}-\breve{\boldsymbol{\beta}}_{n,(1)})\\
&\quad-(\tilde{\boldsymbol{\phi}}_{(1)}-\tilde{\mathbf{N}}_{(1)}\breve{\boldsymbol{\theta}}_{n,(1)})\}+(pe'_{\lambda_{1n}}(|\breve{\beta}_{1,n}|)\mathrm{sgn}(\breve{\beta}_{1,n}),\ldots,pe'_{\lambda_{1n}}(|\breve{\beta}_{p_0,n}|)\mathrm{sgn}(\breve{\beta}_{p_0,n}))^{\top}=\mathbf{0},
\end{aligned}$$

and

$$\begin{aligned}
&\frac{\partial Q_n^*((\breve{\boldsymbol{\gamma}}_{n,(1)}^{\top},\mathbf{0}^{\top})^{\top},(\breve{\boldsymbol{\beta}}_{n,(1)}^{\top},\mathbf{0}^{\top})^{\top},(\breve{\boldsymbol{\theta}}_{n,(1)}^{\top},\mathbf{0}^{\top})^{\top})}{\partial\boldsymbol{\theta}_{(1)}}\\
&=(nT^*)^{-1}\tilde{\mathbf{N}}_{(1)}^{\top}\mathbf{W}\boldsymbol{\Omega}^{-1}\mathbf{W}^{\top}\{\Delta\boldsymbol{\varepsilon}+\Delta\mathbf{Y}_{(1)}^d(\boldsymbol{\gamma}_{(1)}-\breve{\boldsymbol{\gamma}}_{n,(1)})-\Delta\mathbf{X}_{(1)}(\boldsymbol{\beta}_{(1)}-\breve{\boldsymbol{\beta}}_{n,(1)})-(\tilde{\boldsymbol{\phi}}_{(1)}-\tilde{\mathbf{N}}_{(1)}\breve{\boldsymbol{\theta}}_{n,(1)})\}\\
&\quad+\left(pe'_{\lambda_{2n}}(\|\breve{\boldsymbol{\theta}}_{1,n}\|_{\mathrm{N}})(\mathbf{N}_1{}^{\top}\mathbf{N}_1\breve{\boldsymbol{\theta}}_{1,n})^{\top}/\|\breve{\boldsymbol{\theta}}_{1,n}\|_{\mathrm{N}},\ldots,pe'_{\lambda_{2n}}(\|\breve{\boldsymbol{\theta}}_{q_0,n}\|_{\mathrm{N}})(\mathbf{N}_{q_0}{}^{\top}\mathbf{N}_{q_0}\breve{\boldsymbol{\theta}}_{q_0,n})^{\top}/\|\breve{\boldsymbol{\theta}}_{q_0,n}\|_{\mathrm{N}}\right)=0.
\end{aligned}$$

Recognizing that the Taylor expansion at γ_s satisfies

$$pe'_{\lambda_{1n}}(|\breve{\gamma}_{s,n}|)=pe'_{\lambda_{1n}}(|\gamma_s|)+pe''_{\lambda_{1n}}(|\gamma_s|)\mathrm{sgn}(\gamma_s)(\breve{\gamma}_{s,n}-\gamma_s)+o_p(\breve{\gamma}_{s,n}-\gamma_s),$$

and noting that $b_n \to 0$ with $\lambda_{\max} \to 0$, we obtain

$$(pe'_{\lambda_{1n}}(|\breve{\gamma}_{1,n}|)\mathrm{sgn}(\breve{\gamma}_{1,n}), \ldots, pe'_{\lambda_{1n}}(|\breve{\gamma}_{d_0,n}|)\mathrm{sgn}(\breve{\gamma}_{d_0,n}))^\top = o_p(\breve{\boldsymbol{\gamma}}_{n,(1)} - \boldsymbol{\gamma}_{(1)}).$$

Similarly, we can show that

$$(pe'_{\lambda_{1n}}(|\breve{\beta}_{1,n}|)\mathrm{sgn}(\breve{\beta}_{1,n}), \ldots, pe'_{\lambda_{1n}}(|\breve{\beta}_{p_0,n}|)\mathrm{sgn}(\breve{\beta}_{p_0,n}))^\top = o_p(\breve{\boldsymbol{\beta}}_{n,(1)} - \boldsymbol{\beta}_{(1)}) \quad \text{and}$$

$$\left(pe'_{\lambda_{2n}}(\|\breve{\boldsymbol{\theta}}_{1,n}\|_{\mathrm{N}})(\mathbf{N}_1{}^\top\mathbf{N}_1\breve{\boldsymbol{\theta}}_{1,n})^\top/\|\breve{\boldsymbol{\theta}}_{1,n}\|_{\mathrm{N}}, \ldots, pe'_{\lambda_{2n}}(\|\breve{\boldsymbol{\theta}}_{q_0}\|_{\mathrm{N}})(\mathbf{N}_{q_0}{}^\top\mathbf{N}_{q_0}\breve{\boldsymbol{\theta}}_{q_0})^\top/\|\breve{\boldsymbol{\theta}}_{q_0}\|_{\mathrm{N}}\right)^\top$$
$$= o_p(\breve{\boldsymbol{\theta}}_{n,(1)} - \boldsymbol{\theta}_{(1)}).$$

These yield

$$(nT^*)^{-1}(\Delta\mathbf{Y}^d_{(1)}, \Delta\mathbf{X}_{(1)})^\top\mathbf{W}\boldsymbol{\Omega}^{-1}\mathbf{W}^\top\Big\{\Delta\boldsymbol{\varepsilon} + (\Delta\mathbf{Y}^d_{(1)}, \Delta\mathbf{X}_{(1)})((\boldsymbol{\gamma}_{(1)} - \breve{\boldsymbol{\gamma}}_{n,(1)})^\top, (\boldsymbol{\beta}_{(1)} - \breve{\boldsymbol{\beta}}_{n,(1)})^\top)^\top$$
$$+ (\tilde{\boldsymbol{\phi}} - \tilde{\mathbf{N}}_{(1)}\boldsymbol{\theta}_{(1)}) + \tilde{\mathbf{N}}_{(1)}(\boldsymbol{\theta}_{(1)} - \breve{\boldsymbol{\theta}}_{n,(1)})\Big\} + o_p\Big\{((\boldsymbol{\gamma}_{(1)} - \breve{\boldsymbol{\gamma}}_{n,(1)})^\top, (\boldsymbol{\beta}_{(1)} - \breve{\boldsymbol{\beta}}_{n,(1)})^\top)^\top\Big\} = 0, \tag{A.7}$$

and

$$(nT^*)^{-1}\tilde{\mathbf{N}}^\top_{(1)}\mathbf{W}\boldsymbol{\Omega}^{-1}\mathbf{W}^\top\Big\{\Delta\boldsymbol{\varepsilon} + (\Delta\mathbf{Y}^d_{(1)}, \Delta\mathbf{X}_{(1)})((\boldsymbol{\gamma}_{(1)} - \breve{\boldsymbol{\gamma}}_{n,(1)})^\top, (\boldsymbol{\beta}_{(1)} - \breve{\boldsymbol{\beta}}_{n,(1)})^\top)^\top$$
$$+ (\tilde{\boldsymbol{\phi}}_{(1)} - \tilde{\mathbf{N}}_{(1)}\boldsymbol{\theta}_{(1)}) + \tilde{\mathbf{N}}_{(1)}(\boldsymbol{\theta}_{(1)} - \breve{\boldsymbol{\theta}}_{n,(1)})\Big\} + o_p(\boldsymbol{\theta}_{(1)} - \breve{\boldsymbol{\theta}}_{n,(1)}) = 0. \tag{A.8}$$

For notational simplicity, write

$$\boldsymbol{\zeta}_n = ((\breve{\boldsymbol{\gamma}}_{n,(1)} - \boldsymbol{\gamma}_{(1)})^\top, (\breve{\boldsymbol{\beta}}_{n,(1)} - \boldsymbol{\beta}_{(1)})^\top)^\top,$$
$$\boldsymbol{\Pi}_n = (nT^*)^{-1}\tilde{\mathbf{N}}^\top_{(1)}\mathbf{W}\boldsymbol{\Omega}^{-1}\mathbf{W}^\top\tilde{\mathbf{N}}_{(1)},$$

and

$$\boldsymbol{\Phi}_n = (nT^*)^{-1}\tilde{\mathbf{N}}^\top_{(1)}\mathbf{W}\boldsymbol{\Omega}^{-1}\mathbf{W}^\top(\Delta\mathbf{Y}^d_{(1)}, \Delta\mathbf{X}_{(1)}).$$

Note that the solution of $\breve{\boldsymbol{\theta}}_{n,(1)} - \boldsymbol{\theta}_{(1)}$ in Eq. (A.8) has the form

$$\breve{\boldsymbol{\theta}}_{n,(1)} - \boldsymbol{\theta}_{(1)} = -(\boldsymbol{\Pi}_n + o_p(1))^{-1}\boldsymbol{\Phi}_n\boldsymbol{\zeta}_n + (\boldsymbol{\Pi}_n + o_p(1))^{-1}(nT^*)^{-1}\tilde{\mathbf{N}}^\top_{(1)}\mathbf{W}\boldsymbol{\Omega}^{-1}\mathbf{W}^\top\{\Delta\boldsymbol{\varepsilon} + (\tilde{\boldsymbol{\phi}}_{(1)} - \tilde{\mathbf{N}}_{(1)}\boldsymbol{\theta}_{(1)})\}.$$

Substituting the above into (A.7), we obtain

$$(nT^*)^{-1}(\Delta\mathbf{Y}^d_{(1)}, \Delta\mathbf{X}_{(1)})^\top\mathbf{W}\boldsymbol{\Omega}^{-1}\mathbf{W}^\top\{(\Delta\mathbf{Y}^d_{(1)}, \Delta\mathbf{X}_{(1)}) - \tilde{\mathbf{N}}_{(1)}(\boldsymbol{\Pi}_n + o_p(1))^{-1}\boldsymbol{\Phi}_n\}\boldsymbol{\zeta}_n + o_p(\boldsymbol{\zeta}_n)$$
$$= (nT^*)^{-1}(\Delta\mathbf{Y}^d_{(1)}, \Delta\mathbf{X}_{(1)})^\top\mathbf{W}\boldsymbol{\Omega}^{-1}\mathbf{W}^\top\Big[\Delta\boldsymbol{\varepsilon} + (\tilde{\boldsymbol{\phi}}_{(1)} - \tilde{\mathbf{N}}_{(1)}\boldsymbol{\theta}_{(1)}) - (nT^*)^{-1}\tilde{\mathbf{N}}_{(1)}$$
$$\times(\boldsymbol{\Pi}_n + o_p(1))^{-1}\tilde{\mathbf{N}}^\top_{(1)}\mathbf{W}\boldsymbol{\Omega}^{-1}\mathbf{W}^\top\{\Delta\boldsymbol{\varepsilon} + (\tilde{\boldsymbol{\phi}}_{(1)} - \tilde{\mathbf{N}}_{(1)}\boldsymbol{\theta}_{(1)})\}\Big]. \tag{A.9}$$

Some calculations show that

$$\boldsymbol{\Phi}_n^\top\boldsymbol{\Pi}_n^{-1}\tilde{\mathbf{N}}^\top_{(1)}\mathbf{W}\boldsymbol{\Omega}^{-1}\mathbf{W}^\top\{(\Delta\mathbf{Y}^d_{(1)}, \Delta\mathbf{X}_{(1)}) - \tilde{\mathbf{N}}_{(1)}(\boldsymbol{\Pi}_n + o_p(1))^{-1}\boldsymbol{\Phi}_n\} = 0$$

and

$$\boldsymbol{\Phi}_n^\top\boldsymbol{\Pi}_n^{-1}\tilde{\mathbf{N}}^\top_{(1)}\mathbf{W}\boldsymbol{\Omega}^{-1}\mathbf{W}^\top\Big[\Delta\boldsymbol{\varepsilon} + (\tilde{\boldsymbol{\phi}}_{(1)} - \tilde{\mathbf{N}}_{(1)}\boldsymbol{\theta}_{(1)}) - (nT^*)^{-1}\tilde{\mathbf{N}}_{(1)}(\boldsymbol{\Pi}_n + o_p(1))^{-1}\tilde{\mathbf{N}}^\top_{(1)}$$
$$\cdot\mathbf{W}\boldsymbol{\Omega}^{-1}\mathbf{W}^\top\{\Delta\boldsymbol{\varepsilon} + (\tilde{\boldsymbol{\phi}}_{(1)} - \tilde{\mathbf{N}}_{(1)}\boldsymbol{\theta}_{(1)})\}\Big] = 0.$$

Write $\boldsymbol{\Psi} = (\Delta\mathbf{Y}^d_{(1)}, \Delta\mathbf{X}_{(1)}) - \tilde{\mathbf{N}}_{(1)}(\boldsymbol{\Pi}_n + o_p(1))^{-1}\boldsymbol{\Phi}_n$, and note that $\boldsymbol{\Psi}^\top\mathbf{W}\boldsymbol{\Omega}^{-1}\mathbf{W}^\top\tilde{\mathbf{N}}_{(1)} = 0$. Eq. (A.9) implies that

$$\{(nT^*)^{-1}\boldsymbol{\Psi}^\top\mathbf{W}\boldsymbol{\Omega}^{-1}\mathbf{W}^\top\boldsymbol{\Psi} + o_p(1)\}\sqrt{nT^*}\boldsymbol{\zeta}_n$$
$$= \frac{1}{\sqrt{nT^*}}\boldsymbol{\Psi}^\top\mathbf{W}\boldsymbol{\Omega}^{-1}\mathbf{W}^\top\Delta\boldsymbol{\varepsilon} + \frac{1}{\sqrt{nT^*}}\boldsymbol{\Psi}^\top\mathbf{W}\boldsymbol{\Omega}^{-1}\mathbf{W}^\top(\tilde{\boldsymbol{\phi}}_{(1)} - \tilde{\mathbf{N}}_{(1)}\boldsymbol{\theta}_{(1)})$$
$$\stackrel{def}{=} \boldsymbol{\nu}_1 + \boldsymbol{\nu}_2.$$

Along the same lines of the proof for Theorem 1, it can be shown that $\boldsymbol{\nu}_1 \to_D N(\mathbf{0}, \sigma^2\boldsymbol{\Gamma}_{(1)}\boldsymbol{\Sigma}^{-1}_{(1)}\boldsymbol{\Gamma}^\top_{(1)})$ and $\boldsymbol{\nu}_2 = o_p(1)$, where $\boldsymbol{\Sigma}_{(1)}$ and $\boldsymbol{\Gamma}_{(1)}$ are defined in Theorem 6. This completes the proof. ■

References

Ahn, S.C., Schmidt, P., 1995. Efficient estimation of models for dynamic panel data. J. Econometrics 68, 5–27.
Ai, C., You, J., Zhou, Y., 2014. Estimation of fixed effects panel data partially linear additive regression models. Econom. J. 17, 83–106.
Arellano, M., Bond, S.R., 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. Rev. Econom. Stud. 58, 277–297.
Arellano, M., Bover, O., 1995. Another look at the instrumental variable estimation of error-components models. J. Econometrics 68, 29–51.
Baglan, D., 2010. Efficient estimation of a partially linear dynamic panel data model with fixed effects: application to unemployment dynamics in the US. In: 15th International Conference on Panel Data, Bonn, Germany. http://apps.olin.wustl.edu/MEGConference/Files/pdf/2010/96 (accessed 25.04.2014).
Baltagi, B.H., Li, D., 2002. Series estimation of partially linear panel data models with fixed effects. Ann. Econ. Finance 3, 103–116.
Blundell, R., Bond, S., 1998. Initial conditions and moment restrictions in dynamic panel data models. J. Econometrics 87 (1), 115–143.
Blundell, R., Bond, S., Windmeijer, F., 2001. Estimation in dynamic panel data models: Improving on the performance of the standard GMM estimator. In: Baltagi, B.H., Fomby, T.B., Hill, R.C. (Eds.), Advances in Econometrics, 15: Nonstationary Panels, Panel Cointegration and Dynamic Panels. Emerald Group Publishing Limited, USA, pp. 53–91.
Bun, M.J., Windmeijer, F., 2010. The weak instrument problem of the system GMM estimator in dynamic panel data models. Econom. J. 13, 95–126.
Carroll, R., Maity, A., Mammen, E., Yu, K., 2009. Efficient semiparametric marginal estimation for the partially linear additive model for longitudinal/clustered data. Stat. Biosci. 1, 10–31.
Cornwell, C., Trumbull, W., 1994. Estimating the economic model of crime with panel data. Rev. Econ. Stat. 76, 360–366.
de Boor, C., 2001. A Practical Guide to Splines. Springer, New York, rev. ed..
Everaert, G., 2013. Orthogonal to backward mean transformation for dynamic panel data models. Econom. J. 16, 179–221.
Fan, Y., Li, Q., 1996. Consistent model specification tests: Omitted variables and semiparametric functional forms. Econometrica 64, 865–890.
Fan, J., Li, R., 2001. Variable selection via penalized likelihood. J. Amer. Statist. Assoc. 95, 1348–1360.
Fan, J., Peng, H., Huang, T., 2005. Semilinear high-dimensional model for normalization of microarray data: A theoretical analysis and partial consistency. J. Amer. Statist. Assoc. 100, 781–813.
Fan, Y., Li, Q., 2003. A kernel-based method for estimation additive partially linear models. Statist. Sinica 13, 739–762.
Gouriéroux, C., Phillips, P.C.B., Yu, J., 2010. Indirect inference for dynamic panel models. J. Econometrics 157 (1), 68–77.
Han, C., Phillips, P.C.B., 2010. GMM estimation for dynamic panels with fixed effects and strong instruments at unity. Econometric Theory 26, 119–151.
Hansen, L., 1982. Large sample properties of generalized method of moments estimations. Econometrica 50, 1029–1054.
Härdle, W., Huet, S., Mammen, E., Sperlich, S., 2004. Bootstrap inference in semiparametric generalized additive models. Econometric Theory 20 (02), 265–300.
Hastie, T.J., Tibshirani, R.J., 1990. Generalized Additive Models. Chapman and Hall, London.
Henderson, D.J., Carroll, R.J., Li, Q., 2008. Nonparametric estimation and testing of fixed effects panel data models. J. Econometrics 144, 257–275.
Huang, J.Z., Shen, H., 2004. Functional coefficient regression models for nonlinear time series: A polynomial spline approach. Scand. J. Statist. 31, 515–534.
Huang, J.Z., Wu, C.O., Zhou, L., 2004. Polynomial spline estimation and inference for varying coefficient models with longitudinal data. Statist. Sinica 14 (3), 763–788.
Huang, J.Z., Yang, L., 2004. Identification of nonlinear additive autoregressive models. J. R. Stat. Soc. Ser. B 66, 463–477.
Lee, Y., 2012. Bias in dynamic panel models under time series misspecification. J. Econometrics 169, 54–60.
Lee, L.F., Yu, J., 2014. Efficient GMM estimation of spatial dynamic panel data models with fixed effects. J. Econometrics 180, 174–197.
Li, Q., 2000. Efficient estimation of additive partially linear models. Internat. Econom. Rev. 41, 1073–1098.
Linton, O., Nielsen, J., 1995. A kernel method of estimating structured nonparametric regression based on marginal integration. Biometrika 82, 93–100.
Newey, W.K., 1994. The asymptotic variance of semiparametric estimators. Econometrica 62, 1349–1382.
Nickell, S., 1981. Biases in dynamic models with fixed effects. Econometrica 49 (6), 1417–1426.
Su, L., Ullah, A., 2006. Profile likelihood estimation of partially linear panel data models with fixed effects. Econom. Lett. 92, 75–81.
Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. J. R. Stat. Soc. Ser. B 58, 267–288.
Wang, H., Li, R., Tsai, C.L., 2007. Tuning parameter selectors for the smoothly clipped absolute deviation method. Biometrika 94, 553–568.
Wang, L., Yang, L., 2007. Spline-backfitted kernel smoothing of nonlinear additive autoregression model. Ann. Statist. 35 (6), 2474–2503.
You, J., Zhou, Y., Zhou, X., 2011. Series estimation in partially linear in-slide rregression models. Scand. J. Statist. 38, 89–107.