UNIVERSITY OF CAMBRIDGE
Faculty of Economics

JANEWAY INSTITUTE CAMBRIDGE

CAMBRIDGE WORKING PAPERS IN ECONOMICS
JANEWAY INSTITUTE WORKING PAPERS

GMM Estimation for High–Dimensional Panel Data Models

Tingting Cheng
Nankai University

Chaohua Dong
Zhongnan Uni of Economics and Law

Jiti Gao
Monash University

Oliver Linton
University of Cambridge

Abstract

In this paper, we study a class of high dimensional moment restriction panel data models with interactive effects, where factors are unobserved and factor loadings are nonparametrically unknown smooth functions of individual characteristics variables. We allow the dimension of the parameter vector and the number of moment conditions to diverge with sample size. This is a very general framework and includes many existing linear and nonlinear panel data models as special cases. In order to estimate the unknown parameters, factors and factor loadings, we propose a sieve-based generalized method of moments estimation method and we show that under a set of simple identification conditions, all those unknown quantities can be consistently estimated. Further we establish asymptotic distributions of the proposed estimators. In addition, we propose tests for over-identification, specification of factor loading functions, and establish their large sample properties. Moreover, a number of simulation studies are conducted to examine the performance of the proposed estimators and test statistics in finite samples. An empirical example on stock return prediction is studied to demonstrate the usefulness of the proposed framework and corresponding estimation methods and testing procedures.

Reference Details

2245	Cambridge Working Papers in Economics
2221	Janeway Institute Working Paper Series
Published	27 July 2022
Key Words	Generalized method of moments, High dimensional moment model, Interactive effect, Over-identification issue, Panel data, Sieve method
JEL-codes	C13, C14, C23
Websites	www.econ.cam.ac.uk/cwpe www.janeway.econ.cam.ac.uk/working-papers

GMM Estimation for High–Dimensional Panel Data Models

TINGTING CHENG[‡] and CHAOHUA DONG[†] and JITI GAO[*] and OLIVER LINTON[⋆]

[‡] School of Finance, Nankai University, China

[†] Zhongnan University of Economics and Law, China

[*]Monash University, Australia

[⋆]University of Cambridge, UK

July 27, 2022

Abstract

In this paper, we study a class of high dimensional moment restriction panel data models with interactive effects, where factors are unobserved and factor loadings are nonparametrically unknown smooth functions of individual characteristics variables. We allow the dimension of the parameter vector and the number of moment conditions to diverge with sample size. This is a very general framework and includes many existing linear and nonlinear panel data models as special cases. In order to estimate the unknown parameters, factors and factor loadings, we propose a sieve-based generalized method of moments estimation method and we show that under a set of simple identification conditions, all those unknown quantities can be consistently estimated. Further we establish asymptotic distributions of the proposed estimators. In addition, we propose tests for over-identification, specification of factor loading functions, and establish their large sample properties. Moreover, a number of simulation studies are conducted to examine the performance of the proposed estimators and test statistics in finite samples. An empirical example on stock return prediction is studied to demonstrate the usefulness of the proposed framework and corresponding estimation methods and testing procedures.

Keywords: Generalized method of moments, High dimensional moment model, Interactive effect, Over-identification issue, Panel data, Sieve method

JEL classification: C13, C14, C23

1 Introduction

In this paper, we consider a class of high dimensional moment restriction panel data models with interactive effects. Here, the high dimension relates to the dimension of regressors and the number of restrictions; the factors are unobservable and the factor loadings are nonparametrically unknown functions of individual characteristics. Such a setting is very general and includes many existing models as special cases, for example, the linear panel data model with interactive effects in Bai (2009), Bai & Li (2014), Moon & Weidner (2015), Lu & Su (2016), Bai & Liao (2017), Moon & Weidner (2017), nonlinear panel data models with additive individual and time effects in Fernández-Val & Weidner (2016), dynamic panel Logit models with fixed effects in Honoré & Weidner (2020), Kitazawa (2021) and nonlinear panel data models with interactive effects in Chen et al. (2021). We propose the generalized method of moments (GMM) coupled with a sieve approximation to estimate all unknown quantities in a large dimensional moment vector.

Suppose that

$$\mathbb{E}[m(W_{it}, X_{it}'\beta, \lambda(V_i)' f_t)] = 0 \tag{1}$$

for $i = 1, \cdots, N$, $t = 1, \cdots, T$, where $m(\cdot)$ is a $q \times 1$ vector of known functions, W_{it} contains the dependent variable and possible instrumental variables, X_{it} is a $p \times 1$ vector of regressors, $\beta = (\beta_1, \cdots, \beta_p)'$ is a $p \times 1$ vector of unknown slope coefficients, f_t is a $r \times 1$ vector of unknown factors, and the factor loadings $\lambda(V_i) = (\lambda_1(V_{i1}), \cdots, \lambda_r(V_{ir}))'$ are unknown functions of individual characteristic variables $V_i = (V_{i1}, \cdots, V_{ir})'$, which are time invariant. The setting of factor loadings gives researchers and practitioners the maximal flexibility and, on top of that, we shall discuss in Section 2 the case where the characteristics are allowed to vary over time. The setting for the loading functions is also proposed in Connor et al. (2012), Ma et al. (2021) and Dong, Gao & Peng (2021) etc. For example, Connor et al. (2012) allow that factor betas in the Fama–French model to depend on security characteristics, such as size, value, momentum and own-volatility factors. We follow their work by considering unobservable factors due to popularity and wide applicability. Here we consider the case where both the dimension of X_{it}, p, and the dimension of $m(\cdot)$, q, are large, that is, $p = p(N, T) \to \infty$ and $q = q(N, T) \to \infty$, as $N \to \infty$ and $T \to \infty$. In addition, we also allow the number of factors r to be large. The dependence of the dimensionality (q, p and r) on the sample size is suppressed for the sake of simplicity.

We are interested to estimate the unknown parameter β, unknown function $\lambda(\cdot)$ vector and f_t given (W_{it}, X_{it}, V_i). To deal with the unknown functions we shall use the sieve method (Chen & Shen 1998, Chen 2007). Our estimation strategy is to first use the sieve method to approximate the unknown function $\lambda(V_i)$ by a linear combination of basis functions in some suitable function space; in this way, the unknown quantities in (1) are fully parameterized. Then we are able to estimate β, $\lambda(\cdot)$ and f_t simultaneously. It is noteworthy that since factors are unobserved, a set of simple identification conditions are proposed that is quite weak and general, and enables us to identify the factors and their loadings from the estimation above.

Moreover, we propose a test for the over-identification in the GMM framework before the proposed test is utilised for specification testing of loading functions against local deviations.

As is well known, a genesis of moment restrictions is conditional moment restrictions (Ai & Chen 2003, 2007, Chen & Pouzo 2012). To illustrate the usefulness of model (1), we give the following examples, both of which are derived from conditional moment restrictions.

Example 1.1 (Conditional moment restrictions): Suppose that $\rho(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t)$ is a known J-dimensional vector of generalized residual functions and Z_i is a sub-vector of V_i. Then the unknown quantities (β, λ, f) can be determined by a conditional moment restriction $\mathbb{E}[\rho(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t)|Z_i] = 0$, almost surely. In fact, the GMM estimation in Arellano & Carrasco (2003), Breitung & Lechner (1995) and Breitung & Lechner (1998) are based on similar conditional moment conditions. Let $\Phi_k(z)$ be a sequence of vector of functions that can approximate any square integrable function of Z in some sense arbitrarily as $k \to \infty$. Then the conditional restriction implies a set of unconditional moment restrictions $\mathbb{E}[\rho(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t) \otimes \Phi_k(Z_i)] = 0$, where the symbol "$\otimes$" denotes the Kronecker product. Denote $m(W_{it}, Z_i, X_{it}'\beta, \lambda(V_i)'f_t) = \rho(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t) \otimes \Phi_k(Z_i)$. Notice that the dimension of $m(\cdot)$ is Jk which increases with k. Therefore, (β, λ, f) can be solved from the above unconditional moment conditions by GMM.

Example 1.2 (Binary Response Model): Let Y_{it} be a binary outcome assuming either 0 or 1. Suppose we have a binary panel data model of the form,

$$P(Y_{it} = 1|X_{it}, V_i, f_t) = F(\beta' X_{it} + \lambda(V_i)'f_t), \; i = 1, \cdots, N, \; t = 1, \cdots, T, \tag{2}$$

where $\beta, X_{it} \in \mathbb{R}^p$, $V_i \in \mathbb{R}$ and $F(\cdot)$ is a known cumulative distribution function, e.g., the standard normal or standard logistic distribution. Model (2) includes Example 1 in Fernández-Val & Weidner (2016) and Example 2 in Chen et al. (2021) as special cases. As mentioned in Fernández-Val & Weidner (2016), in a labor economics application, Y can be an indicator for female labor force participation and X can include fertility indicators and other socio-economic characteristics. A sieve-based GMM method outlined below in Section 2 can be used to estimate the unknown loading function $\lambda(v)$, factor f_t, and the parameter vector β. Suppose that the function $\lambda(v)$ can be approximated arbitrarily in some sense by a linear combination of k known functions denoted as $\widetilde{\Phi}_k(v)$, that is $\lambda(v) - \widetilde{\Phi}_k(v)'\alpha$ goes to zero in some sense as $k \to \infty$. Then under certain identification conditions, we can estimate β, α and f_t from the first order conditions for maximum likelihood estimation. We will consider this model in our simulation study.

The contributions of this paper can be summarized as follows. First, to the best of our knowledge, this is probably among the first to consider high dimensional moment restriction panel data models with interactive effects, which substantially generalizes the existing panel data models in the current literature. Second, we propose sieve-based GMM estimators for unknown parameters, factors and loadings. First, we use a sieve method to approximate unknown loading functions and then apply a high–dimensional GMM procedure. Third, we propose procedures to test the validity of moment conditions as well as specifications of loading functions. Fourth, under a semiparametric factor structure setting, our identification conditions fully employ the functional information. Last but not least, simulation and empirical studies further demonstrate the advantages and usefulness of this new model in comparison with some natural competitors.

The rest of the paper is organized as follows. Section 2 describes the estimation procedure. Sections 3 and 4 provide the asymptotic theories for the proposed estimators and test statistics. The simulation results and empirical study are presented in Sections 5 and 6, respectively. Section 7 concludes. Appendix A lists the necessary lemmas before they are used in the proofs of the main theorems in Appendix B. The proofs of the lemmas listed in Appendix A and that of Theorem 4.1 are given in an online supplemental document. Throughout this paper, $\|\cdot\|$ denotes Euclidean norm for a vector or Frobenius norm for a matrix, or the norm in function space that would not arise any ambiguity in the context; I_r denotes an identity matrix of dimension r; the operators $\to_P$ and $\to_D$ denote convergence in probability and in distribution, respectively.

2 Assumptions and estimation

In this section, we first describe a Hilbert function space wherein the unknown loading function $\lambda(v)$ admits an infinite orthogonal series representation. Then with the help of certain identification conditions, we develop a sieve-based GMM estimation procedure for unknown parameters and factors and their loadings.

2.1 Sieve estimation for factor loadings

Let $\lambda(v) = (\lambda_1(v_1), \cdots, \lambda_r(v_r))'$ be a vector of unknown functions, where $\lambda_\ell(v_\ell)$, $\ell = 1, 2, \cdots, r$, are univariate functions defined on $\mathbb{V}$, $\mathbb{V} \subset \mathbb{R}$. Let $\lambda_\ell(\cdot)$ belong to a Hilbert space $L^2(\mathbb{V}, \pi(w)) = \left\{g(w) : \int_{\mathbb{V}} g^2(w)\pi(w)dw < \infty\right\}$, where $\pi(w)$ is a user-chosen density function on $\mathbb{V}$. As usual, we define the norm $\|g\|_{L^2} = \left(\int_{\mathbb{V}} g^2(w)\pi(w)dw\right)^{1/2}$ and the inner product $\langle g_1, g_2\rangle = \int_{\mathbb{V}} g_1(w)g_2(w)\pi(w)dw$ for functions in the space. Throughout this study, for the r-vector of functions $\lambda(\cdot)$, its norm is defined as $\|\lambda\|_{L^2} = \left(\sum_{\ell=1}^r \|\lambda_\ell\|_{L^2}^2\right)^{1/2}$.

A sequence $\{\phi_j(\cdot), j \geq 0\}$ in Hilbert space is called orthogonal if $\langle \phi_i, \phi_j \rangle = 0$ for all $i \neq j$, and further orthonormal if $\|\phi_i\| = 1$ for all i. A complete orthonormal sequence forms a basis in Hilbert space; the theory about Hilbert space can be found in standard textbooks on functional analysis.

Assumption 2.1. *Suppose that $\{\phi_j(\cdot), j \geq 0\}$ is a basis for $L^2(\mathbb{V}, \pi(w))$, that is, $\langle \phi_i, \phi_j \rangle = \delta_{ij}$ the Kronecker delta and $\{\phi_j(\cdot), j \geq 0\}$ is complete.*

The existence of orthonormal basis is guaranteed by the Hilbert theory that enables us to have an infinite orthogonal series expansion for $\lambda_\ell(\cdot) \in L^2(\mathbb{V}, \pi(w))$:

$$\lambda_\ell(v_\ell) = \sum_{j=0}^{\infty} \alpha_{\ell j} \phi_j(v_\ell), \text{ where } \alpha_{\ell j} = \langle \lambda_\ell, \phi_j \rangle, \ \ell = 1, \cdots, r. \tag{3}$$

By the Parsevel equality, $\|\lambda_\ell(v_\ell)\|_{L^2}^2 = \sum_{j=0}^{\infty} \alpha_{\ell j}^2$, which implies the attenuation of the coefficients. For a positive integer k, define the partial sum $\lambda_\ell^{(k)}(v_\ell) = \sum_{j=0}^{k-1} \alpha_{\ell j} \phi_j(v_\ell) = \alpha_\ell' \Phi_k(v_\ell)$ as a truncated series, in which $\alpha_\ell = (\alpha_{\ell 0}, \cdots, \alpha_{\ell, k-1})'$ and $\Phi_k(v_\ell) = (\phi_0(v_\ell), \cdots, \phi_{k-1}(v_\ell))'$.

Let $\lambda^{(k)}(v) = (\lambda_1^{(k)}(v_1), \cdots, \lambda_r^{(k)}(v_r))' = \widetilde{\Phi}_k(v)'\alpha$, where $\widetilde{\Phi}_k(v) = \text{diag}(\Phi_k(v_1), \cdots, \Phi_k(v_r))$ and $\alpha = (\alpha_1', \cdots, \alpha_r')'$ is a $kr \times 1$ vector, containing all the coefficients in the truncation series. Define $\gamma_\ell^{(k)}(v_\ell) = \sum_{j=k}^{\infty} \alpha_{\ell j} \phi_j(v_\ell)$, the residue after truncation, and $\gamma^{(k)}(v) = (\gamma_1^{(k)}(v_1), \cdots, \gamma_r^{(k)}(v_r))'$. It is easy to see that $\lambda(v) = \lambda^{(k)}(v) + \gamma^{(k)}(v)$. It follows that $\lambda^{(k)}(v) \to \lambda(v)$ as $k \to \infty$, in some sense.

2.2 Sieve-based GMM estimation

When factors are unobserved, the lack of identification of factors f_t and factor loadings λ_i is well known in the literature as they enter the model in a multiplicative way (see for example Bai 2009, Bai & Li 2014). In this paper, we impose the following identification conditions (IC):

Identification conditions (IC)

(1) $\lambda_\ell(v), \ell = 1, 2, \cdots, r$, *has unitary norm*;
(2) *If* $\phi_0(v) = 1$, $\int_{\mathbb{V}} \lambda_\ell(v)\phi_0(v)\pi(v)dv = 0$ for $2 \leq \ell \leq r$;
(3) $f_{t\ell} > 0$ *for* $\ell = 1, 2, \cdots, r$ and $t = 1, \cdots, T$.

The first IC is common, which is also used by Connor et al. (2012) and Dong, Gao & Peng (2021). It is intuitive and reasonable since the unknown function can be rescaled to have unitary norm. The second IC is employed to normalize the additive terms such that the model can get rid of extra constants, which is standard in additive model; see Assumption C.2 in Dong & Linton (2018) and references therein. The third IC is not as restrictive as it appears, since the positivity of $f_{t\ell}$ can always be imposed using the multiplicative factor structure. In contrast to the identifcation condition in Bai (2009), this set of IC is simple and utilizes the proposed loading and factor structure to identify the factors and their loadings.

The parameter space for model (1) is defined as

$$\Theta = \{(b, g_\ell, s_t) : b \in \mathbb{R}^p; g_\ell \in L^2(\mathbb{V}, \pi(w)), 1 \le \ell \le r; s_t \in \mathbb{R}^r, 1 \le t \le T\},$$

which contains the true parameter (β, λ, F) as an interior point, where $F = (f_1, \cdots, f_T)$. For $b \in \mathbb{R}^p$ and $g \in L^2(\mathbb{V}, \pi(w))$, we define

$$\|(b, g)\| = \left(\|b\|^2 + \|g\|_{L^2}^2\right)^{1/2}.$$

Assumption 2.2. *Suppose that B_{1N} and B_{2N} are two sequences of positive numbers diverging with N, such that β in model (1) is in $\Theta_{1N} = \{b \in \mathbb{R}^p : \|b\| \le B_{1N}\}$ and for sufficiently large N, $\lambda^{(k)}(v)'f_t$ is included in $\Theta_{2N} = \{\Phi_k^r(v)'d_t : \|d_t\| \le B_{2N}\}$.*

The assumption above, coupled with the sieve approach, approximates the parameter space Θ by $\Theta_{1N} \otimes \Theta_{2N}$ as $N, T \to \infty$. The bounds on the parameters facilitate the solution of nonlinear optimization defined below, while, on the other hand, B_{1N} is necessary to take into account the divergency of the dimension of β, and B_{2N} may be mild divergent since the vector $\Phi_k(\cdot)$ of basis functions has norm equal to $O(\sqrt{k})$. See Dong, Linton & Peng (2021) for more details on the discussion on $\|\Phi_k(\cdot)\|$.

Recall $\lambda(v) = \lambda^{(k)}(v) + \gamma^{(k)}(v)$ in Section 2.1. Let $f_t'\lambda^{(k)}(v) = f_t'\widetilde{\Phi}_k(v)'\alpha = \Phi_k^r(v)'D_t$, where $\Phi_k^r(v) = (\Phi_k(v_1)', \cdots, \Phi_k(v_r)')'$ and $D_t = (f_{t1}\alpha_1', \cdots, f_{tr}\alpha_r')'$. Then we rewrite the moment condition (1) as

$$\mathbb{E}\left[m\left(W_{it}, X_{it}'\beta, \Phi_k^r(V_i)'D_t + f_t'\gamma^{(k)}(V_i)\right)\right] = 0, \tag{4}$$

in which $\gamma^{(k)}(V_i)$ is negligible for large values of k.

Our estimation strategy is that we first estimate β and D_t simultaneously for each given t, and then we obtain the estimators of f_t and α by virtue of the identification conditions, and finally estimate $\lambda(v)$ by $\widehat{\lambda}(v) = \widetilde{\Phi}_k(v)'\widehat{\alpha}$. Specifically, for each given t, we estimate β and D_t by

$$(\widehat{\beta}_t, \widehat{D}_t) = \arg\min_{b \in \mathbb{R}^p, d_t \in \mathbb{R}^{kr}} \|M_{Nt}(b, d_t)\|^2, \tag{5}$$

$$\text{subject to } \|b\| \le B_{1N} \text{ and } \|d_t\| \le B_{2N},$$

where $M_{Nt}(b, d_t) = \frac{1}{\sqrt{q}}\frac{1}{N}\sum_{i=1}^N m(W_{it}, X_{it}'b, \Phi_k^r(V_i)'d_t)$.

The reason for including q in the function $M_{Nt}(b, d_t)$ is to take into account the divergence of the dimension of $m(\cdot)$. Note that the subscript t in $\widehat{\beta}_t$ indicates that we only use the information at time t to estimate β. This strategy localizes a global parameter. We then define an estimator for β by $\overline{\widehat{\beta}} = \sum_{t=1}^T w_t\widehat{\beta}_t$, where w_t is a chosen weight. Here we simply set $w_t = 1/T$ and by simulation study, we show that this works well numerically. The averaging converts the localized parameter back to global, and more importantly this accelerates the rate as shown in Theorem 3.2, as if we estimate β using global information.

Write $\widehat{D}_t$ as $\widehat{D}_t = (\widehat{D}_{t1}', \cdots, \widehat{D}_{tr}')'$. Then it is easy to see that $\widehat{D}_{t\ell}$, for $1 \le \ell \le r$, is the estimator of

$f_{t\ell}\alpha_\ell$. When the truncation parameter k is large, we have $\|f_{t\ell}\alpha_\ell\|^2 = f_{t\ell}^2(\|\lambda_\ell(v_\ell)\|^2 + o(1)) = f_{t\ell}^2(1 + o(1))$, where the first equality is due to Parseval equality in Hilbert space and the second equality is due to the identification condition that $\lambda_\ell(\cdot)$ has unitary norm. Thus, by some regular conditions, $\|\widehat{D}_{t\ell}\|^2 - f_{t\ell}^2 = o_P(1)$. This implies that we can use $\|\widehat{D}_{t\ell}\|$ as an estimator for $|f_{t\ell}|$ (and hence $f_{t\ell}$ under the third IC). Consequently, we can further estimate α_ℓ by $\widehat{\alpha}_{t\ell} = \frac{\widehat{D}_{t\ell}}{\|\widehat{D}_{t\ell}\|} = \frac{S_\ell \widehat{D}_t}{\|S_\ell \widehat{D}_t\|}$, where S_ℓ denotes a selection matrix drawing the corresponding sub–vectors from the parent vector. We define this estimator by $\widehat{\alpha}_{t\ell}$ to indicate that it is obtained using the information at time t. Then, the estimators of $f_{t\ell}$ and $\lambda_\ell(v_\ell)$, for $1 \le \ell \le r$, are given by

$$\widehat{f}_{t\ell} = \|\widehat{D}_{t\ell}\|, \quad \overline{\widehat{\lambda}}_\ell(v_\ell) = \Phi_k(v_\ell)' \frac{1}{T}\sum_{t=1}^{T} \widehat{\alpha}_{t\ell}.$$

Also, we define $\widehat{\lambda}_{t\ell}(v_\ell) = \Phi_k(v_\ell)'\widehat{\alpha}_{t\ell}$ as the estimator of $\lambda_\ell(v_\ell)$ where the subscript t indicates that we only use the information at time t to estimate $\lambda_\ell(v_\ell)$.

It should be noted that we can generalize the model by allowing the variable of characteristics V_i to vary with time, that is, each of the factor loading is a function of V_{it}. We may then use sieve method to estimate $\lambda(V_{it})$ by $\widetilde{\Phi}_k(V_{it})'\alpha$ in a similar way, where α can be estimated using the same method as in (5). For notational simplicity, we focus on the case where the variable of characteristics is only indexed by i.

The notation used so far assumes a fully balanced panel data set. It should also be noted that in applications the set of individuals may be allowed to vary over the time sample. For example, the set of equities with full records over a reasonably long sample period is a small subset of the full data set. In this case, we may assume that the observations are unbalanced in the sense that in time period t, we only observe n_t individuals. Then we replace the current definition of $M_{Nt}(b, d_t)$ by $M_{Nt}(b, d_t) = \frac{1}{\sqrt{q}}\frac{1}{n_t}\sum_{i=1}^{n_t} m(W_{it}, X_{it}'b, \Phi_k^r(V_i)'d_t)$ and the rest of the estimation procedures are similar to the balanced case.

3 Asymptotic theory

In this section, we will establish the consistency and asymptotic normality for the estimators of β, factors f_t and factor loading function $\lambda(v)$.

3.1 Consistency

Before establishing the asymptotic theory, we first introduce some necessary assumptions.

Assumption 3.1. *Suppose that*

(a) *Denote $W_t = (W_{1t}, \cdots, W_{Nt})'$ and $X_t = (X_{1t}, \cdots, X_{Nt})'$, and suppose that $\{(W_t, X_t, f_t), 1 \le t \le T\}$ is strictly stationary and α–mixing; denote similarly $W_i = (W_{i1}, \cdots, W_{iT})'$ and $X_i = (X_{i1}, \cdots, X_{iT})'$ and suppose that $\{(W_i, X_i, V_i), 1 \le i \le N\}$ is identically distributed across i.*

Denote $\alpha_{ij}(|t-s|)$ as the α-mixing coefficient between $(W_{it}, X_{it}, V_i, f_t)$ and $(W_{js}, X_{js}, V_j, f_s)$. Let $\sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^\infty \big(\alpha_{ij}(t)\big)^{\delta/(4+\delta)} = O(N)$ and $\sum_{i=1}^N \sum_{j=1}^N \big(\alpha_{ij}(0)\big)^{\delta/(4+\delta)} = O(N)$, where $\delta > 0$ is chosen such that $\mathbb{E}\big[\|m(W_{it}, X_{it}'\beta, \lambda_i' f_t)\|^{2+\delta/2}\big] < \infty$, $\mathbb{E}\big[\|X_{it}\|^{2+\delta/2}\big] < \infty$ and $\mathbb{E}\big[\|f_t\|^{2+\delta/2}\big] < \infty$.

(b) *The density $f_V(v)$ of V_1 satisfies $c\pi(v) \le f_V(v) \le C\pi(v)$ on the support $\mathbb{V}$ of V_1 for some constants $C \ge c > 0$, where $\pi(v)$ is the density function involved in the Hilbert space at Section 2.1.*

(c) *The function $m(\cdot,\cdot,\cdot)$ is continuous in the second and third arguments.*

(d) *$q(N,T) - p(N,T) \ge kr$.*

Assumption 3.2. *Suppose that for each (N, t), there are unique $(\lambda(\cdot)' f_t$ and $\beta \in \mathbb{R}^p$ such that model (1) is satisfied. In other words, for any $\delta > 0$, there is an $\epsilon > 0$ such that*

$$\inf_{(b,d_t)\in\Theta, \|(b-\beta, \Phi_k^r(V_i)'d_t - \lambda(V_i)'f_t)\| \ge \delta} \frac{1}{q} \|\mathbb{E} m(W_{it}, X_{it}' b, \Phi_k^r(V_i)' d_t)\|^2 > \epsilon.$$

Assumption 3.3. *Suppose that for each (N, T), there is a measurable positive function $A(W, X, V)$ such that*

$$\begin{aligned} &q^{-1/2}\|m(W, X'b_1, g_1(V)'s_1) - m(W, X'b_2, g_2(V)'s_2)\| \\ \le &A(W, X, V)\big[\|b_1 - b_2\| + \|g_1(V)'s_1 - g_2(V)'s_2\|\big] \end{aligned}$$

for any $(b_1, g_1, s_1), (b_2, g_2, s_2) \in \Theta$, where (W, X, V) is any realization of (W_{it}, X_{it}, V_i) and the function A satisfies that $\mathbb{E}[A^2(W_{it}, X_{it}, V_i)] < \infty$.

We have the following comments on the assumptions. Assumption 3.1(a) requires stationarity for the data along time series dimension and identical distribution over cross sectional dimension. The identical distribution requirement here simplifies the presentation and some of the calculations, although it is possible to relax it to allow heterogeneity in the cross section. We use the α-mixing coefficient to measure the relationship between $(W_{it}, X_{it}, V_i, f_t)$ and $(W_{js}, X_{js}, V_j, f_s)$, which can capture both cross-sectional dependence and serial dependence. Particularly, $\alpha_{ij}(0)$ only measures the cross-sectional dependence between $(W_{it}, X_{it}, V_i, f_t)$ and $(W_{jt}, X_{jt}, V_j, f_t)$. This set-up is in the same spirit as Assumption A2 of Chen et al. (2012) and Assumption C of Bai (2009), and the entire Assumption 3.1 (a) is quite common in the literature, see Dong et al. (2015) and Liu (2020). The rest of the assumptions are similar to those in Dong, Gao & Linton (2021). Assumption 3.1 (b) is about the relation of the densities of the variable V and the function space, which is widely used in the literature. For the case of compact support for V, we can simply set $\pi(v) = 1$ and the density $f_V(v)$ is bounded away from zero and above from infinity. Here we also allow for unbounded support for V provided that the density π is chosen appropriately. Assumption 3.1(c) imposes continuity condition on the $m(\cdot)$ function and commonly used moment conditions satisfy this. Assumption 3.1 (d) allows for possible over-identification of the parameter vector in the moment conditions.

Moreover, Assumption 3.2 is necessary as it assumes a uniqueness condition in GMM framework for all unknown parameters. The involvement of q is due to the same reason as in the formulation of $M_{Nt}(b, d_t)$, which takes into account the divergent dimension of $m(\cdot)$. Assumption 3.3 is a kind of Lipschitz condition that guarantees the approximation $m(W_{it}, X_{it}'\beta, f_t'\widetilde{\Phi}_k(V_i)'\alpha)$ to $m(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t)$ because

$$
\begin{aligned}
&\|m(W_{it}, X_{it}'\beta, f_t'\widetilde{\Phi}_k(V_i)'\alpha) - m(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t)\| \\
&\leq A(W_{it}, X_{it}, V_i)\left[\|\lambda(V_i) - \widetilde{\Phi}_k(V_i)'\alpha\|\|f_t\|\right] = O_P(1)\|\gamma^{(k)}(v)\|\|f_t\| = o_P(1).
\end{aligned}
$$

Moreover, since $\mathbb{E}m(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t) = 0$, we have $\|\mathbb{E}m(W_{it}, X_{it}'\beta, f_t'\widetilde{\Phi}_k(V_i)'\alpha)\| = o(1)$.

With the above assumptions, we establish the following theorem.

Theorem 3.1. *(Consistency). Let Assumptions 2.1, 2.2, 3.1–3.3 hold and $B_{1N}^2 + B_{2N}^2 = o(N)$. Then we have*

1. *As $(N, T) \to (\infty, \infty)$, $\|(\overline{\widehat{\beta}} - \beta, \overline{\widehat{\lambda}}(v) - \lambda(v))\| \to_P 0$.*
2. *As $N \to \infty$, for any given t, $\|\widehat{f}_t - f_t\| \to_P 0$.*

The proof of Theorem 3.1 is given in Appendix B.

3.2 Asymptotic normality

As the dimension of β diverges with the sample size (N, T), we may not be able to establish a limit distribution for $\overline{\widehat{\beta}}$ directly. Instead, we consider some finite dimensional transformations of β. See some examples for the functionals in Dong, Gao & Linton (2021) and the reference therein.

Let $\mathscr{L}$ be a transformation from $\mathbb{R}^p \mapsto \mathbb{R}^\eta$ with $\eta \geq 1$ fixed. Here we consider the limit distributions of $\mathscr{L}(\widehat{\beta}) - \mathscr{L}(\beta)$, $\widehat{\lambda}(v)$ and $\widehat{f}_t$. To this end, we impose the following assumptions.

Assumption 3.4. *Suppose that each element function m_j of the $m(\cdot,\cdot,\cdot)$ is differentiable with respect to its second and third arguments up to the second order. The second derivative functions satisfy a Lipschitz* *condition in a neighborhood of the* $(\beta, \lambda' f)$*:*

$$|\partial^{(u)} m_j(W, X'\beta, \lambda(V)'f) - \partial^{(u)} m_j(W, X'b, g(V)'s)| \leq B_j(W, X'\beta, \lambda(V)'f)(\|b-\beta\| + \|\lambda' f - g's\|)^\tau$$

for some $\tau \in (0,1]$*, where* u *is two dimensional multiple index with* $|u| = 2$*,* $\partial^{(u)}$ *stands for the partial derivative of the function with respect to the second and third arguments and* B_j *are positive functions such that* $\max_{1\leq j\leq q} \mathbb{E}[B_j(W, X'\beta, \lambda(V)'f)^2] < \infty$.

Let $\frac{\partial}{\partial u} m(\cdots)$ and $\frac{\partial}{\partial w}(\cdots)$ denote the partial derivatives of $m(v,u,w)$ with respect to respectively its arguments u and w.

Assumption 3.5. *Suppose that for any* $i = 1, 2, \cdots, N,\ t = 1, 2, \cdots, T$*,*

(a) $\mathbb{E}\|m(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t)\|^2 = O(q)$*,* $\mathbb{E}\|X_{it}\|^2 = O(p)$*,* $\mathbb{E}\|\Phi_k(V_i)\|^2 = O(k)$ *and* $\mathbb{E}\|f_t\|^2 = O(r)$*;*

(b) $\mathbb{E}\|\frac{\partial}{\partial u} m(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t)\|^2 = O(q)$*, and* $\mathbb{E}\|\frac{\partial}{\partial w} m(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t)\|^2 = O(q)$*;*

(c) $\mathbb{E}\|\frac{\partial}{\partial u} m(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t) \otimes X_{it}\|^2 = O(pq)$*, and* $\mathbb{E}\|\frac{\partial}{\partial w} m(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t) \otimes \Phi_k^r(V_i)\|^2 = O(krq)$*;*

(d) $\mathbb{E}\|\frac{\partial^2}{\partial u^2} m(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t) \otimes X_{it}X_{it}'\|^2 = O(p^2 q)$*, and*
$\mathbb{E}\|\frac{\partial^2}{\partial w^2} m(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t) \otimes \Phi_k^r(V_i)\Phi_k^r(V_i)')\|^2 = O(k^2 r^2 q)$*);*

(e) $\mathbb{E}\|\frac{\partial^2}{\partial u \partial w} m(W_{it}, X_{it}'\beta, \alpha'\widetilde{\Phi}_k(V_i)f_t) \otimes X_{it}\Phi_k^r(V_i)'\|^2 = O(pkrq)$*.*

Assumption 3.6. *Suppose that*

(a) $\|\gamma^{(k)}(V)\|^2 p^2 = o(1),\ (NT)^{-1} p^2 q = o(1)$*;*

(b) $\|\gamma^{(k)}(V)\|^2 k^2 r^2 = o(1),\ (NT)^{-1} k^2 r^2 q = o(1)$*.*

Assumption 3.7. *The partial derivatives of* $m(v,u,w)$ *satisfy*

(a) $q^{-\frac{1}{2}}\|\frac{\partial}{\partial u} m(W, X'b_1, g_1(V)'s_1) - \frac{\partial}{\partial u} m(W, X'b_2, g_2(V)'s_2)\| \leq A_1(W,X,V)[\|b_1 - b_2\| + \|g_1(V)'s_1 - g_2(V)'s_2\|]$*, where* $\mathbb{E}[A_1(W,X,V)^2] < \infty$ *and* $\mathbb{E}[A_1(W,X,V)^2\|X\|^2] = O(p)$*.*

(b) $q^{-\frac{1}{2}}\|\frac{\partial}{\partial w} m(W, X'b_1, g_1(V)'s_1) - \frac{\partial}{\partial w} m(W, X'b_2, g_2(V)'s_2)\| \leq A_2(W,X,V)[\|b_1 - b_2\| + \|g_1(V)'s_1 - g_2(V)'s_2\|]$*, where* $\mathbb{E}[A_2(W,X,V)^2] < \infty$ *and* $\mathbb{E}[A_2(W,X,V)^2\|\Phi_k^r(V)\|^2] = O(kr)$*.*

Assumption 3.8. *The transformation* $\mathscr{L}$ *possesses continuous second partial derivatives and the Hessian matrix of each component* $\mathscr{L}_j$ *of* $\mathscr{L}$ *has uniformly bounded eigenvalues in a neighborhood of* β*; moreover, the first partial derivative of* $\mathscr{L}$ *at* β*,* $\partial\mathscr{L}(\beta)$ *has full rank.*

Assumption 3.4 is a standard assumption on the smoothness of moment functions. We impose the Lipschitz condition for the components of the m function to facilitate the approximation of the Hessian matrix within a neighbourhood of the true parameter. Assumption 3.5 imposes conditions on the second moments of the $m(\cdot)$ function. Since the dimension p of X is diverging and it is reasonable to allow that the second moment diverges too. Similarly, we assume $E\|f_1\|^2 = O(r)$. $E\|\Phi_k(V)\|^2 = O(k)$ can be satisfied for many orthogonal sequences, such as the orthogonal trigonometric polynomials and orthogonal algebraic polynomials on bounded interval with uniform density. We impose similar conditions for the norm of the first partial derivatives of $m(\cdot)$ function. Assumption 3.6 imposes certain conditions on the relation of truncation parameter k, dimension p of X, sample size N, T and the number of factors r. This normally ensures that the orthogonal series expansions for the unknown functions of the factor loading vector converge with certain rates. Assumption 3.7 is similar to Assumption 3.3 but this is for partial derivatives. Assumption 3.8 guarantees that we can approximate $\mathscr{L}(\widehat{\beta})$ by a linear form $\mathscr{L}(\beta)+\partial\mathscr{L}(\beta)'(\widehat{\beta}-\beta)$. All Assumptions 3.4-3.8 are commonly encountered in the literature, see Dong, Gao & Linton (2021).

Before we present the main theorem of asymptotic normality for our proposed estimators, we introduce the following notation. Define

$$\begin{aligned}
\Xi_m &= \mathbb{E}\left[m(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)m(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)'\right] \text{ for any } i=1,2\cdots,N, t=1,2,\cdots,T,\\
\Delta_x &= \mathbb{E}\frac{\partial}{\partial u}m(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)'\otimes X_{it}, \text{ for any } i=1,2\cdots,N, t=1,2,\cdots,T,\\
\Delta_k &= \mathbb{E}\frac{\partial}{\partial w}m(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)'\otimes \Phi_k^r(V_i), \text{ for any } i=1,2\cdots,N, t=1,2,\cdots,T,\\
\Xi_{ij,ts} &= \mathbb{E}\left[m(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)m(W_{js},X_{js}'\beta,\lambda(V_j)'f_s)'\right].
\end{aligned}$$

Theorem 3.2. *(Asymptotic normality). Let Assumptions 2.1, 2.2, 3.1–3.8 hold.*

1. *As $N\to\infty$, for any given t, we have $\sqrt{N}\left(\mathscr{L}(\widehat{\beta}_t)-\mathscr{L}(\beta)\right)\to_D N\left(0,\Sigma_{\beta_0}\right)$, where*

$$\begin{aligned}
\Sigma_{\beta_0} =&\partial\mathscr{L}(\beta)'\left(\Delta_x\Delta_x'\right)^{-1}\Delta_x\Xi_m\Delta_x'\left(\Delta_x\Delta_x'\right)^{-1}\partial\mathscr{L}(\beta)\\
&+\lim_N\frac{1}{N}\sum_{i\neq j}\partial\mathscr{L}(\beta)'\left(\Delta_x\Delta_x'\right)^{-1}\Delta_x\Xi_{ij,11}\Delta_x'\left(\Delta_x\Delta_x'\right)^{-1}\partial\mathscr{L}(\beta).
\end{aligned}$$

2. *As $(N,T)\to(\infty,\infty)$, we obtain $\sqrt{NT}\left(\mathscr{L}(\overline{\widehat{\beta}})-\mathscr{L}(\beta)\right)\to_D N\left(0,\Sigma_\beta\right)$, where*

$$\Sigma_\beta=\lim_{N,T}\frac{1}{NT}\sum_{i,j,t,s}\partial\mathscr{L}(\beta)'\left(\Delta_x\Delta_x'\right)^{-1}\Delta_x\Xi_{ij,ts}\Delta_x'\left(\Delta_x\Delta_x'\right)^{-1}\partial\mathscr{L}(\beta).$$

3. *As $N \to \infty$, for any given t and $1 \le \ell \le r$, we have $\sqrt{N}\left(\widehat{f_{t\ell}} - f_{t\ell}\right) \to_D N\left(0, \Sigma_{f\ell}\right)$, where*

$$\begin{aligned}\Sigma_{f\ell} =& \alpha_\ell' S_\ell \left(\Delta_k \Delta_k'\right)^{-1} \Delta_k \Xi_m \Delta_k' \left(\Delta_k \Delta_k'\right)^{-1} S_\ell' \alpha_\ell \\ &+ \lim_N \frac{1}{N} \sum_{i \neq j} \alpha_\ell' S_\ell \left(\Delta_k \Delta_k'\right)^{-1} \Delta_k \Xi_{ij,11} \Delta_k' \left(\Delta_k \Delta_k'\right)^{-1} S_\ell' \alpha_\ell.\end{aligned}$$

4. *As $N \to \infty$, for any given t and $1 \le \ell \le r$, we have $\frac{\sqrt{N}}{||\Phi_k(V)||}(\widehat{\lambda}_{t\ell}(V) - \lambda_\ell(V)) \to_D N\left(0, \Sigma_{\lambda,t\ell}\right)$, where*

$$\begin{aligned}\Sigma_{\lambda,t\ell} =& \frac{1}{||\Phi_k(V)||^2 f_{t\ell}^2} \Big[\Phi_k'(V) S_\ell \left(\Delta_k \Delta_k'\right)^{-1} \Delta_k \Xi_m \Delta_k' \left(\Delta_k \Delta_k'\right)^{-1} S_\ell' \Phi_k(V) \\ &+ \lim_N \frac{1}{N} \sum_{i \neq j} \Phi_k'(V) S_\ell \left(\Delta_k \Delta_k'\right)^{-1} \Delta_k \Xi_{ij,11} \Delta_k' \left(\Delta_k \Delta_k'\right)^{-1} S_\ell' \Phi_k(V)\Big].\end{aligned}$$

5. *As $(N, T) \to (\infty, \infty)$ and $1 \le \ell \le r$, we have $\frac{\sqrt{NT}}{||\Phi_k(V)||}(\overline{\widehat{\lambda}}_\ell(V) - \lambda_\ell(V)) \to_D N(0, \Sigma_{\lambda\ell})$, where*

$$\Sigma_{\lambda\ell} = \lim_{N,T} \frac{1}{NT} \sum_{i,j,t,s} \frac{1}{||\Phi_k(V)||^2 f_{t\ell} f_{s\ell}} \Phi_k'(V) S_\ell \left(\Delta_k \Delta_k'\right)^{-1} \Delta_k \Xi_{ij,ts} \Delta_k' \left(\Delta_k \Delta_k'\right)^{-1} S_\ell' \Phi_k(V).$$

The proof of Theorem 3.2 is given in Appendix B. Note that in the special case where $\Xi_{ij,ts} = 0$ for either $i \neq j$ or $t \neq s$, all the terms associated with the weak cross-sectional dependence and serial correlations of error terms disappear. Thus, the covariance matrices involved in Theorem 3.2 reduce to

$$\Sigma_{\beta_0} = \partial \mathscr{L}(\beta)' \left(\Delta_x \Delta_x'\right)^{-1} \Delta_x \Xi_m \Delta_x' \left(\Delta_x \Delta_x'\right)^{-1} \partial \mathscr{L}(\beta), \quad \Sigma_\beta = \Sigma_{\beta_0},$$
$$\Sigma_{f\ell} = \alpha_\ell' S_\ell \left(\Delta_k \Delta_k'\right)^{-1} \Delta_k \Xi_m \Delta_k' \left(\Delta_k \Delta_k'\right)^{-1} S_\ell' \alpha_\ell,$$
$$\Sigma_{\lambda,t\ell} = \frac{1}{||\Phi_k(V)||^2 f_{t\ell}^2} \Phi_k'(V) S_\ell \left(\Delta_k \Delta_k'\right)^{-1} \Delta_k \Xi_m \Delta_k' \left(\Delta_k \Delta_k'\right)^{-1} S_\ell' \Phi_k(V), \quad \Sigma_{\lambda\ell} = \lim_T \frac{1}{T} \sum_{t=1}^T \Sigma_{\lambda,t\ell}.$$

Here the estimators $\mathscr{L}(\widehat{\beta}_t)$ and $\widehat{\lambda}_{t\ell}$ have slow rates because they are constructed using information only at t (see assertions 1 and 4); while after "globalization" (average over t), they enjoy fast rates (see assertions 2 and 5). All these are comparable with the parametric and nonparametric literatures. For example, the covariance matrix Σ_β has the same form as those in standard GMM framework, except that here we have transformation $\mathscr{L}$, and Δ_x and $\Xi_{ij,ts}$ have diverging dimensions. From Theorem 3.2, we can see that the convergence rate of $\mathscr{L}(\overline{\widehat{\beta}})$ is $(NT)^{-1/2}$, which is consistent with the literature that $\widehat{\beta} = O_P((NT)^{-1/2})$ (Bai (2009)). On the other hand, the convergence rates of $\overline{\widehat{\lambda}}_\ell(V)$ and $\widehat{f_{t\ell}}$ are $||\Phi_k(V)||(NT)^{-1/2}$ and $N^{-1/2}$, respectively, which are in line with the results from Dong, Gao & Peng (2021).

To make statistical inference by Theorem 3.2, we need to provide a consistent estimator for each of the above covariance matrices. Toward this end, let $\widehat{m}_{it} \equiv m(W_{it}, X_{it}'\widehat{\beta}, \widehat{\lambda}(V_i)'\widehat{f}_t)$, $\widehat{m}_{u,it} \equiv m_u(W_{it}, X_{it}'\widehat{\beta}, \widehat{\lambda}(V_i)'\widehat{f}_t) \otimes X_{it}$ and $\widehat{m}_{w,it} \equiv m_w(W_{it}, X_{it}'\widehat{\beta}, \widehat{\lambda}(V_i)'\widehat{f}_t) \otimes \Phi_k^r(V_i)$, where $m_u(\cdots)$ and $m_w(\cdots)$ are the partial derivatives of m function $m(v,u,w)$ with respect to respectively its arguments u and w. We propose to estimate Δ_x, Δ_k, Ξ_m, $\Xi_{11,ts}(i=j, t\neq s)$, $\Xi_{ij,11}(i\neq j, t=s)$, respectively, by

$$\widehat{\Delta}_x = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\widehat{m}_{u,it}, \qquad \widehat{\Delta}_k = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\widehat{m}_{w,it}, \qquad \widehat{\Xi}_m = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\widehat{m}_{it}\widehat{m}'_{it},$$
$$\widehat{\Xi}_{11,ts} = \frac{1}{N}\sum_{i=1}^{N}\widehat{m}_{it}\widehat{m}'_{is}, \qquad \widehat{\Xi}_{ij,11} = \frac{1}{T}\sum_{t=1}^{T}\widehat{m}_{it}\widehat{m}'_{jt}.$$

The estimation of Σ_{β_0} suffers from heteroskedasticity and cross-sectional correlation (HAC), so that its sample analog is not consistent. We are about to use Assumption 3.1(a) to estimate $N^{-1}\sum_{i\neq j}\Xi_{ij,11}$. To this end, we follow the spirit of Ma et al. (2021) to consider a kernel-based robust estimator that accounts for HAC. Similar ideas have also been used in Phillips (1998) to accommodate serial correlations in the residuals when constructing robust t ratios. Specifically, the quantity $N^{-1}\sum_{i\neq j}\Xi_{ij,11}$ is estimated by a kernel-based approach, and we thus have the estimator of Σ_{β_0} as follows,

$$\begin{aligned}\widehat{\Sigma}_{\beta_0} =& \partial\mathscr{L}(\widehat{\beta})'\left(\widehat{\Delta}_x\widehat{\Delta}'_x\right)^{-1}\widehat{\Delta}_x\widehat{\Xi}_m\widehat{\Delta}'_x\left(\widehat{\Delta}_x\widehat{\Delta}'_x\right)^{-1}\partial\mathscr{L}(\widehat{\beta})\\ &+\partial\mathscr{L}(\widehat{\beta})'\left(\widehat{\Delta}_x\widehat{\Delta}'_x\right)^{-1}\widehat{\Delta}_x\left[\lim_N\frac{1}{N}\sum_{i\neq j}K^*\left(\frac{i-j}{M}\right)\widehat{\Xi}_{ij}\right]\widehat{\Delta}'_x\left(\widehat{\Delta}_x\widehat{\Delta}'_x\right)^{-1}\partial\mathscr{L}(\widehat{\beta}),\end{aligned}$$

where M trims the sample autocovariances and acts as a truncation lag, $K^*(u)$ is a symmetric kernel weighting function satisfying $K^*(0)=1$, and $|K^*(u)|\leq 1$. As shown in Ma et al. (2021) and Kiefer & Vogelsang (2005), the consistency of $\widehat{\Sigma}_{\beta_0}$ can be obtained under the condition that $M\to\infty$ and $M/N\to 0$ as $N\to\infty$.

Another key quantity to estimate is $(NT)^{-1}\sum_{i,j,t,s}\Xi_{ij,ts}$, which is more complicated than that of $N^{-1}\sum_{i\neq j}\Xi_{ij,11}$. For notational simplicity, we let $G=(NT)^{-1}\sum_{i,j,t,s}\Xi_{ij,ts}$. Following Thompson (2011), we estimate G by

$$\widehat{G} = V_{unit}+V_{time,0}-V_{white,0}+\sum_{\tau=1}^{L_T}(V_{time,\tau}+V'_{time,\tau})-\sum_{\tau=1}^{L_T}(V_{white,\tau}+V'_{white,\tau}), \tag{6}$$

where $V_{unit}=(NT)^{-1}\sum_i\widehat{c}_i\widehat{c}'_i$, $V_{time,\tau}=(NT)^{-1}\sum_t\widehat{s}_t\widehat{s}'_{t+\tau}$, $V_{white,\tau}=(NT)^{-1}\sum_t\sum_i\widehat{m}_{it}\widehat{m}'_{i,t+\tau}$, $\widehat{c}_i=\sum_t\widehat{m}_{it}$ and $\widehat{s}_t=\sum_i\widehat{m}_{it}$. Due to the mixing condition imposed in Assumption 3.1, we use the information from the observations where τ (i.e., $|t-s|$) is small and L_T is a sequence diverging with T but with a smaller order. It is also easy to see that $V_{white,0}=\widehat{\Xi}_m$. To estimate $\Sigma_{\lambda\ell}$, we replace $\frac{1}{f_{t\ell}f_{s\ell}}\Xi_{ij,ts}$ by $\frac{1}{\widehat{f}_{t\ell}\widehat{f}_{s\ell}}\widehat{m}_{it}\widehat{m}'_{js}$.

We then estimate Σ_β, $\Sigma_{f\ell}$, $\Sigma_{\lambda,t\ell}$ and $\Sigma_{\lambda\ell}$ by $\widehat{\Sigma}_\beta$, $\widehat{\Sigma}_{f\ell}$, $\widehat{\Sigma}_{\lambda,t\ell}$, $\widehat{\Sigma}_{\lambda\ell}$, respectively, by replacing the unknown quantities with their corresponding estimators.

Corollary 3.1. *Under Assumptions 3.4-3.8, we have* $\partial\mathscr{L}(\widehat{\beta})=\partial\mathscr{L}(\beta)+o_P(1)$, $\widehat{\Delta}_x=\Delta_x+o_P(1)$, $\widehat{\Delta}_k=$

$\Delta_k + o_P(1)$, $\widehat{\Xi}_m = \Xi_m + o_P(1)$, $\widehat{\Xi}_{11,ts} = \Xi_{11,ts} + o_P(1)$, *and* $\widehat{\Xi}_{ij,11} = \Xi_{ij,11} + o_P(1)$ *as* $N, T \to \infty$. *Consequently,* $\widehat{\Sigma}_\beta = \Sigma_\beta + o_P(1)$, $\widehat{\Sigma}_{f\ell} = \Sigma_{f\ell} + o_P(1)$, $\widehat{\Sigma}_{\lambda,t\ell} = \Sigma_{\lambda,t\ell} + o_P(1)$, *and* $\widehat{\Sigma}_{\lambda\ell} = \Sigma_{\lambda\ell} + o_P(1)$.

The proof of Corollary 3.1 is given in Appendix B.

4 Hypothesis testing

This section proposes two testing procedures with the first one being on over–identification testing, while the second one on parametric specification testing for loading functions. Both test statistics proposed make use of the deviation of the moment condition from the null to test against the alternative.

4.1 Test of over-identification

In this section, we aim to test the validity of the following moment conditions, which is crucial and often called the test of over–identification in the GMM literature. The null and alternative hypotheses are given as follows:

$$\begin{aligned} H_{01}: &\quad \mathbb{E}[m(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t)] = 0 \text{ for some } (\beta, \lambda, F) \in \Theta, \\ H_{11}: &\quad \mathbb{E}[m(W_{it}, X_{it}'b, g(V_i)'s_t)] \neq 0 \text{ for any } (b, g, S) \in \Theta, \end{aligned}$$

where Θ is define in Section 2.

For $b \in \mathbb{R}^p$, $d_t \in \mathbb{R}^{kr}$, $d = (d_1', \cdots, d_T')'$ and any $c \in \mathbb{R}^q$ such that $\|c\| = 1$, define

$$L_{NT}(b, d; c) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} L_{Nt}(b, d_t; c),$$

where $L_{Nt}(b, d_t; c) = \frac{1}{D_{Nt}(b, d_t; c)} \sum_{i=1}^{N} c'm(W_{it}, X_{it}'b, \Phi_k^r(V_i)'d_t)$, in which

$$D_{Nt}(b, d_t; c) = \sqrt{\sum_{i=1}^{N} [c'm(W_{it}, X_{it}'b, \Phi_k^r(V_i)'d_t)]^2}.$$

Recall that $D_t = (f_{t1}\alpha_1', \cdots, f_{tr}\alpha_r')'$. Under H_{01}, we can show that $(\overline{\widehat{\beta}}, \widehat{D}_t)$ are consistent estimators of (β, D_t) by Theorem 3.1. Define $\widehat{D} = (\widehat{D}_1', \cdots, \widehat{D}_T')'$. Then we can use the statistic $L_{NT}(\overline{\widehat{\beta}}, \widehat{D}; c)$ to test H_{01} against H_{11} as shown below. To establish an asymptotic distribution for the statistic, we first introduce some necessary assumptions.

Assumption 4.1. *Let* $\overline{m}^*_{Nt}(\overline{\widehat{\beta}}, \widehat{D}_t; c) = o_P(1)$ *when* $N \to \infty$*, where*

$$\overline{m}^*_{Nt}(b, d_t; c) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \mathbb{E}[c'm(W_{it}, X_{it}'b, \Phi_k^r(V_i)'d_t)]$$

for $(b, d_t) \in \Theta$ *and* c *such that* $\|c\| = 1$.

Assumption 4.2. *Let (i)* $p^2 q = o(NT)$ *and* $k^2 r^2 q = o(N)$*; and (ii)* $\|\gamma^{(k)}(v)\|^2 = o(q^{-1})$.

Due to the moment condition $\mathbb{E}[m(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t)] = 0$ and Assumption 3.3, for large values of k, $\mathbb{E}[m(W_{it}, X_{it}'\beta, \Phi_k^r(V_i)'D_t)] = o(1)$. So Assumption 4.1 requires that when (b, d_t) approaches (β, D_t), $\mathbb{E}[m(W_{it}, X_{it}'b, \Phi_k^r(V_i)'d_t)]$ goes to zero quickly. This assumption is in the same spirit as Assumption 3.2, but this is a sample version and the decay of the expectation needs a certain rate. Assumption 4.2 presents some technical requirements. Assumption 4.2 (i) requires a certain relationship between those diverging parameters and Assumption 4.2 (ii) imposes a decay rate for the norm of residue $\|\gamma^{(k)}(v)\|^2$, which can be easily satisfied given a certain smoothness of loading functions.

Theorem 4.1. *Suppose that there is no zero function in the vector* $m(\cdot, \cdot, \cdot)$. *Let Assumptions 2.1-2.2, 3.1-3.7 and 4.1-4.2 hold. For any* $c \in \mathbb{R}^q$ *such that* $\|c\| = 1$, *under* H_{01} *we have as* $(N, T) \to (\infty, \infty)$

$$L_{NT}(\overline{\widehat{\beta}}, \widehat{D}; c) \to_D N(0, 1).$$

Theorem 4.1 establishes the asymptotic normality of the proposed test statistic which makes the statistical inference feasible when factors are unobserved. The proof of Theorem 4.1 is given in Appendix B.

Theorem 4.2. *Suppose that the eigenvalues of* $\mathbb{E}[m(W_{it}, X_{it}'b, g(V_i)'s_t)m(W_{it}, X_{it}'b, g(V_i)'s_t)']$ *are bounded away from zero and infinity uniformly in* N, T *and* $(b, g, S) \in \Theta$. *Under* H_{11}, *suppose further that there exists a positive sequence* ζ_{NT} *such that* $\inf_{(b,g,S)\in\Theta} \|\mathbb{E}[m(W_{it}, X_{it}'b, g(V_i)'s_t)]\| \geq \zeta_{NT}$ *and* $\liminf_{N,T\to\infty} \sqrt{NT}\zeta_{NT} = \infty$. *Then for any vectors* b *and* d, *there exists some* $c^* \in \mathbb{R}^q$ *such that* $\|c^*\| = 1$ *and* $L_{NT}(b, d; c^*) \to_P \infty$, *as* $(N, T) \to (\infty, \infty)$.

This theorem establishes the consistency of the proposed test statistic under very general conditions. For example, in the special case of $\zeta_{NT} = \zeta$, the condition $\liminf_{N,T\to\infty} \sqrt{NT}\zeta_{NT} = \infty$ is automatically satisfied. Here, we allow for $\zeta_{NT} \to 0$ with a rate slower than $(NT)^{-1/2}$ that extends the literature where $\zeta_{NT} = \zeta$ a constant. Though the theorem only claims the existence of c^*, the proof given in Appendix B shows that how c^* is constructed.

4.2 Parametric specification for the loading functions

In practice, we are also interested in testing the parametric forms of loading functions. Therefore, in this section, we consider the following null and alternative hypotheses

$$H_{02}: \ \lambda_\ell(v_\ell; \theta_0) = g_\ell(v_\ell, \theta_0) = \sum_{j=0}^{k_0} \phi_j(v_\ell)\theta_{\ell j,0} \text{ such that } \mathbb{E}[m(W_{it}, X_{it}'\beta, \lambda(V_i; \theta_0)'f_t)] = 0,$$

$$H_{12}: \ \lambda_\ell(v_\ell;\theta_1) = g_\ell(v_\ell,\theta_1) + \Delta_{NT,\ell}(v_\ell) \text{ such that } \mathbb{E}[m(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t)] \neq 0,$$

where $\theta_0 = \left(\theta_{1,0}', \cdots, \theta_{r,0}'\right)' \in \mathbb{R}^{r(k_0+1)}$ where $\theta_{\ell,0} = (\theta_{\ell 0,0}, \cdots, \theta_{\ell k_0,0})'$ is a vector of unknown parameters, $\ell = 1, \cdots, r$, and k_0 is a fixed positive integer, θ_1 is defined similarly, $\phi_j(v_\ell)$ is the same orthonormal sequence as defined in Section 2.1, $\Delta_{NT,\ell}(v)$ is an unknown function going to zero for every v when $(N,T) \to (\infty,\infty)$ for us to study a sequence of local alternatives under H_{12}.

It is noteworthy that the formulation of the above parametric form of $g_\ell(v_\ell,\theta_0)$ is natural because it belongs to a particular subspace in $L^2(V,\pi(v))$. With this form of $g_\ell(v_\ell,\theta_0)$, our estimation procedure can be implemented directly without orthogonal expansion.

For $b \in \mathbb{R}^p$, $\theta \in \mathbb{R}^{r(k_0+1)}$, $\widetilde{f}_t \in \mathbb{R}^r$, and any $c \in \mathbb{R}^q$ such that $\|c\| = 1$, define

$$l_{NT}(b,\theta,\widetilde{f};c) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} l_{Nt}(b,\theta,\widetilde{f}_t;c),$$

where $\widetilde{f} \equiv (\widetilde{f}_1, \cdots, \widetilde{f}_T)'$, $l_{Nt}(b,\theta,\widetilde{f}_t;c) = \frac{1}{D_{Nt}(b,\theta,\widetilde{f}_t;c)} \sum_{i=1}^{N} c' m(W_{it}, X_{it}'b, \lambda(V_i;\theta)'\widetilde{f}_t)$, θ is included in $\lambda(\cdot)$, and $D_{Nt}(b,\theta,\widetilde{f}_t;c) = \sqrt{\sum_{i=1}^{N}[c'm(W_{it}, X_{it}'b, \lambda(V_i;\theta)'\widetilde{f}_t]^2}$.

Under H_{02} we estimate (β,θ,f_t) by $(\widehat{\overline{\beta}}, \widehat{\theta}, \widehat{f}_t)$ by the procedure in Section 2.2. We establish an asymptotic normality under H_{02} in Theorem 4.3, and then establish an asymptotic consistency under H_{12}. To show the consistency of the proposed test statistic under H_{12}, we impose the following assumption on the local deviation $\Delta_{NT}(v) = (\Delta_{NT,1}(v), \cdots, \Delta_{NT,r}(v))$.

Assumption 4.3. *Let $\Delta_{NT}(v) = \delta_{NT}\Delta(v)$, $\Delta(v) = (\Delta_1(v), \cdots, \Delta_r(v))$, satisfy the following conditions:*

(i) Suppose that $\min_{\widetilde{f}_t \in A} \|\Delta(v)'\widetilde{f}_t\| = c_{\min} > 0$, where A is a compact parameter set including all $\widetilde{f}_t$.

(ii) Let $\delta_{NT} \to 0$ and $\sqrt{NT}\epsilon_{NT} \to \infty$ as $(N,T) \to (\infty,\infty)$, where ϵ_{NT} is given by Assumption 3.2 corresponding to $c_{\min}\delta_{NT}$.

In Assumption 4.3 (i) the quantity $c_{\min}$, together with δ_{NT}, signifies the strength of signal of the moment condition under H_{12} deviated from that under H_{02}. $c_{\min}$ can be obtained by further calculating

$$\|\Delta(v)'\widetilde{f}_t\|^2 = \widetilde{f}_t' \int \Delta(v)\Delta(v)'\pi(v)dv\widetilde{f}_t \geq \lambda_{\min}(\Delta)\|\widetilde{f}_t\|^2 \geq \lambda_{\min}(\Delta) \min_{\widetilde{f}_t \in A} \|\widetilde{f}_t\|^2.$$

Thus, a sufficient condition for the existence of $c_{\min}$ is that the matrix $\int \Delta(v)\Delta(v)'\pi(v)dv$ has positive minimum eigenvalue and the set A deviates from the origin with a certain distance.

Assumption 4.3 (ii) is a consequence of Assumption 3.2, viz., given the deviation of the argument in moment function, $c_{\min}\delta_{NT}$, the moment function in norm has infimum ϵ_{NT} and we require this can not be too small to fulfil the detection.

We then have Theorem 4.3 below.

Theorem 4.3. *(i) Let Assumptions 2.1-2.2, 3.1-3.7 and 4.1-4.2(i) hold. Under H_{02}, we have as $(N,T) \to (\infty,\infty)$,*

$$l_{NT}(\overline{\widehat{\beta}}, \widehat{\theta}, \widehat{f}_t; c) \to_D N(0,1)$$

for any q-vector c such that $\|c\| = 1$.

(ii) If, in addition, Assumption 4.3 is satisfied, then under H_{12}, for any vectors b, θ and $\widetilde{f}_t$, there exists some $c^* \in \mathbb{R}^q$, $\|c^*\| = 1$, such that $l_{NT}(b, \theta, \widetilde{f}_t; c^*) \to_P \infty$, as $(N,T) \to (\infty,\infty)$.

This theorem justifies the use of the test statistic $l_{NT}(\overline{\widehat{\beta}}, \widehat{\theta}, \widehat{f}_t; c)$ for the specification testing of the loading functions. We have now established the asymptotic normality and consistency of the proposed test statistic under very general conditions. The proof of Theorem 4.3 are given in Appendix B.

5 Simulation results

In this section, we examine the finite sample performance of our proposed estimation procedure and test statistics using the following linear and nonlinear examples.

5.1 In-sample estimation evaluation

Example 5.1: Linear model

Suppose that we have the following linear regression model, which is quite general and includes many existing linear panel data models as special cases, such as the model considered by Bai (2009), Moon & Weidner (2015) and Bai & Liao (2017). Specifically, we consider a model with the form of

$$Y_{it} = X_{it}'\beta + \lambda(V_i)'f_t + e_{it},$$

where X_{it} are generated from a multivariate standard normal distribution $X_{it} \sim N(0, I_p)$, p is the dimension of X_{it} and $e_{it} \sim N(0,1)$. Here we set $\beta = (0.4, 0.5, 0, \cdots, 0)' \in \mathbb{R}^p$. Assume that we have two common factors, $f_{t\ell}$ for $\ell = 1,2$ and we generate $f_{t\ell}$ from Uniform$(2,3)$. The corresponding factor loadings are $\lambda(V) = (\lambda_1(V_1), \lambda_2(V_2))'$ with $V_\ell \sim$ Uniform$(0,1)$. For simplicity, we let $\lambda_1(v) = \lambda_2(v) = \sqrt{2}\cos(\pi v)$. It is easy to check that the above data generating process for factors and loadings satisfy the identification conditions outlined in Section 2.2. Our objective is to estimate β, $\lambda(\cdot)$ and f_t together with our proposed sieve-based GMM estimation method.

Suppose that $\mathbb{E}[e_{it}|V_i] = 0$. Let $\phi_0(v) = 1$, and for all $j \geq 0$, $\phi_j(v) = \sqrt{2}\cos(\pi j v)$. Then $\{\phi_j(v)\}$ is an orthonormal basis in the Hilbert space $L^2[0,1]$. Let $m(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t) = (Y_{it} - X_{it}'\beta - \lambda(V_i)'f_t)\Phi_q(V_i)$ where $W_{it} = (Y_{it}, V_i)$ and $\Phi_q(V_i) = (\phi_0(V_i), \phi_1(V_i), \cdots, \phi_{q-1}(V_i))'$. Then by law of iterated expectations, we have $\mathbb{E}[m(W_{it}, X_{it}'\beta, \lambda'(V_i)f_t)] = 0$. Note that the dimension of m function is q which is diverging. Thus the parameters $(\beta, \lambda(\cdot), f_t)$ can be solved from the above unconditional moment restrictions using the estimation procedure outlined in Section 2. Specifically, define $(\widehat{\beta}, \widehat{D}_t) = \arg\min \|M_{Nt}(b, d_t)\|^2$, where $M_{Nt}(b, d_t) = \frac{1}{\sqrt{q}} \frac{1}{N} \sum_{i=1}^{N} m(W_{it}, X_{it}' b, \Phi_k^r(V_i)' d_t)$. Then by the identification conditions outlined in Section 2.2, we can get $\widehat{f}_t$ and $\widehat{\alpha}$ separately from $\widehat{D}_t$. Further, $\widehat{\lambda}_1(V_1) = \Phi_k(V_1)'\widehat{\alpha}_1$ and $\widehat{\lambda}_2(V_2) = \Phi_k(V_2)'\widehat{\alpha}_2$, where $\widehat{\alpha}_1$ is the first k elements in $\widehat{\alpha}$ and $\widehat{\alpha}_2$ are the remaining k elements in $\widehat{\alpha}$.

We consider different combinations of (N, T) and for each pair of (N, T), $M = 100$ replications are executed. In this study, we assume the truncation parameter $k = (NT)^{1/5}$, $p = (NT)^{1/5}$, $\nu = 2$, $q = p + kr + \nu$. Note that the choices of these parameters may not be the optimal ones, but they satisfy all the requirements of our assumptions.

To examine the performance of the estimator $\overline{\widehat{\beta}}$ (recall that $\overline{\widehat{\beta}} = \frac{1}{T}\sum_{t=1}^{T} \widehat{\beta}_t$), we compute bias and median which are defined as follows.

$$\text{Bias}(\beta) = \|\beta - \overline{\widehat{\beta}}^*\|, \quad \text{Med}(\beta) = \text{median}(\|\beta - \overline{\widehat{\beta}}^1\|, \cdots, \|\beta - \overline{\widehat{\beta}}^M\|), \tag{7}$$

where the superscript s indicates the sth replication, $\overline{\widehat{\beta}}^*$ is the average of $\overline{\widehat{\beta}}^s$ for $s = 1, 2, \cdots, M$ over Monte Carlo replications.

To examine the performance of the estimators of loading functions, $\overline{\widehat{\lambda}}_\ell(v_\ell)$ (recall that $\overline{\widehat{\lambda}}_\ell(v_\ell) = \Phi_k(v_\ell)' \frac{1}{T}\sum_{t=1}^{T} \widehat{\alpha}_{t\ell}$), we compute bias, standard deviation (Std), root mean squared errors (RMSE) as follows.

$$\begin{aligned}
&\text{Bias}(\lambda_\ell) = \frac{1}{MN}\sum_{s=1}^{M}\sum_{i=1}^{N}\left(\overline{\widehat{\lambda}}_\ell^s(V_i) - \lambda_\ell^s(V_i)\right), \quad \text{Std}(\lambda_\ell) = \left(\frac{1}{MN}\sum_{s=1}^{M}\sum_{i=1}^{N}\left(\overline{\widehat{\lambda}}_\ell^s(V_i) - \overline{\widehat{\lambda}}_\ell^*(V_i)\right)^2\right)^{1/2} \\
&\text{RMSE}(\lambda_\ell) = \left(\frac{1}{MN}\sum_{s=1}^{M}\sum_{i=1}^{N}\left(\overline{\widehat{\lambda}}_\ell^s(V_i) - \lambda_\ell^s(V_i)\right)^2\right)^{1/2},
\end{aligned} \tag{8}$$

where $\overline{\widehat{\lambda}}_\ell^*(V_i)$ is the average of $\overline{\widehat{\lambda}}_\ell^s(V_i)$ over Monte Carlo replications.

Similarly, to examine the performance of $\widehat{f}_\ell$, $\ell = 1, 2$, we compute

$$\begin{aligned}
&\text{Bias}(f_\ell) = \frac{1}{MT}\sum_{s=1}^{M}\sum_{t=1}^{T}\left(\widehat{f}_{t\ell}^s - f_{t\ell}^s\right), \quad \text{Std}(f_\ell) = \left(\frac{1}{MT}\sum_{s=1}^{M}\sum_{t=1}^{T}\left(\widehat{f}_{t\ell}^s - \overline{\widehat{f}}_{t\ell}\right)^2\right)^{1/2} \\
&\text{RMSE}(f_\ell) = \left(\frac{1}{MT}\sum_{s=1}^{M}\sum_{t=1}^{T}\left(\widehat{f}_{t\ell}^s - f_{t\ell}^s\right)^2\right)^{1/2},
\end{aligned} \tag{9}$$

where $\overline{\widehat{f}}_{t\ell}$ is the average of $\widehat{f}_{t\ell}^s$ over Monte Carlo replications.

The results are presented in Table 1. We can find that (1) the bias and median of $\overline{\widehat{\beta}}$ are reasonably small for all sample sizes. Note that due to the divergence of the dimension, it might not make sense to compare the bias and median of $\widehat{\widehat{\beta}}$ for different sample sizes; (2) with the increase of N, standard deviations and root mean squared errors of $\widehat{f}_\ell$ decreases while it does not change much with the increase of T. This is consistent with our theory that the convergence rate of $\widehat{f}_\ell$ is $O(N^{-1/2})$; (3) with the increase of N and T, root mean squared errors of $\overline{\overline{\lambda}}(v)$ decreases. However, we notice that root mean squared errors of $\overline{\overline{\lambda}}(v)$ decrease much faster with the increase of N than with the increase of T. This is probably due to the fact that the function $\lambda(V_i)$ varies with i rather than t, so the increase of N could provide more information for the estimation.

Example 5.2: Nonlinear model

In this example, we consider a nonlinear panel data model below.

Let ε_{it} satisfy $Y_{it} = I\left[X_{it}'\beta + \lambda(V_i)'f_t - \varepsilon_{it} > 0\right]$. We then define a binary panel data model of the form:

$$P(Y_{it} = 1|X_{it}, V_i, f_t) = F(X_{it}'\beta + \lambda(V_i)'f_t),\ i = 1, \cdots, N,\ t = 1, \cdots, T, \tag{10}$$

where $\beta, X_{it} \in \mathbb{R}^p$, $V_i \in \mathbb{R}^2$, and $F(u) = \exp(u)/(1 + \exp(u))$ is the cumulative distribution function of ε_{it}. We assume that regressors X_{it} are generated from a standard multivariate normal distribution, that is, $X_{it} \sim N(0, I_p)$. Here we set $\beta = (0.4, 0.5, 0, \cdots, 0)' \in \mathbb{R}^p$. Assume that we have two common factors $f_{t\ell}$ for $\ell = 1, 2$ and we generate $f_{t\ell}$ from Uniform(0.5, 1.5). The corresponding factor loadings are $\lambda(V) = (\lambda_1(V_1), \lambda_2(V_2))'$ with $V_i \sim i.i.d.N(0_r, I_r)$ and $r = 2$. For simplicity, we set $\lambda_1(v) = \lambda_2(v) = \frac{\sqrt{2}}{\pi^{1/4}} v \exp(-v^2/2)$. We can see that the generation of loadings and factors satisfies the identification conditions in Section 2.2. The sieve-based GMM method outlined in Section 2 can be used to estimate the unknown function $\lambda(V)$, unknown factor f_t and the parameter vector β.

Based on model (10), we can get the log likelihood function

$$\ln\left[\prod_{i=1}^{N}\prod_{t=1}^{T} F^{Y_{it}}(\beta' X_{it} + \lambda(V_i)'f_t)\left(1 - F(\beta' X_{it} + \lambda(V_i)'f_t)\right)^{1-Y_{it}}\right].$$

Let $\Phi_k(v) = (p_0(v), \cdots, p_{k-1}(v))'$, where $\{p_j(x), j \geq 0\}$ is the sequence of Hermite functions that forms an orthonormal basis in $L^2(\mathbb{R})$. By the sieve estimation of $\lambda(v)$, we can approximate the log likelihood function by the following quantity Q_{NT} defined as

$$\begin{aligned} Q_{NT} &= \ln\prod_{i=1}^{N}\prod_{t=1}^{T} F^{Y_{it}}(X_{it}'\beta + f_t'\widetilde{\Phi}_k(V_i)'\alpha)\left(1 - F(X_{it}'\beta + f_t'\widetilde{\Phi}_k(V_i)'\alpha)\right)^{1-Y_{it}}, \\ &= \ln\prod_{i=1}^{N}\prod_{t=1}^{T} F^{Y_{it}}(X_{it}'\beta + \Phi_k^r(V_i)'D_t)\left(1 - F(X_{it}'\beta + \Phi_k^r(V_i)'D_t)\right)^{1-Y_{it}} = \sum_{t=1}^{T} Q_{Nt}, \end{aligned}$$

where $Q_{Nt} = \sum_{i=1}^{N} \ln\left[F^{Y_{it}}(X_{it}'\beta + \Phi_k^r(V_i)'D_t)\left(1 - F(X_{it}'\beta + \Phi_k^r(V_i)'D_t)\right)^{1-Y_{it}}\right]$.

At each given t, to maximize Q_{Nt}, we derive the first order conditions as follows.

$$\frac{\partial Q_{Nt}}{\partial \beta} = \sum_{i=1}^{N} \frac{(Y_{it} - F(X_{it}'\beta + \Phi_k^r(V_i)'D_t))F^{(1)}(X_{it}'\beta + \Phi_k^r(V_i)'D_t)}{F(X_{it}'\beta + \Phi_k^r(V_i)'D_t)(1 - F(X_{it}'\beta + \Phi_k^r(V_i)'D_t))} X_{it} = 0$$
$$\frac{\partial Q_{Nt}}{\partial D_t} = \sum_{i=1}^{N} \frac{(Y_{it} - F(X_{it}'\beta + \Phi_k^r(V_i)'D_t))F^{(1)}(X_{it}'\beta + \Phi_k^r(V_i)'D_t)}{F(X_{it}'\beta + \Phi_k^r(V_i)'D_t)(1 - F(X_{it}'\beta + \Phi_k^r(V_i)'D_t))} \Phi_k^r(V_i) = 0,$$

where $F^{(1)}(u) = \frac{\partial F(u)}{\partial u}$. It is easy to see that the above first order conditions are equivalent as $\frac{1}{N}\frac{\partial Q_{Nt}}{\partial \beta} = 0$ and $\frac{1}{N}\frac{\partial Q_{Nt}}{\partial D_t} = 0$. This can be regarded as the sample version of the following moment conditions with $m(\cdot) = (m_1'(\cdot), m_2'(\cdot))'$:

$$\mathbb{E}[m_1(Y_{it}, X_{it}'\beta, \Phi_k^r(V_i)'D_t)] = \mathbb{E}\left[\frac{(Y_{it} - F(X_{it}'\beta + \Phi_k^r(V_i)'D_t))F^{(1)}(X_{it}'\beta + \Phi_k^r(V_i)'D_t)}{F(X_{it}'\beta + \Phi_k^r(V_i)'D_t)(1 - F(X_{it}'\beta + \Phi_k^r(V_i)'D_t))} X_{it}\right] = 0,$$
$$\mathbb{E}[m_2(Y_{it}, X_{it}'\beta, \Phi_k^r(V_i)'D_t)] = \mathbb{E}\left[\frac{(Y_{it} - F(X_{it}'\beta + \Phi_k^r(V_i)'D_t))F^{(1)}(X_{it}'\beta + \Phi_k^r(V_i)'D_t)}{F(X_{it}'\beta + \Phi_k^r(V_i)'D_t)(1 - F(X_{it}'\beta + \Phi_k^r(V_i)'D_t))} \Phi_k^r(V_i)\right] = 0.$$

Accordingly, define $M_{Nt}(\beta, D_t) = \left(\frac{1}{N}\frac{\partial Q_{Nt}}{\partial \beta'}, \frac{1}{N}\frac{\partial Q_{Nt}}{\partial D_t'}\right)'$ and $(\widehat{\beta}, \widehat{D}_t) = \arg\min \|M_{Nt}(b, d_t)\|^2$. Then by the identification conditions outlined in Section 2.2, we can get $\widehat{f}_t$ and $\widehat{\alpha}$. Further, $\widehat{\lambda}_1(V_1) = \Phi_k(V_1)'\widehat{\alpha}_1$ and $\widehat{\lambda}_2(V_2) = \Phi_k(V_2)'\widehat{\alpha}_2$, where $\widehat{\alpha}_1$ is the first k elements in $\widehat{\alpha}$ and $\widehat{\alpha}_2$ are the remaining k elements in $\widehat{\alpha}$.

We explore different values of (N, T) and for each pair of (N, T), we do $M = 100$ replications. In this study, we assume the truncation parameter $k = (NT)^{1/5}$, $p = (NT)^{1/5}$, $q = p + kr$ (exactly identified). To examine the performance of our estimators, we compute the bias and median for the estimator of β given by (7). We also compute the bias, standard deviation, root mean squared errors for the estimator of $\lambda(V)$ and f_t given by (8) and (9), respectively.

The results are presented in Table 2 that reveals quite similar findings as Example 5.1. Namely, (1) the bias and median of $\overline{\widehat{\beta}}$ are reasonably small for all sample sizes. Note that due to the divergence of the dimension, it might not make sense to compare the bias and median of $\overline{\widehat{\beta}}$ for different sample sizes; (2) with the increase of N, standard deviations and root mean squared errors of $\widehat{f}_\ell$ decreases while it does not change much with the increase of T, which is consistent with our theory that the convergence rate of $\widehat{f}_\ell$ is $O(N^{-1/2})$; (3) with the increase of N and T, root mean squared errors of $\overline{\widehat{\lambda}}(v)$ decreases. However, we notice that root mean squared errors of $\overline{\widehat{\lambda}}(v)$ decrease much faster with the increase of N than with the increase of T.

Table 1: Simulation results in Example 5.1.

(T,N)	$\overline{\widehat{\beta}}$		$\overline{\widehat{\lambda}}_1$			$\overline{\widehat{\lambda}}_2$			$\widehat{f}_1$			$\widehat{f}_2$		
	Bias	Med	Bias	Std	RMSE	Bias	Std	RMSE	Bias	Std	RMSE	Bias	Std	RMSE
(50,50)	0.0027	0.0344	-0.0000	0.0089	0.0091	0.0002	0.0099	0.0101	0.0044	0.1528	0.1535	0.0068	0.1448	0.1454
(50,100)	0.0028	0.0234	-0.0000	0.0058	0.0059	0.0002	0.0070	0.0071	0.0038	0.1039	0.1045	0.0052	0.1083	0.1091
(50,200)	0.0020	0.0156	0.0001	0.0043	0.0044	0.0000	0.0040	0.0040	0.0038	0.0738	0.0743	0.0015	0.0736	0.0740
(100,50)	0.0025	0.0256	-0.0000	0.0055	0.0059	0.0002	0.0063	0.0066	0.0047	0.1536	0.1546	0.0067	0.1438	0.1448
(100,100)	0.0008	0.0165	0.0001	0.0041	0.0042	0.0002	0.0042	0.0044	0.0010	0.1030	0.1035	0.0024	0.1062	0.1068
(100,200)	0.0008	0.0117	0.0000	0.0028	0.0028	0.0000	0.0027	0.0028	0.0008	0.0748	0.0751	0.0011	0.0749	0.0752
(200,50)	0.0015	0.0168	-0.0000	0.0040	0.0044	0.0001	0.0044	0.0048	0.0038	0.1521	0.1529	0.0037	0.1418	0.1426
(200,100)	0.0011	0.0122	0.0000	0.0030	0.0032	0.0002	0.0027	0.0030	0.0017	0.1031	0.1036	0.0028	0.1078	0.1084
(200,200)	0.0009	0.0083	0.0001	0.0021	0.0022	0.0000	0.0020	0.0020	0.0012	0.0741	0.0745	0.0002	0.0748	0.0752
(50,50)	0.0027	0.0344	-0.0000	0.0089	0.0091	0.0002	0.0099	0.0101	0.0044	0.1528	0.1535	0.0068	0.1448	0.1454
(100,50)	0.0025	0.0256	-0.0000	0.0055	0.0059	0.0002	0.0063	0.0066	0.0047	0.1536	0.1546	0.0067	0.1438	0.1448
(200,50)	0.0015	0.0168	-0.0000	0.0040	0.0044	0.0001	0.0044	0.0048	0.0038	0.1521	0.1529	0.0037	0.1418	0.1426
(50,100)	0.0028	0.0234	-0.0000	0.0058	0.0059	0.0002	0.0070	0.0071	0.0038	0.1039	0.1045	0.0052	0.1083	0.1091
(100,100)	0.0008	0.0165	0.0001	0.0041	0.0042	0.0002	0.0042	0.0044	0.0010	0.1030	0.1035	0.0024	0.1062	0.1068
(200,100)	0.0011	0.0122	0.0000	0.0030	0.0032	0.0002	0.0027	0.0030	0.0017	0.1031	0.1036	0.0028	0.1078	0.1084
(50,200)	0.0020	0.0156	0.0001	0.0043	0.0044	0.0000	0.0040	0.0040	0.0038	0.0738	0.0743	0.0015	0.0736	0.0740
(100,200)	0.0008	0.0117	0.0000	0.0028	0.0028	0.0000	0.0027	0.0028	0.0008	0.0748	0.0751	0.0011	0.0749	0.0752
(200,200)	0.0009	0.0083	0.0001	0.0021	0.0022	0.0000	0.0020	0.0020	0.0012	0.0741	0.0745	0.0002	0.0748	0.0752

Table 2: Simulation results in Example 5.2.

(T,N)	$\overline{\widehat{\beta}}$		$\overline{\widehat{\lambda}}_1$			$\overline{\widehat{\lambda}}_2$			$\widehat{f}_1$			$\widehat{f}_2$		
	Bias	Med	Bias	Std	RMSE	Bias	Std	RMSE	Bias	Std	RMSE	Bias	Std	RMSE
(50,50)	0.1232	0.1386	0.0062	0.0441	0.1232	-0.0515	0.0355	0.1300	0.4992	0.6095	0.7900	0.4933	0.5744	0.7597
(50,100)	0.0778	0.0894	0.0125	0.0362	0.1013	-0.0272	0.0305	0.0720	0.2924	0.4445	0.5352	0.3066	0.4778	0.5695
(50,200)	0.0447	0.0557	0.0189	0.0263	0.0583	-0.0192	0.0244	0.0612	0.1829	0.3420	0.3901	0.1670	0.3392	0.3806
(100,50)	0.1245	0.1378	0.0039	0.0328	0.1286	-0.0487	0.0260	0.1183	0.4959	0.5865	0.7710	0.4711	0.5810	0.7506
(100,100)	0.0759	0.0820	0.0149	0.0244	0.0951	-0.0289	0.0237	0.0772	0.2882	0.4516	0.5391	0.3133	0.4737	0.5705
(100,200)	0.0429	0.0477	0.0211	0.0205	0.0587	-0.0189	0.0190	0.0603	0.1893	0.3473	0.3978	0.1618	0.3355	0.3752
(200,50)	0.1257	0.1307	0.0058	0.0217	0.1227	-0.0491	0.0178	0.1203	0.4977	0.5904	0.7750	0.4725	0.5693	0.7423
(200,100)	0.0802	0.0845	0.0138	0.0185	0.0914	-0.0287	0.0156	0.0726	0.2904	0.4475	0.5366	0.3136	0.4728	0.5695
(200,200)	0.0417	0.0443	0.0206	0.0140	0.0560	-0.0191	0.0128	0.0580	0.1881	0.3452	0.3955	0.1610	0.3416	0.3803
(50,50)	0.1232	0.1386	0.0062	0.0441	0.1232	-0.0515	0.0355	0.1300	0.4992	0.6095	0.7900	0.4933	0.5744	0.7597
(100,50)	0.1245	0.1378	0.0039	0.0328	0.1286	-0.0487	0.0260	0.1183	0.4959	0.5865	0.7710	0.4711	0.5810	0.7506
(200,50)	0.1257	0.1307	0.0058	0.0217	0.1227	-0.0491	0.0178	0.1203	0.4977	0.5904	0.7750	0.4725	0.5693	0.7423
(50,100)	0.0778	0.0894	0.0125	0.0362	0.1013	-0.0272	0.0305	0.0720	0.2924	0.4445	0.5352	0.3066	0.4778	0.5695
(100,100)	0.0759	0.0820	0.0149	0.0244	0.0951	-0.0289	0.0237	0.0772	0.2882	0.4516	0.5391	0.3133	0.4737	0.5705
(200,100)	0.0802	0.0845	0.0138	0.0185	0.0914	-0.0287	0.0156	0.0726	0.2904	0.4475	0.5366	0.3136	0.4728	0.5695
(50,200)	0.0447	0.0557	0.0189	0.0263	0.0583	-0.0192	0.0244	0.0612	0.1829	0.3420	0.3901	0.1670	0.3392	0.3806
(100,200)	0.0429	0.0477	0.0211	0.0205	0.0587	-0.0189	0.0190	0.0603	0.1893	0.3473	0.3978	0.1618	0.3355	0.3752
(200,200)	0.0417	0.0443	0.0206	0.0140	0.0560	-0.0191	0.0128	0.0580	0.1881	0.3452	0.3955	0.1610	0.3416	0.3803

5.2 Finite sample evaluation of specification testing

Next, we examine the performance of the test statistic l_{NT} for specification testing of loading functions. We consider the following two examples that correspond to Examples 5.1 and 5.2, respectively.

Example 5.3: Linear model

We consider a model with the form of

$$Y_{it} = X_{it}'\beta + \lambda(V_i)'f_t + e_{it},$$

where X_{it} are generated from a multivariate standard normal distribution $X_{it} \sim N(0, I_p)$, $p = 3$ is the dimension of X_{it} and $e_{it} \sim N(0,1)$. Here we set $\beta = (0.4, 0.5, 0.3)'$. Assume that we have two common factors, $f_{t\ell}$ for $\ell = 1,2$ and we generate $f_{t\ell}$ from Uniform$(2,3)$. The corresponding factor loadings are $\lambda(V) = (\lambda_1(V_1), \lambda_2(V_2))'$ with $V_\ell \sim$ Uniform$(0,1)$. We use the same moment conditions as Example 5.1 to estimate the unknown quantities.

In order to examine the size and power of the proposed statistic l_{NT} for specification of loading functions, we consider the following situations. For each situation, we calculate the percentage (or rejecting frequency) for which the proposed test statistic l_{NT} rejected the corresponding null hypothesis at 1%, 5% and 10% nominal levels among 100 Monte Carlo simulations.

In Situation I, the null hypothesis is $H_0 : \lambda_1(v) = \lambda_2(v) = 0$, and the alternative is $H_1 : \lambda_1(v) = \lambda_2(v) = \tau\sqrt{\log(NT)/\sqrt{NT}}\frac{1}{1+v^2}$ with τ taking values of $0.04, 0.05$ and 0.1, respectively.

In Situation II, the null hypothesis is $H_0 : \lambda_1(v) = \lambda_2(v) = \theta_0\cos(\pi v)$ with $\theta_0 = \sqrt{2}$; and the alternative is $H_1 : \lambda_1(v) = \lambda_2(v) = \theta_1\cos(\pi v) + \tau\sqrt{\log(NT)/\sqrt{NT}}\frac{1}{1+v^2}$ with $\theta_1 = \sqrt{2}$ and τ taking values of $0.04, 0.05$ and 0.1, respectively.

The sizes and power values for Situation I are reported in Table 3. From the first three columns containing the size values, it is readily seen that almost all the sizes fluctuate reasonably around the given significance levels. Overall, the actual sizes are quite close to the nominal sizes. From the rest nine columns in the right, we can see that although the local departure function $\Delta_{NT}(v) = \delta_{NT}\Delta(v)$ is asymptotically negligible as δ_{NT} approaches zero, the power values are quite satisfactory. It is clear that in most cases the power increases as the sample size increases. Also as we expected, with the increase of τ, the power increases rapidly. The sizes and power values for Situation II are reported in Table 4. The observations are similar as those in Table 3.

Example 5.4 : Nonlinear model

In this example, we consider the same binary panel data model as Example 5.2, that is,

$$P(Y_{it} = 1|X_{it}, V_i, f_t) = F(X_{it}'\beta + \lambda(V_i)'f_t),\ i = 1, \cdots, N,\ t = 1, \cdots, T. \tag{11}$$

Table 3: Rejecting frequency in Situation I of Example 5.3 for testing the specification of loading functions at 1%, 5% and 10% nominal levels.

	Size			Power								
				$\tau = 0.04$			$\tau = 0.05$			$\tau = 0.1$		
(N,T)	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
(50, 50)	0.10	0.07	0.01	0.48	0.31	0.13	0.55	0.48	0.24	0.99	0.98	0.85
(50, 100)	0.11	0.09	0.00	0.53	0.46	0.23	0.72	0.55	0.42	1.00	1.00	0.96
(50, 200)	0.07	0.03	0.00	0.78	0.71	0.45	0.91	0.85	0.68	1.00	1.00	1.00
(100, 50)	0.08	0.04	0.01	0.55	0.39	0.15	0.71	0.59	0.34	1.00	0.99	0.98
(100, 100)	0.11	0.03	0.01	0.70	0.61	0.41	0.84	0.78	0.59	1.00	1.00	1.00
(100, 200)	0.12	0.03	0.00	0.84	0.76	0.51	0.97	0.93	0.78	1.00	1.00	1.00
(200, 50)	0.11	0.05	0.02	0.71	0.52	0.32	0.86	0.72	0.47	1.00	1.00	0.99
(200, 100)	0.08	0.06	0.02	0.73	0.58	0.37	0.93	0.84	0.55	1.00	1.00	1.00
(200, 200)	0.11	0.03	0.00	0.91	0.85	0.61	0.98	0.96	0.86	1.00	1.00	1.00

Table 4: Rejecting frequency in Situation II of Example 5.3 for testing the specification of loading functions at 1%, 5% and 10% nominal levels.

	Size			Power								
				$\tau = 0.04$			$\tau = 0.05$			$\tau = 0.1$		
(N,T)	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
(50,50)	0.07	0.02	0.01	0.36	0.23	0.10	0.55	0.38	0.16	0.99	0.96	0.86
(50,100)	0.05	0.02	0.00	0.60	0.41	0.18	0.71	0.66	0.32	1.00	1.00	0.99
(50,200)	0.03	0.00	0.00	0.75	0.62	0.37	0.88	0.85	0.57	1.00	1.00	1.00
(100,50)	0.03	0.01	0.00	0.54	0.43	0.17	0.71	0.59	0.37	1.00	1.00	0.99
(100,100)	0.08	0.03	0.00	0.77	0.65	0.42	0.91	0.85	0.63	1.00	1.00	1.00
(100,200)	0.02	0.01	0.00	0.86	0.76	0.56	1.00	0.97	0.78	1.00	1.00	1.00
(200,50)	0.09	0.04	0.00	0.78	0.63	0.30	0.92	0.84	0.56	1.00	1.00	1.00
(200,100)	0.07	0.05	0.00	0.87	0.74	0.47	0.95	0.92	0.75	1.00	1.00	1.00
(200,200)	0.07	0.01	0.00	0.96	0.91	0.78	1.00	1.00	0.94	1.00	1.00	1.00

We assume that regressors X_{it} are generated from a standard multivariate normal distribution, that is, $X_{it} \sim N(0, I_p)$. Here we set $\beta = (0.4, 0.5, 0.3)'$. Assume that we have two common factors $f_{t\ell}$ for $\ell = 1,2$ and we generate $f_{t\ell}$ from Uniform(0.5, 1.5). The corresponding factor loadings are $\lambda(V) = (\lambda_1(V_1), \lambda_2(V_2))'$ with $V_i \sim i.i.d.N(0_r, I_r)$ and $r = 2$. We estimate the unknown quantities based on the same set of moment conditions as those in Example 5.2. In order to examine the size and power of the proposed statistic l_{NT} for specification of loading functions, we consider the following situations. For each situation, we calculate the percentage (or rejecting frequency) for which the proposed test statistic l_{NT} rejected the corresponding null hypothesis at 1%, 5% and 10% nominal levels among 100 Monte Carlo simulations.

In Situation I, the null hypothesis is $H_0 : \lambda_1(v) = \lambda_2(v) = 0$, and the alternative is $H_1 : \lambda_1(v) = \lambda_2(v) = \tau \sqrt{\log(NT)/\sqrt{NT}} \frac{1}{1+v^2}$ with τ taking values of $0.15, 0.20$ and 0.25, respectively.

In Situation II, the null hypothesis is $H_0 : \lambda_1(v) = \lambda_2(v) = \theta_0 v \exp(-v^2/2)$ with $\theta_0 = \sqrt{2}\pi^{-1/4}$; and the alternative is $H_1 : \lambda_1(v) = \lambda_2(v) = \theta_1 v \exp(-v^2/2) + \tau \sqrt{\log(NT)/\sqrt{NT}} \frac{1}{1+v^2}$ with $\theta_1 = \sqrt{2}\pi^{-1/4}$ and τ taking values of $0.15, 0.20$ and 0.25, respectively.

The sizes and power values for Situation I are reported in Table 5, while the corresponding results for Situation II are presented in Table 6. We can find that for both situations, almost all the sizes fluctuate reasonably around the given significance levels. Overall, the power increases as the sample size increases. Also, consistent with our expectation, the power becomes larger when τ gets larger, which indicates that power increases with the increase of local departures.

Table 5: Rejecting frequency in Situation I of Example 5.4 for testing the specification of loading functions at 1%, 5% and 10% nominal levels.

	Size			Power								
				$\tau = 0.15$			$\tau = 0.20$			$\tau = 0.25$		
(N,T)	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
(50,50)	0.09	0.08	0.03	0.52	0.38	0.14	0.71	0.64	0.38	0.82	0.77	0.57
(50,100)	0.06	0.03	0.00	0.60	0.49	0.22	0.82	0.73	0.51	0.95	0.94	0.74
(50,200)	0.04	0.03	0.01	0.87	0.80	0.58	0.98	0.96	0.88	1.00	1.00	0.98
(100,50)	0.10	0.03	0.00	0.69	0.56	0.34	0.95	0.84	0.61	0.98	0.95	0.92
(100,100)	0.11	0.03	0.00	0.87	0.80	0.49	0.99	0.95	0.83	1.00	0.99	0.99
(100,200)	0.06	0.02	0.00	0.96	0.91	0.76	1.00	0.99	0.96	1.00	1.00	1.00
(200,50)	0.08	0.05	0.02	0.89	0.84	0.63	0.97	0.95	0.89	1.00	1.00	0.98
(200,100)	0.14	0.06	0.01	0.94	0.89	0.82	1.00	0.99	0.95	1.00	1.00	1.00
(200,200)	0.08	0.04	0.01	0.99	0.99	0.95	1.00	1.00	1.00	1.00	1.00	1.00

Table 6: Rejecting frequency in Situation II of Example 5.4 for testing the specification of loading functions at 1%, 5% and 10% nominal levels.

	Size			Power								
				$\tau = 0.15$			$\tau = 0.20$			$\tau = 0.25$		
(N,T)	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
(50,50)	0.15	0.06	0.01	0.40	0.34	0.10	0.66	0.52	0.29	0.79	0.68	0.47
(50,100)	0.12	0.04	0.02	0.53	0.43	0.13	0.78	0.64	0.40	0.91	0.84	0.61
(50,200)	0.37	0.13	0.01	0.80	0.64	0.41	0.97	0.86	0.66	1.00	1.00	0.88
(100,50)	0.07	0.05	0.00	0.72	0.57	0.33	0.86	0.82	0.61	0.99	0.95	0.82
(100,100)	0.18	0.07	0.01	0.80	0.72	0.50	0.94	0.88	0.77	1.00	0.99	0.94
(100,200)	0.18	0.11	0.02	0.95	0.90	0.75	1.00	0.98	0.92	1.00	1.00	1.00
(200,50)	0.09	0.04	0.00	0.84	0.72	0.44	0.97	0.95	0.82	1.00	0.99	0.98
(200,100)	0.15	0.07	0.00	0.93	0.87	0.68	0.98	0.96	0.94	1.00	1.00	0.99
(200,200)	0.15	0.08	0.01	0.99	0.97	0.90	1.00	1.00	0.99	1.00	1.00	1.00

6 Empirical study on stock return prediction

In this section, we apply our proposed sieve-based GMM estimation method to predict excess stock return in U.S. In what follows, we first describe the data and then present the in-sample and out-of-sample results.

6.1 Data and model

During the last three decades, a lot of research[1] has been done on studying the relationship between individual stock returns and security characteristics since the seminal studies of Fama & French (1992,

[1]We acknowledge that there is an emerging line of research using machine learning methods in asset pricing. To mention a few, Fan et al. (2022) develop new structural nonparametric methods for estimating conditional asset pricing models using deep neural networks. Gu et al. (2020) employ a comprehensive set of machine learning tools to predict monthly individual stock returns using firm specific and common predictors. Chen et al. (2019) apply deep neural networks to estimate a nonlinear asset pricing model for U.S. equity data. More relevant studies can also be found in Rossi (2018), Gu et al. (2021), Feng et al. (2018), and Leippold et al. (2021). Here we do not attempt to beat those associated with machine learning methods in stock return prediction. Instead, we show the empirical applicability and relevance of our proposed sieve-based GMM procedure in real data.

1993), who first use size and value factors to model excess stock returns. In addition to the Fama-French size and value characteristics, Connor et al. (2012) and following-up studies (see Fan et al. 2016, Ma et al. 2021, for example) found that momentum and own-volatility characteristics are at least as important as size and value in explaining equity return co-movements. Therefore, in this study, we consider four characteristic variables, which are size, value, momentum and own-volatility. We use the same method as described in Section 5.1 in Connor et al. (2012) to construct size, value, momentum and own-volatility and following their paper, all four characteristics are standardized each month to have zero mean and unit variance. Meanwhile, we follow Ando & Bai (2017) to include lagged returns as regressors.

The data used in this section are from the Center for Research in Security Prices (CRSP), which includes the complete monthly return records for 164 non-financial S&P 500 companies from 2000 to 2018. Throughout the empirical analysis, we use returns in excess of the risk-free return, treating the monthly Treasury bill return from CRSP as the risk-free return.

Suppose that we have the following linear regression model

$$Y_{it} = X_{it}'\beta + \lambda(V_i)'f_t + e_{it}, \tag{12}$$

where Y_{it} denotes the excess return, $X_{it} = (Y_{i,t-1}, \cdots, Y_{i,t-5})'$, $V_i = (V_{i1}, \cdots, V_{i4})'$ denotes the time series average of size, value, momentum and own-volatility characteristics. The loading function $\lambda(V_i) = (\lambda_1(V_{i1}), \cdots, \lambda_4(V_{i4}))'$ and factors f_t are unknown. The set up of factors and loading functions $\lambda(V_i)'f_t$ follows Connor et al. (2012). Here we assume that the loading function $\lambda_\ell(v) \in L^2(\mathbb{R})$, for $1 \leq \ell \leq 4$. Denote $\Phi_k(v) = (p_0(v), \cdots, p_{k-1}(v))'$, where $\{p_j(v), j \geq 0\}$ be the sequence of Hermite functions that forms an orthonormal basis in $L^2(\mathbb{R})$, which can be used to approximate the unknown loading functions.

It is easy to see that (β, λ, f) can be determined by a conditional moment restriction that $\mathbb{E}[e_{it}|V_i] = 0$. By Example 1.1 in the introduction section and Example 5.1 in the simulation study, we know that this conditional moment restriction implies that

$$\mathbb{E}[m(Y_{it}, X_{it}'\beta, \lambda'(V_i)f_t)] = 0, \tag{13}$$

where $m(Y_{it}, X_{it}'\beta, \lambda(V_i)'f_t) = (Y_{it} - X_{it}'\beta - \lambda(V_i)'f_t)\Psi_q(V_i)$ and $\Psi_q(V_i)$ is a basis vector of multivariate functions whose combinations can approximate any square integrable functions of V in some sense arbitrarily as $q \to \infty$. Precisely, as V_i is a 4-dimensional vector, we construct $\Psi_q(V_i)$ by the tensor product $\{p_{j_1}(V_{i1})p_{j_2}(V_{i2})p_{j_3}(V_{i3})p_{j_4}(V_{i4})\}$ for $j_1, j_2, j_3, j_4 = 0, \cdots, q_0$, which is an orthogonal basis system to expand any square integrable functions of V. In this study, we simply set $q_0 = 2$ and consequently $q = (q_0 + 1)^4 = 81$, which is a large number of moment conditions. Note that the choice of q_0 satisfies the theoretical requirement in preceding sections and it works satisfactorily well as shown by the in-sample and out-of-sample forecasting results below. Therefore, by the estimation procedure in Section

2.2, (β, λ, f) can be estimated from $\mathbb{E}[m(Y_{it}, X_{it}'\beta, \lambda'(V_i)f_t)] = 0$ for $i = 1, \cdots, N$ and $t = 1, \cdots, T$. Specifically, for each given t, we first obtain estimates of (β, D_t) by minimizing the objective function

$$M_{Nt}(b, d_t) = \frac{1}{\sqrt{q}} \frac{1}{N} \sum_{i=1}^{N} m(Y_{it}, X_{it}'b, \Phi_k^r(V_i)'d_t),$$

Then by our proposed identification conditions that loading functions have unit norm and factors are assumed to be positive, we can get $\widehat{\alpha}$ and $\widehat{f}_t$ separately from $\widehat{D}_t$. Further we regard $\widehat{\overline{\beta}}$, $\widehat{\lambda}(V_i) = \widetilde{\Phi}_k(V_i)'\widehat{\alpha}$ and $\widehat{f}_t$ as the estimates of $\beta, \lambda(\cdot)$ and f_t, respectively.

For comparison purposes, we consider the following models:

- Traditional linear regression given by

$$Y_{it} = X_{it}'\beta_{ols} + e_{it}, \tag{14}$$

  where we use ordinary least squares estimation method to estimate unknown parameters β_{ols}.

- Fama-French three factor models[2]

$$Y_{it} = \alpha_i + \beta_{1i} MKT_t + \beta_{2i} SMB_t + \beta_{3i} HML_t + e_{it}, \tag{15}$$

  where MKT_t, SMB_t, HML_t denote the Fama-French three factors, which are the Market excess return (MKT) factor, the Small-Minus-Big (SMB) size factor and the High-Minus-Low (HML) value factor at time t, respectively.

We will evaluate the in-sample and out-of-sample performance of our method and these two competitors below.

6.2 In-sample estimation and specification testing

In this section, we use the whole sample covering 2000:01-2018:12 to evaluate the in-sample performance of (13), (14) and (15) in terms of mean squared errors given by

$$\text{MSE} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it} - \widehat{y}_{it})^2,$$

where y_{it} is the observed excess return for stock i at time t and $\widehat{y}_{it}$ is the corresponding estimate.

Before presenting the main results, a key issue for the implementation of the estimation procedure is the determination of the truncation parameters in the orthogonal expansions. However, in practice,

[2]We also consider Fama-French-Carhart four factor model and the results are qualitatively similar. We omit the results for brevity.

there is no universal guide for the choice of such parameters. In this study, we shall choose the truncation parameters by minimizing the commonly used leave-one-unit-out mean squared errors (MSE) criterion defined below over a candidate set $[2, 8]$.

$$\widetilde{\mathrm{MSE}}(k) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{(N-1)T}\sum_{j=1, j\neq i}^{N}\sum_{t=1}^{T}(y_{jt} - \widehat{y}_{jt}^{(-i)}(k))^2,$$

where $\widehat{y}_{it}^{(-i)}(k)$ is the prediction for y_{it} without using the information of stock i and with truncation parameter being k. The results of $\widetilde{\mathrm{MSE}}(k)$ are summarized in the following table, from which we can find that the best choice is $\widehat{k} = 4$.

Table 7: The results of $\widetilde{\mathrm{MSE}}(k)$ when k is over the set $[2, 8]$.

k	2	3	4	5	6	7	8
$\widetilde{\mathrm{MSE}}(\times 100)$	0.5342	0.5264	0.5177	0.5262	0.5715	1.4380	1.5739

We further obtain the mean squared errors of models (13), (14) and (15), which are 0.0048, 0.0070 and 0.0050, respectively. We can see that the performance of (13) and (15) are similar, and both models outperform the traditional linear regression (14). For example, by using our proposed sieve-based GMM method, the accuracy has increased by almost 31% (i.e. $(0.0070-0.0048)/0.0070$) compared with (14) in terms of the mean squared errors. The parameter estimates and the corresponding standard errors are reported in Table 8,[3] in which we find that the parameter estimates in models (13) and (14) are quite different. This suggests that including loading functions and unobserved factors in the model is very influential.

Table 8: Parameter estimates in models (13) and (14). The corresponding standard errors are shown in parentheses.

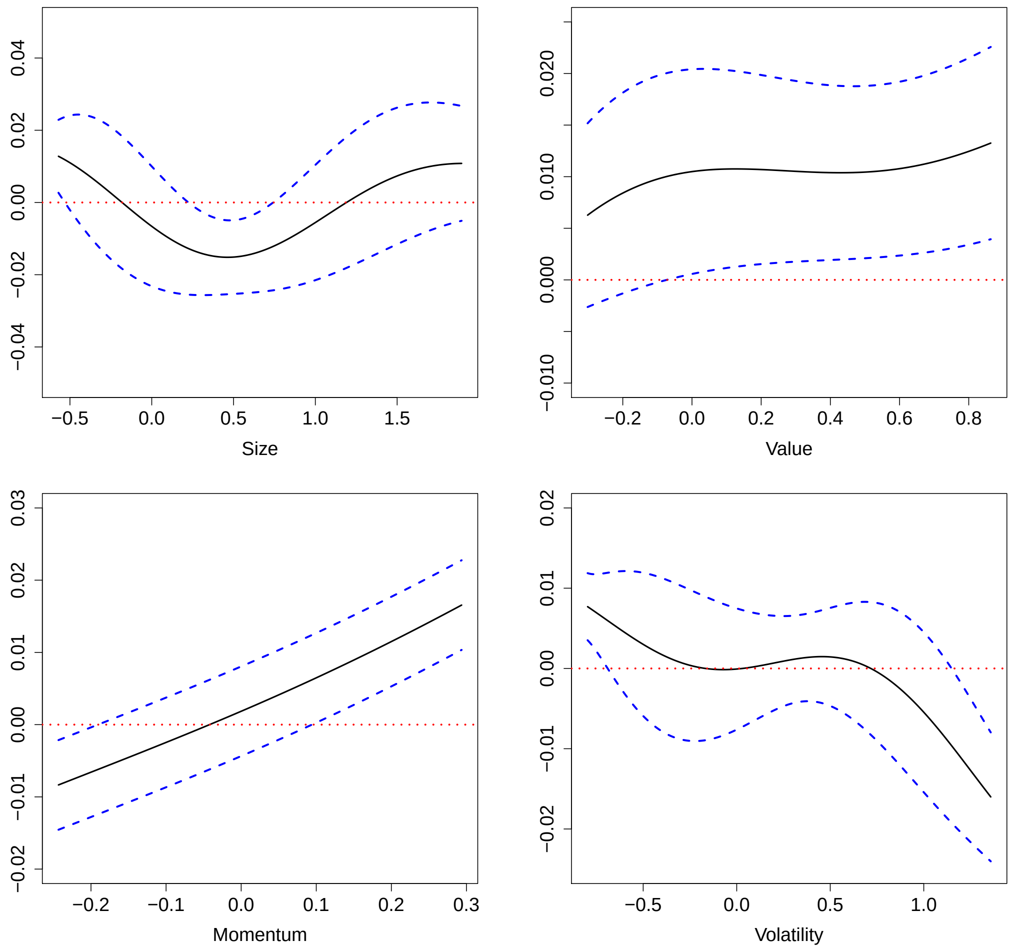
Parameter estimates $\widehat{\beta}$ in (13)				
−0.0193	−0.0092	−0.0215	−0.0125	−0.0027
(0.0013)	(0.0012)	(0.0009)	(0.0014)	(0.0014)
Parameter estimates $\widehat{\beta}_{ols}$ in (14)				
0.0079	−0.0269	0.0290	0.0349	0.0039
(0.0052)	(0.0052)	(0.0052)	(0.0051)	(0.0051)

We plot the estimated loading functions $\overline{\widehat{\lambda}}_{\ell}(v)$, $1 \leq \ell \leq 4$, along with the 95% confidence intervals

[3] Since the parameter estimates in Fama-French three factor model varies with stocks, so we did not report the results here.

in Figure 1, where the intervals are constructed based on the asymptotic variance in assertion 5 in Theorem 3.2 and we replace the unknown quantities by corresponding estimators. We can see that the loading functions or characteristic-beta functions of size and volatility show strong nonlinear pattern while the characteristic-beta functions of value and momentum are close to linear. This illustrates the advantage of the semiparametric approach we adopted.

Figure 1: The estimated loading functions and 95% confidence intervals. In each plot, the solid line displays the estimated loading function and the dashed lines display the corresponding 95% confidence interval.



Moreover, we conduct the following tests for the specification of loading functions $\lambda(v)$ using the test statistic $l_{NT}(\overline{\widehat{\beta}}, \widehat{\theta}, \widehat{f}_t; c)$ in Section 4.2. We first test whether all the four loading functions are jointly zero, that is, we consider the null hypothesis

$$H_0 : \lambda_1(v) = \lambda_2(v) = \lambda_3(v) = \lambda_4(v) = 0, \text{ for } v \in \mathbb{V}.$$

Under H_0, our model (12) becomes a fully parametric model and the unknown parameter β can be estimated by OLS. Then based on moment conditions (13), we obtain the test statistic, which is 5.4080.

This indicates that at least one of the loading functions is significantly different with zero under 5% level of significance.

Further, we consider the following four tests to separately check whether a given loading function is zero.

$$H_{0\ell} : \lambda_\ell(v) = 0, \text{ for } v \in \mathbb{V}, \text{ versus } H_{1\ell} : \lambda_\ell(v) \neq 0, \text{ for } v \in \mathbb{V},$$

where $\ell = 1, 2, 3, 4$. For illustration, suppose that $\ell = 1$, then under H_{01}, we have $\lambda_1(v) = 0$ and there is no restrictions on the rest three loading functions. We employ the estimation procedure outlined in Section 2.2 to estimate β, $\lambda_2(v)$, $\lambda_3(v)$, $\lambda_4(v)$ and the corresponding factors f_2, f_3 and f_4. Then based on moment conditions (13), we obtain the test statistic 2.3497. Similarly, following the same procedure, under $H_{0\ell}$, $\ell = 2, 3, 4$, the test statistics are respectively, 2.1126, 2.0251 and 1.9941. The results suggest that all the loading functions are significantly different with zero under 5% level of significance.

6.3 Out-of-sample prediction performance

In this study, we employ an expansive window scheme to conduct the out-of-sample evaluation of (13), (14) and (15). The estimation sample always starts from the first observation and additional observations are used as they become available. Specifically, for the first window, we estimate model based on the observations $\{y_{it}\}_{i=1}^{N}{}_{t=1}^{n}$. At the point x_n, we predict $\{y_{i,n+1}\}_{i=1}^{N}$ with the estimated value denoted as $\{\widehat{y}_{i,n+1}\}_{i=1}^{N}$. Then we expand the first window to include observations $\{y_{it}\}_{i=1}^{N}{}_{t=1}^{n+1}$ to predict $\{y_{i,n+2}\}_{i=1}^{N}$. Repeat the above procedure until the forecast for all the stocks at the last period (2018 December) is made. As the selection of out-of-sample period is always somewhat arbitrary, in this study, we consider three out-of-sample periods,

- **Period 1:** 2005 June - 2018 December (a long out-of-sample period)
- **Period 2:** 2008 February - 2009 March (Global Financial Crisis).
- **Period 3:** 2010 June - 2018 December (a more recent out-of-sample period).

Regarding the choice of truncation parameters, we use the same criterion (i.e. leave-one-unit-out mse) as that used in the above in-sample estimation. One may select the truncation parameters for each expanding sample (because it may differ across different periods) to gain higher forecast accuracy, but will be at the cost of a significant increase in computation time. We then evaluate the out-of-sample performance of (13), (14) and (15) by computing the out-of-sample mean squared forecasting errors defined as

$$\text{MSFE} = \frac{1}{NR} \sum_{r=1}^{R} \sum_{i=1}^{N} (y_{i,n+r} - \widehat{y}_{i,n+r})^2$$

where $\widehat{y}_{i,n+r}$ is the predicted return for stock i in the r-th window, $y_{i,n+r}$ is the corresponding observed return, n is the initial window size and R is the total number of expansive windows.

Let MSFE_{gmm}, MSFE_{ols} and MSFE_{FF3} denote the mean squared forecasting errors produced by (13), (14) and (15), respectively. The results are presented in Table 9. Consistent with the in-sample estimation results, models (13) and (15) have similar performance while in all three out-of-sample periods, the mean squared forecasting errors produced by those two models are much smaller than model (14) by ordinary least squares method.

To show the advantages of our model under this particular structure, we further compute the relative accuracy of our method against model (14), which we call "$\text{Efficiency}_{gmm/ols}$" as follows.

$$\text{Efficiency}_{gmm/ols} = \frac{\text{MSFE}_{ols} - \text{MSFE}_{gmm}}{\text{MSFE}_{ols}}.$$

Table 9: Prediction results for three out-of-sample periods

OOS	MSFE_{gmm}	MSFE_{ols}	MSFE_{FF3}	$\text{Efficiency}_{gmm/ols}$
Period 1	0.0057	0.0060	0.0057	0.0483
Period 2	0.0148	0.0157	0.0127	0.0604
Period 3	0.0038	0.0045	0.0042	0.1569

7 Conclusions

In this paper, we have considered a class of high dimensional moment restriction panel data models with interactive effects, where the dimension of parameter vector and the number of moment conditions are diverging with sample size. We assume that the common factors are unobserved and the factor loadings are unknown smooth functions of individual characteristic variables. Such model framework is very general and includes many existing linear and nonlinear panel data models as special cases, such as the linear panel data model with interactive effect and binary panel data models. We have proposed a sieve-based generalized method of moments estimation to estimate the unknown parameters, factors and loading functions. In addition, we have established asymptotic theory for the proposed estimators.

Moreover, we have proposed two test statistics for the over-identification, specification of loading functions and established their large sample properties, respectively. Our simulation results further showed that the proposed estimation methods and test procedures perform very well in finite samples. Finally, we have demonstrated the advantages of the proposed method by applying it to forecast stock return prediction.

Acknowledgement

Tingting Cheng was supported financially by the National Natural Science Foundation of China under Grant No. 72173068 and 71803091, Social Science Development Fund of Nankai University (ZB22BZ0321) and the Fundamental Research Funds for the Central Universities (ZB22000101). Chaohua Dong thanks the financial support from National Natural Science Foundation of China under Grant No. 72073143 and 71671143. Jiti Gao was supported by the Australian Research Council Discovery Grants Program for its support under Grant number: DP170104421.

Appendix A: Lemmas

This section provides all technical lemmas and some notation used for the theoretical derivations, while the proofs of these lemmas are postponed to the supplementary material of the paper.

Lemma A.1. Under Assumptions 2.1, 2.2, 3.1-3.3, for each given t, we have
(1) $\|M_{Nt}(\beta, D_t)\|^2 = O_P(\|\gamma^{(k)}(v)\|^2) + O_P(N^{-1})$;
(2) Given $B_{1N}^2 + B_{2N}^2 = o(N)$, $\sup_{\|b\|\le B_{1N}, \|d_t\|\le B_{2N}, \|(b-\beta, d_t - D_t)\|>\delta} \|M_{Nt}(b, d_t)\|^{-2} = O_P(1)$ for each $\delta > 0$, when N is large.

Denote $m(v,u,w) = (m_1(v,u,w), \cdots, m_q(v,u,w))'$. To investigate the limiting distributions, we denote the Score and Hessian functions as follows.

$$S_{Nt}(b,d_t) = \begin{pmatrix} S_{1Nt}(b,d_t) \\ S_{2Nt}(b,d_t) \end{pmatrix} := \begin{pmatrix} \frac{\partial}{\partial b} \\ \frac{\partial}{\partial d_t} \end{pmatrix} \|M_{Nt}(b,d_t)\|^2,$$

$$H_{Nt}(b,d_t) = \begin{pmatrix} H_{11}(b,d_t) \; H_{12}(b,d_t) \\ H_{21}(b,d_t) \; H_{22}(b,d_t) \end{pmatrix} := \begin{pmatrix} \frac{\partial^2}{\partial b \partial b'} & \frac{\partial^2}{\partial b \partial d_t'} \\ \frac{\partial^2}{\partial d_t \partial b'} & \frac{\partial^2}{\partial d_t \partial d_t'} \end{pmatrix} \|M_{Nt}(b,d_t)\|^2.$$

Denote $h_{Nt}(\beta, \lambda' f_t) = \frac{1}{q}\Psi_{Nt}\Psi_{Nt}'$ and $s_{Nt}(\beta, \lambda' f_t) = \frac{1}{q}\Psi_{Nt}\frac{1}{N}\sum_{i=1}^N m(W_{it}, X_{it}'\beta, \lambda(V_i)' f_t)$, where

$$\Psi_{Nt} = \mathbb{E}\begin{pmatrix} \frac{\partial}{\partial u} m(W_{it}, X_{it}'\beta, \lambda(V_i)' f_t)' \otimes X_{it} \\ \frac{\partial}{\partial w} m(W_{it}, X_{it}'\beta, \lambda(V_i)' f_t)' \otimes \Phi_k^r(V_i) \end{pmatrix}_{(p+kr)\times q}.$$

Lemma A.2. Under Assumption 2.1, 2.2, 3.1, 3.3 and 3.5-3.7, (1) $H_{Nt}(\beta, D_t)$ is asymptotically almost surely positive definite; (2) as $N \to \infty$, we have $\|H_{Nt}(\beta, D_t) - h_{Nt}(\beta, \lambda' f_t)\| = o_P(1)$.

Lemma A.3. Under Assumptions 2.1, 2.2, 3.1,3.3, 3.5-3.7, as $N \to \infty$, we have

$$\|S_{Nt}(\beta, D_t) - s_{Nt}(\beta, \lambda' f_t)\| = o_P(1).$$

Appendix B: Proofs of the main results

Proof of Theorem 3.1. By Lemma A.1, we have shown that

$$\|M_{Nt}(\beta, D_t)\|^2 = o_P(1) \tag{16}$$

$$\sup_{\|b\|\le B_{1N}, \|d_t\|\le B_{2N}, \|b-\beta, d_t - D_t\|>\delta} \|M_{Nt}(b, d_t)\|^{-2} = O_P(1), \text{ for each } \delta > 0 \tag{17}$$

Fix $\epsilon > 0$ and $\delta > 0$. Equation (17) means that there exists a large but fixed M such that

$$\limsup P\left(\sup_{\|b\|\le B_{1N}, \|d_t\|\le B_{2N}, \|b-\beta, d_t - D_t\|>\delta} \|M_{Nt}(b, d_t)\|^{-2} > M\right) < \epsilon,$$

and equation (16) implies that $\|M_{Nt}(\widehat{\beta}_t, \widehat{D}_t)\|^2 = \inf_{\|b\|\le B_{1N}, \|d_t\|\le B_{2N}} \|M_{Nt}(b, d_t)\|^2 \le \|M_{Nt}(\beta, D_t)\|^2 = o_P(1)$, which gives $P(\|M_{Nt}(\widehat{\beta}_t, \widehat{D}_t)\|^{-2} > M) \to 1$.

It follows that, with probability of at least $1 - 2\epsilon$ for N large enough,

$$\|M_{Nt}(\widehat{\beta}_t, \widehat{D}_t)\|^{-2} > M \ge \sup_{\|b\|\le B_{1N}, \|d_t\|\le B_{2N}, \|b-\beta, d_t - D_t\|>\delta} \|M_{Nt}(b, d_t)\|^{-2}$$

Therefore, the event $(\widehat{\beta}_t, \widehat{D}_t) \in \{(b, d_t) : \|b\| \le B_{1N}, \|d_t\| \le B_{2N}, \|b-\beta, d_t - D_t\| > \delta\}$ holds with probability at most 2ϵ, $\limsup P(\|(\widehat{\beta}_t - \beta, \widehat{D}_t - D_t)\| > \delta) < 2\epsilon$. As ϵ and δ are arbitrarily chosen, we have $\|(\widehat{\beta}_t - \beta, \widehat{D}_t - D_t)\| \to_P 0$. Since $\widehat{\beta}_t - \beta = o_P(1)$ for any given t, it is straightforward to get that

$$\left\|\overline{\widehat{\beta}} - \beta\right\| = \left\|\frac{1}{T}\sum_{t=1}^{T}\widehat{\beta}_t - \beta\right\| = \frac{1}{T}\sum_{t=1}^{T}\left\|\widehat{\beta}_t - \beta\right\| \le \max_{1\le t\le T}\left\|\widehat{\beta}_t - \beta\right\| = o_P(1) \tag{18}$$

Meanwhile, we have, for $1 \le \ell \le r$,

$$|\widehat{f}_{t\ell}| - |f_{t\ell}| = \|S_\ell \widehat{D}_t\| - \|S_\ell D_t\| \le \|S_\ell(\widehat{D}_t - D_t)\| = o_P(1)$$

$$\widehat{\alpha}_{t\ell} - \alpha_{t\ell} = \frac{S_\ell \widehat{D}_t}{\|S_\ell \widehat{D}_t\|} - \frac{S_\ell D_t}{\|S_\ell D_t\|} = o_P(1).$$

Therefore, $\|\widehat{f}_t - f_t\|^2 \to_P 0$ and

$$\begin{aligned}\|\widehat{\lambda}_\ell(v) - \lambda_\ell(v)\|^2 &= \int \|\widehat{\lambda}_\ell(v) - \lambda_\ell(v)\|^2 \pi(v) dv = \int \|\Phi_k(v)'(\widehat{\alpha}_{t\ell} - \alpha_{t\ell}) + \gamma_\ell^{(k)}(v)\|^2 \pi(v) dv \\ &= \|\widehat{\alpha}_{t\ell} - \alpha_{t\ell}\|^2 + \|\gamma_\ell^{(k)}(v)\|^2 \to_P 0,\end{aligned}$$

as $N, k \to \infty$, by the orthogonality of the basis sequence. Thus it is straightforward to show that $\|\overline{\widehat{\lambda}}_\ell(v) - \lambda_\ell(v)\|^2 = o_P(1)$. The proof is finished.

Proof of Theorem 3.2.

(1) Write

$$\sqrt{N}\left(\mathscr{L}(\widehat{\beta}_t) - \mathscr{L}(\beta)\right)$$

$$
\begin{aligned}
&=\sqrt{N}\partial\mathscr{L}(\beta)'(\widehat{\beta}_t-\beta)\\
&=-\sqrt{N}\partial\mathscr{L}(\beta)'h_{11}(\beta,\lambda' f_t)^{-1}s_{1Nt}(\beta,\lambda' f_t)+o_P(1)\\
&=-\sqrt{N}\partial\mathscr{L}(\beta)'\left(\frac{1}{q}\Delta_x\Delta_x'\right)^{-1}\left(\frac{1}{q}\Delta_x\frac{1}{N}\sum_{i=1}^{N}m(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)\right)+o_P(1)\\
&=-\partial\mathscr{L}(\beta)'\left(\Delta_x\Delta_x'\right)^{-1}\Delta_x\frac{1}{\sqrt{N}}\sum_{i=1}^{N}m(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)+o_P(1)\\
&\xrightarrow{D}N\left(0,\Sigma_{\beta_0}\right)
\end{aligned}
$$

where the first equality is obtained by the linearity of Fréchet derivative and ignoring the higher order term; second equality is based on the first order Taylor expansion and ignoring higher order terms and further apply the results in Lemma A.2 and Lemma A.3; the last step follows from the standard central limit theorem and

$$
\begin{aligned}
\Sigma_{\beta_0}=&\partial\mathscr{L}(\beta)'\left(\Delta_x\Delta_x'\right)^{-1}\Delta_x\Xi_m\Delta_x'\left(\Delta_x\Delta_x'\right)^{-1}\partial\mathscr{L}(\beta)\\
&+\lim_N\frac{1}{N}\sum_{i\neq j}\partial\mathscr{L}(\beta)'\left(\Delta_x\Delta_x'\right)^{-1}\Delta_x\Xi_{ij,11}\Delta_x'\left(\Delta_x\Delta_x'\right)^{-1}\partial\mathscr{L}(\beta).
\end{aligned}
$$

(2) Following similar procedure, we can obtain the asymptotic distribution for $\overline{\widehat{\beta}}$ by

$$
\begin{aligned}
&\sqrt{NT}\left(\mathscr{L}(\overline{\widehat{\beta}})-\mathscr{L}(\beta)\right)\\
&=\sqrt{NT}\partial\mathscr{L}(\beta)'(\overline{\widehat{\beta}}-\beta)\\
&=-\sqrt{NT}\partial\mathscr{L}(\beta)'\frac{1}{T}\sum_{t=1}^{T}(\widehat{\beta}_t-\beta)\\
&=-\sqrt{NT}\partial\mathscr{L}(\beta)'\frac{1}{T}\sum_{t=1}^{T}h_{11}(\beta,\lambda' f_t)^{-1}s_{1Nt}(\beta,\lambda' f_t)+o_P(1)\\
&=-\sqrt{NT}\partial\mathscr{L}(\beta)'\left(\Delta_x\Delta_x'\right)^{-1}\Delta_x\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}m(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)+o_P(1)\\
&=-\partial\mathscr{L}(\beta)'\left(\Delta_x\Delta_x'\right)^{-1}\Delta_x\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\sum_{i=1}^{N}m(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)+o_P(1)\\
&\xrightarrow{D}N\left(0,\Sigma_{\beta}\right),
\end{aligned}
$$

where $\Sigma_\beta=\lim_{N,T}\frac{1}{NT}\sum_{i,j,t,s}\partial\mathscr{L}(\beta)'\left(\Delta_x\Delta_x'\right)^{-1}\Delta_x\Xi_{ij,ts}\Delta_x'\left(\Delta_x\Delta_x'\right)^{-1}\partial\mathscr{L}(\beta)$.

(3) We have

$$
\begin{aligned}
&\sqrt{N}\left(\widehat{f_{t\ell}}-f_{t\ell}\right)\\
&=\frac{\sqrt{N}}{\widehat{f_{t\ell}}+f_{t\ell}}\left(\widehat{f_{t\ell}}^{2}-f_{t\ell}^2\right)
\end{aligned}
$$

$$
\begin{aligned}
&=\frac{\sqrt{N}}{\widehat{f_{t\ell}}+f_{t\ell}}\left(||\widehat{D_{t\ell}}||^2-f_{t\ell}^2\right)\\
&=\frac{\sqrt{N}}{\widehat{f_{t\ell}}+f_{t\ell}}\left(||D_{t\ell}||^2+2D_{t\ell}'\left(\widehat{D_{t\ell}}-D_{t\ell}\right)-{f_{t\ell}}^2\right)+o_P(1)\\
&=\frac{\sqrt{N}}{\widehat{f_{t\ell}}+f_{t\ell}}2f_{t\ell}\alpha_\ell' S_\ell\left(\widehat{D_t}-D_t\right)+o_P(1)\\
&=\sqrt{N}\alpha_\ell' S_\ell\left(\widehat{D_t}-D_t\right)+o_P(1)\\
&=\alpha_\ell' S_\ell\left(\Delta_k\Delta_k'\right)^{-1}\Delta_k\frac{1}{\sqrt{N}}\sum_{i=1}^N m(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)+o_P(1)\\
&\xrightarrow{D} N\left(0,\Sigma_{f\ell}\right),
\end{aligned}
$$

where the last equality is obtained by using the first order Taylor expansion with ignoring higher order terms and applying the results in Lemma A.2 and Lemma A.3; the last step follows from the standard central limit theorem and

$$
\Sigma_{f\ell}=\alpha_\ell' S_\ell\left(\Delta_k\Delta_k'\right)^{-1}\Delta_k\Xi_m\Delta_k'\left(\Delta_k\Delta_k'\right)^{-1}S_\ell'\alpha_\ell+\lim_N\frac{1}{N}\sum_{i\neq j}\alpha_\ell' S_\ell\left(\Delta_k\Delta_k'\right)^{-1}\Delta_k\Xi_{ij,11}\Delta_k'\left(\Delta_k\Delta_k'\right)^{-1}S_\ell'\alpha_\ell.
$$

(4) Write

$$
\begin{aligned}
&\frac{\sqrt{N}}{||\Phi_k(V)||}(\widehat{\lambda}_{t\ell}(V)-\lambda_\ell(V))\\
=&\frac{\sqrt{N}}{||\Phi_k(V)||}\Phi_k'(V)(\widehat{\alpha}_{t\ell}-\alpha_\ell)+o_P(1)\\
=&\frac{\sqrt{N}}{||\Phi_k(V)||}\Phi_k'(V)\left(\frac{S_\ell\widehat{D_t}}{||S_\ell\widehat{D_t}||}-\frac{S_\ell D_t}{||S_\ell D_t||}\right)+o_P(1)\\
=&\frac{\sqrt{N}}{||\Phi_k(V)||}\Phi_k'(V)\left(\frac{S_\ell\widehat{D_t}}{||S_\ell\widehat{D_t}||}-\frac{S_\ell D_t}{||S_\ell\widehat{D_t}||}+\frac{S_\ell D_t}{||S_\ell\widehat{D_t}||}-\frac{S_\ell D_t}{||S_\ell D_t||}\right)+o_P(1)\\
=&\frac{\sqrt{N}}{||\Phi_k(V)||}\Phi_k'(V)\frac{1}{||S_\ell\widehat{D_t}||}S_\ell(\widehat{D_t}-D_t)+\frac{\sqrt{N}}{||\Phi_k(V)||}\Phi_k'(V)\frac{S_\ell D_t(||S_\ell D_t||-||S_\ell\widehat{D_t}||)}{||S_\ell\widehat{D_t}||||S_\ell D_t||}+o_P(1)\\
=&\frac{1}{||\Phi_k(V)||f_{t\ell}\sqrt{N}}\Phi_k'(V)S_\ell\left(\Delta_k\Delta_k'\right)^{-1}\Delta_k\sum_{i=1}^N m(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)+o_P(1)\\
&\xrightarrow{D} N\left(0,\Sigma_{\lambda,t\ell}\right)
\end{aligned}
$$

where the second equality we use the identification condition, the rest steps are obtained by similar procedure as those in (2) of this theorem and the covariance can be obtained by

$$
\Sigma_{\lambda,t\ell}=\frac{1}{||\Phi_k(V)||^2f_{t\ell}^2}\Phi_k'(V)S_\ell\left(\Delta_k\Delta_k'\right)^{-1}\Delta_k\Xi_m\Delta_k'\left(\Delta_k\Delta_k'\right)^{-1}S_\ell'\Phi_k(V)
$$

$$+\lim_{N}\frac{1}{N}\sum_{i\neq j}\frac{1}{||\Phi_k(V)||^2 f_{t\ell}^2}\Phi_k'(V)S_\ell\left(\Delta_k\Delta_k'\right)^{-1}\Delta_k\Xi_{ij,11}\Delta_k'\left(\Delta_k\Delta_k'\right)^{-1}S_\ell'\Phi_k(V).$$

(5) By similar procedure, we obtain the asymptotic distribution for $\overline{\overline{\lambda}}_\ell(V)=\Phi_k(V)\left(\frac{1}{T}\sum_{t=1}^T\widehat{\alpha}_{t\ell}\right)$.

$$\begin{aligned}
&\frac{\sqrt{NT}}{||\Phi_k(V)||}(\overline{\overline{\lambda}}_\ell(V)-\lambda_\ell(V))\\
=&\frac{\sqrt{NT}}{||\Phi_k(V)||}\Phi_k'(V)(\frac{1}{T}\sum_{t=1}^T\widehat{\alpha}_{t\ell}-\alpha_\ell)+o_P(1)\\
=&\frac{\sqrt{NT}}{||\Phi_k(V)||}\Phi_k'(V)\frac{1}{T}\sum_{t=1}^T\left(\frac{S_\ell\widehat{D_t}}{||S_\ell\widehat{D_t}||}-\frac{S_\ell D_t}{||S_\ell D_t||}\right)+o_P(1)\\
=&\frac{\sqrt{NT}}{||\Phi_k(V)||}\Phi_k'(V)\frac{1}{T}\sum_{t=1}^T\left(\frac{S_\ell\widehat{D_t}}{||S_\ell\widehat{D_t}||}-\frac{S_\ell D_t}{||S_\ell\widehat{D_t}||}+\frac{S_\ell D_t}{||S_\ell\widehat{D_t}||}-\frac{S_\ell D_t}{||S_\ell D_t||}\right)+o_P(1)\\
=&\frac{\sqrt{NT}}{||\Phi_k(V)||}\Phi_k'(V)\frac{1}{T}\sum_{t=1}^T\frac{1}{||S_\ell\widehat{D_t}||}S_\ell(\widehat{D_t}-D_t)+\frac{\sqrt{NT}}{||\Phi_k(V)||}\Phi_k'(V)\frac{1}{T}\sum_{t=1}^T\frac{S_\ell D_t(||S_\ell D_t||-||S_\ell\widehat{D_t}||)}{||S_\ell\widehat{D_t}||||S_\ell D_t||}+o_P(1)\\
=&\frac{1}{||\Phi_k(V)||}\Phi_k'(V)S_\ell\left(\Delta_k\Delta_k'\right)^{-1}\Delta_k\frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T\frac{1}{f_{t\ell}}m(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)+o_P(1)\\
\xrightarrow{D}&N\left(0,\Sigma_{\lambda\ell}\right),
\end{aligned}$$

where $\Sigma_{\lambda\ell}=\lim_{N,T}\frac{1}{NT}\sum_{i,j,t,s}\frac{1}{||\Phi_k(V)||^2 f_{t\ell}f_{s\ell}}\Phi_k'(V)S_\ell\left(\Delta_k\Delta_k'\right)^{-1}\Delta_k\Xi_{ij,ts}\Delta_k'\left(\Delta_k\Delta_k'\right)^{-1}S_\ell'\Phi_k(V)$.

The proof is now finished.

Proof of Corollary 3.1. Since the dimension of $\partial\mathscr{L}(\widehat{\beta})-\partial\mathscr{L}(\beta)$ is fixed, the elementwise consistency ensures the consistency of the vector. Observe that Assumption 3.8 stipulates the linearity of $\mathscr{L}(\cdot)$. Thus, it follows immediately from Theorem 3.2 that $\partial\mathscr{L}(\widehat{\beta})=\partial\mathscr{L}(\beta)+o_P(1)$.

Similarly, due to Assumptions 3.4-3.7, $\widehat{\Delta}_x=\Delta_x+o_P(1)$, $\widehat{\Delta}_k=\Delta_k+o_P(1)$, $\widehat{\Xi}_m=\Xi_m+o_P(1)$, $\widehat{\Xi}_{11,ts}=\Xi_{11,ts}+o_P(1)$, and $\widehat{\Xi}_{ij,11}=\Xi_{ij,11}+o_P(1)$ as $N,T\to\infty$. We can also show the consistency of $\widehat{G}$ following the steps outlined in Appendix A in Thompson (2011).

Thus, $\widehat{\Sigma}_\beta=\Sigma_\beta+o_P(1)$, $\widehat{\Sigma}_{f\ell}=\Sigma_{f\ell}+o_P(1)$, $\widehat{\Sigma}_{\lambda,t\ell}=\Sigma_{\lambda,t\ell}+o_P(1)$, and $\widehat{\Sigma}_{\lambda\ell}=\Sigma_{\lambda\ell}+o_P(1)$. holds immediately.

Proof of Theorem 4.1. We will prove Theorem 4.1 in two steps. In the first step, we show that $L_{Nt}(\overline{\widehat{\beta}},\widehat{D}_t;c)\to_D N(0,1)$ for a given t. In the second step, we show $L_{NT}(\overline{\widehat{\beta}},\widehat{D};c)\to_D N(0,1)$ by conventional central limit theorem.

By conventional central limit theorem, we have

$$\left(\sum_{i=1}^N[c'm(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)]^2\right)^{-\frac{1}{2}}\sum_{i=1}^N c'm(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)\to_D N(0,1),$$

as $N \to \infty$ for any $c \in \mathbb{R}^q$ such that $\|c\| = 1$.

Thus, to prove Theorem 4.1, it is sufficient to show

$$L_{Nt}(\overline{\widehat{\beta}}, \widehat{D}_t; c) = \left(\sum_{i=1}^{N}[c'm(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t)]^2\right)^{-\frac{1}{2}} \sum_{i=1}^{N} c'm(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t) + o_p(1).$$

To this end, we will show

$$\frac{1}{N}D_{Nt}(\overline{\widehat{\beta}}, \widehat{D}_t; c)^2 - \frac{1}{N}\sum_{i=1}^{N}[c'm(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t)]^2 = o_p(1),$$
$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} c'm(W_{it}, X_{it}'\overline{\widehat{\beta}}, \Phi_k^r(V_i)'\widehat{D}_t) - \frac{1}{\sqrt{N}}\sum_{i=1}^{N} c'm(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t) = o_p(1).$$

The proof is lengthy and we leave the detailed proof in the supplementary material of the paper.

Proof of Theorem 4.2. It is easy to see that for any (b, d_t) and c with $\|c\| = 1$,

$$\begin{aligned}\frac{1}{\sqrt{N}}D_{Nt}(b, d_t; c) &= \left(\mathbb{E}[c'm(W_{it}, X_{it}'b, \Phi_k^r(V_i)'d_t)]^2\right)^{1/2} + o_P(1)\\ &= \left(c'\mathbb{E}[m(W_{it}, X_{it}'b, \Phi_k^r(V_i)'d_t)m(W_{it}, X_{it}'b, \Phi_k^r(V_i)'d_t)']c\right)^{1/2} + o_P(1),\end{aligned}$$

which is bounded away from zero and infinity in probability by our condition.

Therefore, to prove this theorem, it suffices to show that there exists some c^* with $\|c^*\| = 1$ such that

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T} c^{*\prime}m(W_{it}, X_{it}'b, \Phi_k^r(V_i)'d_t) \to_P \infty$$

as $N, T \to \infty$ for any $b \in \mathbb{R}^p$ and $d_t \in \mathbb{R}^{kr}$.

By the law of large numbers, we have

$$\begin{aligned}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T} c^{*\prime}m(W_{it}, X_{it}'b, \Phi_k^r(V_i)'d_t) &= \sqrt{NT}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} c^{*\prime}m(W_{it}, X_{it}'b, \Phi_k^r(V_i)'d_t)\\ &= \sqrt{NT}\left(\mathbb{E}[c^{*\prime}m(W_{it}, X_{it}'b, \Phi_k^r(V_i)'d_t)] + o_P(1)\right).\end{aligned}$$

Let $c^* = \mathbb{E}[m(W_{it}, X_{it}'b, \Phi_k^r(V_i)'d_t)]/\|\mathbb{E}[m(W_{it}, X_{it}'b, \Phi_k^r(V_i)'d_t)]\|$. Then

$$\begin{aligned}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T} c^{*\prime}m(W_{it}, X_{it}'b, \Phi_k^r(V_i)'d_t) &= \sqrt{NT}\left(\|\mathbb{E}[m(W_{it}, X_{it}'b, \Phi_k^r(V_i)'d_t)]\| + o_P(1)\right)\\ &\geq \sqrt{NT}\left(\inf_{(b,g,s)\in\Theta}\|\mathbb{E}[m(W_{it}, X_{it}'b, s_t'g(V_i))]\| + o_P(1)\right)\\ &\geq \sqrt{NT}(\zeta_{NT} + o_P(1)) \to_P \infty,\end{aligned}$$

as $N, T \to \infty$. Then we finish the proof of this theorem.

Proof of Theorem 4.3. The asymptotic normality of $l_{NT}(\overline{\widehat{\beta}}, \widehat{\theta}, \widehat{f}_t; c)$ in the first part follows similarly from the proof of Theorem 4.1. The reason is straightforward. Here the null hypothesis is $H_{02} : \lambda(v) = g(v, \theta_0) = \widetilde{\Phi}_k(v)'\theta_0$. Thus, without orthogonal expansion, we have the parametric form of $g(v, \theta_0)$ as a combination of basis functions that allows to directly apply the estimation procedure outlined in Section 2.2 to estimate β, θ_0 and f_t.

Now, let's prove the second part. Under H_{12}, $\lambda(V_i) = g(V_i, \theta_1) + \Delta_{NT}(V_i)$. Notice that

$$\begin{aligned}
&\left|\sum_{i=1}^{N}\{[c'm(W_{it}, X_{it}'b, g(V_i, \theta_1)'\widetilde{f}_t + \Delta_{NT}(V_i)'\widetilde{f}_t)]^2 - [c'm(W_{it}, X_{it}'b, g(V_i, \theta_1)'\widetilde{f}_t)]^2\}\right| \\
\leq &\sum_{i=1}^{N}\{|c'[m(W_{it}, X_{it}'b, g(V_i, \theta_1)'\widetilde{f}_t + \Delta_{NT}(V_i)'\widetilde{f}_t) - m(W_{it}, X_{it}'b, g(V_i, \theta_1)'\widetilde{f}_t)| \\
&\qquad \times |c'[m(W_{it}, X_{it}'b, g(V_i, \theta_1)'\widetilde{f}_t + \Delta_{NT}(V_i)'\widetilde{f}_t) + m(W_{it}, X_{it}'b, g(V_i, \theta_1)'\widetilde{f}_t)]| \\
\leq &q^{1/2}\sum_{i=1}^{N} A(W_{it}, X_{it}, V_i)|\Delta_{NT}(V_i)'\widetilde{f}_t| \times [2\|m(W_{it}, X_{it}'b, g(V_i, \theta_1)'\widetilde{f}_t)\| + q^{1/2}A(W_{it}, X_{it}, V_i)|\Delta_{NT}(V_i)'\widetilde{f}_t|] \\
= &q^{1/2}\delta_{NT}\sum_{i=1}^{N} A(W_{it}, X_{it}, V_i)|\Delta(V_i)'\widetilde{f}_t| \times [2\|m(W_{it}, X_{it}'b, g(V_i, \theta_1)'\widetilde{f}_t)\| + q^{1/2}\delta_{NT}A(W_{it}, X_{it}, V_i)\|\Delta(V_i)'\widetilde{f}_t\|] \\
= &2q^{1/2}\delta_{NT}\sum_{i=1}^{N} A(W_{it}, X_{it}, V_i)\|m(W_{it}, X_{it}'b, g(V_i, \theta_1)'\widetilde{f}_t)\Delta(V_i)'\widetilde{f}_t\| + q\delta_{NT}^2\sum_{i=1}^{N} A(W_{it}, X_{it}, V_i)^2|\Delta(V_i)'\widetilde{f}_t|^2 \\
= &O(q\delta_{NT}N),
\end{aligned}$$

uniformly in t by Assumptions 3.1 and 3.3. Thus,

$$\frac{1}{Nq}\sum_{i=1}^{N}[c'm(W_{it}, X_{it}'b, g(V_i, \theta_1)'\widetilde{f}_t + \Delta_{NT}(V_i)\widetilde{f}_t)]^2 = \frac{1}{Nq}\sum_{i=1}^{N}[c'm(W_{it}, X_{it}'b, g(V_i, \theta_1)'\widetilde{f}_t)]^2 + o_P(1),$$

uniformly in t, in view of $\delta_{NT} = o(1)$. This further gives $D_{Nt} = O_P(\sqrt{Nq})$ uniformly in t.

Accordingly, to fulfill the proof, we need to show $\frac{1}{\sqrt{NTq}}\sum_{t=1}^{T}\sum_{i=1}^{N} c^{*\prime}m(W_{it}, X_{it}'b, \lambda(V_i)'\widetilde{f}_t) \to_P \infty$ for some $c^* : \|c^*\| = 1$ given any $b \in \mathbb{R}^p$, $\theta \in \widetilde{\Theta}$, $\widetilde{f}_t \in \mathbb{R}^r$.

Notice that

$$\begin{aligned}
\frac{1}{\sqrt{NTq}}\sum_{t=1}^{T}\sum_{i=1}^{N} c'm(W_{it}, X_{it}'b, \lambda(V_i)'\widetilde{f}_t) &= \frac{\sqrt{NT}}{\sqrt{q}}\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N} c'm(W_{it}, X_{it}'b, \lambda(V_i)'\widetilde{f}_t) \\
&= \frac{\sqrt{NT}}{\sqrt{q}}[c'\mathbb{E}m(W_{it}, X_{it}'b, \lambda(V_i)'\widetilde{f}_t) + o_P(1)].
\end{aligned}$$

Let $c^* = \mathbb{E}m(W_{it}, X_{it}'b, \lambda(V_i)'\widetilde{f}_t)/\|\mathbb{E}m(W_{it}, X_{it}'b, \lambda(V_i)'\widetilde{f}_t)\|$. Then,

$$\frac{1}{\sqrt{NT}\sqrt{q}}\sum_{t=1}^{T}\sum_{i=1}^{N} c^{*\prime}m(W_{it}, X_{it}'b, \lambda(V_i)'\widetilde{f}_t) = \frac{\sqrt{NT}}{\sqrt{q}}[\|\mathbb{E}m(W_{it}, X_{it}'b, \lambda(V_i)'\widetilde{f}_t)\| + o_P(1)]$$

Moreover, under H_{12} and by Assumption 4.3, $\Delta_{NT}(v) = \delta_{NT}\Delta(v)$ where $\delta_{NT} \to 0$ as $N, T \to \infty$. Then by Assumption 3.2, we have

$$\begin{aligned}
&\frac{1}{\sqrt{q}}\|\mathbb{E}[m(W_{it}, X_{it}'b, \lambda(V_i)'\widetilde{f}_t)]\| \\
=&\frac{1}{\sqrt{q}}\|\mathbb{E}[m(W_{it}, X_{it}'b, g(V_i, \theta_1)'\widetilde{f}_t + \delta_{NT}\Delta(V_i)'\widetilde{f}_t)]\| \\
\geq& \inf_{\widetilde{f}_t:\|[\widetilde{\lambda}(V_i)-\lambda(V_i)]'\widetilde{f}_t\|\geq\delta_{NT}\|\Delta(V_i)'\widetilde{f}_t\|} \frac{1}{\sqrt{q}}\|\mathbb{E}[m(W_{it}, X_{it}'b, \widetilde{\lambda}(V_i)'\widetilde{f}_t)]\| \\
\geq& \inf_{b,\widetilde{f}_t:\|b-\beta\|+\|[\widetilde{\lambda}(V_i)-\lambda(V_i)]'\widetilde{f}_t\|\geq\delta_{NT}c_{\min}} \frac{1}{\sqrt{q}}\|\mathbb{E}[m(W_{it}, X_{it}'b, \widetilde{\lambda}(V_i)'\widetilde{f}_t)]\| \\
\geq&\epsilon_{NT}.
\end{aligned}$$

Thus, under H_{12} we have

$$\begin{aligned}
&\frac{1}{\sqrt{NT}\sqrt{q}}\sum_{t=1}^{T}\sum_{i=1}^{N} c^{*\prime} m(W_{it}, X_{it}'b, \lambda(V_i)'\widetilde{f}_t) \\
=&\frac{\sqrt{NT}}{\sqrt{q}}[\|\mathbb{E}m(W_{it}, X_{it}'b, g(V_i, \theta_1)'\widetilde{f}_t + \delta_{NT}\Delta(V_i)'f_t)\| + o_P(1)] \\
\geq&\sqrt{NT}[\epsilon_{NT} + o_P(q^{-1/2})] \to_P \infty,
\end{aligned}$$

when $N, T \to \infty$. Then we finish the proof of this theorem.

References

Ai, C. & Chen, X. (2003), 'Efficient estimation of models with conditional moment restrictions containing unknown functions', *Econometrica* **71**, 1795–1843.

Ai, C. & Chen, X. (2007), 'Estimation of possibly misspecified semiparametric conditional moment restriction models with different conditioning variables', *Journal of Econometrics* **141**, 5–43.

Ando, T. & Bai, J. (2017), 'Clustering huge number of financial time series: A panel data approach with high-dimensional predictors and factor structures', *Journal of the American Statistical Association* **112**(519), 1182–1198.

Arellano, M. & Carrasco, R. (2003), 'Binary choice panel data models with predetermined variables', *Journal of Econometrics* **115**(1), 125–157.

Bai, J. (2009), 'Panel data models with interactive fixed effects', *Econometrica* **77**(4), 1229–1279.

Bai, J. & Li, K. (2014), 'Theory and methods of panel data models with interactive effects', *The Annals of Statistics* **42**(1), 142–170.

Bai, J. & Liao, Y. (2017), 'Inferences in panel data with interactive effects using large covariance matrices', *Journal of Econometrics* **200**(1), 59–78.

Breitung, J. & Lechner, M. (1995), GMM-estimation of nonlinear models on panel data, Technical report.

Breitung, J. & Lechner, M. (1998), Alternative gmm methods for nonlinear panel data models, Technical report, SFB 373 Discussion Paper.

Chen, J., Gao, J. & Li, D. (2012), 'Semiparametric trending panel data models with cross-sectional dependence', *Journal of Econometrics* **171**(1), 71–85.

Chen, L., Pelger, M. & Zhu, J. (2019), 'Deep learning in asset pricing', *Available at SSRN 3350138* .

Chen, M., Fernández-Val, I. & Weidner, M. (2021), 'Nonlinear factor models for network and panel data', *Journal of Econometrics* **220**(2), 296–324.

Chen, X. (2007), *Handbook of Econometrics*, Vol. 6B, Elsevier, Amsterdam: North Holland, chapter Large sample sieve estimation of semi-parametric models, pp. 5550–5588.

Chen, X. & Pouzo, D. (2012), 'Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals', *Econometrica* **80**, 277–321.

Chen, X. & Shen, X. (1998), 'Sieve extremum estimates for weakly dependent data', *Econometrica* **66**, 289–314.

Connor, G., Hagmann, M. & Linton, O. (2012), 'Efficient semiparametric estimation of the fama–french model and extensions', *Econometrica* **80**(2), 713–754.

Dong, C., Gao, J. & Linton, O. (2021), 'High dimensional semiparametric moment restriction models', *Journal of Econometrics* **Forthcoming**.

Dong, C., Gao, J. & Peng, B. (2015), 'Semiparametric single-index panel data models with cross-sectional dependence', *Journal of Econometrics* **188**(1), 301–312.

Dong, C., Gao, J. & Peng, B. (2021), 'Varying-coefficient panel data models with nonstationarity and partially observed factor structure', *Journal of Business & Economic Statistics* **39**(3), 700–711.

Dong, C. & Linton, O. (2018), 'Additive nonparametric models with time variable and both stationary and nonstationary regressors', *Journal of Econometrics* **207**(1), 212–236.

Dong, C., Linton, O. & Peng, B. (2021), 'A weighted sieve estimator for nonparametric time series models with nonstationary variables', *Journal of Econometrics* **222**, 909–932.

Fama, E. F. & French, K. R. (1992), 'The cross-section of expected stock returns', *the Journal of Finance* **47**(2), 427–465.

Fama, E. F. & French, K. R. (1993), 'Common risk factors in the returns on stocks and bonds', *Journal of Financial Economics* **33**(1), 3–56.

Fan, J., Ke, Z. T., Liao, Y. & Neuhierl, A. (2022), 'Structural deep learning in conditional asset pricing', *Available at SSRN 4117882* .

Fan, J., Liao, Y. & Wang, W. (2016), 'Projected principal component analysis in factor models', *Annals of statistics* **44**(1), 219.

Feng, G., He, J. & Polson, N. G. (2018), 'Deep learning for predicting asset returns', *arXiv preprint arXiv:1804.09314* .

Fernández-Val, I. & Weidner, M. (2016), 'Individual and time effects in nonlinear panel models with large n, t', *Journal of Econometrics* **192**(1), 291–312.

Gu, S., Kelly, B. T. & Xiu, D. (2021), 'Autoencoder asset pricing models', *Journal of Econometrics* **222**(1), 429–450.

Gu, S., Kelly, B. & Xiu, D. (2020), 'Empirical asset pricing via machine learning', *The Review of Financial Studies* **33**(5), 2223–2273.

Honoré, B. E. & Weidner, M. (2020), 'Moment conditions for dynamic panel logit models with fixed effects', *arXiv preprint arXiv:2005.05942* .

Kiefer, N. M. & Vogelsang, T. J. (2005), 'A new asymptotic theory for heteroskedasticity-autocorrelation robust tests', *Econometric Theory* pp. 1130–1164.

Kitazawa, Y. (2021), 'Transformations and moment conditions for dynamic fixed effects logit models', *Journal of Econometrics* **Forthcoming**.

Leippold, M., Wang, Q. & Zhou, W. (2021), 'Machine learning in the Chinese stock market', *Journal of Financial Economics* .

Liu, F. (2020), 'Nonparametric time-varying panel data models with heterogeneity', *Available at SSRN 3743529* .

Lu, X. & Su, L. (2016), 'Shrinkage estimation of dynamic panel data models with interactive fixed effects', *Journal of Econometrics* **190**(1), 148–175.

Ma, S., Linton, O. & Gao, J. (2021), 'Estimation and inference in semiparametric quantile factor models', *Journal of Econometrics* **222**(1), 295–323.

Moon, H. R. & Weidner, M. (2015), 'Linear regression for panel with unknown number of factors as interactive fixed effects', *Econometrica* **83**(4), 1543–1579.

Moon, H. R. & Weidner, M. (2017), 'Dynamic linear panel regression models with interactive fixed effects', *Econometric Theory* **33**(1), 158–195.

Phillips, P. C. B. (1998), 'New tools for understanding spurious regressions', *Econometrica* **66**(6), 1299–1325.

Rossi, A. G. (2018), Predicting stock market returns with machine learning, Technical report, Working paper.

Thompson, S. B. (2011), 'Simple formulas for standard errors that cluster by both firm and time', *Journal of Financial Economics* **99**(1), 1–10.

Supplementary material to
"GMM Estimation for High–Dimensional Panel Data Models"

TINGTING CHENG[‡] and CHAOHUA DONG[†] and JITI GAO[*] and OLIVER LINTON[⋆]

[‡] School of Finance, Nankai University, China

[†] Zhongnan University of Economics and Law, China

[*]Monash University, Australia

[⋆]University of Cambridge, UK

July 27, 2022

Appendix C

This supplementary material contains the proofs of Lemmas A.1–A.3 and Theorem 4.1 in Appendix C below.

Proof of Lemma A.1

(1) We have

$$\begin{aligned}\|M_{Nt}(\beta,D_t)\|^2 &= \left\| \frac{1}{\sqrt{q}}\frac{1}{N}\sum_{i=1}^{N} m(W_{it},X_{it}'\beta,\Phi_k^r(V_i)'D_t)\right\|^2 \\ &= \frac{1}{q}\sum_{l=1}^{q}\left[\frac{1}{N}\sum_{i=1}^{N} m_l(W_{it},X_{it}'\beta,\Phi_k^r(V_i)'D_t)\right]^2,\end{aligned}$$

where $m(\cdot,\cdot,\cdot)=(m_1(\cdot,\cdot,\cdot),\cdots,m_q(\cdot,\cdot,\cdot))'$.

It is easy to show that

$$\begin{aligned}&\mathbb{E}\left[\|M_{Nt}(\beta,D_t)\|^2\right] = \frac{1}{q}\sum_{l=1}^{q}\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N} m_l(W_{it},X_{it}'\beta,\Phi_k^r(V_i)'D_t)\right]^2 \\ &= \frac{1}{q}\sum_{l=1}^{q}\left[\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}m_l(W_{it},X_{it}'\beta,\Phi_k^r(V_i)'D_t)\right]^2 + \frac{1}{q}\sum_{l=1}^{q}\mathrm{Var}\left[\frac{1}{N}\sum_{i=1}^{N} m_l(W_{it},X_{it}'\beta,\Phi_k^r(V_i)'D_t)\right] \\ &= \frac{1}{q}\sum_{l=1}^{q}\left[\mathbb{E}m_l(W_{it},X_{it}'\beta,\Phi_k^r(V_i)'D_t)\right]^2 \\ &+ \frac{1}{q}\sum_{l=1}^{q}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathrm{cov}\left[m_l(W_{it},X_{it}'\beta,\Phi_k^r(V_i)'D_t), m_l(W_{jt},X_{jt}'\beta,\Phi_k^r(V_j)'D_t)\right]\end{aligned}$$

$$= I_1 + \frac{1}{q}\sum_{l=1}^{q}\frac{1}{N^2}I_2,$$

where $I_1 = \frac{1}{q}\|\mathbb{E}m(W_{it}, X_{it}'\beta, \Phi_k^r(V_i)'D_t)\|^2$ and
$I_2 = \sum_{i=1}^N\sum_{j=1}^N \operatorname{cov}\left[m_l(W_{it}, X_{it}'\beta, \Phi_k^r(V_i)'D_t), m_l(W_{jt}, X_{jt}'\beta, \Phi_k^r(V_j)'D_t)\right]$.

By Assumption 3.3, we can further show that

$$\begin{aligned}
I_1 &= \frac{1}{q}\|\mathbb{E}m(W_{it}, X_{it}'\beta, \Phi_k^r(V_i)'D_t)\|^2 \\
&= \frac{1}{q}\|\mathbb{E}\left[m(W_{it}, X_{it}'\beta, \Phi_k^r(V_i)'D_t) - 0\right]\|^2 \\
&= \frac{1}{q}\|\mathbb{E}\left[m(W_{it}, X_{it}'\beta, \Phi_k^r(V_i)'D_t) - m(W_{it}, X_{it}'\beta, \Phi_k^r(V_i)'D_t + \gamma^{(k)}(V_i)'f_t)\right]\|^2 \\
&\le \left(\mathbb{E}\left[A(W_{it}, X_{it}, V_i)\|\gamma^{(k)}(V_i)\|\|f_t\|\right]\right)^2 \\
&\le \mathbb{E}\left[A(W_{it}, X_{it}, V_i)\right]^2 \mathbb{E}\|\gamma^{(k)}(V_i)\|^2\mathbb{E}\|f_t\|^2 \\
&= o_p(\|\gamma^{(k)}(v)\|^2).
\end{aligned}$$

Meanwhile, by applying Lemma A.1 in Gao (2007) and Assumption 3.1, we have

$$\begin{aligned}
|I_2| &= \sum_{i=1}^N\sum_{j=1}^N\left|\operatorname{cov}\left[m_l(W_{it}, X_{it}'\beta, \Phi_k(V_i)'\alpha' f_t), m_l(W_{jt}, X_{jt}'\beta, \Phi_k(V_j)'\alpha' f_t)\right]\right| \\
&\le C\sum_{i=1}^N\sum_{j=1}^N \alpha_{ij}(0)^{\delta/(4+\delta)}\left(\mathbb{E}\left[\left|m_l(W_{it}, X_{it}'\beta, \Phi_k(V_i)'\alpha' f_t)\right|^{2+\delta/2}\right]\right)^{2/(4+\delta)} \\
&\left(\mathbb{E}\left[\left|m_l(W_{jt}, X_{jt}'\beta, \Phi_k(V_j)'\alpha' f_t)\right|^{2+\delta/2}\right]\right)^{2/(4+\delta)} \\
&\le CN. \qquad (1)
\end{aligned}$$

Thus $I_2 = O(N)$, and $\|M_{Nt}(\beta, D_t)\|^2 = O_P(\|\gamma^{(k)}(v)\|^2) + O_P(N^{-1})$.

(2) It is easy to see that

$$\begin{aligned}
&M_{Nt}(b, d_t) - \frac{1}{\sqrt{q}}\mathbb{E}m(W_{it}, X_{it}'b, \Phi_k^r(V_i)'d_t) \\
&= \frac{1}{\sqrt{q}}\frac{1}{N}\sum_{i=1}^N\left[m(W_{it}, X_{it}'b, \Phi_k^r(V_i)'d_t) - \mathbb{E}m(W_{it}, X_{it}'b, \Phi_k^r(V_i)'d_t)\right].
\end{aligned}$$

Then we have that

$$\begin{aligned}&\mathbb{E}\left\| M_{Nt}(b,d_t) - \frac{1}{\sqrt{q}}\mathbb{E}m(W_{it}, X_{it}'b, \Phi_k^r(V_i)'d_t)\right\|^2 \\ &= \frac{1}{q}\sum_{l=1}^{q}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathrm{cov}\Big[m_l(W_{it}, X_{it}'b, \Phi_k^r(V_i)'d_t), m_l(W_{jt}, X_{jt}'b, \Phi_k^r(V_j)'d_t)\Big].\end{aligned}$$

Further, by applying Lemma A.1 in Gao (2007), we have

$$\begin{aligned}&\sum_{i=1}^{N}\sum_{j=1}^{N}\Big|\mathrm{cov}\Big[m_l(W_{it}, X_{it}'b, \Phi_k^r(V_i)'d_t), m_l(W_{jt}, X_{jt}'b, \Phi_k^r(V_j)'d_t)\Big]\Big| \\ &\leq C\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_{ij}(0)^{\delta/(4+\delta)}\left(\mathbb{E}\Big[\Big|m_l(W_{it}, X_{it}'b, \Phi_k^r(V_i)'d_t)\Big|^{2+\delta/2}\Big]\right)^{2/(4+\delta)} \\ &\left(\mathbb{E}\left[\Big|m_l(W_{jt}, X_{jt}'b, \Phi_k^r(V_j)'d_t)\Big|^{2+\delta/2}\right]\right)^{2/(4+\delta)},\end{aligned}$$

where by Assumptions 3.1 and 3.3, we have

$$\begin{aligned}&\mathbb{E}\Big[\Big|m_l(W_{it}, X_{it}'b, \Phi_k^r(V_i)'d_t)\Big|^{2+\delta/2}\Big] \\ &\leq \mathbb{E}\Big[\Big|m_l(W_{it}, X_{it}'\beta, \Phi_k^r(V_i)'D_t)\Big|^{2+\delta/2}\Big] \\ &+ \mathbb{E}\left[A(W_{it}, X_{it}, V_i)\right]^{2+\delta/2}\Big[\|b-\beta\|^{2+\delta/2} + \mathbb{E}[\Phi_k^r(V_i)'d_t - \Phi_k^r(V_i)'D_t]^{2+\delta/2}\Big] \\ &= O(B_{1N}^{2+\delta/2} + B_{2N}^{2+\delta/2}),\end{aligned}$$

and thus, $\left(\mathbb{E}\Big[\Big|m_l(W_{it}, X_{it}'b, \Phi_k^r(V_i)'d_t)\Big|^{2+\delta/2}\Big]\right)^{2/(4+\delta)} = O(B_{1N} + B_{2N})$.

Therefore,

$$\begin{aligned}&\sum_{i=1}^{N}\sum_{j=1}^{N}\Big|\mathrm{cov}\Big[m_l(W_{it}, X_{it}'b, \Phi_k^r(V_i)'d_t), m_l(W_{jt}, X_{jt}'b, \Phi_k^r(V_j)'d_t)\Big]\Big| \\ &\leq C\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_{ij}(0)^{\delta/(4+\delta)}\left(\mathbb{E}\Big[\Big|m_l(W_{it}, X_{it}'b, \Phi_k^r(V_i)'d_t)\Big|^{2+\delta/2}\Big]\right)^{2/(4+\delta)} \\ &\left(\mathbb{E}\left[\Big|m_l(W_{jt}, X_{jt}'b, \Phi_k^r(V_j)'d_t)\Big|^{2+\delta/2}\right]\right)^{2/(4+\delta)} \\ &= O(N(B_{1N}^2 + B_{2N}^2)).\end{aligned}$$

Then we have that

$$
\begin{aligned}
&\mathbb{E}\left\|M_{Nt}(b,d_t)-\frac{1}{\sqrt{q}}\mathbb{E}m(W_{it},X_{it}'b,\Phi_k^r(V_i)'d_t)\right\|^2\\
&=\frac{1}{q}\sum_{l=1}^{q}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathrm{cov}\Big[m_l(W_{it},X_{it}'b,\Phi_k^r(V_i)'d_t),m_l(W_{jt},X_{jt}'b,\Phi_k^r(V_j)'d_t)\Big]\\
&=O(N^{-1}(B_{1N}^2+B_{2N}^2)).
\end{aligned}
$$

Then by the triangle inequality, we have that

$$
\begin{aligned}
&\left|\|M_{Nt}(b,d_t)\|-\frac{1}{\sqrt{q}}\|\mathbb{E}m(W_{it},X_{it}'b,\Phi_k^r(V_i)'d_t)\|\right|\\
&\leq\|M_{Nt}(b,d_t)-\frac{1}{\sqrt{q}}\mathbb{E}m(W_{it},X_{it}'b,\Phi_k^r(V_i)'d_t)\|\\
&=O_P\left(N^{-1/2}(B_{1N}+B_{2N})\right),
\end{aligned}
$$

which implies that $\|M_{Nt}(b,d_t)\|=\frac{1}{\sqrt{q}}\|\mathbb{E}m(W_{it},X_{it}'b,\Phi_k^r(V_i)'d_t)\|+O_P\left(N^{-1/2}(B_{1N}+B_{2N})\right)$, that is, $\|M_{Nt}(b,d_t)\|^2=\frac{1}{q}\|\mathbb{E}m(W_{it},X_{it}'b,\Phi_k^r(V_i)'d_t)\|^2+O_P\left(N^{-1}(B_{1N}^2+B_{2N}^2)\right)$.

For any $\delta>0$, let N be large (so k is large) such that $\delta>\|\gamma^{(k)}(v)\|$. Moreover, by Assumption 3.2, there exist an $\epsilon>0$ such that

$$
\inf_{(b,d_t)\in\Theta,\|(b-\beta,\Phi_k^r(V_i)'d_t-\lambda'f_t)\|\geq\delta}\frac{1}{q}\|\mathbb{E}m(W_{it},X_{it}'b,\Phi_k^r(V_i)'d_t)\|^2>\epsilon.
$$

In addition, we have $\|\Phi_k^r(v)'d_t-\lambda(v)'f_t\|^2=\sum_{\ell=1}^r\|\Phi_k(v)'s_{t\ell}a_\ell-\Phi_k(v)'f_{t\ell}\alpha_\ell\|^2+\|\gamma^{(k)}(v)\|^2=\|d_t-D_t\|^2+\|\gamma^{(k)}(v)\|^2$ by the orthogonality of the basis sequence.

Then we can show that

$$
\begin{aligned}
&\inf_{\|(b-\beta,d_t-D_t)\|\geq\delta}\frac{1}{q}\|\mathbb{E}m(W_{it},X_{it}'b,\Phi_k^r(V_i)'d_t)\|^2\\
&=\inf_{\|b-\beta\|^2+\|d_t-D_t\|^2\geq\delta^2}\frac{1}{q}\|\mathbb{E}m(W_{it},X_{it}'b,\Phi_k^r(V_i)'d_t)\|^2\\
&\geq\inf_{\|b-\beta\|^2+\|d_t-D_t\|^2\geq\delta^2-\|\gamma^{(k)}(v)\|^2}\frac{1}{q}\|\mathbb{E}m(W_{it},X_{it}'b,\Phi_k^r(V_i)'d_t)\|^2\\
&\geq\inf_{\|b-\beta\|^2+\|\Phi_k^r(v)'d_t-\lambda(v)'f\|^2\geq\delta^2}\frac{1}{q}\|\mathbb{E}m(W_{it},X_{it}'b,\Phi_k^r(V_i)'d_t)\|^2\\
&\geq\inf_{\|b-\beta,\Phi_k^r(V_i)'d_t-\lambda'f_t\|\geq\delta}\frac{1}{q}\|\mathbb{E}m(W_{it},X_{it}'b,\Phi_k^r(V_i)'d_t)\|^2>\epsilon.
\end{aligned}
$$

Then we finish the proof of Lemma A.1.

Proof of Lemma A.2

Write $H_{Nt}(\beta, D_t) = \widetilde{H}_{Nt}(\beta, D_t) + \Delta_{Nt}(\beta, D_t)$, where $\widetilde{H}_{Nt}(\beta, D_t)$ is a symmetric 2-by-2 block matrix with blocks:

$$\begin{aligned}
\widetilde{H}_{11}(\beta, D_t) &= \frac{1}{q}\sum_{l=1}^{q}\left(\frac{1}{N}\sum_{j=1}^{N}\frac{\partial}{\partial u}m_l(W_{jt}, X_{jt}'\beta, \Phi_k^r(V_j)'D_t)X_{jt}\right) \\
&\quad\times\left(\frac{1}{N}\sum_{i=1}^{N}\frac{\partial}{\partial u}m_l(W_{it}, X_{it}'\beta, \Phi_k^r(V_i)'D_t)X_{it}\right)' \\
\widetilde{H}_{12}(\beta, D_t) &= \frac{1}{q}\sum_{l=1}^{q}\left(\frac{1}{N}\sum_{j=1}^{N}\frac{\partial}{\partial u}m_l(W_{jt}, X_{jt}'\beta, \Phi_k^r(V_j)'D_t)X_{jt}\right) \\
&\quad\times\left(\frac{1}{N}\sum_{i=1}^{N}\frac{\partial}{\partial w}m_l(W_{it}, X_{it}'\beta, \Phi_k^r(V_i)'D_t)\Phi_k^r(V_i)\right)' \\
\widetilde{H}_{22}(\beta, D_t) &= \frac{1}{q}\sum_{l=1}^{q}\left(\frac{1}{N}\sum_{j=1}^{N}\frac{\partial}{\partial w}m_l(W_{jt}, X_{jt}'\beta, \Phi_k^r(V_j)'D_t)\Phi_k^r(V_j)\right) \\
&\quad\times\left(\frac{1}{N}\sum_{i=1}^{N}\frac{\partial}{\partial w}m_l(W_{it}, X_{it}'\beta, \Phi_k^r(V_i)'D_t)\Phi_k^r(V_i)\right)'
\end{aligned}$$

and $\widetilde{H}_{21}(\beta, D_t) = \widetilde{H}_{12}(\beta, D_t)'$.

$\Delta_{Nt}(\beta, D_t)$ has blocks as follows.

$$\begin{aligned}
\Delta_{11}(\beta, D_t) &= \frac{1}{q}\sum_{l=1}^{q}\left(\frac{1}{N}\sum_{i=1}^{N}m_l(W_{it}, X_{it}'\beta, \Phi_k^r(V_i)'D_t)\right) \\
&\quad\times\left(\frac{1}{N}\sum_{j=1}^{N}\frac{\partial^2}{\partial u^2}m_l(W_{jt}, X_{jt}'\beta, \Phi_k^r(V_j)'D_t)X_{jt}X_{jt}'\right) \\
\Delta_{12}(\beta, D_t) &= \frac{1}{q}\sum_{l=1}^{q}\left(\frac{1}{N}\sum_{i=1}^{N}m_l(W_{it}, X_{it}'\beta, \Phi_k^r(V_i)'D_t)\right) \\
&\quad\times\left(\frac{1}{N}\sum_{j=1}^{N}\frac{\partial^2}{\partial u\partial w}m_l(W_{jt}, X_{jt}'\beta, \Phi_k^r(V_j)'D_t)X_{jt}\Phi_k^r(V_j)'\right) \\
\Delta_{22}(\beta, D_t) &= \frac{1}{q}\sum_{l=1}^{q}\left(\frac{1}{N}\sum_{i=1}^{N}m_l(W_{it}, X_{it}'\beta, \Phi_k^r(V_i)'D_t)\right) \\
&\quad\times\left(\frac{1}{N}\sum_{j=1}^{N}\frac{\partial^2}{\partial w^2}m_l(W_{jt}, X_{jt}'\beta, \Phi_k^r(V_j)'D_t)\Phi_k^r(V_j)\Phi_k^r(V_j)'\right),
\end{aligned}$$

and $\Delta_{21}(\beta, D_t) = \Delta_{12}(\beta, D_t)'$.

To prove that $H_{Nt}(\beta, D_t)$ is almost surely positive definite, we shall show that (i) $\widetilde{H}_{Nt}(\beta, D_t)$ is almost surely positive definite; (ii) $\|\Delta(\beta, D_t)\| = o_P(1)$.

To prove (i), for any vectors $\mathbf{Q}_1 \in \mathbb{R}^p$ and $\mathbf{Q}_2 \in \mathbb{R}^{kr}$, where at least one of them is not zero, we have

$$\begin{aligned}
&(\mathbf{Q}_1', \mathbf{Q}_2')\widetilde{H}_{Nt}(\beta, D_t)(\mathbf{Q}_1', \mathbf{Q}_2')' \\
&= \frac{1}{q}\sum_{l=1}^{q}\Big(\frac{1}{N}\sum_{j=1}^{N}\frac{\partial}{\partial u}m_l(W_{jt}, X_{jt}'\beta, \Phi_k^r(V_j)'D_t)\mathbf{Q}_1'X_{jt}\Big)^2 \\
&+ \frac{1}{q}\sum_{l=1}^{q}\Big(\frac{1}{N}\sum_{j=1}^{N}\frac{\partial}{\partial w}m_l(W_{jt}, X_{jt}'\beta, \Phi_k^r(V_j)'D_t)\mathbf{Q}_2'(\Phi_k^r(V_j))\Big)^2 \\
&+ \frac{2}{q}\sum_{l=1}^{q}\Big(\frac{1}{N}\sum_{j=1}^{N}\frac{\partial}{\partial u}m_l(W_{jt}, X_{jt}'\beta, \Phi_k^r(V_j)'D_t)\mathbf{Q}_1'X_{jt}\Big)\Big(\frac{1}{N}\sum_{j=1}^{N}\frac{\partial}{\partial w}m_l(W_{jt}, X_{jt}'\beta, \Phi_k^r(V_j)'D_t)\mathbf{Q}_2'(\Phi_k^r(V_j))\Big) \\
&= \frac{1}{q}\sum_{l=1}^{q}\Big[\frac{1}{N}\sum_{j=1}^{N}\frac{\partial}{\partial u}m_l(W_{jt}, X_{jt}'\beta, \Phi_k^r(V_j)'D_t)\mathbf{Q}_1'X_{jt} + \frac{1}{N}\sum_{j=1}^{N}\frac{\partial}{\partial w}m_l(W_{jt}, X_{jt}'\beta, \Phi_k^r(V_j)'D_t)\mathbf{Q}_2'(\Phi_k^r(V_j))\Big]^2
\end{aligned}$$

which is almost surely positive. Hence, $\widetilde{H}_{Nt}(\beta, D_t)$ is almost surely positive definite.

To prove (ii), it suffices to prove the result for each block.

By the triangle inequality and Cauchy-Schwarz inequality, we have

$$\begin{aligned}
\|\Delta_{11}(\beta, D_t)\|^2 &\leq \frac{1}{q}\sum_{l=1}^{q}\left(\frac{1}{N}\sum_{i=1}^{N}m_l(W_{it}, X_{it}'\beta, \Phi_k^r(V_i)'D_t)\right)^2 \\
&\quad \times \frac{1}{q}\sum_{l=1}^{q}\left\|\frac{1}{N}\sum_{j=1}^{N}\frac{\partial^2}{\partial u^2}m_l(W_{jt}, X_{jt}'\beta, \Phi_k^r(V_j)'D_t)X_{jt}X_{jt}'\right\|^2 \\
&= \|M_{Nt}(\beta, D_t)\|^2\frac{1}{q}\sum_{l=1}^{q}\left\|\frac{1}{N}\sum_{j=1}^{N}\frac{\partial^2}{\partial u^2}m_l(W_{jt}, X_{jt}'\beta, \Phi_k^r(V_j)'D_t)X_{jt}X_{jt}'\right\|^2.
\end{aligned}$$

By Lemma A.1, we know that $\|M_{Nt}(\beta, D_t)\|^2 = O_P(\|\gamma^{(k)}(v)\|^2) + O_P(N^{-1})$. Meanwhile,

$$\begin{aligned}
&\frac{1}{q}\sum_{l=1}^{q}\left\|\frac{1}{N}\sum_{j=1}^{N}\frac{\partial^2}{\partial u^2}m_l(W_{jt}, X_{jt}'\beta, \Phi_k^r(V_j)'D_t)X_{jt}X_{jt}'\right\|^2 \\
&\leq \frac{2}{q}\sum_{l=1}^{q}\left\|\mathbb{E}\frac{\partial^2}{\partial u^2}m_l(W_{1t}, X_{1t}'\beta, \Phi_k^r(V_1)'D_t)X_{1t}X_{1t}'\right\|^2 \\
&+ \frac{2}{q}\sum_{l=1}^{q}\left\|\frac{1}{N}\sum_{j=1}^{N}\left(\frac{\partial^2}{\partial u^2}m_l(W_{jt}, X_{jt}'\beta, \Phi_k^r(V_j)'D_t)X_{jt}X_{jt}' - \mathbb{E}\frac{\partial^2}{\partial u^2}m_l(W_{jt}, X_{jt}'\beta, \Phi_k^r(V_j)'D_t)X_{jt}X_{jt}'\right)\right\|^2.
\end{aligned} \tag{2}$$

By Assumption 3.5, we have for any given t, $\frac{1}{q}\sum_{l=1}^{q}\left\|\mathbb{E}\frac{\partial^2}{\partial u^2}m_l(W_{1t},X_{1t}'\beta,\Phi_k^r(V_1)'D_t)\otimes X_{11}X_{11}'\right\|^2=O(p^2)$.

By Assumption 3.1, we have

$$
\begin{aligned}
&\frac{1}{q}\sum_{l=1}^{q}\mathbb{E}\left\|\frac{1}{N}\sum_{j=1}^{N}\left(\frac{\partial^2}{\partial u^2}m_l(W_{jt},X_{jt}'\beta,\Phi_k^r(V_j)'D_t)X_{jt}X_{jt}'-\mathbb{E}\frac{\partial^2}{\partial u^2}m_l(W_{jt},X_{jt}'\beta,\Phi_k^r(V_j)'D_t)X_{jt}X_{jt}'\right)\right\|^2\\
&=\frac{1}{q}\sum_{l=1}^{q}\frac{1}{N^2}\mathbb{E}\left\|\sum_{j=1}^{N}\left(\frac{\partial^2}{\partial u^2}m_l(W_{jt},X_{jt}'\beta,\Phi_k^r(V_j)'D_t)X_{jt}X_{jt}'-\mathbb{E}\frac{\partial^2}{\partial u^2}m_l(W_{jt},X_{jt}'\beta,\Phi_k^r(V_j)'D_t)X_{jt}X_{jt}'\right)\right\|^2\\
&=\frac{1}{q}\sum_{l=1}^{q}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathrm{cov}\left[\frac{\partial^2}{\partial u^2}m_l(W_{it},X_{it}'\beta,\Phi_k^r(V_i)'D_t)X_{it}X_{it}',\frac{\partial^2}{\partial u^2}m_l(W_{jt},X_{jt}'\beta,\Phi_k^r(V_j)'D_t)X_{jt}X_{jt}'\right]\\
&=O(N^{-1}p^2), \qquad (3)
\end{aligned}
$$

The last equality is obtained by the same procedure as (1) combined with $\mathbb{E}\|X_{11}\|^2=O(p)$ in Assumption 3.5.

Then we have $\|\Delta_{11}(\beta,D_t)\|^2=o_P(1)$.

Similarly, we have

$$
\|\Delta_{12}(\beta,D_t)\|^2\le\|M_{Nt}(\beta,D_t)\|^2\frac{1}{q}\sum_{l=1}^{q}\left\|\frac{1}{N}\sum_{j=1}^{N}\frac{\partial^2}{\partial u\partial w}m_l(W_{jt},X_{jt}'\beta,\Phi_k^r(V_j)'D_t)X_{jt}\Phi_k^r(V_j)'\right\|^2
$$

in which

$$
\frac{1}{q}\sum_{l=1}^{q}\mathbb{E}\left\|\frac{1}{N}\sum_{j=1}^{N}\frac{\partial^2}{\partial u\partial w}m_l(W_{jt},X_{jt}'\beta,\Phi_k^r(V_j)'D_t)X_{jt}\Phi_k^r(V_j)'\right\|^2=O(pkr)+O(N^{-1}pkr),
$$

by the same procedure of (2). This implies that $\|\Delta_{12}(\beta,D_t)\|^2=o_P(1)$.

Similarly

$$
\|\Delta_{22}(\beta,D_t)\|^2\le\|M_{Nt}(\beta,D_t)\|^2\frac{1}{q}\sum_{l=1}^{q}\left\|\frac{1}{N}\sum_{j=1}^{N}\frac{\partial^2}{\partial w^2}m_l(W_{jt},X_{jt}'\beta,\Phi_k^r(V_j)'D_t)\Phi_k^r(V_j)\Phi_k^r(V_j)'\right\|^2,
$$

where

$$
\frac{1}{q}\sum_{l=1}^{q}\mathbb{E}\left\|\frac{1}{N}\sum_{j=1}^{N}\frac{\partial^2}{\partial w^2}m_l(W_{jt},X_{jt}'\beta,\Phi_k^r(V_j)'D_t)\Phi_k^r(V_j)\Phi_k^r(V_j)'\right\|^2=O(k^2r^2)+O(N^{-1}k^2r^2),
$$

implying that $\|\Delta_{22}(\beta,D_t)\|^2=o_P(1)$.

Now we show the result (2) in Lemma A.2. Since $\|H_{Nt}(\beta,D_t)-h_{Nt}(\beta,\lambda'f_t)\|\le\|\Delta_{Nt}(\beta,D_t)\|+$

$\|\widetilde{H}_{Nt}(\beta,D_t) - h_{Nt}(\beta,\lambda' f_t)\| = o_P(1) + \|\widetilde{H}_{Nt}(\beta,D_t) - h_{Nt}(\beta,\lambda' f_t)\|$. It is sufficient to show that $\|\widetilde{H}_{Nt}(\beta,D_t) - h_{Nt}(\beta,\lambda' f_t)\| = o_P(1)$. In what follows, we will show the results in block-sense.

$$
\begin{aligned}
&\widetilde{H}_{11}(\beta,D_t) - h_{11}(\beta,\lambda' f_t) \\
&= \frac{1}{q}\sum_{l=1}^{q}\left(\frac{1}{N}\sum_{j=1}^{N}\frac{\partial}{\partial u}m_l(W_{jt},X_{jt}'\beta,\Phi_k^r(V_j)'D_t)X_{jt}\right)\left(\frac{1}{N}\sum_{i=1}^{N}\frac{\partial}{\partial u}m_l(W_{it},X_{it}'\beta,\Phi_k^r(V_i)'D_t)X_{it}\right)' \\
&\quad - \frac{1}{q}\sum_{l=1}^{q}\left(\mathbb{E}\frac{\partial}{\partial u}m_l(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)X_{it}\right)\left(\mathbb{E}\frac{\partial}{\partial u}m_l(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)X_{it}\right)' \\
&= \frac{1}{q}\sum_{l=1}^{q}\frac{1}{N}\sum_{j=1}^{N}\left(\frac{\partial}{\partial u}m_l(W_{jt},X_{jt}'\beta,\Phi_k^r(V_j)'D_t)X_{jt} - \mathbb{E}\frac{\partial}{\partial u}m_l(W_{jt},X_{jt}'\beta,\lambda(V_j)'f_t)X_{jt}\right) \\
&\quad \times \left(\frac{1}{N}\sum_{i=1}^{N}\frac{\partial}{\partial u}m_l(W_{it},X_{it}'\beta,\Phi_k^r(V_i)'D_t)X_{it}\right)' \\
&\quad + \frac{1}{q}\sum_{l=1}^{q}\left(\mathbb{E}\frac{\partial}{\partial u}m_l(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)X_{it}\right) \\
&\quad \times \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\partial}{\partial u}m_l(W_{it},X_{it}'\beta,\Phi_k^r(V_i)'D_t)X_{it} - \mathbb{E}\frac{\partial}{\partial u}m_l(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)X_{it}\right)' \\
&= I_1 + I_2,
\end{aligned}
$$

where

$$
\begin{aligned}
I_1 &= \frac{1}{q}\sum_{l=1}^{q}\frac{1}{N}\sum_{j=1}^{N}\left(\frac{\partial}{\partial u}m_l(W_{jt},X_{jt}'\beta,\Phi_k^r(V_j)'D_t)X_{jt} - \frac{\partial}{\partial u}m_l(W_{jt},X_{jt}'\beta,\lambda(V_j)'f_t)X_{jt}\right) \\
&\quad \times \left(\frac{1}{N}\sum_{i=1}^{N}\frac{\partial}{\partial u}m_l(W_{it},X_{it}'\beta,\Phi_k^r(V_i)'D_t)X_{it}\right)' \\
&\quad + \frac{1}{q}\sum_{l=1}^{q}\frac{1}{N}\sum_{j=1}^{N}\left(\frac{\partial}{\partial u}m_l(W_{jt},X_{jt}'\beta,\lambda(V_j)'f_t)X_{jt} - \mathbb{E}\frac{\partial}{\partial u}m_l(W_{jt},X_{jt}'\beta,\lambda(V_j)'f_t)X_{jt}\right) \\
&\quad \times \left(\frac{1}{N}\sum_{i=1}^{N}\frac{\partial}{\partial u}m_l(W_{it},X_{it}'\beta,\Phi_k^r(V_i)'D_t)X_{it}\right)'.
\end{aligned}
$$

By Cauchy-Schwarz inequality, we have

$$
\|I_1\|^2 \le \frac{1}{q}\sum_{l=1}^{q}\left\|\frac{1}{N}\sum_{j=1}^{N}\left(\frac{\partial}{\partial u}m_l(W_{jt},X_{jt}'\beta,\Phi_k^r(V_j)'D_t)X_{jt} - \frac{\partial}{\partial u}m_l(W_{jt},X_{jt}'\beta,\lambda(V_j)'f_t)X_{jt}\right)\right\|^2
$$

$$
\begin{aligned}
&\times \frac{1}{q}\sum_{l=1}^{q}\left\|\frac{1}{N}\sum_{i=1}^{N}\frac{\partial}{\partial u}m_l(W_{it},X_{it}'\beta,\Phi_k^r(V_i)'D_t)X_{it}\right\|^2\\
&+\frac{1}{q}\sum_{l=1}^{q}\left\|\frac{1}{N}\sum_{j=1}^{N}\left(\frac{\partial}{\partial u}m_l(W_{jt},X_{jt}'\beta,\lambda(V_j)'f_t)X_{jt}-\mathbb{E}\frac{\partial}{\partial u}m_l(W_{jt},X_{jt}'\beta,\lambda(V_j)'f_t)X_{jt}\right)\right\|^2\\
&\times \frac{1}{q}\sum_{l=1}^{q}\left\|\frac{1}{N}\sum_{i=1}^{N}\frac{\partial}{\partial u}m_l(W_{it},X_{it}'\beta,\Phi_k^r(V_i)'D_t)X_{it}\right\|^2\\
&= I_{11}\times I_{13}+I_{12}\times I_{13}.
\end{aligned}
$$

Following the procedure of deriving (2), I_{11} has the same order in probability as

$$
\begin{aligned}
&\frac{1}{q}\sum_{l=1}^{q}\left\|\mathbb{E}\left(\frac{\partial}{\partial u}m_l(W_{11},X_{11}'\beta,\Phi_k^r(V_1)'D_1)-\frac{\partial}{\partial u}m_l(W_{11},X_{11}'\beta,\lambda(V_1)'f_1)\right)X_{11}\right\|^2\\
&=\frac{1}{q}\left\|\mathbb{E}\left(\frac{\partial}{\partial u}m(W_{11},X_{11}'\beta,\Phi_k^r(V_1)'D_1)-\frac{\partial}{\partial u}m(W_{11},X_{11}'\beta,\lambda(V_1)'f_1)\right)\otimes X_{11}\right\|^2\\
&\le \mathbb{E}\left[A(W_{11},X_{11},V_1)^2\|X_{11}\|^2\right]\mathbb{E}\left\|\gamma^{(k)}(V_1)^2\right\| = O_P(\|\gamma^{(k)}(V_1)\|^2 p),
\end{aligned} \tag{4}
$$

by Assumption 3.5 and 3.7.

For I_{12}, by Assumption 3.1, we have

$$
\begin{aligned}
\mathbb{E}[I_{12}] &= \frac{1}{q}\sum_{l=1}^{q}\mathbb{E}\left\|\frac{1}{N}\sum_{j=1}^{N}\left(\frac{\partial}{\partial u}m_l(W_{jt},X_{jt}'\beta,\lambda(V_j)'f_t)X_{jt}-\mathbb{E}\frac{\partial}{\partial u}m_l(W_{jt},X_{jt}'\beta,\lambda(V_j)'f_t)X_{jt}\right)\right\|^2\\
&=\frac{1}{q}\sum_{l=1}^{q}\frac{1}{N^2}\mathbb{E}\left\|\sum_{j=1}^{N}\left(\frac{\partial}{\partial u}m_l(W_{jt},X_{jt}'\beta,\lambda(V_j)'f_t)X_{jt}-\mathbb{E}\frac{\partial}{\partial u}m_l(W_{jt},X_{jt}'\beta,\lambda(V_j)'f_t)X_{jt}\right)\right\|^2\\
&=\frac{1}{q}\sum_{l=1}^{q}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\operatorname{cov}\left[\frac{\partial}{\partial u}m_l(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)X_{it},\frac{\partial}{\partial u}m_l(W_{jt},X_{jt}'\beta,\lambda(V_j)'f_t)X_{jt}\right]\\
&=O(N^{-1}p),
\end{aligned}
$$

where the last equality is obtained by the same procedure as (1) combined with $\mathbb{E}\|X_{11}\|^2 = O(p)$ in Assumption 3.5.

Moreover, following the procedure of deriving (2), I_{13} has the same order in probability as

$$
\frac{1}{q}\sum_{l=1}^{q}\left\|\mathbb{E}\frac{\partial}{\partial u}m_l(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)X_{it}\right\|^2
$$

$$
\begin{aligned}
&+\frac{1}{q}\sum_{l=1}^{q}\left\|\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[\frac{\partial}{\partial u}m_l(W_{it},X_{it}'\beta,\Phi_k^r(V_i)'D_t)-\frac{\partial}{\partial u}m_l(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)\right]X_{it}\right\|^2\\
&=\frac{1}{q}\left\|\mathbb{E}\frac{\partial}{\partial u}m(W_{11},X_{11}'\beta,\lambda(V_1)'f_1)\otimes X_{11}\right\|^2\\
&+\frac{1}{q}\sum_{l=1}^{q}\left\|\mathbb{E}\left[\frac{\partial}{\partial u}m_l(W_{11},X_{11}'\beta,\Phi_k^r(V_1)'D_1)-\frac{\partial}{\partial u}m_l(W_{11},X_{11}'\beta,\lambda(V_1)'f_1)\right]X_{11}\right\|^2\\
&=O(p)+\frac{1}{q}\left\|\mathbb{E}\left[\frac{\partial}{\partial u}m(W_{11},X_{11}'\beta,\Phi_k^r(V_1)'D_1)-\frac{\partial}{\partial u}m(W_{11},X_{11}'\beta,\lambda(V_1)'f_1)\right]\otimes X_{11}\right\|^2\\
&=O(p)+\left(\mathbb{E}[A(W_{11},X_{11},V_1)|\gamma^{(k)}(v)|\|X_{11}\|]\right)^2=O(p)+O\left(\|\gamma^{(k)}(v)\|^2p\right),
\end{aligned}
$$

which implying that $\|I_1\|^2=O_P(N^{-1}p^2)+O_P\left(\|\gamma^{(k)}(v)\|^2p^2\right)=o_P(1)$.

Next, we consider I_2. By Cauchy-Schwarz inequality, we have

$$
\begin{aligned}
\|I_2\|^2 &\le \frac{1}{q}\sum_{l=1}^{q}\left\|\mathbb{E}\frac{\partial}{\partial u}m_l(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)X_{it}\right\|^2\\
&\times\frac{1}{q}\sum_{l=1}^{q}\left\|\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\partial}{\partial u}m_l(W_{it},X_{it}'\beta,\Phi_k^r(V_i)'D_t)X_{it}-\mathbb{E}\frac{\partial}{\partial u}m_l(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)X_{it}\right)\right\|^2\\
&\le 2\frac{1}{q}\left\|\mathbb{E}\frac{\partial}{\partial u}m(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)\otimes X_{it}\right\|^2\\
&\times\frac{1}{q}\left\|\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\partial}{\partial u}m(W_{it},X_{it}'\beta,\Phi_k^r(V_i)'D_t)-\frac{\partial}{\partial u}m(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)\right)\otimes X_{it}\right\|^2\\
&+2\frac{1}{q}\left\|\mathbb{E}\frac{\partial}{\partial u}m(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)\otimes X_{it}\right\|^2\\
&\times\frac{1}{q}\sum_{l=1}^{q}\left\|\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\partial}{\partial u}m_l(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)X_{it}-\mathbb{E}\frac{\partial}{\partial u}m_l(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)X_{it}\right)\right\|^2\\
&=2I_{21}(I_{22}+I_{23}).
\end{aligned}
$$

By Assumption 3.5, $I_{21}=O(p)$. Following the procedure of deriving (2), I_{22} has the same order in probability as

$$
\begin{aligned}
&\frac{1}{q}\left\|\mathbb{E}\left(\frac{\partial}{\partial u}m(W_{11},X_{11}'\beta,\Phi_k^r(V_1)'D_1)-\frac{\partial}{\partial u}m(W_{11},X_{11}'\beta,\lambda(V_1)'f_1)\right)\otimes X_{11}\right\|^2\\
&\le\left(\mathbb{E}[A(W_{11},X_{11},V_1)\|\gamma^{(k)}(V_1)\|\|X_{11}\|]\right)^2=O(p)\|\gamma^{(k)}(v)\|^2,
\end{aligned}
$$

using Assumptions 3.5 and 3.7. Meanwhile, by Assumption 3.1, we have

$$
\begin{aligned}
\mathbb{E}[I_{23}] &= \frac{1}{q}\sum_{l=1}^{q}\mathbb{E}\left\|\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\partial}{\partial u}m_l(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)X_{it}-\mathbb{E}\frac{\partial}{\partial u}m_l(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)X_{it}\right)\right\|^2 \\
&= \frac{1}{q}\sum_{l=1}^{q}\frac{1}{N^2}\mathbb{E}\left\|\sum_{i=1}^{N}\left(\frac{\partial}{\partial u}m_l(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)X_{it}-\mathbb{E}\frac{\partial}{\partial u}m_l(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)X_{it}\right)\right\|^2 \\
&= \frac{1}{q}\sum_{l=1}^{q}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathrm{cov}\left[\frac{\partial}{\partial u}m_l(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)X_{it},\frac{\partial}{\partial u}m_l(W_{jt},X_{jt}'\beta,\lambda(V_j)'f_t)X_{jt}\right] \\
&= O(N^{-1}p).
\end{aligned} \tag{5}
$$

where the last equality is obtained by the same procedure as (1) combined with $\mathbb{E}\|X_{11}\|^2 = O(p)$ in Assumption 3.5.

Therefore, $\|I_2\|^2 = O_P(N^{-1}p^2)+O_P(\|\gamma^{(k)}(v)\|^2p^2) = o_P(1)$. Thus $\widetilde{H}_{11}(\beta,D_t)-h_{11}(\beta,\lambda'f_t) = o_P(1)$. In addition,

$$
\begin{aligned}
&\widetilde{H}_{12}(\beta,D_t)-h_{12}(\beta,\lambda'f_t) \\
&= \frac{1}{q}\sum_{l=1}^{q}\left(\frac{1}{N}\sum_{j=1}^{N}\frac{\partial}{\partial u}m_l(W_{jt},X_{jt}'\beta,\Phi_k^r(V_j)'D_t)X_{jt}\right)\left(\frac{1}{N}\sum_{i=1}^{N}\frac{\partial}{\partial w}m_l(W_{it},X_{it}'\beta,\Phi_k^r(V_i)'D_t)\Phi_k^r(V_i)\right)' \\
&- \frac{1}{q}\sum_{l=1}^{q}\left(\mathbb{E}\frac{\partial}{\partial w}m_l(W_{11},X_{11}'\beta,\lambda(V_1)'f_1)X_{11}\right)\left(\mathbb{E}\frac{\partial}{\partial u}m_l(W_{11},X_{11}'\beta,\lambda(V_1)'f_1)\Phi_k^r(V_1)\right)' \\
&= \frac{1}{q}\sum_{l=1}^{q}\left(\frac{1}{N}\sum_{j=1}^{N}\frac{\partial}{\partial u}m_l(W_{jt},X_{jt}'\beta,\Phi_k^r(V_j)'D_t)X_{jt}-\mathbb{E}\frac{\partial}{\partial w}m_l(W_{11},X_{11}'\beta,\lambda(V_1)'f_1)X_{11}\right) \\
&\times\left(\frac{1}{N}\sum_{i=1}^{N}\frac{\partial}{\partial w}m_l(W_{it},X_{it}'\beta,\Phi_k^r(V_i)'D_t)\Phi_k^r(V_i)\right)' \\
&+ \frac{1}{q}\sum_{l=1}^{q}\left(\mathbb{E}\frac{\partial}{\partial w}m_l(W_{11},X_{11}'\beta,\lambda(V_1)'f_1)X_{11}\right) \\
&\times\left(\frac{1}{N}\sum_{i=1}^{N}\frac{\partial}{\partial w}m_l(W_{it},X_{it}'\beta,\Phi_k^r(V_i)'D_t)\Phi_k^r(V_i)-\mathbb{E}\frac{\partial}{\partial u}m_l(W_{11},X_{11}'\beta,\lambda(V_1)'f_1)\Phi_k^r(V_1)\right)' \\
&= I_3+I_4.
\end{aligned}
$$

Similar to I_1, we have $\|I_3\|^2 = O_P(N^{-1}pkr)+O_P(\|\gamma^{(k)}(v)\|^2pkr) = o_P(1)$ and $\|I_4\|^2 = O_P(N^{-1}pkr)+O_P(\|\gamma^{(k)}(v)\|^2pkr) = o_P(1)$. We then have $\widetilde{H}_{12}(\beta,D_t)-h_{12}(\beta,\lambda'f_t) = o_P(1)$.

$$
\begin{aligned}
&\widetilde{H}_{22}(\beta,D_t)-h_{22}(\beta,\lambda'f_t) \\
&= \frac{1}{q}\sum_{l=1}^{q}\left(\frac{1}{N}\sum_{j=1}^{N}\frac{\partial}{\partial w}m_l(W_{jt},X_{jt}'\beta,\Phi_k^r(V_j)'D_t)\Phi_k^r(V_j\right)\left(\frac{1}{N}\sum_{i=1}^{N}\frac{\partial}{\partial w}m_l(W_{it},X_{it}'\beta,\Phi_k^r(V_i)'D_t)\Phi_k^r(V_i)\right)'
\end{aligned}
$$

$$
\begin{aligned}
&-\frac{1}{q}\sum_{l=1}^{q}\left(\mathbb{E}\frac{\partial}{\partial w}m_l(W_{11},X_{11}'\beta,\lambda(V_1)'f_1)\Phi_k^r(V_1)\right)\left(\mathbb{E}\frac{\partial}{\partial w}m_l(W_{11},X_{11}'\beta,\lambda(V_1)'f_1)\Phi_k^r(V_1)\right)'\\
&=\frac{1}{q}\sum_{l=1}^{q}\left(\frac{1}{N}\sum_{j=1}^{N}\frac{\partial}{\partial w}m_l(W_{jt},X_{jt}'\beta,\Phi_k^r(V_j)'D_t)\Phi_k^r(V_j)-\mathbb{E}\frac{\partial}{\partial w}m_l(W_{11},X_{11}'\beta,\lambda(V_1)'f_1)\Phi_k^r(V_1)\right)\\
&\times\left(\frac{1}{N}\sum_{i=1}^{N}\frac{\partial}{\partial w}m_l(W_{it},X_{it}'\beta,\Phi_k^r(V_i)'D_t)\Phi_k^r(V_i)\right)'\\
&+\frac{1}{q}\sum_{l=1}^{q}\left(\mathbb{E}\frac{\partial}{\partial w}m_l(W_{11},X_{11}'\beta,\lambda(V_1)'f_1)\Phi_k^r(V_1)\right)\\
&\times\left(\frac{1}{N}\sum_{i=1}^{N}\frac{\partial}{\partial w}m_l(W_{it},X_{it}'\beta,\Phi_k^r(V_i)'D_t)\Phi_k^r(V_i)-\mathbb{E}\frac{\partial}{\partial w}m_l(W_{11},X_{11}'\beta,\lambda(V_1)'f_1)\Phi_k^r(V_1)\right)'\\
&=I_5+I_6.
\end{aligned}
$$

Similar to I_1, we have $\|I_5\|^2=O_P(N^{-1}k^2r^2)+O_P(\|\gamma^{(k)}(v)\|^2k^2r^2)=o_P(1)$ and $\|I_6\|^2=O_P(N^{-1}k^2r^2)+O_P(\|\gamma^{(k)}(v)\|^2k^2r^2)=o_P(1)$ by Assumption 3.6. We then have $\widetilde{H}_{22}(\beta,D_t)-h_{22}(\beta,\lambda'f_t)=o_P(1)$.

This finishes the proof of Lemma A.2.

Denote $s_{Nt}(\beta,\lambda'f_t)=(s_{1Nt}(\beta,\lambda'f_t)',s_{2Nt}(\beta,\lambda'f_t)')'$, where

$$
\begin{aligned}
s_{1Nt}(\beta,\lambda'f_t)&=\frac{1}{q}\frac{1}{N}\sum_{l=1}^{q}\sum_{i=1}^{N}m_l(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)\mathbb{E}\frac{\partial}{\partial u}m_l(W_{jt},X_{jt}'\beta,\lambda(V_i)'f_t)X_{jt}\\
&=\left[\frac{1}{q}\mathbb{E}\left(\frac{\partial}{\partial u}m(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)'\otimes X_{it}\right)\right]\frac{1}{N}\sum_{i=1}^{N}m(W_{it},X_{it}'\beta,\lambda(V_i)'f_t),\\
s_{2Nt}(\beta\lambda'f_t)&=\frac{1}{q}\frac{1}{N}\sum_{l=1}^{q}\sum_{i=1}^{N}m_l(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)\mathbb{E}\frac{\partial}{\partial w}m_l(W_{jt},X_{jt}'\beta,\lambda(V_i)'f_t)\Phi_k^r(V_j)\\
&=\left[\frac{1}{q}\mathbb{E}\left(\frac{\partial}{\partial w}m(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)'\otimes\Phi_k^r(V_i)\right)\right]\frac{1}{N}\sum_{i=1}^{N}m(W_{it},X_{it}'\beta,\lambda(V_i)'f_t).
\end{aligned}
$$

Hence, we have $s_{Nt}(\beta,\lambda'f_t)=\frac{1}{q}\Psi_{Nt}\frac{1}{N}\sum_{i=1}^{N}m(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)$, where

$$
\Psi_{Nt}=\mathbb{E}\begin{pmatrix}\frac{\partial}{\partial u}m(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)'\otimes X_{it}\\ \frac{\partial}{\partial w}m(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)'\otimes\Phi_k^r(V_i)\end{pmatrix}_{(p+kr)\times q}. \tag{6}
$$

Proof of Lemma A.3

To prove the Lemma A.3, it is sufficient to show that $\|S_{1Nt}(\beta,D_t)-s_{1Nt}(\beta,\lambda'f_t)\|=o_P(1)$ and $\|S_{2Nt}(\beta,D_t)-s_{2Nt}(\beta,\lambda'f_t)\|=o_P(1)$.

$$
S_{1Nt}(\beta,D_t)-s_{1Nt}(\beta,\lambda'f_t)
$$

$$
\begin{aligned}
&= \frac{1}{q}\frac{1}{N}\sum_{l=1}^{q}\sum_{i=1}^{N}\left[m_l(W_{it}, X_{it}'\beta, \Phi_k^r(V_i)'D_t) - m_l(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t)\right] \\
&\times \frac{1}{N}\sum_{j=1}^{N}\frac{\partial}{\partial u}m_l(W_{jt}, X_{jt}'\beta, \Phi_k^r(V_j)'D_t)X_{jt} \\
&+ \frac{1}{q}\frac{1}{N}\sum_{l=1}^{q}\sum_{i=1}^{N}m_l(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t) \\
&\times \frac{1}{N}\sum_{j=1}^{N}\left(\frac{\partial}{\partial u}m_l(W_{jt}, X_{jt}'\beta, \Phi_k^r(V_j)'D_t) - \frac{\partial}{\partial u}m_l(W_{jt}, X_{jt}'\beta, \lambda(V_i)'f_t)\right)X_{jt} \\
&+ \frac{1}{q}\frac{1}{N}\sum_{l=1}^{q}\sum_{i=1}^{N}m_l(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t) \\
&\times \frac{1}{N}\sum_{j=1}^{N}\left(\frac{\partial}{\partial u}m_l(W_{jt}, X_{jt}'\beta, \lambda(V_i)'f_t)X_{jt} - \mathbb{E}\frac{\partial}{\partial u}m_l(W_{jt}, X_{jt}'\beta, \lambda(V_i)'f_t)X_{jt}\right) \\
&= I_1 + I_2 + I_3,
\end{aligned}
$$

where

$$
\begin{aligned}
\|I_1\|^2 &\leq \frac{1}{q}\sum_{l=1}^{q}\left(\frac{1}{N}\sum_{i=1}^{N}\left[m_l(W_{it}, X_{it}'\beta, \Phi_k^r(V_i)'D_t) - m_l(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t)\right]\right)^2 \\
&\times \frac{1}{q}\sum_{l=1}^{q}\left\|\frac{1}{N}\sum_{j=1}^{N}\frac{\partial}{\partial u}m_l(W_{jt}, X_{jt}'\beta, \Phi_k^r(V_j)'D_t)X_{jt}\right\|^2 \\
&= I_{11} \times I_{12}.
\end{aligned} \tag{7}
$$

Following the procedure of deriving (2), we can observe that I_{11} has the same order in probability as

$$
\begin{aligned}
&\frac{1}{q}\sum_{l=1}^{q}\mathbb{E}\left(\left[m_l(W_{it}, X_{it}'\beta, \Phi_k^r(V_i)'D_t) - m_l(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t)\right]\right)^2 \\
&= \frac{1}{q}\mathbb{E}\left\|\left[m_l(W_{it}, X_{it}'\beta, \Phi_k^r(V_i)'D_t) - m_l(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t)\right]\right\|^2 \\
&\leq \mathbb{E}|A(W_{11}, X_{11}, V_1)|^2\|\gamma^{(k)}(V_1)\|^2\|f_1\|^2 \\
&= O(\|\gamma^{(k)}(V_1)\|^2).
\end{aligned}
$$

Similarly, I_{12} has the same order in probability as

$$
\frac{1}{q}\mathbb{E}\left\|\frac{\partial}{\partial u}m(W_{jt}, X_{jt}'\beta, \Phi_k^r(V_j)'D_t) \otimes X_{jt}\right\|^2 = O(p).
$$

Hence, $I_1 = o_P(1)$ by Assumption 3.6.

For I_2, we have

$$
\begin{aligned}
\|I_2\|^2 &\leq \frac{1}{q}\sum_{l=1}^{q}\left(\frac{1}{N}\sum_{i=1}^{N} m_l(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t)\right)^2 \\
&\times \frac{1}{q}\sum_{l=1}^{q}\left\|\frac{1}{N}\sum_{j=1}^{N}\left(\frac{\partial}{\partial u} m_l(W_{jt}, X_{jt}'\beta, \Phi_k^r(V_j)'D_t)X_{jt} - \frac{\partial}{\partial u} m_l(W_{jt}, X_{jt}'\beta, \lambda(V_i)'f_t)X_{jt}\right)\right\|^2 \\
&= I_{21} \times I_{22}.
\end{aligned}
$$

By Assumption 3.1 and moment condition (1), we have

$$
\begin{aligned}
\mathbb{E}\|I_{21}\| &= \frac{1}{q}\sum_{l=1}^{q}\mathbb{E}\left(\frac{1}{N}\sum_{i=1}^{N} m_l(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t)\right)^2 \\
&= \frac{1}{q}\sum_{l=1}^{q}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}\Big[m_l(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t), m_l(W_{jt}, X_{jt}'\beta, \lambda(V_j)'f_t)\Big] \\
&= \frac{1}{q}\sum_{l=1}^{q}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathrm{cov}\Big[m_l(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t), m_l(W_{jt}, X_{jt}'\beta, \lambda(V_j)'f_t)\Big] \\
&= O(N^{-1}),
\end{aligned} \tag{8}
$$

where the last equality is obtained by the same procedure as (1).

Meanwhile, by (4), we have $I_{22} = O(\|\gamma^{(k)}(V_1)\|^2 p)$. Therefore $I_2 = o_P(1)$.

By (5) and (8), it is obvious that

$$
\begin{aligned}
\|I_3\|^2 &\leq \frac{1}{q}\sum_{l=1}^{q}\left(\frac{1}{N}\sum_{i=1}^{N} m_l(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t)\right)^2 \\
&\times \frac{1}{q}\sum_{l=1}^{q}\left\|\frac{1}{N}\sum_{j=1}^{N}\left(\frac{\partial}{\partial u} m_l(W_{jt}, X_{jt}'\beta, \lambda(V_i)'f_t)X_{jt} - \mathbb{E}\frac{\partial}{\partial u} m_l(W_{jt}, X_{jt}'\beta, \lambda(V_i)'f_t)X_{jt}\right)\right\|^2 \\
&= O_P(N^{-1})O_P(N^{-1}p) = o_P(1).
\end{aligned}
$$

This finished the proof of $\|S_{1Nt}(\beta, D_t) - s_{1Nt}(\beta, \lambda' f_t)\| = o_P(1)$.

Next, we show that $\|S_{2Nt}(\beta, D_t) - s_{2Nt}(\beta, \lambda' f_t)\| = o_P(1)$.

$$
S_{2Nt}(\beta, D_t) - s_{2Nt}(\beta, \lambda' f_t)
$$

$$
\begin{aligned}
&= \frac{1}{q}\frac{1}{N^2}\sum_{l=1}^{q}\sum_{i=1}^{N} m_l(W_{it}, X_{it}'\beta, \Phi_k^r(V_i)'D_t) \\
&\times \sum_{j=1}^{N}\frac{\partial}{\partial w} m_l(W_{jt}, X_{jt}'\beta, \Phi_k^r(V_j)'D_t)\Phi_k^r(V_j) \\
&- \frac{1}{q}\frac{1}{N}\sum_{l=1}^{q}\sum_{i=1}^{N} m_l(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t)\mathbb{E}\frac{\partial}{\partial w} m_l(W_{jt}, X_{jt}'\beta, \lambda(V_i)'f_t)\Phi_k^r(V_j) \\
&= \frac{1}{q}\frac{1}{N^2}\sum_{l=1}^{q}\sum_{i=1}^{N}\big(m_l(W_{it}, X_{it}'\beta, \Phi_k^r(V_i)'D_t) - m_l(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t)\big) \\
&\times \sum_{j=1}^{N}\frac{\partial}{\partial w} m_l(W_{jt}, X_{jt}'\beta, \Phi_k^r(V_j)'D_t)\Phi_k^r(V_j) \\
&+ \frac{1}{q}\frac{1}{N^2}\sum_{l=1}^{q}\sum_{i=1}^{N} m_l(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t) \\
&\times \sum_{j=1}^{N}\left(\frac{\partial}{\partial w} m_l(W_{jt}, X_{jt}'\beta, \Phi_k^r(V_j)'D_t) - \frac{\partial}{\partial w} m_l(W_{jt}, X_{jt}'\beta, \lambda(V_j)'f_t)\right)\Phi_k^r(V_j) \\
&+ \frac{1}{q}\frac{1}{N}\sum_{l=1}^{q}\sum_{i=1}^{N} m_l(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t) \\
&\times \left(\sum_{j=1}^{N}\frac{\partial}{\partial w} m_l(W_{jt}, X_{jt}'\beta, \lambda(V_j)'f_t)\Phi_k^r(V_j) - \mathbb{E}\frac{\partial}{\partial w} m_l(W_{jt}, X_{jt}'\beta, \lambda(V_i)'f_t)\Phi_k^r(V_j)\right) \\
&= I_4 + I_5 + I_6.
\end{aligned}
$$

By Cauchy-Schwarz inequality, we have

$$
\begin{aligned}
&\|I_4\|^2 \le \frac{1}{q}\sum_{l=1}^{q}\left(\frac{1}{N}\sum_{i=1}^{N}\big(m_l(W_{it}, X_{it}'\beta, \Phi_k^r(V_i)'D_t) - m_l(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t)\big)\right)^2 \\
&\times \frac{1}{q}\sum_{l=1}^{q}\left\|\frac{1}{N}\sum_{j=1}^{N}\frac{\partial}{\partial w} m_l(W_{jt}, X_{jt}'\beta, \Phi_k^r(V_j)'D_t)\Phi_k^r(V_j)\right\|^2 \\
&\le 2\frac{1}{q}\sum_{l=1}^{q}\left(\frac{1}{N}\sum_{i=1}^{N}\big(m_l(W_{it}, X_{it}'\beta, \Phi_k^r(V_i)'D_t) - m_l(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t)\big)\right)^2 \\
&\times \frac{1}{q}\sum_{l=1}^{q}\left\|\frac{1}{N}\sum_{j=1}^{N}\left(\frac{\partial}{\partial w} m_l(W_{jt}, X_{jt}'\beta, \Phi_k^r(V_j)'D_t) - \frac{\partial}{\partial w} m_l(W_{jt}, X_{jt}'\beta, \lambda(V_j)'f_t)\right)\Phi_k^r(V_j)\right\|^2 \\
&+ 2\frac{1}{q}\sum_{l=1}^{q}\left(\frac{1}{N}\sum_{i=1}^{N}\big(m_l(W_{it}, X_{it}'\beta, \Phi_k^r(V_i)'D_t) - m_l(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t)\big)\right)^2
\end{aligned}
$$

$$\times \frac{1}{q}\sum_{l=1}^{q}\left\|\frac{1}{N}\sum_{j=1}^{N}\frac{\partial}{\partial w}m_l(W_{jt},X_{jt}'\beta,\lambda(V_j)'f_t)\Phi_k(V_j)\right\|^2,$$

which by Assumption 3.7, the second term is the leading term and following the procedure of deriving (2), it has the same order as

$$\begin{aligned}
&\frac{1}{q}\sum_{l=1}^{q}\left(\mathbb{E}\left(m_l(W_{it},X_{it}'\beta,\Phi_k^r(V_i)'D_t)-m_l(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)\right)\right)^2\\
&\times \frac{1}{q}\sum_{l=1}^{q}\left\|\mathbb{E}\frac{\partial}{\partial w}m_l(W_{jt},X_{jt}'\beta,\lambda(V_j)'f_t)\Phi_k^r(V_j)\right\|^2\\
&=\frac{1}{q}\left(\mathbb{E}\left(m(W_{it},X_{it}'\beta,\Phi_k^r(V_i)'D_t)-m(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)\right)\right)^2\\
&\times \frac{1}{q}\left\|\mathbb{E}\frac{\partial}{\partial w}m_l(W_{jt},X_{jt}'\beta,\lambda(V_j)'f_t)\Phi_k^r(V_j)\right\|^2\\
&\leq \left|\mathbb{E}[A(W_{11},X_{11},V_1)\gamma^{(k)}(V_1)]\right|^2 O(kr)\leq O_P(\|\gamma^{(k)}(V)\|^2 kr)=o_P(1),
\end{aligned}$$

by Assumptions 3.6 and 3.7 as $N\to\infty$.

It is straightforward to see that $I_5=o_P(1)$ by Assumptions 3.6 and 3.7. Next, we have

$$\begin{aligned}
&\|I_6\|^2\leq \frac{1}{q}\sum_{l=1}^{q}\left(\frac{1}{N}\sum_{i=1}^{N}m_l(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)\right)^2\\
&\times \frac{1}{q}\sum_{l=1}^{q}\left\|\frac{1}{N}\sum_{j=1}^{N}\frac{\partial}{\partial w}m_l(W_{jt},X_{jt}'\beta,\lambda(V_j)'f_t)\Phi_k^r(V_j)-\mathbb{E}\frac{\partial}{\partial w}m_l(W_{jt},X_{jt}'\beta,\lambda(V_i)'f_t)\Phi_k^r(V_j)\right\|^2\\
&=I_{61}\times I_{62},
\end{aligned}$$

where $I_{61}=I_{21}=O(N^{-1})$ and

$$\begin{aligned}
\mathbb{E}[I_{62}]&=\frac{1}{q}\sum_{l=1}^{q}\frac{1}{N^2}\mathbb{E}\left\|\sum_{j=1}^{N}\frac{\partial}{\partial w}m_l(W_{jt},X_{jt}'\beta,\lambda(V_j)'f_t)\Phi_k^r(V_j)-\mathbb{E}\frac{\partial}{\partial w}m_l(W_{jt},X_{jt}'\beta,\lambda(V_i)'f_t)\Phi_k^r(V_j)\right\|^2\\
&=\frac{1}{q}\sum_{l=1}^{q}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathrm{cov}\left[\frac{\partial}{\partial w}m_l(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)\Phi_k(V_i),\frac{\partial}{\partial w}m_l(W_{jt},X_{jt}'\beta,\lambda(V_j)'f_t)\Phi_k^r(V_j)\right]\\
&=O(N^{-1}kr),
\end{aligned}$$

where the last equality is obtained by the same procedure as (1) combined with $\mathbb{E}\|\Phi_k(V_1)\|^2=O(k)$ and $\mathbb{E}\|f_1\|^2=O(r)$ in Assumption 3.5. Then by Assumption 3.6, $\|I_6\|^2=o_P(N^{-1}kr)=o_P(1)$.

This finished the proof of Lemma A.3.

Proof of Theorem 4.1.

We will prove Theorem 4.1 in two steps. In the first step, we show that $L_{Nt}(\overline{\widehat{\beta}}, \widehat{D}_t; c) \to_D N(0,1)$ for a given t. In the second step, we show $L_{NT}(\overline{\widehat{\beta}}, \widehat{D}; c) \to_D N(0,1)$ by conventional central limit theorem.

By conventional central limit theorem, we have

$$\left(\sum_{i=1}^{N}[c'm(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t)]^2\right)^{-\frac{1}{2}} \sum_{i=1}^{N} c'm(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t) \to_D N(0,1),$$

as $N \to \infty$ for any $c \in \mathbb{R}^q$ such that $\|c\| = 1$.

Thus, to prove Theorem 4.1, it is sufficient to show

$$L_{Nt}(\overline{\widehat{\beta}}, \widehat{D}_t; c) = \left(\sum_{i=1}^{N}[c'm(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t)]^2\right)^{-\frac{1}{2}} \sum_{i=1}^{N} c'm(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t) + o_p(1).$$

To this end, we will show

$$\frac{1}{N} D_{Nt}(\overline{\widehat{\beta}}, \widehat{D}_t; c)^2 - \frac{1}{N}\sum_{i=1}^{N}[c'm(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t)]^2 = o_p(1), \quad \text{and}$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} c'm(W_{it}, X_{it}'\overline{\widehat{\beta}}, \Phi_k^r(V_i)'\widehat{D}_t) - \frac{1}{\sqrt{N}}\sum_{i=1}^{N} c'm(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t) = o_p(1).$$

Notice that

$$\begin{aligned}
&\frac{1}{N} D_{Nt}(\overline{\widehat{\beta}}, \widehat{D}_t; c)^2 = \frac{1}{N}\sum_{i=1}^{N}[c'm(W_{it}, X_{it}'\overline{\widehat{\beta}}, \Phi_k^r(V_i)'\widehat{D}_t)]^2 \\
&= \frac{1}{N}\sum_{i=1}^{N}[c'm(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t)]^2 \\
&+ \frac{1}{N}\sum_{i=1}^{N}\left\{[c'm(W_{it}, X_{it}'\overline{\widehat{\beta}}, \Phi_k^r(V_i)'\widehat{D}_t)]^2 - [c'm(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t)]^2\right\}.
\end{aligned}$$

We will show that the second term is $o_P(1)$.

From Theorem 3.2, it is easy to see that $\|\overline{\widehat{\beta}} - \beta\|^2 = O_P(p/NT)$ and $\|\widehat{D}_t - D_t\|^2 = O_P(kr/N)$.

By the first order Taylor expansion, Cauchy-Schwarz inequality, Assumption 3.5 and Assumption 4.2, we have that

$$\frac{1}{N}\sum_{i=1}^{N}\left|[c'm(W_{it}, X_{it}'\overline{\widehat{\beta}}, \Phi_k^r(V_i)'\widehat{D}_t)]^2 - [c'm(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t)]^2\right|$$

$$
\begin{aligned}
&\leq \frac{1}{N}\sum_{i=1}^{N}|c'[m(W_{it},X_{it}'\overline{\widehat{\beta}},\Phi_k^r(V_i)'\widehat{D}_t)-m(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)]|^2\\
&+\ 2\frac{1}{N}\sum_{i=1}^{N}|c'[m(W_{it},X_{it}'\overline{\widehat{\beta}},\Phi_k^r(V_i)'\widehat{D}_t)-m(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)]||c'm(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)|\\
&\leq \frac{2}{N}\sum_{i=1}^{N}\left|c'\frac{\partial m(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)}{\partial u}(\overline{\widehat{\beta}}-\beta)'X_{it}\right|^2\\
&+\frac{2}{N}\sum_{i=1}^{N}\left|c'\frac{\partial m(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)}{\partial w}(\Phi_k^r(V_i)'\widehat{D}_t-\lambda(V_i)'f_t)\right|^2\\
&+\frac{2}{N}\sum_{i=1}^{N}\left|c'\frac{\partial m(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)}{\partial u}(\overline{\widehat{\beta}}-\beta)'X_{it}\right||c'm(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)|\\
&+\frac{2}{N}\sum_{i=1}^{N}\left|c'\frac{\partial m(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)}{\partial w}(\Phi_k^r(V_i)'\widehat{D}_t-\lambda(V_i)'f_t)\right||c'm(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)|\\
&\leq \|\overline{\widehat{\beta}}-\beta\|^2\frac{2}{N}\sum_{i=1}^{N}\left\|\frac{\partial m(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)}{\partial u}\otimes X_{it}\right\|^2\\
&+\|\widehat{D}_t-D_t\|^2\frac{4}{N}\sum_{i=1}^{N}\left\|\frac{\partial m(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)}{\partial w}\otimes\Phi_k^r(V_i)\right\|^2\\
&+\frac{4}{N}\sum_{i=1}^{N}\left\|\frac{\partial m(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)}{\partial w}\right\|^2\|\gamma^{(k)}(V_i)\|^2\|f_t\|^2\\
&+2\|\overline{\widehat{\beta}}-\beta\|\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{\partial m(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)}{\partial u}\otimes X_{it}\right\|^2\right)^{\frac{1}{2}}\\
&\quad\times\left(c'\frac{1}{N}\sum_{i=1}^{N}m((W_{it},X_{it}'\beta,\lambda(V_i)'f_t)m((W_{it},X_{it}'\beta,\lambda(V_i)'f_t)'c\right)^{\frac{1}{2}}\\
&+2\left(\frac{1}{N}\sum_{i=1}^{N}\left|c'\frac{\partial m(W_{it},X_{it}'\beta,\lambda(V_i)'f_t)}{\partial w}\right|^2(\Phi_k^r(V_i)'\widehat{D}_t-\lambda(V_i)'f_t)^2\right)^{1/2}\\
&\quad\times\left(c'\frac{1}{N}\sum_{i=1}^{N}m((W_{it},X_{it}'\beta,\lambda(V_i)'f_t)m((W_{it},X_{it}'\beta,\lambda(V_i)'f_t)'c\right)^{\frac{1}{2}}\\
&=\|\overline{\widehat{\beta}}-\beta\|^2O_p(qp)+\|\widehat{D}_t-D_t\|^2O_p(qkr)+\sup_v\|\gamma^{(k)}(v)\|^2O_p(q)\\
&\quad+\|\overline{\widehat{\beta}}-\beta\|O_p(\sqrt{qp})+\|\widehat{D}_t-D_t\|O_p(\sqrt{qkr})+\sup_v\|\gamma^{(k)}(v)\|O_p(\sqrt{q})=o_p(1).
\end{aligned}
$$

Define $\nu_{Nt}(b,g,s;c)=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}c'(m(W_{it},X_{it}'b,g(V_i)'s_t)-E[m(W_{it},X_{it}'b,g(V_i)'s_t)])$ for any $c\in\mathbb{R}^q$ such that $\|c\|=1$ and $(b,g,s)\in\Theta$.

Given Theorem 3.2, we will show that $\nu_{Nt}(\overline{\widehat{\beta}},\widehat{\lambda},\widehat{f};c)-\nu_{Nt}(\beta,\lambda,f;c)=o_P(1)$.

By the first order Taylor expansion, we have that

$$\begin{aligned}
&m(W_{it}, X_{it}'b, g(V_i)'s_t) - m(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t) \\
&= \frac{\partial m(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t)}{\partial u}(b-\beta)'X_{it} + \frac{\partial m(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t)}{\partial w}(g(V_i)'s_t - \lambda(V_i)'f_t) \qquad (9)
\end{aligned}$$

for all (b, g, s) in the neighborhood of (β, λ, f), where g has the form $\widetilde{\Phi}_k(V)'a$.

Therefore,

$$\begin{aligned}
&P\left(\sup_{\|(b,g,s)-(\beta,\lambda,f)\|<\delta} |\nu_{Nt}(b,g,s;c) - \nu_{Nt}(\beta,\lambda,f;c)| > \eta\right) \\
&\leq P\left(\sup_{\|(b,g,s)-(\beta,\lambda,f)\|<\delta} \left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N} c'\left[\frac{\partial m}{\partial u}(b-\beta)'X_{it} - \mathbb{E}\frac{\partial m}{\partial u}(b-\beta)'X_{it}\right]\right| > \frac{\eta}{2}\right) \\
&+ P\left(\sup_{\|(b,g,s)-(\beta,\lambda,f)\|<\delta} \left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N} c'\left[\frac{\partial m}{\partial w}(g(V_i)'s_t - \lambda(V_i)'f_t) - \mathbb{E}\frac{\partial m}{\partial w}(g(V_i)'s_t - \lambda(V_i)'f_t)\right]\right| > \frac{\eta}{2}\right) \\
&\leq P\left(\sup_{\|(b,g,s)-(\beta,\lambda,f)\|<\delta} \left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}[c'\frac{\partial m}{\partial u}X_{it} - \mathbb{E}c'\frac{\partial m}{\partial u}X_{it}]'(b-\beta)\right| > \frac{\eta}{2}\right) \\
&+ P\left(\sup_{\|(b,g,s)-(\beta,\lambda,f)\|<\delta} \left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}[c'\frac{\partial m}{\partial w}\Phi_k^r(V_i) - \mathbb{E}c'\frac{\partial m}{\partial w}\Phi_k^r(V_i)]'(d_t - D_t)\right| > \frac{\eta}{4}\right) \\
&+ P\left(\sup_{\|(b,g,s)-(\beta,\lambda,f)\|<\delta} \left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}[c'\frac{\partial m}{\partial w}f_t'\gamma^{(k)}(V_i) - \mathbb{E}c'\frac{\partial m}{\partial w}f_t'\gamma^{(k)}(V_i)]\right| > \frac{\eta}{4}\right) \\
&\leq P\left(\sup_{\|(b,g,s)-(\beta,\lambda,f)\|<\delta} \left\|\frac{1}{\sqrt{Np}}\sum_{i=1}^{N}[c'\frac{\partial m}{\partial u}X_{it} - \mathbb{E}c'\frac{\partial m}{\partial u}X_{it}]\right\| \|\sqrt{p}(b-\beta)\| > \frac{\eta}{2}\right) \\
&+ P\left(\sup_{\|(b,g,s)-(\beta,\lambda,f)\|<\delta} \left\|\frac{1}{\sqrt{Nkr}}\sum_{i=1}^{N}[c'\frac{\partial m}{\partial w}\Phi_k^r(V_i) - \mathbb{E}c'\frac{\partial m}{\partial w}\Phi_k^r(V_i)]\right\| \|\sqrt{kr}(d_t - D_t)\| > \frac{\eta}{4}\right) \\
&+ P\left(\sup_{\|(b,g,s)-(\beta,\lambda,f)\|<\delta} \left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}[c'\frac{\partial m}{\partial w}f_t'\gamma^{(k)}(V_i) - \mathbb{E}c'\frac{\partial m}{\partial w}f_t'\gamma^{(k)}(V_i)]\right| > \frac{\eta}{4}\right) \\
&= I_{1Nt} + I_{2Nt} + I_{3Nt}.
\end{aligned}$$

By central limit theorem, we have

$$\begin{aligned}
\frac{1}{\sqrt{Np}}\sum_{i=1}^{N}\left[c'\frac{\partial m}{\partial u}X_{it} - \mathbb{E}c'\frac{\partial m}{\partial u}X_{it}\right] &= O_p(1), \\
\frac{1}{\sqrt{Nkr}}\sum_{i=1}^{N}\left[c'\frac{\partial m}{\partial w}\Phi_k^r(V_i) - \mathbb{E}c'\frac{\partial m}{\partial w}\Phi_k^r(V_i)\right] &= O_p(1).
\end{aligned}$$

If $\|\sqrt{p}(b-\beta)\|$ and $\|\sqrt{kr}(d_t - D_t)\|$ are sufficiently small, then $I_{1Nt} < \frac{\varepsilon}{3}$ and $I_{2Nt} < \frac{\varepsilon}{3}$. Meanwhile, by Assumption 4.2, we have $\sqrt{q}\sup_v \|\gamma^{(k)}(v)\| = o(1)$, we have $I_{3Nt} < \frac{\varepsilon}{3}$.

Therefore, when N are large, for a given t, we have $P(|\nu_{Nt}(\overline{\widehat{\beta}}, \widehat{\lambda}, \widehat{f}; c) - \nu_{Nt}(\beta, \lambda, f; c)| > \eta) < \varepsilon$ for any given $\varepsilon, \eta > 0$. Recall that $\widehat{\lambda}(V_i)\widehat{f}_t = \Phi_k^r(V_i)'\widehat{D}_t$, thus $m(W_{it}, X_{it}'\overline{\widehat{\beta}}, \widehat{\lambda}(V_i)\widehat{f}_t) = m(W_{it}, X_{it}'\overline{\widehat{\beta}}, \Phi_k^r(V_i)'\widehat{D}_t)$. Furthermore,

$$\begin{aligned}
&\frac{1}{\sqrt{N}}\sum_{i=1}^{N} c'[m(W_{it}, X_{it}'\overline{\widehat{\beta}}, \Phi_k^r(V_i)'\widehat{D}_t) - m(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t)] \\
=&\nu_{NT}(\overline{\widehat{\beta}}, \widehat{\lambda}, \widehat{f}; c) - \nu_{NT}(\beta, \lambda, f; c) + \overline{m}_{NT}^*(\overline{\widehat{\beta}}, \widehat{D}_t; c).
\end{aligned}$$

By Assumption 4.1, we have

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} c'm(W_{it}, X_{it}'\overline{\widehat{\beta}}, \Phi_k^r(V_i)'\widehat{D}_t) - \frac{1}{\sqrt{N}}\sum_{i=1}^{N} c'm(W_{it}, X_{it}'\beta, \lambda(V_i)'f_t) = o_p(1).$$

Then we finish the proof.

References

Gao, J. (2007), *Nonlinear Time Series: Semi– and Non–Parametric Methods*, Chapman and Hall/CRC.