Department of Econometrics and Business Statistics

An Integrated Panel Data Approach to Modelling Economic Growth

Guohua Feng, Jiti Gao and Bin Peng

March 2019

An Integrated Panel Data Approach
to Modelling Economic Growth

Guohua Feng*, Jiti Gao♯ and Bin Peng†

*University of North Texas, ♯Monash University and †University of Bath

March 18, 2019

Abstract

Empirical growth analysis has three major problems — variable selection, parameter heterogeneity and cross-sectional dependence — which are addressed independently from each other in most studies. The purpose of this study is to propose an integrated framework that extends the conventional linear growth regression model to allow for parameter heterogeneity and cross-sectional error dependence, while simultaneously performing variable selection. We also derive the asymptotic properties of the estimator under both low and high dimensions, and further investigate the finite sample performance of the estimator through Monte Carlo simulations. We apply the framework to a dataset of 89 countries over the period from 1960 to 2014. Our results reveal some cross-country patterns not found in previous studies (e.g., "middle income trap hypothesis", "natural resources curse hypothesis", "religion works via belief, not practice", etc.).

Keywords: Growth Regressions, Variable Selection, Parameter Heterogeneity, Cross-Sectional Dependence

JEL classification: C23, O47

* Department of Economics, University of North Texas, Denton, TX 76201, USA. Email: Guohua.Feng@unt.edu

♯ Department of Econometrics and Business Statistics, Monash University, Caulfield East, Victoria 3145, Australia. Email: Jiti.Gao@monash.edu

† Department of Economics, University of Bath, Bath BA2 7JP, UK. Email: B.Peng2@bath.ac.uk

1 Introduction

Following the seminal works of Kormendi and Meguire (1985) and Barro (1991), a vast amount of studies in the empirical growth literature have attempted to identify salient determinants of economic growth. A main tool used by these studies is "cross-country growth regressions" — that is, to regress observed GDP growth on a plethora of possible explanatory variables that could possibly affect growth across countries. Excellent surveys of these studies and their role in the broader context of economic growth theory are provided in Durlauf and Quah (1999), Temple (1999) and Durlauf et al. (2005).

Despite the vast amount of research, the literature has identified a number of problems with conventional growth regressions, among which three deserve particular attention. The first problem is determining what variables to be included in growth regressions. This problem arises because of the nature of growth theories: although a plethora of growth theories have been proposed to identify factors that affect growth, these theories are open-ended in the sense that the validity of one causal theory of growth does not imply the falsity of another (Brock and Durlauf, 2001). In words of Durlauf et al. (2008), "a given body of candidate growth theories defines a space of possible models rather than a single specification". From an empirical perspective, this problem stems from the fact that the number of potential explanatory variables is large (over 140 identified in Durlauf et al., 2005) relative to the number of countries with enough data availability, rendering the all-inclusive regression computationally infeasible (Sala-I-Martin et al., 2004; Durlauf et al., 2005). In dealing with this problem some studies have resorted to simply "trying" combinations of variables which could be potentially important determinants of growth and report the results of their preferred specification. However, as noted by Leamer (1983) and Sala-I-Martin et al. (2004) such "data-mining" could lead to spurious inference.

The second problem with conventional growth analysis is that most empirical growth studies assume that the parameters of growth regressions are identical across countries. This assumption complies with the classical Solow model (Mankiw et al., 1992), which assumes that all countries share an identical aggregate Cobb-Douglas production function. However, an increasing number of studies (e.g., Durlauf and Johnson, 1995; Durlauf et al., 2001; Salimans, 2012) have suggested that the parameters are heterogeneous across countries. These studies, though using different econometric methods, all suggest that the assumption of a single linear growth model that applies to all countries is inappropriate. For example, Durlauf and Johnson (1995) employ a regression tree analysis to show that a cross-sectional regression using the Summers and Heston (1991) data appears to provide

support for several distinct regimes in which aggregate production functions vary among countries according to their level of development, while Durlauf et al. (2001), employing a varying coefficient growth model, also find strong evidence of parameter heterogeneity across countries.

The third problem is that few studies in the empirical growth literature allow for cross-sectional dependence of individual countries. Panel data econometrics has recently seen an increasing interest in models with unobserved time-varying heterogeneity caused by latent common shocks influencing all units, possibly to a different degree. This type of heterogeneity introduces cross-sectional dependence to individual countries, which, when neglected, can lead to biased estimates and spurious inference (Pesaran, 2006; Bai, 2009). In the context of cross-country growth analysis, the problem of cross-sectional dependence seems particularly salient due to the omnipresence of common global shocks (such as global financial crises and world oil price shocks) that affect all countries through trade and financial linkages (Chudik et al., 2017). Durlauf and Quah (1999) discuss the possibility of cross-sectional dependence in a Lucas (1993) growth model with human capital spillovers. They find that these spillovers markedly change the dynamics of convergence and the authors call for the modelling of cross-country interactions in empirical convergence analysis.

The three aforementioned problems have received more or less individual attention in the growth literature. For example, Durlauf et al. (2001) address the problem of parameter heterogeneity using a varying coefficient growth model, but do not deal with the problems of variable selection and cross-sectional dependence; both Sala-I-Martin et al. (2004) and Moral-Benito (2012) select growth determinants using Bayesian averaging, but do not account for parameter heterogeneity and cross-sectional dependence.

The main goal of this study is to propose an integrated framework that is capable of dealing with parameter heterogeneity and cross-sectional dependence, while simultaneously performing variable selection. Specifically, parameter heterogeneity is allowed for by permitting the coefficients to vary across countries according to a country's initial conditions, while cross-sectional dependence is accounted for via a factor structure. We then propose a least absolute shrinkage and selection operator (LASSO) estimator to select growth determinants, establish the associated asymptotic results, and further verify our asymptotic results through extensive simulations, which constitutes another contribution of this paper. We apply this framework to a new dataset of 89 countries over the period 1960-2014. Our findings broadly support the more "optimistic" conclusion of Sala-I-Martin (1997), that is, some variables are important regressors for explaining cross-

country growth patterns. Moreover, our empirical results also provide support to some important hypotheses in the growth literature, e.g., "middle income trap hypothesis", "natural resources curse hypothesis", "religion works via belief, not practice", etc.

The rest of the paper is organized as follows. Section 2 explains how to extend the canonical cross-country growth regression to account for the aforementioned issues. Section 3 describes a procedure for estimating the extended growth regression model in Section 2, and presents the associated asymptotic properties. Section 4 describes the data. The empirical results are presented in Section 5. Section 6 concludes. Due to space limitations, preliminary lemmas, proofs of the main theorems and Monte Carlo simulations, together with auxiliary tables and figures, are presented in the supplementary Appendix A of this paper. The proofs of the preliminary lemmas are presented in the supplementary Appendix B of this paper, which can be found at the authors' website (https://papers.ssrn.com/sol3/cf_dev/AbsByAuth.cfm?per_id=646779).

2 A Varying Coefficient Growth Regression Model with Factor Structure and Sparsity

A generic representation of the canonical cross-country growth regression is

$$y_{it} = x'_{it}\beta_0 + e_{it}, \tag{2.1}$$

where $i = 1, 2, \ldots, N$ index countries; $t = 1, 2, \ldots, T$ index time; y_{it} is the rate of economic growth; x_{it} represents a set of observable explanatory variables, including those originally suggested by Solow as well as other growth theories, and e_{it} is an error term. Equation (2.1) represents the baseline for much of growth econometrics.

However, as discussed in the Introduction, (2.1) is based on two problematic assumptions. First, it assumes that the parameters (i.e., β_0) are homogeneous across all countries. Second, it assumes that there is no cross-sectional dependence across countries.

To relax the two assumptions, in what follows we extend the conventional cross-country growth regression in (2.1) in two ways. First, in Section 2.1 we allow for parameter heterogeneity by allowing β_0 to vary across countries according to a country's initial conditions. Second, in Section 2.2 we introduce cross-sectional dependence into the model by means of a factor structure.

2.1 Parameter Heterogeneity

Following Durlauf et al. (2001), we allow for parameter heterogeneity by generalizing (2.1) into a varying coefficient model:

$$y_{it} = x'_{it}\beta_0(z_{it}) + e_{it}, \tag{2.2}$$

where z_{it} can be interpreted as some measure of "development" (or initial condition) of a country, and $\beta_0(z) = (\beta_{01}(z), \ldots, \beta_{0p}(z))'$ is a vector of smooth functions that maps the scalar index variable z_{it} into a set of country-specific parameters.

This generalization in (2.2) provides a framework within which one can bridge the gap between cross-country regression models and new growth theories. For instance, if one believes that initial GDP per capita causally affects a country's production technology and growth as in Durlauf et al. (2001), then initial GDP per capita can be introduced as a "development" index. As pointed out by Durlauf and Johnson (1995), (2.2) is compatible both with a model in which economies pass through distinct phases of development towards a unique steady state as well one in which multiple steady states exist.

2.2 Cross-Sectional Error Dependence

Having accounted for parameter heterogeneity, we next introduce the cross-sectional dependence of error terms into (2.2) using a factor structure:

$$e_{it} = \gamma'_{0i}f_{0t} + \varepsilon_{it}, \tag{2.3}$$

where f_{0t} is an $r \times 1$ vector of unobservable common factors, γ_{0i} is an $r \times 1$ vector of factor loadings that capture country-specific responses to the common shocks, and ε_{it} is the idiosyncratic error term. These common factors can be a combination of "strong" factors, such as world oil price shocks, global financial crises, and recessions in major advanced economies; and "weak" factors, such as local spillover effects along channels determined by shared culture heritage, geographic proximity, economic or social interaction (Chudik et al., 2011). Moreover, the components of the factor structure are allowed to drive both economic growth and explanatory variables, thus partially accounting for potential endogeneity of explanatory variables, which is neglected by the traditional approaches to causal interpretation of cross-country empirical analysis.

2.3 The Varying Coefficient Growth Regression Model with Factor Structure and Sparsity

Substituting (2.3) into (2.2) yields the following growth regression model that allows for parameter heterogeneity and cross-sectional dependence

$$y_{it} = x'_{it}\beta_0(z_{it}) + \gamma'_{0i}f_{0t} + \varepsilon_{it}, \ t = 1, 2, \cdots, T, \tag{2.4}$$

The model in (2.4) extends the local Solow growth model investigated in Durlauf et al. (2001) into a panel data context with interactive fixed effects (or factor structure). From an econometric perspective, (2.4) extends the panel data model with interactive fixed effects in Bai (2009) into a varying coefficient context. Some closely related studies include, but are not limited to, Dong et al. (2018) on (2.4) with partially observed factor structure, Feng et al. (2017) on a special case of (2.4) with discrete z_{it} and $\gamma'_{0i}f_{0t}$ being reduced to fixed-effects α_i, Liu et al. (2018) on a time-varying heterogeneous model with $z_{it} = \frac{t}{T}$, and Malikov et al. (2016) on a binary varying-coefficient panel data setting with endogenous selection and fixed-effects. There are some key differences between this paper and the relevant literature. First, it is worth pointing out that these previous studies introduce parameter heterogeneity and cross-sectional dependence in different manners. Second, none of them consider performing variable selection on their varying coefficient models as we will show below particularly in the high-dimensional setting. Finally and most importantly, both model (2.4) and its asymptotic theory in Section 3 below are naturally motivated by the relevant empirical literature in economic growth.

In addition to parameter heterogeneity and cross-sectional dependence, we are also interested in another issue that is prominent in the empirical growth literature — variable selection. This issue is important because (1) the dimension of x_{it} can be very large; and (2) not all elements of x_{it} drive economic growth. In other words, for those factors not driving economic growth, it is reasonable to assume that their associated coefficients are zero, which is called "sparsity" in the literature of high dimensional econometrics.

In order to formally introduce the sparsity to the model (2.4), we assume that there exists an unknown set $\mathcal{A}^\dagger \subseteq \{1, \ldots, p\}$ satisfying that $E|\beta_{0j}(z_{it})|^2 = 0$ if and only if $j \in \mathcal{A}^\dagger$. For notational simplicity, we assume $\mathcal{A}^\dagger = \{p^* + 1, \ldots, p\}$ for an unknown integer p^* satisfying $1 \le p^* < p$. Further, let $\mathcal{A}^* = \{1, \ldots, p^*\}$, $x^*_{it} = (x_{it,1}, \ldots, x_{it,p^*})'$, and $\beta^*_0(z) = (\beta_{01}(z), \ldots, \beta_{0p^*}(z))'$. Throughout this study, we always define the variables or functions corresponding to the sets $\mathcal{A}^*$ and $\mathcal{A}^\dagger$ with super-indices $*$ and $\dagger$ respectively. Thus, identifying growth determinants is equivalent to distinguishing $\mathcal{A}^*$ and $\mathcal{A}^\dagger$, which will be

achieved by a LASSO estimator presented in the following section. Finally, regarding the dimension of regressors, we consider two cases where (1) p is fixed, and (2) p diverges as the sample size increases. We refer to them as the low dimensional (LD) case and the high dimensional (HD) case, respectively. In terms of econometric methodology, both cases with the sparsity setting have not been studied in the literature to the best of our knowledge.

As discussed in the Introduction, failure to perform variable selection may result in spurious inference, failure to allow parameters to differ across countries is inconsistent with the increasing body of research that find cross-country parameter heterogeneity, and failure to account for cross-sectional dependence can lead to biased estimates and spurious inference. These possible consequences thus necessitate an integrated approach to simultaneously addressing the three issues.

In the following section, we introduce a LASSO estimator that is designed specifically for performing variable selection on the extended growth regression model in (2.4). The combination of varying coefficients, factor structure, and the LASSO estimator provides us an integrated framework that is capable of simultaneously addressing the three problems mentioned in Section 1 — variable selection, parameter heterogeneity, and cross-sectional dependence.

3   Estimation

In this section, we describe a procedure for estimating the model in (2.4) and derive the associated asymptotic properties. Specifically, we propose a LASSO estimator to select the appropriate variables, adopt a sieve method to recover the functional components, and employ the principle component analysis (PCA) technique to estimate the unobservable factor structure.

Before proceeding further, it is convenient to introduce some notations. We let $Y_i = (y_{i1}, \ldots, y_{iT})'$, $X_i = (x_{i1}, \ldots, x_{iT})'$, $Z_i = (z_{i1}, \ldots, z_{iT})'$, $\mathcal{E}_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iT})'$, $F_0 = (f_{01}, \ldots, f_{0T})'$, and $\Gamma_0 = (\gamma_{01}, \ldots, \gamma_{0N})'$. $\|\cdot\|$ denotes the Euclidean norm of a vector or the Frobenius norm of a matrix; for a square matrix W, let $\eta_{\min}(W)$ and $\eta_{\max}(W)$ stand for the minimum and maximum eigenvalues of W respectively; $M_W = I_T - P_W$ denotes the orthogonal projection matrix generated by matrix W, where $P_W = W(W'W)^{-1}W'$, and W is a matrix with full column rank.

We adopt the sieve method (e.g., Dong and Linton, 2018) to estimate the functional

component of (2.4). Specifically, assume that $\beta_{0\ell} \in L^2(V_z)$ for $\ell = 1, \ldots, p$, where $L^2(V_z) = \{g \mid \int_{V_z} g^2(z)dz < \infty\}$ is a Hilbert space. Suppose that there exists an orthonormal function sequence $\{h_j(z) \mid j \geq 0\}$ in $L^2(V_z)$ such that $\sup_{z \in V_z} \sup_{j \geq 0} |h_j(z)| < \infty$. Then, for $\forall g(z) \in L^2(V_z)$, we have an orthogonal series expansion $g(z) := g_m(z) + \delta_m(z)$, where $g_m(z) = \sum_{j=0}^{m-1} c_j h_j(z)$, $\delta_m(z) = \sum_{j=m}^{\infty} c_j h_j(z)$, $c_j = \int_{V_z} g(z)h_j(z)dz$, and m is the so-called truncation parameter. By the Parseval equality, the norm can be expressed as $\|g\|_{L^2} = \{\int_{V_z} g^2(z)dz\}^{1/2} = \{\sum_{j=0}^{\infty} c_j^2\}^{1/2}$. For a vector of functions $G(z) = (g_1(z), \ldots, g_d(z))'$, its norm is defined by $\|G\|_{L^2} = \{\sum_{\ell=1}^{d} \|g_\ell\|_{L^2}^2\}^{1/2}$.

Without loss of generality, truncating the expansions of all the elements of $\beta_0(z)$ by the same m gives

$$\beta_0(z) = \beta_{0,m}(z) + \Delta_m(z), \tag{3.1}$$

where $\beta_{0,m}(z) = C_{\beta_0} H_m(z)$, $H_m(z) = (h_0(z), h_1(z) \ldots, h_{m-1}(z))'$, $C_{\beta_0} = (C_{\beta_0}^{*}{}', 0'_{(p-p^*) \times m})'$, and $\Delta_m(z) = (\Delta_m^{*}(z)', 0'_{(p-p^*) \times 1})'$. Thus, the first p^* elements of $\beta_0(z)$ can be expressed by $\beta_0^*(z) = \beta_{0,m}^*(z) + \Delta_m^*(z)$, where $\beta_{0,m}^*(z) = C_{\beta_0}^* H_m(z)$.

We are now ready to rewrite (2.4) as

$$M_{F_0} Y_i = M_{F_0} \phi_i[\beta_{0,m}] + M_{F_0} \phi_i[\Delta_m] + M_{F_0} \mathcal{E}_i$$

by projecting out the factor structure, where $\phi_i[\beta] = (x_{i1}'\beta(z_{i1}), \ldots, x_{iT}'\beta(z_{iT}))'$ for any $p \times 1$ vector of functions $\beta(z)$. The objective function is then defined by

$$Q_\lambda(C_\beta, F) = \sum_{i=1}^{N} (Y_i - \phi_i[\beta_m])' M_F (Y_i - \phi_i[\beta_m]) + \sum_{j=1}^{p} \lambda_j \|C_{\beta,j}\|, \tag{3.2}$$

where $\beta_m(w) = C_\beta H_m(w)$, $C_{\beta,j}$ stands for the jth row of C_β, and $\{\lambda_1, \ldots, \lambda_p\}$ are the regularizers of the coefficient functions and are to be determined by data. The estimators of C_{β_0} and F_0 for both LD and HD cases are always obtained by

$$(\widehat{C}_\beta, \widehat{F}) = \underset{C_\beta, F \in \mathsf{D}_F}{\operatorname{argmin}} Q_\lambda(C_\beta, F), \tag{3.3}$$

where $\mathsf{D}_F = \{F \mid \frac{F'F}{T} = I_r\}$. In what follows, we always partition $\widehat{C}_\beta$, according to the partition $\mathcal{A}^*$ and $\mathcal{A}^\dagger$, as $\widehat{C}_\beta = (\widehat{C}_\beta^{*}{}', \widehat{C}_\beta^{\dagger}{}')'$ wherever necessary.

At this point it is convenient to state some fundamental assumptions that are needed for the derivation of the asymptotic results for both LD and HD cases.

Assumption 1.

1. Let $\mathcal{F}^0_{-\infty}$ and $\mathcal{F}^\infty_\tau$ denote the σ-algebras generated by $\{(x_t, z_t, \varepsilon_t, f_{0t}) \mid t \leq 0\}$ and $\{(x_t, z_t, \varepsilon_t, f_{0t}) \mid t \geq \tau\}$ respectively, where $x_t = (x_{1t}, \ldots, x_{Nt})'$, $z_t = (z_{1t}, \ldots, z_{Nt})'$, $\varepsilon_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Nt})'$. Let $\alpha(\tau) = \sup_{A \in \mathcal{F}^0_{-\infty}, B \in \mathcal{F}^\infty_\tau} |\Pr(A)\Pr(B) - \Pr(AB)|$ be the mixing coefficient.

   (a) $\{X_i, Z_i, \mathcal{E}_i, \gamma_{0i}\}$ is identically distributed over i. $\{x_t, z_t, \varepsilon_t, f_{0t}\}$ is strictly stationary and α-mixing such that for some $\nu_1 > 0$, $E[\|\varepsilon_{11}\| + \|x_{11}\|]^{4+\nu_1} < \infty$, and the mixing coefficient satisfies $\sum_{t=1}^\infty [\alpha(t)]^{\nu_1/(2+\nu_1)} < \infty$.

   (b) $E[\varepsilon_{11}] = 0$, $E[\varepsilon_{11}^2] = \sigma_\varepsilon^2$, and $\{\varepsilon_{it}\}$ is independent of the other variables. Let $E[\varepsilon_{i1}\varepsilon_{j1}] = \sigma_{ij}$ for $i \neq j$, $\sum_{i \neq j} |\sigma_{ij}| = O(N)$, and $\sum_{i,j=1}^N \sum_{t,s=1}^T |E[\varepsilon_{it}\varepsilon_{js}]| = O(NT)$.

2. Let $\left\|\frac{1}{T}F_0'F_0 - \Sigma_f\right\| = O_P\left(\frac{1}{\sqrt{T}}\right)$ and $\left\|\frac{1}{N}\Gamma_0'\Gamma_0 - \Sigma_\gamma\right\| = O_P\left(\frac{1}{\sqrt{N}}\right)$, where Σ_f and Σ_γ are deterministic and positive definite. Moreover, $E\|f_{01}\|^4 < \infty$ and $E\|\gamma_{01}\|^4 < \infty$.

Assumption 2.

1. Suppose that $\sup_{z \in V_z} \|\Delta_m(z)\| = O(m^{-\mu/2})$, and $\eta_{\max}(\frac{1}{NT}\sum_{i=1}^N \mathcal{Z}_i'\mathcal{Z}_i') < \infty$ with probability approaching one, where $\mathcal{Z}_i = (\mathcal{Z}_{i1}, \ldots, \mathcal{Z}_{iT})'$ and $\mathcal{Z}_{it} = H_m(z_{it}) \otimes x_{it}$.

2. Let $\Omega_1(F) = \frac{1}{NT}\sum_{i=1}^N \mathcal{Z}_i'M_F\mathcal{Z}_i$, $\Omega_2(F) = \frac{1}{NT}\sum_{i=1}^N \gamma_{0i}\otimes(M_F\mathcal{Z}_i)$ and $\Omega_3 = \frac{\Gamma_0'\Gamma_0}{NT}\otimes I_T$. Suppose $\inf_{F \in \mathsf{D}_F} \eta_{\min}(\Omega(F)) > 0$, where $\Omega(F) = \Omega_1(F) - \Omega_2'(F)\Omega_3^{-1}\Omega_2(F)$.

Assumption 1 is standard in the literature. The mixing conditions are similar to Assumption C of Bai (2009) and Assumption 3.4 of Fan et al. (2016). In Assumption 2.1, the condition $\sup_{z \in V_z} \|\Delta_m(z)\| = O(m^{-\mu/2})$ is the same as Assumption 3 of Newey (1997), and essentially requires certain smoothness of the elements of $\beta_0(\cdot)$. The condition $\eta_{\max}(\frac{1}{NT}\sum_{i=1}^N \mathcal{Z}_i'\mathcal{Z}_i') < \infty$ of Assumption 2.1 is similar to Assumption 3.1 of Fan et al. (2016). Now, consider a special case where $\{z_{it}\}$, $\{x_{it}\}$ and $\{f_{0t}\}$ are mutually independent, $E[x_{it}] = 0$ and $E[x_{it}x_{it}'] = I_p$. Under this setting, it is easy to see that $\eta_{\max}(\frac{1}{NT}\sum_{i=1}^N \mathcal{Z}_i'\mathcal{Z}_i') < \infty$ holds true for both of the LD and HD cases by some standard analyses. Assumption 2.2 ensures that the estimators given in (3.3) are well defined, and is equivalent to Assumption A of Bai (2009).

Based on the above setting, we move on to investigate the asymptotic results associated with (3.3) under the LD setting.

3.1 Low Dimensional Case

In order to identify $\mathcal{A}^*$ and $\mathcal{A}^\dagger$ and establish the asymptotic distribution, we further make the following assumptions.

Assumption 3.

1. $\frac{m^2}{T} \to 0$ and $\frac{\lambda^*_{\max}}{N^{\frac{3}{4}}T} \to 0$, where $\lambda^*_{\max} = \max\{\lambda_1, \ldots, \lambda_{p^*}\}$.

2. $\frac{N}{T} \to \kappa_0$ and $\frac{\lambda^\dagger_{\min}}{m^{\frac{1}{2}}N^{\frac{7}{8}}T} \to \kappa_1$, where $\lambda^\dagger_{\min} = \min\{\lambda_{p^*+1}, \ldots, \lambda_p\}$, $0 \le \kappa_0 < \infty$, and $\kappa_1 > 0$.

Assumption 4.

1. Suppose that for $t \ge s$, $E[f'_{0t}f_{0s} \mid \mathcal{X}_{Nt}] = a_{ts}$, and $\sum_{t=1}^{T}\sum_{s=1}^{t}|a_{ts}| = O(T)$, where $\mathcal{X}_{Nt} := \{(x_{1t}, z_{1t}), \ldots, (x_{Nt}, z_{Nt})\}$. Moreover, $\frac{NT}{m^{\mu+1}} \to 0$, $\frac{mN}{T} \to 0$, $\frac{T}{N^2} \to 0$ and $\frac{m\lambda^*_{\max}}{\sqrt{NT}} \to 0$.

2. Let $\Sigma^*_{\mathcal{Z}} = E[\mathcal{Z}^*_{11}\mathcal{Z}^{*}_{11}{}']$ and $\Omega_\star = \lim_{(N,T)\to(\infty,\infty)} E[\Psi_1\Psi'_1]$ for $\forall z \in V_z$, where

$$\Psi_1 = \sqrt{\frac{NT}{m}}\left[H'_m(z) \otimes I_{p^*}\right]\Psi_2^{-1}\Sigma^{*}_{\mathcal{Z}}{}^{-1} \cdot \frac{1}{NT}\sum_{i=1}^{N}\left\{\mathcal{Z}'_i + \frac{1}{N}\sum_{j=1}^{N}\mathcal{Z}^{*}_j{}'\gamma'_{0j}\Sigma^{-1}_\gamma\gamma_{0i}\right\}\mathcal{E}_i,$$

$$\Psi_2 = I_{mp^*} - \Sigma^{*}_{\mathcal{Z}}{}^{-1}E[\mathcal{Z}^*_{11}\gamma'_{01}]\Sigma^{-1}_\gamma E[\gamma_{01}\mathcal{Z}^{*}_{11}{}'], \quad \text{and} \quad \mathcal{Z}^*_{it} = H_m(z_{it}) \otimes x^*_{it}.$$

Suppose that for $\forall z \in V_z$, as $(N,T) \to (\infty,\infty)$, $\Psi_1 \to_D N(0, \Omega_\star)$.

The conditions of Assumption 3, though seemingly complicated, can be easily satisfied. For example, let $N = \lfloor T^{b_0} \rfloor$, $m = \lfloor T^{b_1} \rfloor$, $\lambda^*_{\max} = T^{b_2}$, and $\lambda^\dagger_{\min} = T^{b_3}$, where $\lfloor a \rfloor$ means the largest integer part of a real number a. Then Assumption 3 essentially requires that $0 < b_0 \le 1$, $0 < b_1 < \frac{1}{2}$, $b_2 < \frac{3}{4}b_0 + 1$ and $b_3 \ge \frac{b_1}{2} + \frac{7b_0}{8} + 1$.

The current requirements of Assumption 4.1 are in the same spirit as Connor et al. (2012, Eq. 3 and Eq. 20) and Jiang et al. (2017, pp. 21-22). Without this assumption, some other types of conditions would be needed to achieve asymptotic normality. For example, one can require $N/T \to \rho$ with $0 < \rho < \infty$ and establish the normality with biases as in Theorem 3 of Bai (2009). Assumption 4.2 is equivalent to Assumption E of Bai (2009). It is worth mentioning that deriving the rates of convergence in Lemma A.3 and Lemma A.5 of the supplementary Appendix A does not require Assumption 4 at all. For better presentation and in order not to deviate from our main goal, we present these lemmas in the supplementary Appendix A instead of the main text.

Theorem 3.1. *Let Assumptions 1-3 hold.*

1. $\Pr(\|\widehat{C}_\beta^\dagger\| = 0) \to 1$.

2. *Suppose Assumption 4 also holds. Then* $\sqrt{\frac{NT}{m}}(\widehat{\beta}_m^*(z) - \beta_0^*(z)) \to_D N(0, \Omega_\star)$ *for* $\forall z \in V_z$, *as* $(N, T) \to (\infty, \infty)$, *where* $\widehat{\beta}_m^*(z) = \widehat{C}_\beta^* H_m(z)$.

The first result of Theorem 3.1 indicates that we are able to distinguish $\mathcal{A}^*$ and $\mathcal{A}^\dagger$; while the second result of Theorem 3.1 establishes the asymptotic distribution of the coefficient functions associated with the variables which truly drive economic growth.

To complete our discussion on the LD case, we propose the following BIC type criteria in order to select λ practically.

$$\mathrm{BIC}_\lambda = \ln \mathrm{RSS}_\lambda + \mathrm{df}_\lambda \frac{\ln N}{\sqrt[4]{N}}, \tag{3.4}$$

where $\mathrm{RSS}_\lambda = \frac{1}{NT} \sum_{i=1}^N \left(Y_i - \phi_i[\widehat{\beta}_m^\lambda]\right)' M_{\widehat{F}^\lambda} \left(Y_i - \phi_i[\widehat{\beta}_m^\lambda]\right)$, $\widehat{\beta}_m^\lambda(z) = \widehat{C}_\beta^\lambda H_m(z)$, $(\widehat{C}_\beta^\lambda, \widehat{F}^\lambda)$ are obtained by implementing (3.3) using λ as the weight vector, and df_λ is the number of nonzero coefficient functions identified by $\widehat{C}_\beta^\lambda$. The penalty term $\frac{\ln N}{\sqrt[4]{N}}$ of (3.4) is constructed in view of the slow rate documented in Lemma A.2 of the supplementary Appendix A. We select λ by

$$\widehat{\lambda} = \underset{\lambda}{\mathrm{argmin}} \, \mathrm{BIC}_\lambda. \tag{3.5}$$

Further let $S_{\widehat{\lambda}} = \{j \,|\, \|\widehat{C}_{\beta,j}^{\widehat{\lambda}}\| > 0, 1 \le j \le p\}$ indicate the set of relevant variables identified by $\widehat{C}_\beta^{\widehat{\lambda}}$. Then the next result follows.

Theorem 3.2. *Let Assumptions 1-3 hold.* $\Pr(S_{\widehat{\lambda}} = \mathcal{A}^*) \to 1$ *as* $(N, T) \to (\infty, \infty)$.

Again, Assumption 4 is unnecessary for establishing Theorem 3.2.

3.2 High Dimensional Case

In this subsection, we allow the dimension of x_{it} to diverge as the sample size increases. The following assumptions are crucial for deriving the asymptotic results for the HD case.

Assumption 5.

1. $\|\mathcal{E}\|_{\mathrm{sp}} = O_P(\max\{\sqrt{N}, \sqrt{T}\})$, *where* $\|\cdot\|_{\mathrm{sp}}$ *denotes the spectral norm of a matrix and* $\mathcal{E} = (\mathcal{E}_1, \ldots, \mathcal{E}_N)'$;

2. $\frac{p^* \lambda_{\max}^* \sqrt[4]{\xi_{NT}}}{NT} \to 0$, $\left(\frac{\xi_{NT} + mp}{NT} + p^* m^{-\mu}\right)\sqrt{\xi_{NT}} \to \kappa_2$, $\frac{\sqrt[8]{\xi_{NT}} \lambda_{\min}^\dagger}{NT} \to \kappa_3$, where $\xi_{NT} = \min\{N, T\}$, $0 \le \kappa_2 < \infty$ and $\kappa_3 > 0$.

Assumption 5.1 is identical to Assumption iii of Li et al. (2016) and Assumption A.1.v of Lu and Su (2016). Assumption 5.2 further imposes bounds on some parameters, and can be verified in exactly the same way as shown under Assumption 3.

With regard to the selection of λ, we still use the BIC criterion with a minor modification:

$$\mathrm{BIC}_\lambda = \ln \mathrm{RSS}_\lambda + \mathrm{df}_\lambda \Upsilon_{NT}, \tag{3.6}$$

where Υ_{NT} is a penalty term satisfying $\Upsilon_{NT} \to 0$ as $(N, T) \to (\infty, \infty)$; and all other notations are defined in exactly the same way as in the LD case. Select λ by $\widehat{\lambda} = \mathrm{argmin}_\lambda \mathrm{BIC}_\lambda$.

Then the next theorem holds.

Theorem 3.3. *Let Assumptions 1, 2 and 5 hold. As $(N, T) \to (\infty, \infty)$,*

1. $\Pr(\|\widehat{C}_\beta^\dagger\| = 0) \to 1$.

2. *Additionally, let $\Upsilon_{NT} \xi_{NT}^{1/8} \to \kappa_4 > 0$. Then $\Pr(S_{\widehat{\lambda}} = \mathcal{A}^*) \to 1$.*

Once the zero coefficient functions are identified, the rest of the analysis (such as the investigation of the rate of convergence) will be similar to that done in Section 3.1 except that one needs to account for the divergence of both m and p^*. To avoid repetition, we will not present the analysis here again.

In summary, in either of the two cases (LD and HD), when $\Pr(S_{\widehat{\lambda}} = \mathcal{A}^*) \to 1$, all zero coefficient functions can be identified. In the growth regression context, this is equivalent to saying that when $\Pr(S_{\widehat{\lambda}} = \mathcal{A}^*) \to 1$, all variables not driving economic growth can be identified and thus removed from the growth regression. In the meantime, the varying coefficients can be recovered using the sieve method, while the factor structure can be estimated by the PCA technique. Thus, all the three aforementioned issues that are prominent in the empirical growth literature (i.e., variable selection, parameter heterogeneity, and cross-sectional dependence) can be addressed simultaneously within a single, integrated framework. Before moving on to the empirical analysis, we next describe the data employed in this study.

11

4 Data

Of the many variables that have been found to be significantly correlated with growth in the literature, we choose a total of 60 (including the dependent variable, the growth rate of per capita GDP). The choice of these variables is based on previous studies (e.g., Sala-I-Martin et al., 2004; Moral-Benito, 2012) and data availability. Our final dataset covers 89 countries over the period 1960 - 2014. It contains countries in different stages of development and with a wide geographic dispersion. The explanatory variables cover a wide range of factors, including stage of development, social issues, health, geography, politics, education and more. The variable names, their means, and standard deviations are presented in Table Table 1. Table 2 provides a list of the included countries.

A common practice in the literature is to take a five-year simple moving average of both dependent and independent variables[1]. This technique has the advantages of reducing the potential effects of short-term fluctuations and maintaining a high number of time series observations. Despite these advantages, this technique may still suffer from reverse causality or simultaneity, because causality between regressors and growth could go the other way as well or some regressors and growth may be simultaneously determined (e.g., Bils and Klenow, 2000). To mitigate this problem, we deviate from the common practice by measuring dependent and independent variables differently. Specifically, while the dependent variable is measured as a five-year moving average of economic growth, all explanatory variables are measured at the beginning of each five-year period, with the exception of the variables related to war, geography, and terms of trade[2] (Salimans, 2012). These latter explanatory variables are expected to be truly independent of contemporaneous economic growth, and thus also are measured as five-year moving averages (as with the dependent variable). This treatment further alleviates endogeneity, which is already mitigated by the use of multi-factor error structure as discussed in Section 2.

---

[1]Another popular method of looking at annual data in empirical growth literature is to use averaged five-year period data. But, as is stressed by Soto (2003) and Attanasio et al. (2000), the use of n-year averages is not suitable because of the lost of information that it implies, and attempting to use data on averaged five-year periods severely limited the number of observations to draw from in the data.

[2]Specifically, these variables include: fraction spent in war (each five-year period); number of war participation (each five-year period); number of revolutions (each five-year period); coups d'etat and coup attempts within (each five-year period); time of independence; East Asian dummy; African dummy; European dummy; Latin American dummy; British colony dummy; Spanish colony dummy; landlocked country dummy; percentage of land area in Koeppen-Geiger tropics; percentage of land area within 100 km of ice-free coast; terms of trade; and terms of trade growth.

5 Empirical Results

5.1 Choices of the Number of Factors and the Development Index

In Section 3, we assume that the number of factors r is known. In practice, r is unknown and has to be estimated. The main tool for estimating the number of factors of large dimensional datasets is the use of information criteria. In view of the fact that there are 59 observable explanatory variables in our case, we follow Ando and Bai (2017) to choose the number of the factors by minimizing the next criteria function:

$$\text{PIC}(r) = \widehat{\sigma}_\varepsilon^2 \cdot \left(1 + r \cdot \frac{N+T}{NT} \log(NT) \right), \tag{5.1}$$

where $\widehat{\sigma}_\varepsilon^2 = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(y_{it} - x'_{it}\widehat{\beta}_m - \widehat{f}_t'\widehat{\gamma}_i \right)^2$, and for $\forall r$, $\widehat{\beta}_m$, $\widehat{f}_t$ and $\widehat{\gamma}_i$ are the corresponding estimates using the approach of Section 2.

We now turn to the choice of the development index, z_{it}. In Section 2 we have specified a varying coefficient growth regression model capable of capturing parameter heterogeneity by means of a development index. Of the possible development indexes, output and human capital are believed to be the most important ones in previous studies (e.g., Durlauf and Johnson, 1995; Liu and Stengos, 1999; Minier, 2007; Salimans, 2012). Following those studies, we consider four alternative development indexes in log form: (1) initial GDP per capita, (2) initial primary schooling enrolment rate, (3) initial secondary schooling enrolment, and (4) initial higher education enrolment rate.

When choosing among the four alternative development indices, we use the in-sample root mean squared error (RMSE) which is consistent with the criterion function used in estimation. Specifically, for each development index, we first choose the number of factors and select the regressors. Then we run post selection regression as documented in Section A.1 of the supplementary Appendix A without including the weight parameters to calculate the corresponding RMSE. Table 3 presents the chosen number of factors and in-sample RMSE for each of the four development indices. In addition, we also consider the homogeneous parameter growth regression model where β_0 (i.e., the coefficients of growth determinants) is homogeneous across countries, and the results are reported in the first column of Table 3. This table shows that the model with initial GDP per capita as the development index has the lowest in-sample RMSE and thus fits the data best.

In summary, the model with six factors and initial GDP per capita as the development index (i.e., $r = 6$ and z = initial GDP per capita) receives the most support from the

13

data. Hence, in what follows we concentrate on the results obtained from this model.

5.2 Estimates of the Common Factors and Their Associated Loadings

Figure 1 and Figure 2 plot the estimates of the factors identified above and their corresponding loadings respectively. The former shows that all the common factors varies considerably over time with the exception of the first factor which exhibits a relatively small amount of variation during the sample period, while the latter shows that all the factor loadings vary substantially across countries.

We are also interested in the importance of each common factor in explaining the total variance of the error terms e_{it}'s of (2.3). Table 4 shows the proportion of the total variance attributed to each common factor. As the table shows, the first common factor accounts for 87.86% percent of the total variance, and the other five factors account for 7.01%, 1.78%, 1.45%, 0.61%, 0.30% respectively. Overall these six common factors account for 99.01% of the total variance, indicating a fairly parsimonious description of the data.

5.3 Estimates of the Coefficient Functions of Selected Variables

5.3.1 General Findings

Our results are broadly consistent with those of Fernández et al. (2001) and Sala-I-Martin et al. (2004) in that we have identified a number of robust growth determinants (variables) that are also found to be significant in the previous studies. In this sense, our results broadly support the more "optimistic" conclusion of Sala-I-Martin (1997), that is, some variables are important regressors for explaining cross-country growth patterns. Specifically, we have identified 31 robust growth determinants, providing evidentiary support for the canonical neoclassical growth variables; i.e., initial income, investment, and population growth, as well as macroeconomic policies, geography, institutions, religion and ethnic fractionalization. Table 5 reports the estimates of the coefficients of each of the 31 robust growth determinants for ln(initial GDP per capita) = 3.98 (minimum), 5, 6, 7, 8, and 8.81 (maximum), together with their associated 95% bootstrapped confidence intervals (CI)[3]. To see these coefficients more clearly, we also plot them against initial

---

[3]Note that these confidence intervals need to be interpreted carefully. As well understood, one cannot establish the confidence intervals for the estimates under HD case unless certain transformation is further employed (e.g., Huang et al., 2008; Dong et al., 2017). However, if one regards 31 (the number of

GDP per capita in Figure A.5 of the supplementary file.

Despite the similarity, there are at least three differences between the results of this study and those of the previous studies. First, our set of robust growth determinants differs from those identified in the previous studies, in spite of many overlaps between them. Specifically, some variables appear to be robust in our study but not in the previous (such as secondary school enrolment rate and terms of trade growth) or vice versa (such as primary school enrolment rate and fraction GDP of mining). There are at least three possible reasons for this difference: (1) we use a different model specification that allows for both parameter heterogeneity and cross-sectional dependence; (2) we use a different variable selection procedure (i.e., a LASSO estimator); and (3) we use a different dataset that spans a longer time period and covers a slightly different set of countries.

Second, our estimates of the coefficients of the robust growth determinants vary considerably across countries according to their level of development, while those in most previous studies are identical across countries. Specifically, we find that some coefficients have the same sign but different values across different levels of initial GDP per capita (such as civil liberty, terms of trade growth, and percentage of land area in tropics), while other coefficients not only have different signs but also different magnitudes across different levels of initial GDP per capita (such as consumption share of government, life expectancy, military expenditure, and OPEC dummy). These findings suggest that it is inappropriate to apply a growth regression with homogeneous parameters to all countries.

Third, our estimates of the coefficients of the robust growth determinants reveal some cross-country patterns not found in previous studies. Taking the coefficient of initial GDP per capita for example, we find that its estimate is positive for countries with GDP per capita between $1,780 and $2,117 in 1960 U.S. dollars (between $13,166 and $15,665 in 2014 U.S. dollars) while being negative for all other countries. This finding is in accordance with the "middle income trap hypothesis", which refers to countries that have experienced rapid growth and thus quickly reached middle-income status but then failed to overcome that income range to further catch up to the developed countries (Gill and Kharas, 2007). To give another example, our estimate of the oil reserve coefficient increases monotonically with GDP per capita and eventually becomes positive for economies with initial GDP per capital above $2,175 in 1960 U.S. dollars ($16,094 in 2014

selected variables) as a relatively small number, then one can treat our regression as a LD case and employ the same bootstrap procedure as in Su et al. (2015). In order to ensure the validity of the bootstrap procedure, stronger assumptions on the error terms are needed. For example, one can employ the martingale difference type of assumptions (see Assumption A.4 of Su et al., 2015), or simply assume that the error terms are i.i.d. over both i and t. Generally speaking, when the error term exhibits both cross-sectional and serial correlation, the bootstrap results are not reliable or incorrect.

U.S. dollars). This latter finding is consistent with recent studies (e.g., Leite and Weidmann, 1999) which suggest that in developed economies where economic institutions are generally well-developed, natural resources tend to promote economic growth; whereas in developing economies where economic institutions are generally weak, natural resources tend to hamper economic growth. We will discuss these two examples in more details below where it is appropriate.

5.3.2 Specific Findings

We now analyse some of the variables that are "significantly" related to growth in more details. Figure 3.1 presents our estimate of the coefficient of initial GDP per capita. This figure reveals three findings. First, this estimate is negative for most GDP per capita levels, thus being largely consistent with findings in the existing conditional convergence literature as well as previous studies that have employed model averaging methods to growth. Second, the coefficient has an inverse U-shaped relationship with initial GDP per capita. This finding is consistent with those reported by previous studies. For example, Durlauf et al. (2001) find that the coefficient of initial GDP per capita does not exhibit any sort of monotonicity with respect to level of development; Salimans (2012) finds that the coefficient of initial GDP per capita first increases with level of development up to a point and then declines afterwards; and Durlauf and Johnson (1995) find that the coefficient of initial GDP per capita is not monotonic with respect either GDP per capita or literacy rate. Third, the coefficient is positive for countries with GDP per capita between $319 and $1,812 in 1960 U.S. dollars (or between $2,554 and $14,510 in 2014 U.S. dollars), suggesting that some of these countries have been unable to catch up with more developed countries. This finding is in line with the "middle-income trap" hypothesis, which refers to the phenomenon of hitherto rapidly growing economies stagnating at middle-income levels and failing to graduate into the ranks of high-income countries (e.g., Eichengreen et al., 2014). In our sample South Africa and Columbia are two example countries that have never been able leave the "middle-income range" over the entire sample period since their GDP per capita fell into this range at the beginning of the sample period (i.e., 1960).

Figure 3.2 shows our estimate of the coefficient of price for investment goods. Three findings emerge from this figure. First, this estimate is negative for countries with initial GDP per capita up to $543 in 1960 U.S. dollars (or $4,348 in 2014 U.S. dollars), suggesting that for these countries a relative low price of investment goods in the first year of each five-year period is strongly and positively related to subsequent income growth. This

16

finding is not surprising because a low investment price stimulates investment (including investment in machinery and equipment), which further spurs economic growth (De Long and Summers, 1991, 1992). Second, the estimated coefficient falls in absolute value as initial GDP per capita increases, meaning that the marginal effect of investment price on growth is stronger for poor countries than for rich countries. This latter finding is consistent with Temple (1999) who find that the growth-spurring effects of investment is greater for developing countries, because total investment includes machinery embodying well-established technologies and developing countries may be able to take advantage of new and old equipment because they have little of any technology. Third, for countries with initial GDP per capita above \$543 in 1960 U.S. dollars (or \$4,348 in 2014 U.S. dollars) the estimated coefficient of investment goods price is positive but insignificant, because the associated confidence intervals contain zero. This suggests that investment goods price has no growth effects for these countries. A possible reason for this latter finding is that the data from the Penn World Table is not disaggregated enough to distinguish price of equipment investment, which has strong growth effects, and price of other forms of investment, which have little growth effects (De Long and Summers, 1991).

Figure 3.3 shows our estimate of the coefficient of secondary schooling enrolment. As this figure shows, this estimate is positive for most levels of initial GDP per capita. This is not surprising because secondary education is a vital part of a virtuous circle of economic growth within the context of a globalized knowledge economy. Many studies have documented that a large pool of workers with secondary education is indispensable for knowledge spillover to take place and for attracting imports of technologically advanced goods and foreign direct investment (Borensztein et al., 1998; Caselli and Coleman II, 2001). That said, we also note that the estimated coefficient for secondary schooling is negative for middle income countries with initial GDP per capita between \$272 and \$1,192 in 1960 U.S. dollars (or between \$2178 and \$9,545 in 2014 U.S. dollars). This result suggests that a possible reason for the "middle income trap" discussed above is that these middle income countries, unlike other countries, fail to take advantage of the benefits brought by secondary schooling.

We also note from Figure 3.4 that the estimate of the coefficient of higher education is negative for almost all countries with the exception of middle income countries. This finding is consistent with Sala-I-Martin et al. (2004) and Salimans (2012), both of which find that the higher schooling coefficient is negative for most of their sample countries but positive for the rest. It also parallels the concavity argument that the earnings function is concave in education, meaning that returns are higher for lower levels of education

17

(Psacharopoulos, 1994; Psacharopoulos and Patrinos, 2004).

Figure 3.5 shows our estimate of the coefficient of oil reserve. Two findings stand out from this figure. First, this estimate is negative for countries with initial GDP per capita below \$2,373 in 1960 U.S. dollars (or \$19,002 in 2014 U.S. dollars). This finding is consistent with the "natural resources curse hypothesis" (e.g., Sachs and Warner, 2001) and can be explained by the rent-seeking behaviour of countries with large endowments of natural resources. Second, the oil reserve coefficient increases monotonically with GDP per capita and eventually becomes positive for economies with initial GDP per capita above \$2,373 in 1960 U.S. dollars (or \$19,002 in 2014 U.S. dollars). This latter finding accords with recent studies (e.g., Leite and Weidmann, 1999) on the nexus between natural resources and economic growth. Specifically, these studies suggest that the contribution of natural resources to a country's economy does not take place in isolation, but rather in the overall context of the country's economic management and institutions. It is thus the quality and competency of these policies and institutions that will determine whether natural resources can promote economic growth, or whether revenues generated by the sector might impede development. Therefore, in developed economies where economic institutions are generally well-developed, natural resources tend to promote economic growth; whereas in developing economies where economic institutions are generally weak, natural resources tend to hamper economic growth.

Figure 3.6 presents the estimate of the coefficient of terms of trade growth. As this figure shows, our estimate of the terms of trade growth coefficient is positive for all countries, suggesting that growth tends to be faster in countries where the rate of change of terms of trade is higher. This finding is consistent with previous studies (Mendoza, 1995, 1997; Kose and Riezman, 2001; Bleaney and Greenaway, 2001) that find that an improvement in the terms of trade leads to higher levels of investment, and hence long-run economic growth. In addition, this figure shows that the terms of trade growth coefficient increases with GDP per capita, indicating that the marginal effect of terms of trade growth is larger in richer countries than in poorer ones. This latter finding is also consistent with previous studies (e.g., Blattman et al., 2007). Specifically, those studies argue that higher volatility in the terms of trade reduces investment and hence growth because of aversion to risk, and that rich countries with more sophisticated institutions and markets are likely to have cheaper ways to insure against price volatility than poor countries, so terms of trade instability is likely to have a smaller negative impact on rich countries.

Here we note that as in Sala-I-Martin et al. (2004), trade openness (defined as exports

plus imports as a share of GDP) is insignificant, presumably reflecting the crudity of this measure, and perhaps the distinction between opening to international trade generating a one-time step increase in income as factors are reallocated according to comparative advantage versus an ongoing growth impact associated with greater openness.

Figures 3.7-3.9 show the estimates of the coefficients of fraction of Christian, Muslim, and Jewish respectively. These coefficients are negative at all levels of development or nearly all levels of development. This finding is consistent with that of Barro and McCleary (2005) who find that religion works via belief, not practice. They argue that higher church attendance uses up time and resources and eventually runs into diminishing returns. The "religion sector", as they call it, can consume more than it yields.

6 Conclusion

A rigorous cross-country growth regression analysis should simultaneously account for three major problems identified in the literature — variable selection, parameter heterogeneity, and cross-sectional dependence. Though these three problems have received individual attention, little or no research has sought to integrate them into a single, comprehensive framework. The purpose of this study is to fill this void by proposing a new, integrated framework that is capable of dealing with parameter heterogeneity and cross-sectional dependence, while simultaneously performing variable selection. Specifically, parameter heterogeneity is allowed for by means of a varying coefficient growth regression model, while cross-sectional dependence is introduced into the model via a multi-factor structure. For simplicity, we refer to the resulting growth regression model as the "varying coefficient growth regression model with factor structure and sparsity". We then propose a LASSO estimator that is capable of performing variable selection on this model. In addition, we have established the associated asymptotic results for this estimator and further investigate the performance of the estimator by conducting extensive simulations.

We apply the above framework to a new dataset that covers 89 countries over the period from 1960 to 2014. We have identified 31 robust growth determinants, providing evidentiary support for the canonical neoclassical growth variables; i.e., initial income, investment, and population growth, as well as macroeconomic policies, geography, institutions, religion and ethnic fractionalization. Moreover, we find that all the coefficients of the robust growth determinants vary considerably across countries according to their level of development, which reveals some interesting cross-country patterns not found in

previous studies. For example, we find that the coefficient of the initial GDP per capita is positive for countries with GDP per capita between \$319 and \$1,812 in 1960 U.S. dollars (or between \$2,554 and \$14,510 in 2014 U.S. dollars), suggesting that some of these countries have fallen into the so-called "middle income trap". As another example, we find that the oil reserve coefficient increases monotonically with GDP per capita and eventually becomes positive for economies with initial GDP per capita above \$2,373 in 1960 U.S. dollars (or \$19,002 in 2014 U.S. dollars), thus being consistent with recent studies that stress the role of institutions in determining how natural resources affect economic growth.

References

Ando, T. and 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.

Attanasio, O. P., Picci, L. and Scorcu, A. E. (2000), 'Saving, growth, and investment: A macroeconomic analysis using a panel of countries', *Review of Economics and Statistics* **82**(2), 182–211.

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

Barro, R. J. (1991), 'Economic growth in a cross section of countries', *Quarterly Journal of Economics* **106**(2), 407–443.

Barro, R. J. and McCleary, R. M. (2005), 'Which countries have state religions?', *Quarterly Journal of Economics* **120**(4), 1331–1370.

Bils, M. and Klenow, P. J. (2000), 'Does schooling cause growth?', *American Economic Review* **90**(5), 1160–1183.

Blattman, C., Hwang, J. and Williamson, J. G. (2007), 'The impact of the terms of trade on economic development in the periphery, 1870-1939: Volatility and secular change', *Journal of Development Economics* **82**(1), 156–179.

Bleaney, M. and Greenaway, S. D. (2001), 'The impact of terms of trade and real exchange rate volatility on investment and growth in sub-saharan africa', *Journal of Development Economics* **65**(2), 491–500.

Borensztein, E., De Gregorioand, J. and Lee, J. (1998), 'How does foreign direct investment affect economic growth?', *Journal of International Economics* **45**(1), 115–135.

Brock, W. A. and Durlauf, S. N. (2001), 'Discrete choice with social interactions', *Review of Economic Studies* **68**(2), 235–260.

Caselli, F. and Coleman II, W. J. (2001), 'The U.S. structural transformation and regional convergence: A reinterpretation', *Journal of Political Economy* **109**(3), 584–616.

Chudik, A., Mohaddes, K., Pesaran, M. H. and Raissi, M. (2017), 'Is there a debt-threshold effect on output growth?', *Review of Economics and Statistics* **99**(1), 135–150.

Chudik, A., Pesaran, M. H. and Tosetti, E. (2011), 'Weak and strong cross-section dependence and estimation of large panels', *Econometrics Journal* **14**(1), C45–C90.

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

De Long, J. B. and Summers, L. (1991), 'Equipment investment and economic growth', *The Quarterly Journal of Economics* **106**(2), 445–502.

De Long, J. B. and Summers, L. (1992), 'Equipment investment and economic growth: How strong is the nexus?', *Brookings Papers on Economic Activity* **2**, 157–199.

Dong, C., Gao, J. and Linton, O. (2017), High dimensional semiparametric moment restriction models. https://ssrn.com/abstract=3045063.

Dong, C., Gao, J. and Peng, B. (2018), Varying-coefficient panel data models with partially observed factor structure. https://ssrn.com/abstract=3102631.

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

Durlauf, S. N. and Johnson, P. A. (1995), 'Multiple regimes and cross-country growth behaviour', *Journal of Applied Econometrics* **10**(4), 365–384.

Durlauf, S. N., Johnson, P. A. and Temple, J. R. (2005), 'Growth econometrics', *Handbook of Macroeconomics* **1, Part A**, 555–677.

Durlauf, S. N., Kourtellos, A. and Minkin, A. (2001), 'The local solow growth model', *European Economic Review* **45**(4-6), 928–940.

Durlauf, S. N., Kourtellos, A. and Tan, C. M. (2008), 'Are any growth theories robust', *Economic Journal* **118**(527), 329–346.

Durlauf, S. N. and Quah, D. T. (1999), 'The new empirics of economic growth', *Handbook of Macroeconomics* **1, Part A**, 235–308.

Eichengreen, B., Park, D. and Shin, K. (2014), 'Growth slowdowns redux: New evidence on the middle-income trap', *Japan and the World Economy* **32**, 65–84.

Fan, J., Liao, Y. and Wang, W. (2016), 'Projected Principal Component Analysis in Factor Models', *Annals of Statistics* **44**(1), 219–254.

Feng, G., Gao, J., Peng, B. and Zhang, X. (2017), 'A varying-coefficient panel data model with fixed effects: theory and an application to the US commercial banks', *Journal of Econometrics* **196**(1), 68–82.

Fernández, C., Ley, E. and Steel, M. F. J. (2001), 'Model uncertainty in cross-country growth regressions', *Journal of Applied Econometrics* **16**(5), 563–576.

Gill, I. and Kharas, H. (2007), *An East Asian Renaissance : Ideas for Economic Growth*, Washington, DC: World Bank.

Huang, J., Horowitz, J. L. and Ma, S. (2008), 'Asymptotic properties of bridge estimators in sparse high-dimensional regression models', *Annals of Statistics* **36**(2), 587–613.

Jiang, B., Yang, Y., Gao, J. and Hsiao, C. (2017), Recursive estimation in large panel data models: Theory and practice. Working paper available at https://ssrn.com/abstract=2915749.

Kormendi, R. C. and Meguire, P. G. (1985), 'Macroeconomic determinants of growth: Cross-country evidence', *Journal of Monetary Economics* **16**(2), 141–163.

Kose, A. and Riezman, R. (2001), 'Trade shocks and macroeconomic fluctuations in africa', *Journal of Development Economics* **65**(1), 55–80.

Leamer, E. (1983), 'Let's take the con out of econometrics', *American Economic Review* **73**(1), 31–43.

Leite, C. and Weidmann, J. (1999), Does mother nature corrupt? natural resources, corruption, and economic growth. IMF Working paper 99/85.

Li, D., Qian, J. and Su, L. (2016), 'Panel data models with interactive fixed effects and multiple structural breaks', *Journal of the American Statistical Association* **111**(516), 1804–1819.

Liu, F., Gao, J. and Yang, Y. (2018), Nonparametric time-varying panel data models with heterogeneity. https://ssrn.com/abstract=3214046.

Liu, Z. and Stengos, T. (1999), 'Non-linearities in cross-country growth regressions: A semiparametric approach', *Journal of Applied Econometrics* **14**(5), 527–538.

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

Lucas, R. (1993), 'Making a miracle', *Econometrica* **61**(2), 251–72.

Malikov, E., Kumbhakar, S. C. and Sun, Y. (2016), 'Varying coefficient panel data model in the presence of endogenous selectivity and fixed effects', *Journal of Econometrics* **190**(2), 233–251.

Mankiw, N. G., Romer, D. and Weil, D. N. (1992), 'A contribution to the empirics of economic growth', *Quarterly Journal of Economics* **107**(2), 407–437.

Mendoza, E. G. (1995), 'The terms of trade, the real exchange rate, and economic fluctuations', *International Economic Review* **36**(1), 101–137.

Mendoza, E. G. (1997), 'Terms-of-trade uncertainty and economic growth', *Journal of Development Economics* **54**(2), 323–356.

Minier, J. (2007), 'Nonlinearities and robustness in growth regressions', *American Economic Review* **97**(2), 388–392.

Moral-Benito, E. (2012), 'Determinants of economic growth: A bayesian panel data approach', *Review of Economics and Statistics* **94**(2), 566–579.

Newey, W. K. (1997), 'Convergence rates and asymptotic normality for series estimators', *Journal of Econometrics* **79**(1), 147–168.

Pesaran, M. H. (2006), 'Estimation and inference in large heterogeneous panels with a multifactor error structure', *Econometrica* **74**(4), 967–1012.

Psacharopoulos, G. . (1994), 'Returns to investment in education: A global update', *World Development* **22**(9), 1325–1343.

Psacharopoulos, G. and Patrinos, H. A. (2004), 'Returns to investment in education: a further update', *Education Economics* **12**(2), 111–134.

Sachs, J. D. and Warner, A. (2001), 'The curse of natural resources', *European Economic Review* **45**(4-6), 827–838.

Sala-I-Martin, X. (1997), 'I just ran two million regressions', *American Economic Review* **87**(2), 178–183.

Sala-I-Martin, X., Doppelhofer, G. and Miller, R. I. (2004), 'Determinants of long-term growth: A bayesian averaging of classical estimates (bace) approach', *American Economic Review* **94**(4), 813–835.

Salimans, T. (2012), 'Variable selection and functional form uncertainty in cross-country growth regressions', *Journal of Econometrics* **171**(2), 267–280.

Soto, M. (2003), 'Taxing capital flows: an empirical comparative analysis', *Journal of Development Economics* **72**(1), 203–221.

Su, L., Jin, S. and Zhang, Y. (2015), 'Specification test for panel data models with interactive fixed effects', *Journal of Econometrics* **186**(1), 222–244.

Summers, R. and Heston, A. (1991), 'The Penn World Table (Mark 5): an expanded set of international comparisons, 1950-1988', *Quarterly Journal of Economics* **106**(2), 327–368.

Temple, J. (1999), 'The new growth evidence', *Journal of Economic Literature* **37**(1), 112–156.

Table 1: Definitions of All Variables in the Regression

Variables	Description	Formula	Mean	Std
EG	Economic growth rate	$\ln(\text{rgdpo}_t/\text{rgdpo}_{t-1})$	0.0363	0.0649
log(GPC)	log GDP per capita		6.0642	0.9861
csh_g	Government consumption share		0.2074	0.1163
Openness	Openness measure	csh_x + csh_m	-0.0322	0.1274
IP	Investment price, i.e., price level of capital formation		0.4213	0.3220
PGR	Population growth rates	$\ln(\text{pop}_t/\text{pop}_{t-1})$	0.0197	0.0127
Sch_P	Primary school enrollment		0.7396	0.2170
Sch_S	Secondary school enrollment		0.4672	0.3204
Sch_H	Higher education School Enrollment		0.1336	0.1581
LE	Life Expecancy		0.5932	0.1102
PESS	Public education spending share in GDP		0.0399	0.0296
PIS	Public investment share	GFCF - GFCF_PS	0.0713	0.0501
Land	Land area (sq. km / 1,000,000)		0.8719	2.1518
Exports	Percentage of Primary Export	Exports_OM + Exports_ARM	0.1978	0.2085
Mining	Fraction GDP in mining		0.0733	0.0874
Fertility	Fertility rate, total (births per woman)		4.7412	1.9724
Military	Military expenditure share in GDP		0.0295	0.0302
PCS	Public consumption share	GGFCE - PESS - Military	0.0809	0.0501
Malaria	Malaria prevalence: Incidence of malaria (per 1,000 population at risk)		155.1712	245.5145
Inflation	Inflation rate		1.3854	7.0448
Political	Political rights		4.2100	1.9429
Civil	Civil liberties		4.1496	1.6628
Cap	Degree of capitalism		3.2697	1.7275
Trade	Terms of trade		1.3475	2.2807
Tra_Gro	Terms of trade growth		0.0073	0.0974
Locked	Landlocked country dummy (1, yes; 0, no)		0.2697	0.4438
Ind_Year	Time of independence <=1914 = 0; 1915-1945 = 1; 1946-1989 = 2; >= 1990 = 3		1.4382	1.0382
kgatr	Percentage of land area in Koeppen-Geiger tropics		0.4017	0.4205
kgptr	Percentage of population in Koeppen-Geiger tropics		0.3915	0.4217
lcr100km	Percentage of Land area within 100 km of ice-free coast		0.3788	0.3640
pop100cr	Ratio of population within 100 km of ice-free coast/navigable river to total population		0.4520	0.3728
cen_lat	latitude of country centroid		0.1522	0.2197
Bri_Col	British colony dummy (1, yes; 0, no)		0.2584	0.4378
Spa_Col	Spanish colony dummy (1, yes; 0, no)		0.1910	0.3931

Oil_OPEC	Oil-producing country dummy (1, yes; 0, no)		0.0674	0.2508
Gas	proved reserves (cubic meters / 10^12)		1.3789	6.1997
Oil	proved reserves (bbl / 10^9)		4.5521	19.1378
Chris	Percentage of Christian		0.5369	0.3807
Mus	Percentage of Muslim		0.3046	0.3825
Hin	Percentage of Hindu		0.0263	0.1211
Bud	Percentage of Buddhist		0.0410	0.1541
Fol	Percentage of Folk religion		0.0284	0.0628
Oth	Percentage of other religion		0.0037	0.0064
Jew	Percentage of Jewish		0.0019	0.0027
GS	Government spending share of GDP		0.1501	0.0686
Distortion	Real exchange rate distortions		129.6824	35.8479
OO	Outward orientation		-2.7398	0.7542
SIL	Ethnolinguistic fractionalization		0.4886	0.3127
ESP	English-speaking population in percentage		0.1762	0.2692
EA	East Asian dummy		0.0225	0.1482
AF	African dummy		0.4270	0.4947
EU	European dummy		0.1124	0.3158
LA	Latin American dummy		0.1573	0.3641
WarFrac	Fraction spent in war (1960-2014)		0.3265	0.4343
NoWars	No. of war participation (1960-2014)		0.8028	1.2609
Coup	coups d'etat and coup attempts within (1960-2014)		0.1870	0.4964
Revolution	Number of revolutions (1960-2014)		0.1941	0.5038
Pop_Dens	Population Density/1000		0.0812	0.1135
WorkIR	Growth rate of work force	$\ln(\text{WP}_t/\text{WP}_{t-1})$	0.0210	0.0132

rgdpo — Size of economy (GDP in million)

pop — Population (in million)

csh_x — Share of merchandise exports

csh_m — Share of merchandise imports

WP — Fraction population of work force (1-A65-U15)

A65 — Fraction population over 65 years old

U15 — Fraction population under 15 years old

GFCF — Gross fixed capital formation

GFCF_PS — Gross fixed capital formation, private sector

Exports_OM — Percentage of Ores and metals exports

Exports_ARM — Percentage of Agricultural raw materials exports

GGFCE — General government final consumption expenditure share in GDP

Table 2: Sample Countries and Their Associated ISO 3166-1 alpha-3 Codes

AGO	Angola	HND	Honduras	PAK	Pakistan
ALB	Albania	HRV	Croatia	PAN	Panama
ARM	Armenia	HTI	Haiti	PER	Peru
AZE	Azerbaijan	IND	India	PHL	Philippines
BDI	Burundi	IRN	Iran, Islamic Republic of	POL	Poland
BEN	Benin	JAM	Jamaica	PRY	Paraguay
BFA	Burkina Faso	JOR	Jordan	RUS	Russian Federation
BGD	Bangladesh	JPN	Japan	RWA	Rwanda
BGR	Bulgaria	KAZ	Kazakhstan	SDN	Sudan
BLR	Belarus	KEN	Kenya	SEN	Senegal
BRA	Brazil	KGZ	Kyrgyzstan	SLE	Sierra Leone
BWA	Botswana	KHM	Cambodia	SLV	El Salvador
CAF	Central African Republic	LAO	Lao People's Democratic Republic	SWZ	Swaziland
CIV	Côte d'Ivoire	LBN	Lebanon	SYR	Syrian Arab Republic
CMR	Cameroon	LKA	Sri Lanka	TCD	Chad
COG	Congo	LSO	Lesotho	TGO	Togo
COL	Colombia	MDA	Moldova, Republic of	THA	Thailand
DOM	Dominican Republic	MDG	Madagascar	TTO	Trinidad and Tobago
DZA	Algeria	MEX	Mexico	TUN	Tunisia
ECU	Ecuador	MKD	Macedonia	TUR	Turkey
EGY	Egypt	MLI	Mali	TZA	Tanzania, United Republic of
ETH	Ethiopia	MNG	Mongolia	UGA	Uganda
GAB	Gabon	MOZ	Mozambique	UKR	Ukraine
GBR	United Kingdom	MWI	Malawi	URY	Uruguay
GEO	Georgia	MYS	Malaysia	USA	United States
GHA	Ghana	NAM	Namibia	VEN	Venezuela, Bolivarian Republic of
GIN	Guinea	NER	Niger	YEM	Yemen
GMB	Gambia	NIC	Nicaragua	ZAF	South Africa
GNB	Guinea-Bissau	NPL	Nepal	ZWE	Zimbabwe
GTM	Guatemala	OMN	Oman		

Table 3: Comparison among Alternative Development Indexes (z)

	Parametric	z of Varying Coefficient			
		ln(GPC)	School_P	School_S	School_H
RMSE	0.022	0.017	0.019	0.027	0.020
No. of factors	6	6	6	3	5

Table 4: Cumulative Variation of the Residuals Explained by the Factors

No. Factors	1	2	3	4	5	6
Cumulative Variation	87.86%	94.87%	96.65%	98.10%	98.71%	99.01%

Table 5: Estimates of Coefficients at log(GPC)=3.98, 5, 6, 7, 8 and 8.81

	log(GPC)=3.98	log(GPC)=5	log(GPC)=6	log(GPC)=7	log(GPC)=8	log(GPC)=8.81
log(GPC)	-0.0766	-0.0447	0.0117	0.0255	-0.0531	-0.2055
	(-0.1145, -0.0429)	(-0.0618, -0.0305)	(-0.0038, 0.0249)	(0.0080, 0.0469)	(-0.0886, -0.0034)	(-0.2883, -0.0898)
csh_g	0.3341	0.1483	-0.0587	-0.2143	-0.2619	-0.1964
	(0.1077, 0.5534)	(0.0706, 0.2248)	(-0.1110, -0.0107)	(-0.3141, -0.1454)	(-0.5782, 0.0082)	(-0.9416, 0.4733)
IP	-0.1284	-0.0545	-0.0093	0.0160	0.0282	0.0325
	(-0.1971, -0.0621)	(-0.0686, -0.0422)	(-0.0191, 0.0059)	(-0.0046, 0.0406)	(-0.0146, 0.0638)	(-0.0895, 0.1209)
PGR	3.2821	0.5353	0.8485	1.5689	0.6502	-2.2853
	(1.7325, 4.7072)	(0.0426, 0.9486)	(0.4054, 1.2542)	(1.1371, 1.9781)	(-0.1228, 1.3620)	(-3.5441, -0.7206)
School_S	0.1322	0.0823	-0.0420	-0.0147	0.3303	0.9119
	(-0.1505, 0.4845)	(0.0108, 0.1495)	(-0.0780, -0.0081)	(-0.0548, 0.0177)	(0.1989, 0.4201)	(0.5495, 1.1719)
School_H	-1.5869	-0.1175	0.2220	0.0326	-0.2066	-0.1808
	(-2.4454, -0.7631)	(-0.3214, 0.1119)	(0.1562, 0.3020)	(-0.0157, 0.0921)	(-0.2810, -0.1039)	(-0.3986, 0.1401)
LE	0.3148	0.2935	-0.0597	-0.2370	0.1455	1.0373
	(-0.0171, 0.6403)	(0.1811, 0.3938)	(-0.1739, 0.0389)	(-0.3918, -0.0990)	(-0.2875, 0.5411)	(0.0000, 1.9922)
PESS	-1.2399	-0.5737	-0.2476	-0.0158	0.3070	0.7260
	(-2.4950, 0.0107)	(-0.7969, -0.3513)	(-0.4210, -0.0944)	(-0.2163, 0.1716)	(-0.6249, 1.3302)	(-1.5776, 3.4530)
Military	-1.1748	-0.1436	-0.0428	-0.0020	0.6417	1.9197
	(-2.4253, -0.0639)	(-0.3588, 0.1047)	(-0.1667, 0.0719)	(-0.2609, 0.1609)	(0.0134, 1.1085)	(0.1982, 3.4369)
Inflation	-0.0055	0.0006	-0.0011	-0.0031	0.0007	0.0107
	(-0.0094, -0.0012)	(0.0000, 0.0012)	(-0.0016, -0.0005)	(-0.0039, -0.0023)	(-0.0043, 0.0048)	(-0.0020, 0.0209)
Civil	0.0183	0.0050	-0.0004	0.0033	0.0166	0.0340
	(0.0048, 0.0308)	(0.0011, 0.0085)	(-0.0027, 0.0024)	(0.0000, 0.0064)	(0.0027, 0.0282)	(-0.0036, 0.0661)
Tra_Gro	0.0009	0.0304	0.0048	0.0018	0.0796	0.2279
	(-0.1138, 0.1100)	(0.0117, 0.0458)	(-0.0162, 0.0252)	(-0.0248, 0.0351)	(-0.0062, 0.1727)	(-0.0124, 0.4799)
kgatr	-0.3099	-0.0815	-0.0763	-0.1422	-0.1597	-0.0776
	(-0.4367, -0.1693)	(-0.1281, -0.0391)	(-0.1161, -0.0447)	(-0.1887, -0.0969)	(-0.2602, -0.0333)	(-0.3270, 0.2396)
lcr100km	0.6793	0.1930	0.0511	-0.0621	-0.3850	-0.9133
	(0.5005, 0.8498)	(0.1396, 0.2461)	(0.0127, 0.0940)	(-0.1305, -0.0005)	(-0.5977, -0.2168)	(-1.4617, -0.5189)
cen_lat	0.5894	0.0293	0.0360	0.1693	0.0873	-0.3058
	(0.3011, 0.8198)	(-0.0723, 0.1294)	(-0.0315, 0.1114)	(0.0985, 0.2655)	(-0.0526, 0.2459)	(-0.6344, 0.0789)
Spa_Col	-0.5174	-0.1564	-0.0204	0.0548	0.1945	0.4167

|						
	(-0.6748, -0.3222)	(-0.2111, -0.1080)	(-0.0503, 0.0091)	(0.0171, 0.0950)	(0.1040, 0.3131)	(0.1867, 0.6800)
Oil_OPEC	-0.1254	-0.0460	-0.0947	-0.0815	0.1358	0.5247
	(-1.0835, 0.6246)	(-0.2325, 0.1181)	(-0.1496, -0.0423)	(-0.1383, -0.0344)	(-0.0124, 0.2413)	(0.1090, 0.8785)
Oil	-0.0100	-0.0021	-0.0008	-0.0011	0.0008	0.0057
	(-0.0217, 0.0018)	(-0.0051, 0.0009)	(-0.0017, 0.0000)	(-0.0022, -0.0001)	(-0.0010, 0.0023)	(0.0004, 0.0099)
Chris	0.2689	0.1112	-0.0056	-0.1262	-0.2818	-0.4469
	(0.1164, 0.4267)	(0.0650, 0.1711)	(-0.0535, 0.0461)	(-0.2001, -0.0629)	(-0.4824, -0.1333)	(-0.8908, -0.0694)
Mus	-0.0736	0.0130	-0.0292	-0.1685	-0.3744	-0.5692
	(-0.1864, 0.0735)	(-0.0383, 0.0599)	(-0.0774, 0.0195)	(-0.2344, -0.1004)	(-0.5710, -0.2335)	(-1.0085, -0.2474)
Oth	-3.7386	0.6890	-1.1133	2.9016	21.6126	51.3537
	(-13.0771, 4.1928)	(-1.3860, 2.4081)	(-3.0584, 0.7782)	(0.2919, 6.3889)	(11.6269, 32.3911)	(27.3474, 75.2005)
Jew	-32.4452	-10.5477	-2.0163	-0.9826	-2.7502	-3.6114
	(-55.7335, -10.0679)	(-18.6022, -3.8609)	(-7.6269, 3.6243)	(-6.9105, 4.0461)	(-12.1471, 5.8912)	(-27.3730, 18.2231)
GS	-0.2516	-0.2079	-0.0476	-0.0738	-0.5101	-1.2583
	(-0.5709, 0.0529)	(-0.3047, -0.1254)	(-0.1142, 0.0215)	(-0.1676, 0.0172)	(-0.8230, -0.1881)	(-2.1023, -0.4396)
Distortion	0.0231	0.0884	0.1085	0.0482	-0.1148	-0.3280
	(-0.0449, 0.1048)	(0.0580, 0.1187)	(0.0701, 0.1435)	(-0.0064, 0.1025)	(-0.2793, 0.0061)	(-0.6718, -0.0148)
OO	1.0569	4.1811	5.1496	2.3016	-5.4384	-15.5641
	(-2.1588, 4.9579)	(2.7557, 5.6164)	(3.3263, 6.8047)	(-0.2707, 4.8851)	(-13.1848, 0.3040)	(-31.9253, -0.7429)
ESP	0.3819	0.1039	0.0581	0.1010	0.1193	0.0577
	(0.2151, 0.5483)	(0.0575, 0.1465)	(0.0214, 0.0975)	(0.0494, 0.1631)	(-0.0059, 0.2319)	(-0.2260, 0.2967)
EA	0.2924	0.0841	0.0263	-0.1256	-0.5515	-1.1753
	(-0.1574, 0.7975)	(-0.0589, 0.2081)	(-0.0668, 0.1185)	(-0.2975, 0.0293)	(-1.0307, -0.1357)	(-2.2484, -0.2191)
EU	-1.0237	-0.2206	-0.0275	0.0738	0.4780	1.2174
	(-1.7313, -0.3070)	(-0.3896, -0.0438)	(-0.0797, 0.0197)	(0.0169, 0.1225)	(0.3182, 0.6845)	(0.7527, 1.7964)
WarFrac	-0.0044	-0.0035	-0.0120	-0.0130	0.0058	0.0411
	(-0.0372, 0.0351)	(-0.0121, 0.0072)	(-0.0179, -0.0040)	(-0.0208, -0.0036)	(-0.0157, 0.0235)	(-0.0227, 0.0882)
Coup	-0.0190	-0.0004	-0.0028	-0.0070	0.0016	0.0251
	(-0.0287, -0.0089)	(-0.0031, 0.0020)	(-0.0051, -0.0005)	(-0.0102, -0.0038)	(-0.0086, 0.0113)	(-0.0005, 0.0502)
Revolution	-0.0052	0.0044	-0.0016	-0.0066	0.0021	0.0252
	(-0.0198, 0.0090)	(0.0007, 0.0075)	(-0.0042, 0.0011)	(-0.0106, -0.0020)	(-0.0054, 0.0084)	(-0.0034, 0.0464)

Figure 1: Estimates of Common Factors
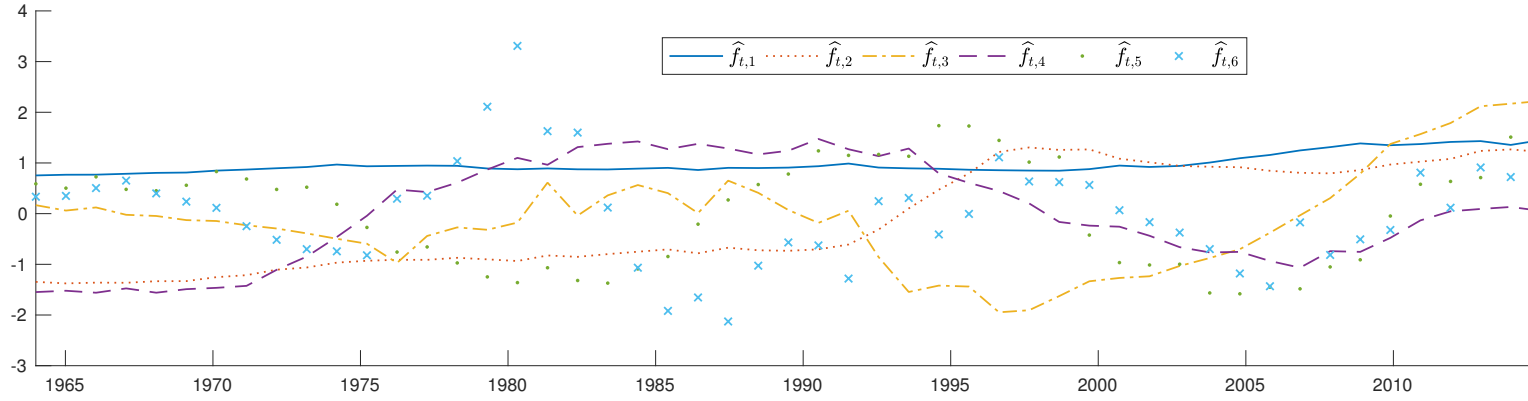
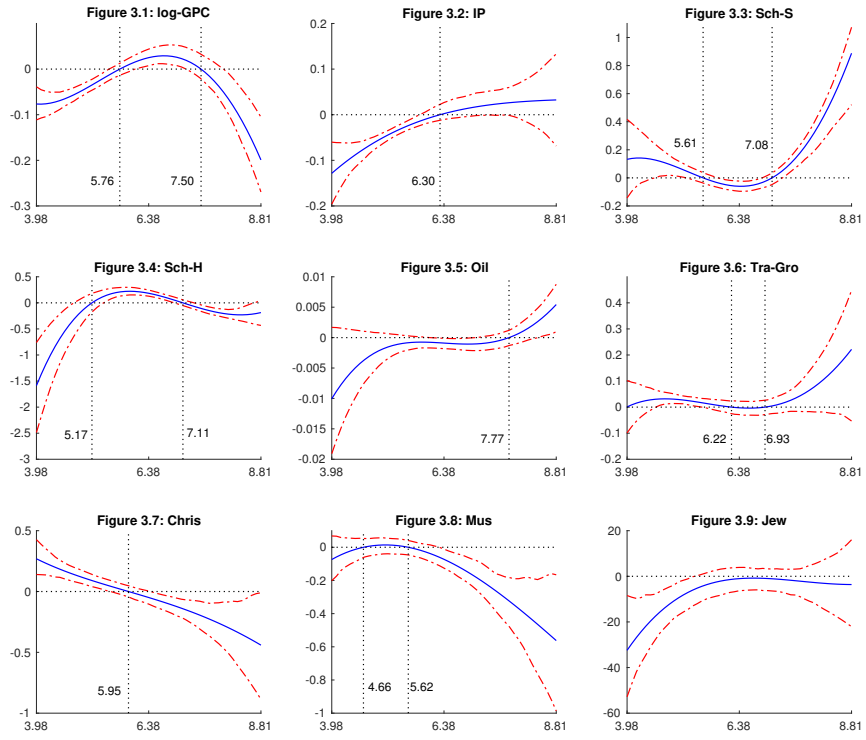$\widehat{f}_{t,j}$ stands for the estimate of the jth factor, where $j = 1 \ldots, 6$.

Figure 2: Estimates of Factor Loadings
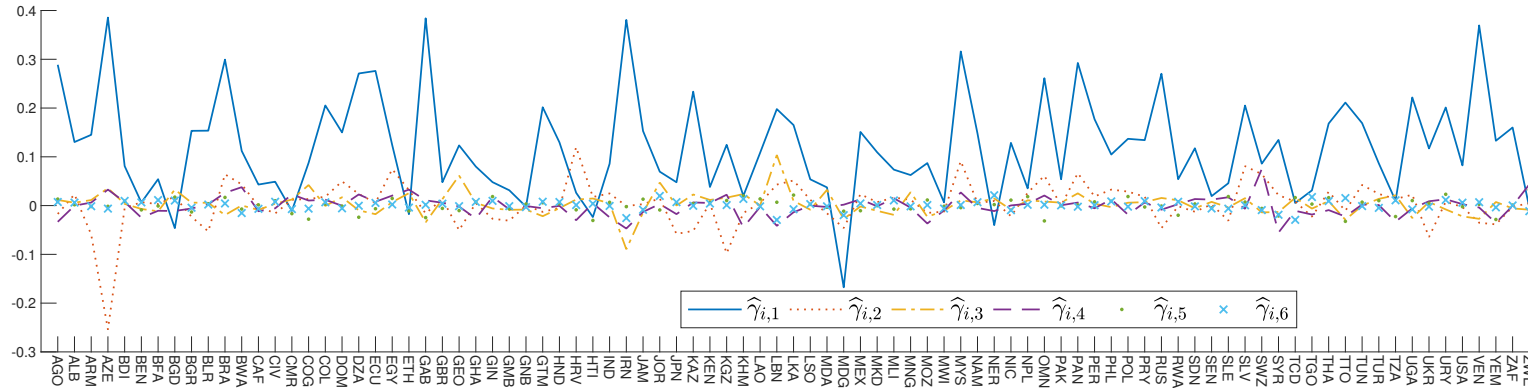
$\widehat{\gamma}_{i,j}$ stands for the estimate of the jth factor loading, where $j = 1 \ldots, 6$.

Figure 3: Estimates of Selected Coefficient Functions

Supplementary Appendix A to
"An Integrated Panel Data Approach to Modelling Economic Growth"

Guohua Feng[*], Jiti Gao[♯] and Bin Peng[†]

[*]University of North Texas, [♯]Monash University and [†]University of Bath

Appendix A is divided into five sections. Section A.1 provides the numerical algorithm. Section A.2 examines the asymptotic results of Section 3 through several simulations. Section A.3 presents the preliminary lemmas and the proofs of the main theorems. Section A.4 explains why our method can partially solve the issue of time trend that has found limited attention in the empirical growth literature. In Section A.5, we provide auxiliary tables and figures of the empirical study.

Recall that in the main text, we have let $\xi_{NT} = \min\{N, T\}$, $\|\cdot\|_{\mathrm{sp}}$ be the spectral norm of a matrix, and $\lfloor a \rfloor$ stand for the largest integer part of a real number a. Here we further define some notations, which will be used throughout this file. Let $\phi_i^*[\beta^*] = (x_{i1}^{*\prime}\beta^*(z_{i1}), \ldots, x_{iT}^{*\prime}\beta^*(z_{iT}))'$, and $\phi_i^\dagger[\beta^\dagger] = (x_{i1}^{\dagger\prime}\beta^\dagger(z_{i1}), \ldots, x_{iT}^{\dagger\prime}\beta^\dagger(z_{iT}))'$, where $\beta^*(\cdot)$ and $\beta^\dagger(\cdot)$ are $p^* \times 1$ and $(p - p^*) \times 1$ respectively. Moreover, $\mathrm{diag}\{A_1, \ldots, A_k\}$ means constructing block diagonal matrix from matrices (or scalars) $A_1, \ldots, A_k$.

A.1 Numerical Implementation

The following procedure essentially combines two algorithms discussed in Bai (2009) and Wang and Xia (2009) together. For each given $\lambda = (\lambda_1, \ldots, \lambda_p)'$, the estimates can be obtained using the following iteration procedure. Let $\widehat{C}_\beta^{(n)}$ and $\widehat{F}^{(n)}$ be the estimates obtained from the $n^{th} \geq 1$ iteration. Then, for the $(n+1)^{th}$ iteration, the estimates are obtained as

$$\text{Sub-step 1:} \quad \mathrm{vec}(\widehat{C}_\beta^{(n+1)}) = \left(\sum_{i=1}^N \mathcal{Z}_i' M_{\widehat{F}^{(n)}} \mathcal{Z}_i + \frac{D_{m,p}^{(n)}}{2}\right)^{-1} \sum_{i=1}^N \mathcal{Z}_i' M_{\widehat{F}^{(n)}} Y_i,$$

$$\text{Sub-step 2:} \quad \frac{1}{NT}\sum_{i=1}^N \left(Y_i - \phi_i[\widehat{\beta}_m^{(n+1)}]\right)\left(Y_i - \phi_i[\widehat{\beta}_m^{(n+1)}]\right)' \widehat{F}^{(n+1)} = \widehat{F}^{(n+1)} V_{NT},$$

where $D_{m,p} = I_m \otimes \mathrm{diag}\left\{\frac{\lambda_1}{\|\widehat{C}_{\beta,1}^{(n)}\|}, \ldots, \frac{\lambda_p}{\|\widehat{C}_{\beta,p}^{(n)}\|}\right\}$; and V_{NT} is a diagonal matrix with the diagonal being the r largest eigenvalues of

$$\frac{1}{NT}\sum_{i=1}^N \left(Y_i - \phi_i[\widehat{\beta}_m^{(n+1)}]\right)\left(Y_i - \phi_i[\widehat{\beta}_m^{(n+1)}]\right)'$$

arranged in descending order. We stop the iteration when the estimates reach certain criteria, say $\|\widehat{C}_\beta^{(n+1)} - \widehat{C}_\beta^{(n)}\| \le \epsilon$. To start the above iteration, we randomly generate $\widehat{F}^{(0)}$, where each element of $\widehat{F}^{(0)}$ follows from $N(0,1)$.

To choose the optimal λ, we follow Wang and Xia (2009) to simplify it as follows:

$$\lambda = \nu \left(\|\bar{C}_{\beta,1}\|^{-1}, \ldots, \|\bar{C}_{\beta,p}\|^{-1} \right)', \tag{A.1}$$

where ν is a scalar, and $\bar{C}_{\beta,j}$ stands for the jth row of the unregularized estimator $\bar{C}_\beta$ (i.e., implementing (3.3) of the main text with $\lambda = 0_{p \times 1}$). With the specification of (A.1), the idea for choosing lambda becomes straightforward. The unregularized estimator $\bar{C}_\beta$ is a consistent estimator. It provides information on how likely each row of C_{β_0} is a zero row. In other words, smaller $\|\bar{C}_{\beta,j}\|$ implies that the jth row of C_{β_0} is more likely to be zero and hence suggests a larger regularizer on $\|C_{\beta,j}\|$. Given (A.1), the selection on the vector λ reduces to the selection on the scalar ν. Finally, we consider the possible value of ν over a sufficiently large interval of the real line. The optimal ν is chosen by minimizing the BIC type criteria proposed in the main text. For the HD case, Υ_{NT} is chosen as $\frac{\ln \xi_{NT}}{\sqrt[8]{\xi_{NT}}}$ in view of the development of Lemma A.7 and Theorem 3.3.

A.2   A Numerical Study

In this section, we examine the performance of the methodology of Section 3 through several simulations. Consider the model (2.4) of the main text. For the factor structure, let $f_{0t} \sim$ i.i.d. $N(0_{r \times 1}, I_r)$ and $\gamma_{0i} \sim$ i.i.d. $N(0.5 \cdot 1_{r \times 1}, I_r)$. In order to generate the regressors and univariate index variable, we firstly generate $v_{it} = 0.5 \cdot v_{i,t-1} + \xi_{it}$, where $\xi_{it} \sim$ i.i.d. $N(0_{p \times 1}, I_p)$. Then let $x_{it} = v_{it} + |\gamma_{0i}' f_{0t}|$, and $z_{it} = |v_{it,1}| + $ i.i.d. $N(0,1)$, where $v_{it,1}$ stands for the first element of v_{it}. By doing so, we generate certain correlation between the regressors and the factor structure, and also introduce some correlation between z_{it} and x_{it}. The error terms are generated as $\varepsilon_t = 0.5 \cdot \varepsilon_{t-1} + \zeta_t$ in which $\zeta_t \sim$ i.i.d. $N(0_{N \times 1}, \Sigma_\zeta)$ and $\Sigma_\zeta = \{0.5^{|i-j|}\}_{N \times N}$, so that the weak cross-sectional dependence among individuals, and serial correlation over time dimension are generated. For both LD and HD cases, the rest settings are as follows:

- **LD Case:** $p^* = 2$, $p = 5$, $r = 3$, and let $\beta_{01}(z) = \exp(-z^2/2) + 0.4$ and $\beta_{02}(z) = z \cdot \exp(-z^2/2) + 0.7$;

- **HD Case:** $p^* = 2 \cdot \lfloor 1.2(NT)^{1/6} \rfloor$, $p = 30$, $r = 3$. For $j = 1, \ldots, p^*$, $\beta_{0j}(z) = \exp(-z^2/2) + 0.4$ when j is odd, and $\beta_{0j}(z) = z \cdot \exp(-z^2/2) + 0.7$ when j is even.

For each dataset generated, we implement the procedure of Section 3 to perform variable selection first. After identifying $\mathcal{A}^*$ and $\mathcal{A}^\dagger$, we implement post-selection estimation (i.e., remove the irrelevant regressors and then implement (3.3) of the main text with $\lambda = 0_{p \times 1}$) for the

coefficient functions. We adopt the Hermite functions of Dong and Linton (2018) as the basis functions, repeat the above procedure 1000 times, and let[1] $N \in \{40, 80, 120\}, T \in \{40, 80, 120\}$, and $m = \lfloor 1.2(NT)^{1/6} \rfloor$.

To evaluate our simulation results, we firstly report two percentages: (1) the percentage of missed true regressors (i.e., false negative rate, FNR); and (2) the percentage of falsely selected noise regressors (i.e., false positive rate, FPR). Secondly, we evaluate the estimates on the components of $\beta_0(\cdot)$. Take $\beta_{01}(\cdot)$ as an example. For the jth replication, we obtain $\widehat{\beta}_{1j}(z)$ for $\forall z$ (given it is not identified as 0; otherwise, we record 0 as the estimate). For $\forall z$, we calculate $\widehat{\beta}_1(z) = \frac{1}{1000} \sum_{j=1}^{1000} \widehat{\beta}_{1j}(z)$, and also record the 95% confidence bands based on $\{\widehat{\beta}_{1j}(z) \,|\, j = 1, \ldots, 1000\}$. We plot these values over a certain range of z. The values of $\beta_{01}(z)$ are plotted in solid black line, and the values of $\widehat{\beta}_1(z)$ are plotted in red dotted line, and the associated 95% confidence bands are plotted in blue dashed curves.

Table A.1 summarizes the FNR and FPR for the LD and HD cases respectively. It is clear that our method proposed in Section 3 works well, as both FNR and FPR are either 0 or very close to 0. It is worth mentioning that although FPR is slightly higher than zero for the HD case, over selecting the regressors will still yield consistent estimation.

Table A.1: FNR & FPR

		FNR			FPR		
	$N \setminus T$	40	80	120	40	80	120
LD	40	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
	40	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
	120	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
HD	40	0.00%	0.00%	0.00%	9.90%	1.10%	1.60%
	40	0.00%	0.00%	0.00%	1.70%	1.90%	1.20%
	120	0.00%	0.00%	0.00%	3.60%	1.00%	0.80%

For both the LD and HD cases, we plot $\beta_{01}(z)$ and $\beta_{02}(z)$ on $[-1, 2]$ in Figures A.1-A.4, as the majority of z_{it}'s lie in this range. Due to similarity, we do not report the estimates of the rest coefficient functions for HD case. It is easy to see that as the sample size increases, the 95% confidence bands become much narrower and the mean estimate approaches the true curve. Also, the estimates from the HD case have wider 95% confidence bands, and seem to be less accurate compared to the LD case as expected.

---

[1]Note that the optimal choice of m may not be the optimal one, but it satisfies all the requirements of our assumptions. Although the optimal choice of truncation parameter and the optimal bandwidth selection have been solved for some cross-sectional models and time series models (e.g., Gao, 2007; Hall et al., 2007) under the low dimensional cases, it is well understood that the question is still open even for the nonparametric panel data model with fixed effects (cf., Chen et al., 2012; Su and Jin, 2012). The question is even more daunting when the factor structure and variable selection procedure get involved.
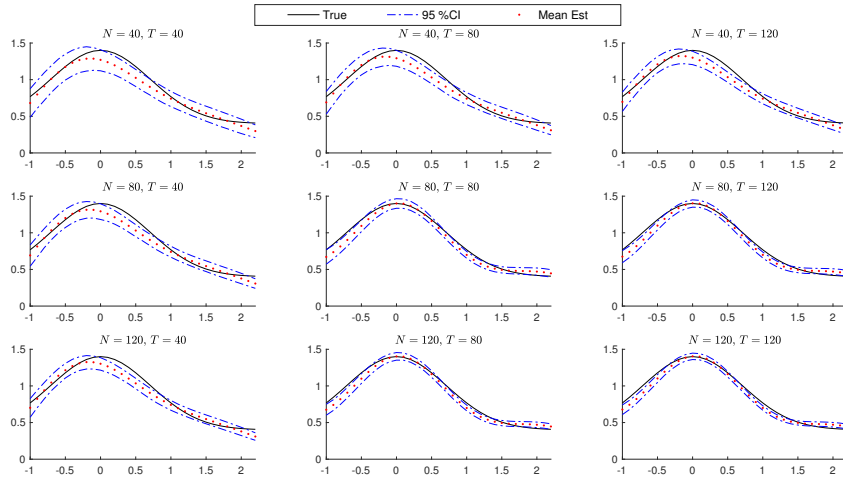
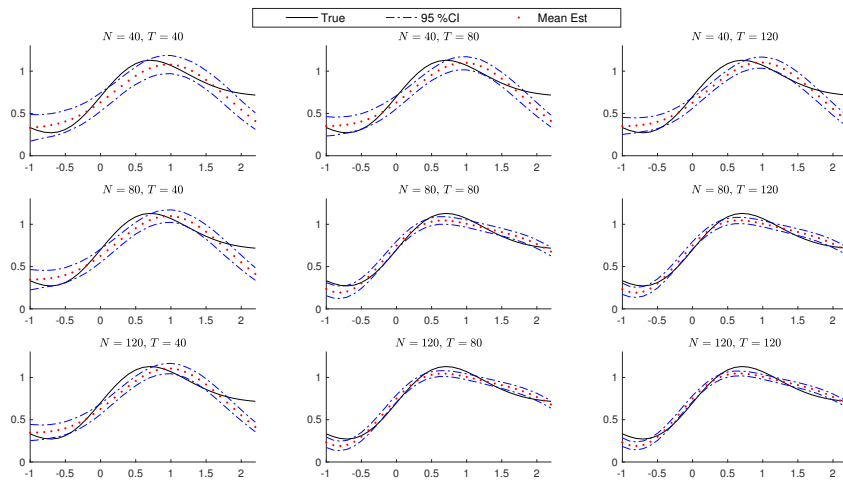Figure A.1: LD: $\beta_{01}(z) = \exp(-z^2/2) + 0.4$
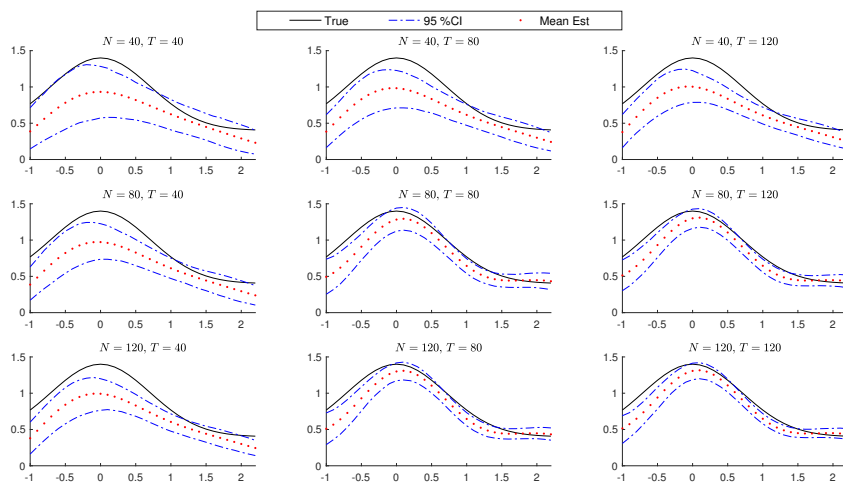

Figure A.2: LD: $\beta_{02}(z) = z\exp(-z^2/2) + 0.7$
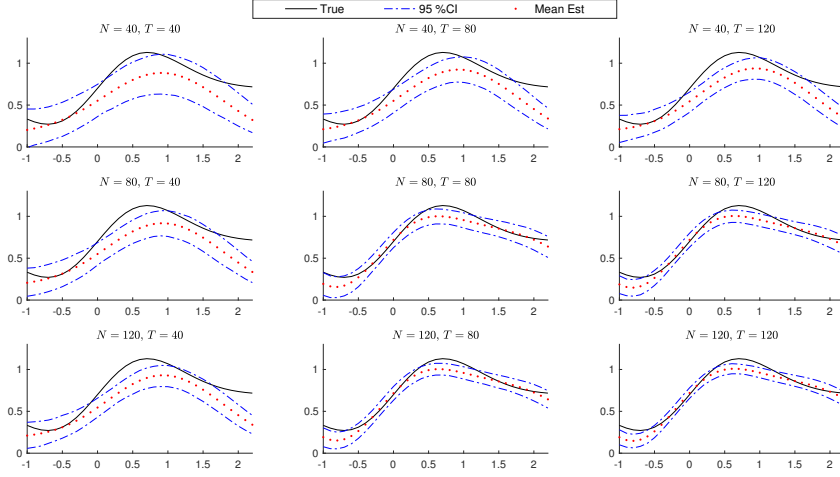

Figure A.3: HD: $\beta_{01}(z) = \exp(-z^2/2) + 0.4$

4

Figure A.4: HD: $\beta_{02}(z) = z \exp(-z^2/2) + 0.7$

A.3 Proofs

Before proving the main theorems, we present the following preliminary lemmas.

Lemma A.1. *Consider two non-singular symmetric matrices A, B with the same dimensions $k \times k$, where k tends to ∞. Suppose that their minimum eigenvalues satisfy that $\eta_{\min}(A) > 0$ and $\eta_{\min}(B) > 0$ uniformly in k. Then $\left\| A^{-1} - B^{-1} \right\| \leq \eta_{\min}^{-1}(A) \cdot \eta_{\min}^{-1} \left\| A - B \right\|$.*

Lemma A.2. *Let Assumptions 1 and 2 hold. As $(N, T) \to (\infty, \infty)$,*

1. $\left\| \frac{1}{NT} \mathcal{E}' \mathcal{E} \right\| = O_P \left(\frac{1}{\sqrt{N}} \right) + O_P \left(\frac{1}{\sqrt{T}} \right)$ *and* $\left\| \frac{1}{NT} \mathcal{E} \mathcal{E}' \right\| = O_P \left(\frac{1}{\sqrt{N}} \right) + O_P \left(\frac{1}{\sqrt{T}} \right)$, *where* $\mathcal{E} = (\mathcal{E}_1, \ldots, \mathcal{E}_N)'$,

2. $\displaystyle \sup_{F \in \mathsf{D}_F} \frac{1}{NT} \sum_{i=1}^{N} \mathcal{E}_i' P_F \mathcal{E}_i = O_P \left(\frac{1}{\sqrt{N}} \right) + O_P \left(\frac{1}{\sqrt{T}} \right)$,

3. $\displaystyle \sup_{F \in \mathsf{D}_F} \left| \frac{1}{NT} \sum_{i=1}^{N} \gamma_{0i}' F_0' M_F \mathcal{E}_i \right| = O_P \left(\frac{1}{\sqrt[4]{N}} \right) + O_P \left(\frac{1}{\sqrt[4]{T}} \right)$,

4. $\displaystyle \sup_{\|C_\beta\| \leq M, F \in \mathsf{D}_F} \left| \frac{1}{NT} \sum_{i=1}^{N} \left(\phi_i[\beta_{0,m}] - \phi_i[\beta_m] \right)' M_F \mathcal{E}_i \right| = O_P \left(\frac{1}{\sqrt[4]{N}} \right) + O_P \left(\frac{1}{\sqrt[4]{T}} \right)$,

5. $\displaystyle \sup_{F \in \mathsf{D}_F} \left| \frac{1}{NT} \sum_{i=1}^{N} \phi_i[\Delta_m]' M_F \phi_i[\Delta_m] \right| = O_P \left(m^{-\mu} \right)$,

6. $\displaystyle \sup_{F \in \mathsf{D}_F} \left| \frac{1}{NT} \sum_{i=1}^{N} \phi_i[\Delta_m]' M_F F_0 \gamma_{0i} \right| = O_P(m^{-\frac{\mu}{2}})$,

7. $\displaystyle \sup_{\|C_\beta\| \leq M, F \in \mathsf{D}_F} \left| \frac{1}{NT} \sum_{i=1}^{N} \phi_i[\Delta_m]' M_F \left\{ \phi_i[\beta_m] - \phi_i[\beta_{0,m}] \right\} \right| = O_P(m^{-\frac{\mu}{2}})$,

5

where M is a sufficiently large constant.

Let $\Pi_{NT}^{-1} = V_{NT}(F_0'\widehat{F}/T)^{-1}(\Gamma_0'\Gamma_0/N)^{-1}$, where V_{NT} is a diagonal matrix with the diagonal being the r largest eigenvalues of

$$\frac{1}{NT}\sum_{i=1}^{N}\left(Y_i - \phi_i[\widehat{\beta}_m]\right)\left(Y_i - \phi_i[\widehat{\beta}_m]\right)'$$

arranged in descending order.

Lemma A.3. *Let Assumptions 1, 2 and 3.1 hold. As $(N,T) \to (\infty,\infty)$,*

1. $\|\widehat{\beta}_m - \beta_0\|_{L^2} = o_P(1)$;

2. $\|P_{\widehat{F}} - P_{F_0}\| = o_P(1)$;

3. $V_{NT} \to_P V$, *where V is an $r \times r$ diagonal matrix consisting of the eigenvalues of $\Sigma_f\Sigma_\gamma$;*

4. $\frac{1}{\sqrt{T}}\|\widehat{F}\Pi_{NT}^{-1} - F_0\| = O_P(\|\widehat{\beta}_m - \beta_0\|_{L^2}) + O_P\left(\frac{1}{\sqrt{N}}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$;

5. $\left\|\frac{1}{T}\widehat{F}'(\widehat{F} - F_0\Pi_{NT})\right\| = O_P(\|\widehat{\beta}_m - \beta_0\|_{L^2}) + O_P\left(\frac{1}{N}\right) + O_P\left(\frac{1}{T}\right)$;

6. $\|P_{\widehat{F}} - P_{F_0}\|^2 = O_P(\|\widehat{\beta}_m - \beta_0\|_{L^2}) + O_P\left(\frac{1}{N}\right) + O_P\left(\frac{1}{T}\right)$.

Lemma A.4. *Let Assumptions 1, 2 and 3 hold. As $(N,T) \to (\infty,\infty)$, $\Pr(\|\widehat{C}_\beta^\dagger\| = 0) \to 1$.*

Lemma A.5. *Let Assumptions 1-3 hold. As $(N,T) \to (\infty,\infty)$,*

$$\|\widehat{C}_\beta^* - C_{\beta_0}^*\| = O_P\left(\sqrt{\frac{m}{NT}}\right) + O_P(m^{-\frac{\mu}{2}}) + O_P\left(\frac{m\lambda_{\max}^*}{NT}\right).$$

Lemma A.6. *Let Assumptions 1, 2 and 5 hold. As $(N,T) \to (\infty,\infty)$,*

1. $\displaystyle\sup_{\|C_\beta\| \le a_0\sqrt{p}, F \in \mathsf{D}_F}\left|\frac{1}{NT}\sum_{i=1}^{N}(\phi_i[\beta_{0,m}] - \phi_i[\beta_m])'M_F\mathcal{E}_i\right| = O_P\left(\sqrt{\frac{p(\xi_{NT} + mp)}{NT}}\right)$;

2. $\displaystyle\sup_{F \in \mathsf{D}_F}\left|\frac{1}{NT}\sum_{i=1}^{N}\phi_i[\Delta_m]'M_F\phi_i[\Delta_m]\right| = O_P\left(p^*m^{-\mu}\right)$;

3. $\displaystyle\sup_{F \in \mathsf{D}_F}\left|\frac{1}{NT}\sum_{i=1}^{N}\phi_i[\Delta_m]'M_FF_0\gamma_{0i}\right| = O_P(\sqrt{p^*}m^{-\frac{\mu}{2}})$;

4. $\displaystyle\sup_{F \in \mathsf{D}_F}\left|\frac{1}{NT}\sum_{i=1}^{N}\phi_i[\Delta_m]'M_F\mathcal{E}_i\right| = O_P(\sqrt{p^*}m^{-\frac{\mu}{2}})$;

5. $\displaystyle\sup_{\|C_\beta\| \le a_0\sqrt{p}, F \in \mathsf{D}_F}\left|\frac{1}{NT}\sum_{i=1}^{N}\phi_i[\Delta_m]'M_F\{\phi_i[\beta_m] - \phi_i[\beta_{0,m}]\}\right| = O_P(\sqrt{pp^*}m^{-\frac{\mu}{2}})$,

where a_0 is a sufficiently large constant.

Lemma A.7. *Let Assumptions 1, 2 and 5 hold. As $(N, T) \to (\infty, \infty)$,*

1. *$\|\widehat{\beta}_m - \beta_0\|_{L^2} = o_P(1)$;*

2. *$\|P_{\widehat{F}} - P_{F_0}\| = o_P(1)$.*

Proof of Theorem 3.1:

(1). The first result follows from Lemma A.4.

(2). Based on the development of Lemma A.5, the definition of $\widehat{\beta}_m^*$ and (3.1), we can write for $\forall z \in V_z$,

$$
\begin{aligned}
\sqrt{\frac{NT}{m}} & \left(\widehat{\beta}_m^*(z) - \beta_0^*(z) \right) \\
= & \sqrt{\frac{NT}{m}} \left[H_m'(z) \otimes I_{p^*} \right] \left[\mathrm{vec}(\widehat{C}_\beta^* - C_{\beta_0}^*) \right] + \sqrt{\frac{NT}{m}} \Delta_m^*(z) \\
= & \sqrt{\frac{NT}{m}} \left[H_m'(z) \otimes I_{p^*} \right] \left[\mathrm{vec}(\widehat{C}_\beta^* - \widehat{C}_\beta^\sharp) + \mathrm{vec}(\widehat{C}_\beta^\sharp - C_{\beta_0}^*) \right] + o_P(1) \\
= & \sqrt{\frac{NT}{m}} \left[H_m'(z) \otimes I_{p^*} \right] A_{1NT}^{-1} \Sigma_{\mathcal{Z}}^{*-1} \cdot \frac{1}{NT} \sum_{i=1}^{N} \left\{ \mathcal{Z}_i^{*\prime} M_{\widehat{F}} + A_{3,i} \right\} \mathcal{E}_i \\
& + \sqrt{\frac{NT}{m}} \left[H_m'(z) \otimes I_{p^*} \right] A_{1NT}^{-1} \Sigma_{\mathcal{Z}}^{*-1} \cdot J_{6NT,1} + O_P\left(\frac{m\|\lambda^*\|}{\sqrt{NT}} \right) + o_P(1) \\
= & \sqrt{\frac{NT}{m}} \left[H_m'(z) \otimes I_{p^*} \right] A_{1NT}^{-1} \Sigma_{\mathcal{Z}}^{*-1} \cdot \frac{1}{NT} \sum_{i=1}^{N} \left\{ \mathcal{Z}_i^{*\prime} M_{\widehat{F}} + A_{3,i} \right\} \mathcal{E}_i + o_P(1) \\
:= & \Lambda_1 + o_P(1),
\end{aligned}
$$

where the second equality follows from $\|\Delta_m(z)\| = O(m^{-\mu/2})$ and the condition $\frac{NT}{m^{\mu+1}} \to 0$; the third equality follows from the above development on $\mathrm{vec}(\widehat{C}_\beta^*) - \mathrm{vec}(\widehat{C}_\beta^\sharp)$ and $\mathrm{vec}(\widehat{C}_\beta^\sharp) - \mathrm{vec}(C_{\beta_0}^*)$, and the fact that $\Sigma_{\mathcal{Z},f}^{*-1}$ reduces to $\Sigma_{\mathcal{Z}}^{*-1}$ using Assumption 4; and the fourth equality follows from (B.22) of Appendix B, $\frac{mN}{T} \to 0$, and $\frac{m\lambda_{\max}^*}{\sqrt{NT}} \to 0$.

We next consider Λ_1 by starting with $\frac{1}{NT} \sum_{i=1}^{N} \mathcal{Z}_i^{*\prime} M_{\widehat{F}} \mathcal{E}_i$.

$$
\begin{aligned}
\frac{1}{NT} \sum_{i=1}^{N} \mathcal{Z}_i^{*\prime} M_{\widehat{F}} \mathcal{E}_i &= \frac{1}{NT} \sum_{i=1}^{N} \mathcal{Z}_i^{*\prime} M_{F_0} \mathcal{E}_i + \frac{1}{NT} \sum_{i=1}^{N} \mathcal{Z}_i^{*\prime} (M_{\widehat{F}} - M_{F_0}) \mathcal{E}_i \\
&= \frac{1}{NT} \sum_{i=1}^{N} \mathcal{Z}_i^{*\prime} M_{F_0} \mathcal{E}_i - \frac{1}{NT} \sum_{i=1}^{N} \mathcal{Z}_i^{*\prime} (P_{\widehat{F}} - P_{F_0}) \mathcal{E}_i \\
&:= D_1 - D_2.
\end{aligned}
$$

Firstly, we shall show $\left\| \sqrt{\frac{NT}{m}} \left[H_m'(z) \otimes I_{p^*} \right] A_{1NT}^{-1} \Sigma_{\mathcal{Z}}^{*-1} D_2 \right\| = o_P(1)$. Let $\mathcal{Z}_{i,j}^*$ be the jth column of $\mathcal{Z}_i^*$, and let $\mathcal{Z}_{it,j}^*$ be the tth element of $\mathcal{Z}_{i,j}^*$. Write

7

$$D_2 = \frac{1}{NT} \sum_{i=1}^{N} \mathcal{Z}_i^{*\prime} \left(\frac{\widehat{F}\widehat{F}'}{T} - P_{F_0} \right) \mathcal{E}_i$$

$$= \frac{1}{NT} \sum_{i=1}^{N} \frac{\mathcal{Z}_i^{*\prime}(\widehat{F} - F_0\Pi_{NT})}{T} \Pi_{NT}' F_0' \mathcal{E}_i + \frac{1}{NT} \sum_{i=1}^{N} \frac{\mathcal{Z}_i^{*\prime}(\widehat{F} - F_0\Pi_{NT})}{T} (\widehat{F} - F_0\Pi_{NT})' \mathcal{E}_i$$

$$+ \frac{1}{NT} \sum_{i=1}^{N} \frac{\mathcal{Z}_i^{*\prime} F_0\Pi_{NT}}{T} (\widehat{F} - F_0\Pi_{NT})' \mathcal{E}_i + \frac{1}{NT} \sum_{i=1}^{N} \frac{\mathcal{Z}_i^{*\prime} F_0}{T} [\Pi_{NT}\Pi_{NT}' - (F_0'F_0/T)^{-1}] F_0' \mathcal{E}_i$$

$$:= D_{21} + D_{22} + D_{23} + D_{24},$$

where the definitions of D_{21} to D_{24} are obvious.

In the following, we let $D_{2\ell,j}$ be the jth row of $D_{2\ell}$ for $\ell = 1, 2, 3, 4$. Thus, for D_{21}, consider

$$\|D_{21,j}\| = \left\| \frac{1}{NT} \sum_{i=1}^{N} \frac{\mathcal{Z}_{i,j}^{*\,\prime}(\widehat{F} - F_0\Pi_{NT})}{T} \Pi_{NT}' F_0' \mathcal{E}_i \right\|$$

$$\leq \left\| \frac{1}{NT} \sum_{i=1}^{N} (\mathcal{E}_i' F_0) \otimes \frac{\mathcal{Z}_{i,j}^{*\,\prime}}{\sqrt{T}} \right\| \cdot \left\| \frac{1}{\sqrt{T}} \operatorname{vec} \left[(\widehat{F} - F_0\Pi_{NT})\Pi_{NT}' \right] \right\|$$

$$= O_P \left(\frac{1}{\sqrt{NT}} \right) \frac{1}{\sqrt{T}} \|\widehat{F} - F_0\Pi_{NT}\|.$$

Summing up over j for $D_{21,j}$, we obtain that $\|D_{21}\| = O_P\left(\sqrt{\frac{m}{NT}}\right) \frac{1}{\sqrt{T}} \|\widehat{F} - F_0\Pi_{NT}\|$.

For D_{22}, write

$$\|D_{22,j}\| = \left\| \frac{1}{NT} \sum_{i=1}^{N} \frac{\mathcal{Z}_{i,j}^{*\,\prime}(\widehat{F} - F_0\Pi_{NT})}{T} (\widehat{F} - F_0\Pi_{NT})' \mathcal{E}_i \right\|$$

$$\leq \left\| \frac{1}{NT} \sum_{i=1}^{N} \mathcal{E}_i' \otimes \mathcal{Z}_{i,j}^{*\,\prime} \right\| \cdot \left\| \frac{1}{T} \operatorname{vec} \left[(\widehat{F} - F_0\Pi_{NT})(\widehat{F} - F_0\Pi_{NT})' \right] \right\|$$

$$= O_P \left(\frac{1}{\sqrt{N}} \right) \frac{1}{T} \|\widehat{F} - F_0\Pi_{NT}\|^2.$$

Summing $D_{22,j}$ up over j, we obtain that $\|D_{22}\| = O_P\left(\sqrt{\frac{m}{N}}\right) \frac{1}{T} \|\widehat{F} - F_0\Pi_{NT}\|^2$.

For D_{23}, write

$$\|D_{23,j}\| = \left\| \frac{1}{NT} \sum_{i=1}^{N} \frac{\mathcal{Z}_{i,j}^{*\,\prime} F_0\Pi_{NT}}{T} (\widehat{F} - F_0\Pi_{NT})' \rceil_i \right\|$$

$$\leq \left\| \frac{1}{NT} \sum_{i=1}^{N} \mathcal{E}_i' \otimes \frac{\mathcal{Z}_{i,j}^{*\,\prime} F_0}{\sqrt{T}} \right\| \cdot \left\| \frac{1}{\sqrt{T}} \operatorname{vec} \left[\Pi_{NT}(\widehat{F} - F_0\Pi_{NT})' \right] \right\|.$$

Note that

$$E \left\| \frac{1}{NT} \sum_{i=1}^{N} \mathcal{E}_i' \otimes \frac{\mathcal{Z}_{i,j}^{*\,\prime} F_0}{\sqrt{T}} \right\|^2$$

$$= \frac{1}{N^2 T^3} E \left\| \sum_{i=1}^{N} \mathcal{E}_i' \otimes \sum_{t=1}^{T} \mathcal{Z}_{it,j}^{*} f_{0t}' \right\|^2 = \frac{1}{N^2 T^3} \sum_{s=1}^{T} E \left\| \sum_{i=1}^{N} \varepsilon_{is} \sum_{t=1}^{T} \mathcal{Z}_{it,j}^{*} f_{0t}' \right\|^2$$

$$= \frac{1}{N^2T^3} \sum_{s=1}^{T} \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} E\left[\left(\sum_{t=1}^{T} \mathcal{Z}_{i_1 t,j}^* f_{0t}'\right)\left(\sum_{t=1}^{T} \mathcal{Z}_{i_2 t,j}^* f_{0t}\right)\right] E[\varepsilon_{i_1 s}\varepsilon_{i_2 s}]$$

$$= \frac{1}{N^2T^2} \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} E\left[\left(\sum_{t=1}^{T} \mathcal{Z}_{i_1 t,j}^* f_{0t}'\right)\left(\sum_{t=1}^{T} \mathcal{Z}_{i_2 t,j}^* f_{0t}\right)\right] \sigma_{i_1 i_2}$$

$$= \frac{1}{N^2T^2} \sum_{t=1}^{T} \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} E\left[\mathcal{Z}_{i_1 t,j}^* \mathcal{Z}_{i_2 t,j}^* E\left[\|f_{0t}\|^2 \mid \mathcal{X}_{Nt}\right]\right] \sigma_{i_1 i_2}$$

$$+ \frac{2}{N^2T^2} \sum_{t_1>t_2} \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} E\left[\mathcal{Z}_{i_1 t_1,j}^* \mathcal{Z}_{i_2 t_2,j}^* E\left[f_{0t_1}' f_{0t_2} \mid \mathcal{X}_{Nt_1}\right]\right] \sigma_{i_1 i_2}$$

$$= \frac{1}{N^2T^2} \sum_{t=1}^{T} \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} E\left[\mathcal{Z}_{i_1 t,j}^* \mathcal{Z}_{i_2 t,j}^*\right] a_{tt}\sigma_{i_1 i_2}$$

$$+ \frac{2}{N^2T^2} \sum_{t_1>t_2} \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} E\left[\mathcal{Z}_{i_1 t_1,j}^* \mathcal{Z}_{i_2 t_2,j}^*\right] a_{t_1 t_2}\sigma_{i_1 i_2}$$

$$\leq O(1) \frac{2}{N^2T^2} \sum_{t_1 \geq t_2} \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} |a_{t_1 t_2}| \cdot |\sigma_{i_1 i_2}| = O(1)\frac{1}{NT},$$

where the fourth equality follows from Assumption 1.2; the sixth equality follows from Assumption 4; and the seventh equality follows from both Assumptions 1.1 and 4.1.

Thus, $\|D_{23,j}\| = O_P\left(\frac{1}{\sqrt{NT}}\right) \frac{1}{\sqrt{T}} \left\|\widehat{F} - F_0\Pi_{NT}\right\|$. Summing $D_{23,j}$ up over j, we obtain that $\|D_{23}\| = O_P\left(\sqrt{\frac{m}{NT}}\right) \frac{1}{\sqrt{T}}\|\widehat{F} - F_0\Pi_{NT}\|$.

Similarly, write for D_{24},

$$\|D_{24,j}\| = \left\| \frac{1}{NT} \sum_{i=1}^{N} \frac{\mathcal{Z}_{i,j}^{*}{}' F_0}{T}[\Pi_{NT}\Pi_{NT}' - (F_0'F_0/T)^{-1}]F_0'\mathcal{E}_i \right\|$$

$$\leq \left\| \frac{1}{NT} \sum_{i=1}^{N} (\mathcal{E}_i' F_0) \otimes \frac{\mathcal{Z}_{i,j}^{*}{}' F_0}{T} \right\| \cdot \left\| \Pi_{NT}\Pi_{NT}' - (F_0'F_0/T)^{-1} \right\|$$

$$= O_P\left(\frac{1}{\sqrt{NT}}\right) \left\| \Pi_{NT}\Pi_{NT}' - (F_0'F_0/T)^{-1} \right\|.$$

Summing $D_{24,j}$ up over j, we obtain that $\|D_{24}\| = O_P\left(\sqrt{\frac{m}{NT}}\right) \left\|\Pi_{NT}\Pi_{NT}' - (F_0'F_0/T)^{-1}\right\|$, where $\left\|\Pi_{NT}\Pi_{NT}' - (F_0'F_0/T)^{-1}\right\| = o_P(1)$ by the development for the fourth result of Lemma A.3.

Based on the analyses of D_{21} to D_{24}, we obtain

$$\sqrt{\frac{NT}{m}}\|D_2\| = O_P(1)\frac{1}{\sqrt{T}}\|\widehat{F} - F_0\Pi_{NT}\| + O_P(1)\left\|\Pi_{NT}\Pi_{NT}' - (F_0'F_0/T)^{-1}\right\|$$

$$+ O_P(1)\sqrt{T} \cdot \frac{1}{T}\|\widehat{F} - F_0\Pi_{NT}\|^2$$

which further gives $\left\|\sqrt{\frac{NT}{m}} [H_m'(z) \otimes I_{d_x}] A_{1NT}^{-1} \Sigma_{\mathcal{Z}}^{*-1} D_2\right\| = o_P(1)$ by Lemma A.3 and the condition $\frac{T}{N^2} \to 0$.

Similarly, we obtain

9

$$\left\| \sqrt{\frac{NT}{m}} \left[H'_m(z) \otimes I_{p^*} \right] A_{1NT}^{-1} \Sigma_{\mathcal{Z}}^{*-1} \frac{1}{NT} \sum_{i=1}^{N} A_{3,i} \mathcal{E}_i \right.$$

$$\left. - \sqrt{\frac{NT}{m}} \left[H'_m(z) \otimes I_{p^*} \right] A_{1NT}^{-1} \Sigma_{\mathcal{Z}}^{*-1} \frac{1}{NT} \sum_{i=1}^{N} \widetilde{A}_{3,i} \mathcal{E}_i \right\| = o_P(1),$$

where $\widetilde{A}_{3,i} = \frac{1}{N} \sum_{j=1}^{N} \mathcal{Z}_j^{*\prime} M_{F_0} \gamma'_{0j} (\Gamma'_0 \Gamma_0 / N)^{-1} \gamma_{0i}$.

Finally, by Assumption 4 and after some simple algebra, we obtain

$$\begin{aligned} \Lambda_1 &= \sqrt{\frac{NT}{m}} \left[H'_m(z) \otimes I_{p^*} \right] \widetilde{A}_{1NT}^{-1} \Sigma_{\mathcal{Z}}^{*-1} \cdot \frac{1}{NT} \sum_{i=1}^{N} \left\{ \mathcal{Z}_i^{*\prime} M_{F_0} + \widetilde{A}_{3,i} \right\} \mathcal{E}_i + o_P(1) \\ &\to_D N(0, \Omega_\star), \end{aligned}$$

where $\widetilde{A}_{1NT} = I_{mp^*} - \Sigma_{\mathcal{Z}}^{*-1} \widetilde{A}_{2NT}$ and $\widetilde{A}_{2NT} = \frac{1}{N^2 T} \sum_{i=1}^{N} \sum_{j=1}^{N} \mathcal{Z}_i^{*\prime} M_{F_0} \mathcal{Z}_j^* \gamma'_{0j} \Sigma_\gamma^{-1} \gamma_{0i}$. The proof is then complete. ∎

Proof of Theorem 3.2:

Recall that we have denoted the set $\mathcal{A}^*$. Before proceeding further, we introduce some variables to facilitate the development. For an arbitrary model S, we say it is under-fitted if it misses at least one variable with a nonzero coefficient[2]; it is over-fitted if S not only includes all relevant variables but also includes at least one redundant regressor (i.e., $\mathcal{A}^* \subset S$ but $\mathcal{A}^* \neq S$). Then, according to whether the model S_λ is under fitted, correctly fitted, or over fitted, we create three mutually exclusive sets A^-, $A^0 = \{\lambda \in \mathbb{R}^p : S_\lambda = \mathcal{A}^*\}$ and $A^+ = \{\lambda \in \mathbb{R}^p : S_\lambda \supset \mathcal{A}^*, S_\lambda \neq \mathcal{A}^*\}$. Suppose that there is a sequence $\{\lambda_{NT}\}$ that ensures the conditions required by Lemma A.4. Let $(\widehat{C}_\beta^{\lambda_{NT}}, \widehat{F}^{\lambda_{NT}})$ denote the estimator obtained by implementing (3.3) using λ_{NT}.

Case 1: Under-fitted model. Without loss of generality, we assume that only one variable is missing, and suppose that the first $p^* - 1$ rows of $\widehat{C}_\beta^\lambda$ are obtained from the under-fitted model and the p^{*th} row of $\widehat{C}_\beta^\lambda$ is a 0 row. Moreover, let $\text{RSS}_0 = \frac{1}{NT} \sum_{i=1}^{N} \left(Y_i - \phi_i[\beta_{0,m}] \right)' M_{F_0} \left(Y_i - \phi_i[\beta_{0,m}] \right)$.

We then write

$$\begin{aligned} \text{RSS}_\lambda - \text{RSS}_0 &= \frac{1}{NT} \sum_{i=1}^{N} \left(Y_i - \phi_i[\widehat{\beta}_m^\lambda] \right)' M_{\widehat{F}^\lambda} \left(Y_i - \phi_i[\widehat{\beta}_m^\lambda] \right) \\ &\quad - \frac{1}{NT} \sum_{i=1}^{N} \left(Y_i - \phi_i[\beta_{0,m}] \right)' M_{F_0} \left(Y_i - \phi_i[\beta_{0,m}] \right) \\ &\geq \rho_1 \| C_{\beta_0, p^*} \|^2 > \frac{\rho_1}{2} \| \beta_{0p^*} \|_{L^2}^2 > 0, \end{aligned}$$

where the first inequality follows from the development given for (B.3) of Appendix B.

Again, using the development given for (B.3) of Appendix B, we have

---

[2]Under-fitted case allows for including redundant regressor.

$$\mathrm{RSS}_{\lambda_{NT}} - \mathrm{RSS}_0$$

$$= \mathrm{vec}(C_{\beta_0} - \widehat{C}_\beta^{\lambda_{NT}})' \frac{1}{NT} \sum_{i=1}^N \mathcal{Z}_i' M_{\widehat{F}^{\lambda_{NT}}} \mathcal{Z}_i \, \mathrm{vec}(C_{\beta_0} - \widehat{C}_\beta^{\lambda_{NT}})$$

$$+ \frac{1}{NT} \mathrm{tr}\left(M_{\widehat{F}^{\lambda_{NT}}} F_0 \Gamma_0' \Gamma_0 F_0' M_{\widehat{F}^{\lambda_{NT}}} \right)$$

$$+ 2 \, \mathrm{vec}(C_{\beta_0} - \widehat{C}_\beta^{\lambda_{NT}})' \frac{1}{NT} \sum_{i=1}^N \mathcal{Z}_i' M_{\widehat{F}^{\lambda_{NT}}} F_0 \gamma_{0i} + o_P(1)$$

$$= o_P(1),$$

where the second equality follows from the development of Lemma A.3.

Thus, we can conclude that $\Pr\left(\inf_{\lambda \in A^-} \mathrm{BIC}_\lambda > \mathrm{BIC}_{\lambda_{NT}} \right) \to 1$.

Case 2: Over-fitted model. Consider $\forall \lambda \in A^+$ and recall that $\widehat{C}_\beta^\lambda$ determines a model S_λ. Under such a model S_λ, we can define another unpenalized estimator as

$$(\check{C}_\beta, \check{F}) = \operatorname*{argmin}_{C_\beta, F} \frac{1}{NT} \sum_{i=1}^N \left(Y_i - \phi_i[\beta_m] \right)' M_F \left(Y_i - \phi_i[\beta_m] \right) \tag{A.1}$$

subject to $F \in \mathsf{D}_F$, where $\|C_{\beta,j}\| = 0$ with $\forall j \notin S_\lambda$. In other words, $(\check{C}_\beta, \check{F})$ is the unpenalized estimator under the model determined by $\widehat{C}_\beta^\lambda$. By definition, we obtain immediately that $\mathrm{RSS}_\lambda \geq \mathrm{RSS}_{S_\lambda}$, where $\mathrm{RSS}_{S_\lambda} = \frac{1}{NT} \sum_{i=1}^N \left(Y_i - \phi_i[\check{\beta}_m] \right)' M_{\check{F}} \left(Y_i - \phi_i[\check{\beta}_m] \right)$.

Write

$$\ln \mathrm{RSS}_{S_\lambda} - \ln \mathrm{RSS}_{\lambda_{NT}} = \ln \left(1 + \frac{\mathrm{RSS}_{S_\lambda} - \mathrm{RSS}_{\lambda_{NT}}}{\mathrm{RSS}_{\lambda_{NT}}} \right)$$

$$\geq -\frac{\mathrm{RSS}_{S_\lambda} - \mathrm{RSS}_{\lambda_{NT}}}{\mathrm{RSS}_{\lambda_{NT}}}.$$

In view of the proof of Lemma A.3, it is easy to see that $\mathrm{RSS}_{\lambda_{NT}}$ converges to a positive constant. With regard to $\mathrm{RSS}_{S_\lambda} - \mathrm{RSS}_{\lambda_{NT}}$, we have

$$\mathrm{RSS}_{S_\lambda} - \mathrm{RSS}_{\lambda_{NT}} = \frac{1}{NT} \sum_{i=1}^N \left(Y_i - \phi_i[\check{\beta}_m] \right)' M_{\check{F}} \left(Y_i - \phi_i[\check{\beta}_m] \right)$$

$$- \frac{1}{NT} \sum_{i=1}^N \left(Y_i - \phi_i[\widehat{\beta}_m^{\lambda_{NT}}] \right)' M_{\widehat{F}^{\lambda_{NT}}} \left(Y_i - \phi_i[\beta_m^{\lambda_{NT}}] \right).$$

By Lemmas A.2 and A.5, it is not hard to see $|\mathrm{RSS}_{S_\lambda} - \mathrm{RSS}_{\lambda_{NT}}| \leq O_P(1) \frac{1}{\sqrt[4]{N}}$, so we can further write

$$\ln \mathrm{RSS}_{S_\lambda} - \ln \mathrm{RSS}_{\lambda_{NT}} \geq -\frac{\mathrm{RSS}_{S_\lambda} - \mathrm{RSS}_{\lambda_{NT}}}{\mathrm{RSS}_{\lambda_{NT}}} \geq -\left| O_P(1) \frac{1}{\sqrt[4]{N}} \right|.$$

We then write

$$\inf_{\lambda \in A^+} \mathrm{BIC}_\lambda - \mathrm{BIC}_{\lambda_{NT}} = \inf_{\lambda \in A^+} \ln \mathrm{RSS}_{S_\lambda} - \ln \mathrm{RSS}_{\lambda_{NT}} + (\mathrm{df}_\lambda - \mathrm{df}_{\lambda_{NT}}) \frac{\ln N}{\sqrt[4]{N}}.$$

By Lemma A.4, we know that $\Pr(\mathrm{df}_{\lambda_{NT}} = p^*) \to 1$. Since $\lambda \in A^+$, we must have that $\Pr(\mathrm{df}_\lambda \geq p^* + 1) \to 1$. Then it is clear that $\Pr\left(\inf_{\lambda \in A^+} \mathrm{BIC}_\lambda > \mathrm{BIC}_{\lambda_{NT}}\right) \to 1$.

Combining Cases 1 and 2, we obtain that $\Pr\left(\inf_{\lambda \in A^- \cup A^+} \mathrm{BIC}_\lambda > \mathrm{BIC}_{\lambda_{NT}}\right) \to 1$. This further indicates that $\Pr(S_{\widehat{\lambda}} = \mathcal{A}^*) \to 1$. The proof is now complete. ∎

Proof of Theorem 3.3:

(1). By (B.26) of Appendix B, and the development of (1) and (5) of Lemma A.6, we now can further conclude that

$$\left| \frac{1}{NT} \sum_{i=1}^N \left(\phi_i[\beta_{0,m}] - \phi_i[\widehat{\beta}_m] \right)' M_{\widehat{F}} \mathcal{E}_i \right| = o_P\left(\sqrt{\frac{\xi_{NT} + mp}{NT}} \right),$$

$$\left| \frac{1}{NT} \sum_{i=1}^N \phi_i[\Delta_m]' M_{\widehat{F}} \left\{ \phi_i\left[\widehat{\beta}_m\right] - \phi_i[\beta_{0,m}] \right\} \right| = o_P(\sqrt{p^*} m^{-\frac{\mu}{2}}),$$

which allows us to improve the rate of (B.26) of Appendix B as follows.

$$\|C_{\beta_0} - \widehat{C}_\beta\|^2 = o_P\left(\sqrt{\frac{\xi_{NT} + mp}{NT}} + \sqrt{p^*} m^{-\frac{\mu}{2}} \right) + O_P\left(\frac{1}{\sqrt[4]{\xi_{NT}}} + \frac{p^* \lambda^*_{\max}}{NT} \right). \qquad \text{(A.2)}$$

Then we can conclude $\|C_{\beta_0} - \widehat{C}_\beta\| = O_P(h_{NT}^{1/2})$ by Assumption 5.2, where $h_{NT} = \frac{1}{\sqrt[4]{\xi_{NT}}}$.

Thus, for a large constant A, $\widehat{C}_\beta$ lies in a ball $\left\{ C_\beta \mid \|C_\beta - C_{\beta_0}\| \leq A h_{NT}^{1/2} \right\}$ with probability approaching 1. Corresponding to $\widehat{C}_\beta^*$ and $\widehat{C}_\beta^\dagger$, construct C_β and U as $C_\beta = (C_\beta^{*\prime}, C_\beta^{\dagger\prime})'$ and $U = (U^{*\prime}, U^{\dagger\prime})'$, where $C_\beta^* = C_{\beta_0}^* + h_{NT}^{1/2} U^*$ and $C_\beta^\dagger = C_{\beta_0}^\dagger + h_{NT}^{1/2} U^\dagger = h_{NT}^{1/2} U^\dagger$ with $\|U\|^2 \leq A^2$. Further define

$$V_\lambda(U^*, U^\dagger, F) = \frac{1}{NT} Q_\lambda(C_\beta, F) = \frac{1}{NT} Q_\lambda((C_{\beta_0}^{*\prime} + h_{NT}^{1/2} U^{*\prime}, h_{NT}^{1/2} U^{\dagger\prime})', F), \qquad \text{(A.3)}$$

so $\widehat{C}_\beta^*$ and $\widehat{C}_\beta^\dagger$ can be obtained by minimizing $V_\lambda(U^*, U^\dagger, F)$ over $\|U\|^2 \leq A^2$ except on an event with probability converging to 0.

It suffices to show that for any $U = (U^*, U^\dagger)$ with $\|U\|^2 \leq A^2$ and $\|U^\dagger\| > 0$, $V_\lambda(U^*, U^\dagger, F) > V_{NT}(U^*, 0_{(p-p^*) \times m}, F)$ with probability converging to 1 regardless the value of F. Recall that some notations used below have been defined in the beginning of the supplementary file. Further denote that $\beta_m^\dagger(\cdot) = C_\beta^\dagger H_m(\cdot)$. Then write

$$V_\lambda(U^*, U^\dagger, F) - V_{NT}(U^*, 0_{(p-p^*) \times m}, F)$$

$$= \frac{1}{NT} \sum_{i=1}^N \Delta\phi_i^*[\beta_m^*]' M_F \Delta\phi_i^*[\beta_m^*] + \frac{1}{NT} \sum_{i=1}^N \Delta\phi_i^\dagger[\beta_m^\dagger]' M_F \Delta\phi_i^\dagger[\beta_m^\dagger]$$

$$+ \frac{1}{NT} \sum_{i=1}^N \gamma_{0i}' F_0' M_F F_0 \gamma_{0i} + \frac{1}{NT} \sum_{i=1}^N \mathcal{E}_i' M_F \mathcal{E}_i$$

$$+ \frac{2}{NT} \sum_{i=1}^N \Delta\phi_i^*[\beta_m^*]' M_F \mathcal{E}_i + \frac{2}{NT} \sum_{i=1}^N \Delta\phi_i^\dagger[\beta_m^\dagger]' M_F \mathcal{E}_i + \frac{2}{NT} \sum_{i=1}^N \gamma_{0i}' F_0' M_F \mathcal{E}_i$$

12

$$+\frac{2}{NT}\sum_{i=1}^{N}\Delta\phi_i^*[\beta_m^*]'M_F F_0\gamma_{0i} + \frac{2}{NT}\sum_{i=1}^{N}\Delta\phi_i^\dagger[\beta_m^\dagger]'M_F F_0\gamma_{0i}$$

$$+\frac{2}{NT}\sum_{i=1}^{N}\Delta\phi_i^*[\beta_m^*]'M_F \phi_i^\dagger[\beta^\dagger] + \sum_{j=1}^{p^*}\frac{\lambda_j}{NT}\|C_{\beta,j}\| + \sum_{j=p^*+1}^{p}\frac{\lambda_j}{NT}\|C_{\beta,j}\|$$

$$-\frac{1}{NT}\sum_{i=1}^{N}\Delta\phi_i^*[\beta_m^*]'M_F\Delta\phi_i^*[\beta_m^*] - \frac{1}{NT}\sum_{i=1}^{N}\gamma_{0i}'F_0'M_F F_0\gamma_{0i} - \frac{1}{NT}\sum_{i=1}^{N}\mathcal{E}_i'M_F\mathcal{E}_i$$

$$-\frac{2}{NT}\sum_{i=1}^{N}\Delta\phi_i^*[\beta_m^*]'M_F\mathcal{E}_i - \frac{2}{NT}\sum_{i=1}^{N}\gamma_{0i}'F_0'M_F\mathcal{E}_i$$

$$-\frac{2}{NT}\sum_{i=1}^{N}\Delta\phi_i^*[\beta_m^*]'M_F F_0\gamma_{0i} - \sum_{j=1}^{p^*}\frac{\lambda_j}{NT}\|C_{\beta,j}\|$$

$$=\frac{1}{NT}\sum_{i=1}^{N}\Delta\phi_i^\dagger[\beta_m^\dagger]'M_F\Delta\phi_i^\dagger[\beta_m^\dagger] + \frac{2}{NT}\sum_{i=1}^{N}\Delta\phi_i^*[\beta_m^*]'M_F\phi_i^\dagger[\beta_m^\dagger]$$

$$+\frac{2}{NT}\sum_{i=1}^{N}\Delta\phi_i^\dagger[\beta_m^\dagger]'M_F\mathcal{E}_i + \frac{2}{NT}\sum_{i=1}^{N}\Delta\phi_i^\dagger[\beta_m^\dagger]'M_F F_0\gamma_{0i} + \sum_{j=p^*+1}^{p}\frac{\lambda_j}{NT}\|C_{\beta,j}\|$$

$$= B_{1NT} + 2B_{2NT} + 2B_{3NT} + 2B_{4NT} + B_{5NT},$$

where the definitions of B_{jNT} for $j = 1, \ldots, 5$ are obvious.

For $B_{1NT} + 2B_{2NT}$, write

$$B_{1NT} + 2B_{2NT} \geq \frac{1}{NT}\sum_{i=1}^{N}\Delta\phi_i^\dagger[\beta_m^\dagger]'M_F\Delta\phi_i^\dagger[\beta_m^\dagger]$$

$$-\frac{1}{NT}\sum_{i=1}^{N}\Delta\phi_i^*[\beta_m^*]'M_F\Delta\phi_i^*[\beta_m^*] - \frac{1}{NT}\sum_{i=1}^{N}\phi_i^\dagger[\beta_m^\dagger]'M_F\phi_i^\dagger[\beta_m^\dagger]$$

$$\geq -\frac{1}{NT}\sum_{i=1}^{N}\Delta\phi_i^*[\beta_m^*]'M_F\Delta\phi_i^*[\beta_m^*]$$

$$\geq -h_{NT}\frac{1}{NT}\sum_{i=1}^{N}U^{*\prime}\mathcal{Z}_i^{*\prime}M_F\mathcal{Z}_i^*U^*$$

$$\geq -h_{NT}\rho_1\|U^*\|^2 \geq -h_{NT}\rho_1 A^2$$

where the third inequality follows by construction.

For $B_{3NT} + B_{4NT}$, it is easy to know that

$$|B_{3NT} + B_{4NT}| \leq \left\{\frac{1}{NT}\sum_{i=1}^{N}\|\Delta\phi_i^\dagger[\beta_m^\dagger]\|^2\right\}^{1/2} \cdot \left\{\frac{1}{NT}\sum_{i=1}^{N}\mathcal{E}_i'M_F\mathcal{E}_i\right\}^{1/2}$$

$$+\left\{\frac{1}{NT}\sum_{i=1}^{N}\|\Delta\phi_i^\dagger[\beta_m^\dagger]\|^2\right\}^{1/2} \cdot \left\{\frac{1}{NT}\sum_{i=1}^{N}\gamma_{0i}'F_0'M_F F_0\gamma_{0i}\right\}^{1/2}$$

$$= O_P(1)h_{NT}^{1/2}.$$

For B_{5NT}, we have

$$B_{5NT} = \sum_{j=p^*+1}^{p} \frac{\lambda_j}{NT} \|C_{\beta,j}\| \geq \frac{\lambda_{\min}^{\dagger}}{NT} \left(\sum_{j=p^*+1}^{p} \|C_{\beta,j}\|^2 \right)^{1/2} = \frac{\lambda_{\min}^{\dagger}}{NT} \|U^{\dagger}\|.$$

In view of the above development and the condition $\frac{\lambda_{\min}^{\dagger}}{NTh_{NT}^{1/2}} \to \kappa_3$ with κ_3 being sufficiently large, the first result follows.

(2). As in the case of Theorem 3.2, we define three mutually exclusive sets A^-, $A^0 = \{\lambda \in \mathbb{R}^p : S_\lambda = \mathcal{A}^*\}$ and $A^+ = \{\lambda \in \mathbb{R}^p : S_\lambda \supset \mathcal{A}^*, S_\lambda \neq \mathcal{A}^*\}$ according to whether the model S_λ is under fitted, correctly fitted, or over fitted respectively. Suppose that there is a sequence $\{\lambda_{NT}\}$ that satisfies he conditions required by the first result of this theorem. Let $(\widehat{C}^{\lambda_{NT}}, \widehat{F}^{\lambda_{NT}})$ denote the estimator obtained by implementing (3.3) using λ_{NT}.

Case 1: Under-fitted model. Without loss of generality, we assume that only one variable is missing, so suppose that the first $p^* - 1$ rows of $\widehat{C}_\beta^\lambda$ are obtained from the under-fitted model and the p^{*th} row of $\widehat{C}_\beta^\lambda$ is a 0 row. Moreover, let $\text{RSS}_0 = \frac{1}{NT} \sum_{i=1}^{N} (Y_i - \phi_i[\beta_{0,m}])' M_{F_0} (Y_i - \phi_i[\beta_{0,m}])$.

We then write

$$\text{RSS}_\lambda - \text{RSS}_0 = \frac{1}{NT} \sum_{i=1}^{N} (Y_i - \phi_i[\widehat{\beta}_m^\lambda])' M_{\widehat{F}^\lambda} (Y_i - \phi_i[\widehat{\beta}_m^\lambda])$$

$$-\frac{1}{NT} \sum_{i=1}^{N} (Y_i - \phi_i[\beta_{0,m}])' M_{F_0} (Y_i - \phi_i[\beta_{0,m}])$$

$$\geq \rho_1 \|C_{\beta_0,p^*}\|^2 > \frac{\rho_1}{2} \|\beta_{0p^*}\|_{L^2}^2 > 0,$$

where the first inequality follows from the development of Lemma A.7. Similarly, we have

$$\text{RSS}_{\lambda_{NT}} - \text{RSS}_0$$

$$= \text{vec}(C_{\beta_0} - \widehat{C}_\beta^{\lambda_{NT}})' \frac{1}{NT} \sum_{i=1}^{N} \mathcal{Z}_i' M_{\widehat{F}^{\lambda_{NT}}} \mathcal{Z}_i \, \text{vec}(C_{\beta_0} - \widehat{C}_\beta^{\lambda_{NT}})$$

$$+\frac{1}{NT} \text{tr} \left(M_{\widehat{F}^{\lambda_{NT}}} F_0 \Gamma_0' \Gamma_0 F_0' M_{\widehat{F}^{\lambda_{NT}}} \right)$$

$$+2 \, \text{vec}(C_{\beta_0} - \widehat{C}_\beta^{\lambda_{NT}})' \frac{1}{NT} \sum_{i=1}^{N} \mathcal{Z}_i' M_{\widehat{F}^{\lambda_{NT}}} F_0 \gamma_{0i} + o_P(1)$$

$$= o_P(1).$$

Thus, we can conclude that $\Pr \left(\inf_{\lambda \in A^-} \text{BIC}_\lambda > \text{BIC}_{\lambda_{NT}} \right) \to 1$.

Case 2: Over-fitted model. Consider $\forall \lambda \in A^+$ and recall that $\widehat{C}_\beta^\lambda$ determines a model S_λ. Under such a model S_λ, we can define another unpenalized estimator as

$$(\check{C}_\beta, \check{F}) = \operatorname*{argmin}_{C_\beta, F} \frac{1}{NT} \sum_{i=1}^{N} (Y_i - \phi_i[\beta_m])' M_F (Y_i - \phi_i[\beta_m]) \tag{A.4}$$

14

subject to $\|C_\beta\| \le a_0\sqrt{p}$ and $F \in \mathsf{D}_F$, where, for $j = 1, \dots, p$, $\|C_{\beta,j}\| = 0$ with $\forall j \notin S_\lambda$. In other words, $(\check{C}_\beta, \check{F})$ is the unpenalized estimator under the model determined by $\widehat{C}_\beta^\lambda$. By definition, we obtain immediately that $\mathrm{RSS}_\lambda \ge \mathrm{RSS}_{S_\lambda}$, where $\mathrm{RSS}_{S_\lambda} = \frac{1}{NT} \sum_{i=1}^N \left(Y_i - \phi_i[\check{\beta}_m] \right)' M_{\check{F}} \left(Y_i - \phi_i[\check{\beta}_m] \right)$.

Write

$$\ln \mathrm{RSS}_{S_\lambda} - \ln \mathrm{RSS}_{\lambda_{NT}} = \ln \left(1 + \frac{\mathrm{RSS}_{S_\lambda} - \mathrm{RSS}_{\lambda_{NT}}}{\mathrm{RSS}_{\lambda_{NT}}} \right)$$
$$\ge -\frac{\mathrm{RSS}_{S_\lambda} - \mathrm{RSS}_{\lambda_{NT}}}{\mathrm{RSS}_{\lambda_{NT}}}.$$

In view of the proof of Lemma A.7, it is easy to see that $\mathrm{RSS}_{\lambda_{NT}}$ converges to a positive constant. As for $\mathrm{RSS}_{S_\lambda} - \mathrm{RSS}_{\lambda_{NT}}$, we obtain

$$\mathrm{RSS}_{S_\lambda} - \mathrm{RSS}_{\lambda_{NT}} = \frac{1}{NT} \sum_{i=1}^N \left(Y_i - \phi_i[\check{\beta}_m] \right)' M_{\check{F}} \left(Y_i - \phi_i[\check{\beta}_m] \right)$$
$$- \frac{1}{NT} \sum_{i=1}^N \left(Y_i - \phi_i[\widehat{\beta}_m^{\lambda_{NT}}] \right)' M_{\widehat{F}^{\lambda_{NT}}} \left(Y_i - \phi_i[\beta_m^{\lambda_{NT}}] \right).$$

Using the development of Lemma A.7, it is not hard to see $|\mathrm{RSS}_{S_\lambda} - \mathrm{RSS}_{\lambda_{NT}}| \le O_P(1) h_{NT}^{1/2}$. Thus, we can further write

$$\ln \mathrm{RSS}_{S_\lambda} - \ln \mathrm{RSS}_{\lambda_{NT}} \ge -\frac{\mathrm{RSS}_{S_\lambda} - \mathrm{RSS}_{\lambda_{NT}}}{\mathrm{RSS}_{\lambda_{NT}}} \ge - \left| O_P(1) h_{NT}^{1/2} \right|.$$

We then write

$$\inf_{\lambda \in A^+} \mathrm{BIC}_\lambda - \mathrm{BIC}_{\lambda_{NT}} = \inf_{\lambda \in A^+} \ln \mathrm{RSS}_{S_\lambda} - \ln \mathrm{RSS}_{\lambda_{NT}} + (\mathrm{df}_\lambda - \mathrm{df}_{\lambda_{NT}}) \Upsilon_{NT}.$$

By the first result of this theorem, we know that $\Pr(\mathrm{df}_{\lambda_{NT}} = p^*) \to 1$. Since $\lambda \in A^+$, we must have $\Pr(\mathrm{df}_\lambda \ge p^* + 1) \to 1$. Given $\Upsilon_{NT} h_{NT}^{-1/2} \to \infty$, it is clear $\Pr\left(\inf_{\lambda \in A^+} \mathrm{BIC}_\lambda > \mathrm{BIC}_{\lambda_{NT}} \right) \to 1$.

Combining Cases 1 and 2, we obtain that $\Pr\left(\inf_{\lambda \in A^- \cup A^+} \mathrm{BIC}_\lambda > \mathrm{BIC}_{\lambda_{NT}} \right) \to 1$. This further indicates that $\Pr(S_{\widehat{\lambda}} = \mathcal{A}^*) \to 1$. The proof is now complete. ∎

A.4 Discussion on Time Trends

As mentioned in Section 1, our primary focus is on proposing an integrated framework to tackle the three issues of variable selection, parameter heterogeneity, and cross-sectional dependence in the context of cross-country growth regressions. We do not attempt to address other interesting topics such as convergence of countries or any further model specification issues. That said, we would like to briefly discuss the issue of time trend that has found limited attention in the empirical growth literature (Eberhardt and Teal, 2011) but can be partially solved under our setting.

15

As pointed out by Eberhardt and Teal (2011), any macro production function is likely to contain at least some countries with certain time trends in the input and output variables, and these time-series properties need to be taken into account in the empirical analysis. The question in fact has somewhat been addressed by recent developments in econometrics. We provide two examples below.

Though macro variables are likely to have certain time trends, they do not have to be as strong as polynomial terms t or t^2 (Dong and Linton, 2018). More often than not, it can simply be captured by a structure like

$$x_{it} = A_i' d_{0t} + B_i' f_{0t} + u_{it}, \tag{A.1}$$

where f_{0t} is the same as that of (2.4) of the main text, d_{0t} includes other unobservable common shocks, both A_i and B_i are the unknown factor loadings, and u_{it} stands for an error term. This structure has been well discussed in Pesaran (2006) and Kapetanios et al. (2011).

In the second case, as in Pedroni (2007) we assume that different countries have different types of time trends. This case has been partially discussed in Chen et al. (2012) and Gao et al. (2018). In particular, Gao et al. (2018) allow the regressors x_{it} to have the following form

$$x_{it} = g_i(t/T) + v_i + u_{it}, \tag{A.2}$$

where $g_i(\cdot)$ is some trending function and varies across i, v_i is an individual effect, and u_{it} stands for an error term. Using (A.2), $g_i(t/T)$ can mimic different types of time trends for each individual country.

For either case of (A.1) and (A.2), our methodology including the estimator and its asymptotic results remains valid with some minor modifications regarding the proof.

A.5 Additional Tables and Figures for the Empirical Results

Figure A.5: Estimates of Coefficient Functions of All Selected Variables
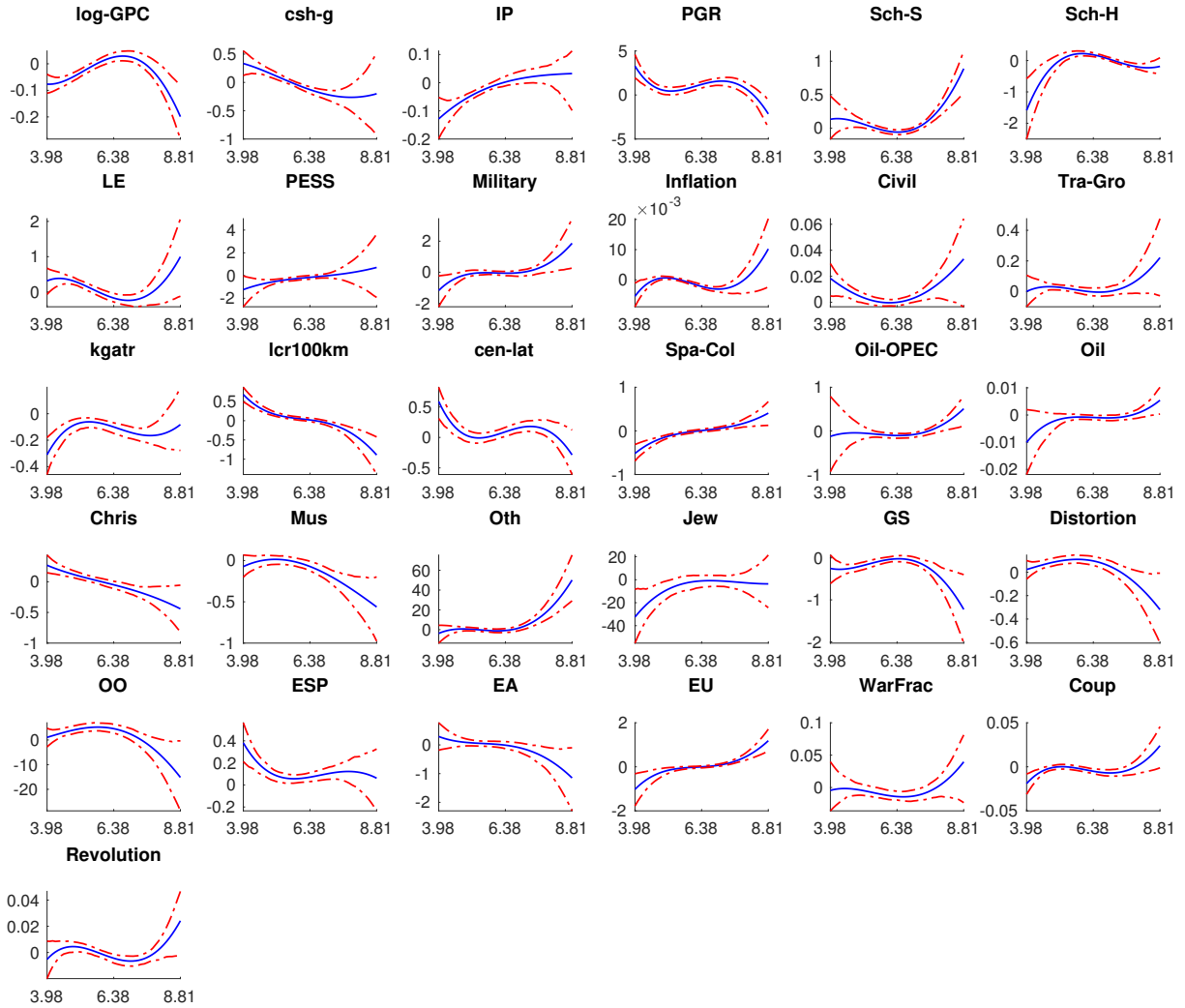
As discussed in the context of Figure 1 in the main text, these confidence intervals need to be interpreted with caution. As well understood, one cannot establish the confidence intervals for the estimates under HD case unless certain transformation is further employed (e.g., Huang et al., 2008; Dong et al., 2017). However, if one regards 31 as a relatively small number after selection, we can then employ a procedure similar to the relevant literature by considering our regression under LD framework. In order to ensure the validity of the bootstrap procedure, stronger assumptions on the error terms are needed. For example, one can employ the martingale difference type of assumptions (see Assumption A.4 of Su et al., 2015), or simply assume that the error terms are i.i.d. over both i and t. Generally speaking, when the error term exhibits both cross-sectional and serial correlation, the bootstrap results are not reliable or incorrect.

References

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

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

Dong, C., Gao, J. and Linton, O. (2017), High dimensional semiparametric moment restriction models. https://ssrn.com/abstract=3045063.

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

Eberhardt, M. and Teal, F. (2011), 'Econometrics for grumblers: A new look at the literature on cross-country growth empirics', *Journal of Economic Surveys* **25**(1), 109–155.

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

Gao, J., Xia, K. and Zhu, H. (2018), 'Heterogeneous panel data models with cross-sectional dependence', *Journal of Econometrics* p. forthcoming.

Hall, P., Li, Q. and Racine, J. S. (2007), 'Nonparametric estimation of regression functions in the presence of irrelevant regressors', *Review of Economics and Statistics* **89**(4), 784–789.

Huang, J., Horowitz, J. L. and Ma, S. (2008), 'Asymptotic properties of bridge estimators in sparse high-dimensional regression models', *Annals of Statistics* **36**(2), 587–613.

Kapetanios, G., Pesaran, M. H. and Yamagata, T. (2011), 'Panels with non-stationary multifactor error structures', *Journal of Econometrics* **160**(2), 326–348.

Pedroni, P. (2007), 'Social capital, barriers to production and capital shares: implications for the importance of parameter heterogeneity from a nonstationary panel approach', *Journal of Applied Econometrics* **22**(2), 326–348.

Pesaran, M. H. (2006), 'Estimation and inference in large heterogeneous panels with a multifactor error structure', *Econometrica* **74**(4), 967–1012.

Su, L. and Jin, S. (2012), 'Sieve estimation of panel data models with cross section dependence', *Journal of Econometrics* **169**(1), 34–47.

Su, L., Jin, S. and Zhang, Y. (2015), 'Specification test for panel data models with interactive fixed effects', *Journal of Econometrics* **186**(1), 222–244.

Wang, H. and Xia, Y. (2009), 'Shrinkage estimation of the varying coefficient', *Journal of the American Statistical Association* **104**(486), 747–757.

Supplementary Appendix B to
"An Integrated Panel Data Approach to Modelling Economic Growth"

Guohua Feng[*], Jiti Gao[♯] and Bin Peng[†]

[*]University of North Texas, [♯]Monash University and [†]University of Bath

Appendix B provides proofs of the preliminary lemmas stated in the supplementary Appendix A.

Proof of Lemma A.1:

Write

$$
\begin{aligned}
\left\| A^{-1} - B^{-1} \right\| = \left\| B^{-1} \left(B - A \right) A^{-1} \right\| &= \left\| \text{vec} \left(B^{-1} \left(B - A \right) A^{-1} \right) \right\| \\
&= \left\| \left(A^{-1} \otimes B^{-1} \right) \text{vec} \left(B - A \right) \right\| \leq \eta_{\min}^{-1} \left(A \otimes B \right) \left\| \text{vec} \left(B - A \right) \right\| \\
&= \eta_{\min}^{-1} \left(A \right) \cdot \eta_{\min}^{-1} \left(B \right) \left\| A - B \right\|.
\end{aligned}
$$

The proof is then complete. ∎

Proof of Lemma A.2:

(1). Firstly, write

$$
\begin{aligned}
\frac{1}{N^2 T^2} E \| \mathcal{E}' \mathcal{E} \|^2 &= E \left\| \frac{1}{NT} \sum_{i=1}^{N} \mathcal{E}_i \mathcal{E}_i' \right\|^2 = \frac{1}{N^2 T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \left(\sum_{i=1}^{N} E[\varepsilon_{it}^2 \varepsilon_{is}^2] + \sum_{i \neq j} E[\varepsilon_{it} \varepsilon_{is} \varepsilon_{jt} \varepsilon_{js}] \right) \\
&= \frac{1}{N^2 T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \left(\sum_{i=1}^{N} E[\varepsilon_{it}^2 \varepsilon_{is}^2] + \sum_{i \neq j} E[(\varepsilon_{it} \varepsilon_{jt} - \sigma_{ij})(\varepsilon_{is} \varepsilon_{js} - \sigma_{ij})] + \sum_{i \neq j} \sigma_{ij}^2 \right) \\
&= \frac{1}{N^2 T^2} \sum_{t=1}^{T} \left(\sum_{i=1}^{N} E[\varepsilon_{it}^4] + \sum_{i \neq j} E[(\varepsilon_{it} \varepsilon_{jt} - \sigma_{ij})^2] \right) \\
&\quad + \frac{1}{N^2 T^2} \sum_{t \neq s} \left(\sum_{i=1}^{N} E[\varepsilon_{it}^2 \varepsilon_{is}^2] + \sum_{i \neq j} E[(\varepsilon_{it} \varepsilon_{jt} - \sigma_{ij})(\varepsilon_{is} \varepsilon_{js} - \sigma_{ij})] \right) + \frac{1}{N^2} \sum_{i \neq j} \sigma_{ij}^2 \\
&= O(1) \frac{1}{N} + O(1) \frac{1}{T}, \tag{B.1}
\end{aligned}
$$

where the fifth equality follows from using the mixing condition on $e_{it} e_{jt}$ across t. Thus, $\frac{1}{NT} \| \mathcal{E}' \mathcal{E} \| = O_P \left(\frac{1}{\sqrt{N}} \right) + O_P \left(\frac{1}{\sqrt{T}} \right)$.

Secondly, note that

$$E\left\|\frac{1}{NT}\mathcal{E}\mathcal{E}'\right\|^2 = \left\{E\left[\frac{1}{NT}\mathcal{E}_i'\mathcal{E}_j\right]^2\right\}_{N\times N} = \sum_{i=1}^N\sum_{j=1}^N\frac{1}{N^2T^2}\sum_{t=1}^T\sum_{s=1}^T E\left[\varepsilon_{it}\varepsilon_{jt}\varepsilon_{is}\varepsilon_{js}\right]$$

$$= \sum_{t=1}^T\sum_{s=1}^T\frac{1}{N^2T^2}\left(\sum_{i=1}^N E[\varepsilon_{it}^2\varepsilon_{is}^2] + \sum_{i\neq j}E[\varepsilon_{it}\varepsilon_{is}\varepsilon_{jt}\varepsilon_{js}]\right) = O\left(\frac{1}{N}\right) + O\left(\frac{1}{T}\right),$$

where the last step follows from (B.1). Thus, $\frac{1}{NT}\|\mathcal{E}\mathcal{E}'\| = O_P\left(\frac{1}{\sqrt{N}}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$.

(2). Write

$$\sup_{F\in\mathsf{D}_F}\frac{1}{NT}\sum_{i=1}^N\mathcal{E}_i'P_F\mathcal{E}_i = \sup_{F\in\mathsf{D}_F}\frac{1}{NT}\mathrm{tr}\left(P_F\mathcal{E}'\mathcal{E}\right) \leq \sup_{F\in\mathsf{D}_F}\frac{r}{NT}\|P_F\|_{\mathrm{sp}}\|\mathcal{E}'\mathcal{E}\|_{\mathrm{sp}}$$

$$\leq \sup_{F\in\mathsf{D}_F}\frac{r}{NT}\|P_F\|_{\mathrm{sp}}\|\mathcal{E}'\mathcal{E}\| = O_P\left(\frac{1}{\sqrt{N}}\right) + O_P\left(\frac{1}{\sqrt{T}}\right),$$

where the first inequality follows from the fact that $|\mathrm{tr}\left(A\right)| \leq \mathrm{rank}\left(A\right)\|A\|_{\mathrm{sp}}$; and the second equality follows from (1) of this lemma.

(3). Write

$$\sup_{F\in\mathsf{D}_F}\left|\frac{1}{NT}\sum_{i=1}^N\gamma_{0i}'F_0'M_F\mathcal{E}_i\right| = \sup_{F\in\mathsf{D}_F}\left|\frac{1}{NT}\mathrm{tr}\left(F_0'M_F\mathcal{E}'\Gamma_0\right)\right| \leq \sup_{F\in\mathsf{D}_F}\frac{r}{NT}\left\|F_0'M_F\mathcal{E}'\Gamma_0\right\|_{\mathrm{sp}}$$

$$\leq \sup_{F\in\mathsf{D}_F}\frac{r}{NT}\|F_0\|_{\mathrm{sp}}\|M_F\|_{\mathrm{sp}}\|\Gamma_0'\mathcal{E}\|_{\mathrm{sp}} = \sup_{F\in\mathsf{D}_F}\frac{r}{NT}\|F_0\|_{\mathrm{sp}}\left\|\Gamma_0'\mathcal{E}\mathcal{E}'\Gamma_0\right\|_{\mathrm{sp}}^{1/2}$$

$$= \sup_{F\in\mathsf{D}_F}\frac{r}{\sqrt{NT}}\|F_0\|_{\mathrm{sp}}\|\Gamma_0\|_{\mathrm{sp}}\left(\frac{1}{NT}\|\mathcal{E}\mathcal{E}'\|\right)^{1/2} = O_P\left(\frac{1}{NT}\|\mathcal{E}\mathcal{E}'\|\right)^{1/2}$$

$$= O_P\left(\frac{1}{\sqrt[4]{N}}\right) + O_P\left(\frac{1}{\sqrt[4]{T}}\right),$$

where the first inequality follows from the fact that $|\mathrm{tr}\left(A\right)| \leq \mathrm{rank}\left(A\right)\|A\|_{\mathrm{sp}}$; the second equality follows from Fact 5.10.18 of Bernstein (2005); and the last equality follows from (1) of this lemma.

(4). Write

$$\frac{1}{NT}\sum_{i=1}^N\left(\phi_i[\beta_{0,m}] - \phi_i[\beta_m]\right)'M_F\mathcal{E}_i$$

$$= \frac{1}{NT}\sum_{i=1}^N\left(\phi_i[\beta_{0,m}] - \phi_i[\beta_m]\right)'\mathcal{E}_i + \frac{1}{NT}\sum_{i=1}^N\left(\phi_i[\beta_{0,m}] - \phi_i[\beta_m]\right)'P_F\mathcal{E}_i$$

$$:= \Lambda_1 + \Lambda_2.$$

For Λ_1, write

$$\sup_{\|C_\beta\|\leq M}\left|\frac{1}{NT}\sum_{i=1}^N\left(\phi_i[\beta_{0,m}] - \phi_i[\beta_m]\right)'\mathcal{E}_i\right| \leq \sup_{\|C_\beta\|\leq M}\|\mathrm{vec}(C_{\beta_0} - C_\beta)\| \cdot \left\|\frac{1}{NT}\sum_{i=1}^N\mathcal{Z}_i'\mathcal{E}_i\right\|$$

2

$$= \sup_{\|C_\beta\| \le M} \|C_{\beta_0} - C_\beta\| \cdot O_P\left(\sqrt{\frac{m}{NT}}\right) = O_P\left(\sqrt{\frac{m}{NT}}\right),$$

where the first equality follows from some standard analysis on the term $\frac{1}{NT}\sum_{i=1}^N \mathcal{Z}_i' \mathcal{E}_i$ using Assumption 1.1; and the last equality follows from $\|C_\beta\| \le M$.

In order to consider Λ_2, let $\Delta b = (\phi_1[\beta_{0,m}] - \phi_1[\beta_m], \ldots, \phi_N[\beta_{0,m}] - \phi_N[\beta_m])$. Note that

$$\sup_{\|C_\beta\| \le M} \frac{1}{NT}\|\Delta b\|^2 = \sup_{\|C_\beta\| \le M} \frac{1}{NT}\sum_{i=1}^N (C_\beta - C_{\beta_0})' \mathcal{Z}_i' \mathcal{Z}_i (C_\beta - C_{\beta_0})$$

$$\le O_P(1) \sup_{\|C_\beta\| \le M} \|C_\beta - C_{\beta_0}\|^2 = O_P(1), \tag{B.2}$$

where the inequality follows from Assumptions 2.1, and $\left\|\frac{1}{NT}\sum_{i=1}^N \mathcal{Z}_i' \mathcal{Z}_i' - \Sigma_{\mathcal{Z}}\right\| = o_P(1)$ by some standard analysis using Assumption 1.1; and the last equality follows from $\|C_\beta\| \le M$.

Then we are able to write

$$\sup_{\|C_\beta\| \le M, F \in \mathsf{D}_F} \left|\frac{1}{NT}\sum_{i=1}^N (\phi_i[\beta_{0,m}] - \phi_i[\beta_m])' P_F \mathcal{E}_i\right| = \sup_{\|C_\beta\| \le M, F \in \mathsf{D}_F} \left|\frac{1}{NT}\mathrm{tr}\left(P_F \mathcal{E}' \Delta b'\right)\right|$$

$$\le \frac{r}{NT} \sup_{\|C_\beta\| \le M, F \in \mathsf{D}_F} \|P_F \mathcal{E}' \Delta b'\|_{\mathrm{sp}} \le \sup_{\|C_\beta\| \le M, F \in \mathsf{D}_F} \frac{r}{NT}\|P_F\|_{\mathrm{sp}}\|\mathcal{E}\|_{\mathrm{sp}}\|\Delta b\|_{\mathrm{sp}}$$

$$= \sup_{\|C_\beta\| \le M, F \in \mathsf{D}_F} r\|P_F\|_{\mathrm{sp}}\left(\frac{1}{NT}\|\mathcal{E}\mathcal{E}'\|\right)^{1/2}\left(\frac{1}{\sqrt{NT}}\|\Delta b\|\right) = O_P\left(\frac{1}{\sqrt[4]{N}}\right) + O_P\left(\frac{1}{\sqrt[4]{T}}\right),$$

where the second equality follows from Fact 5.10.18 of Bernstein (2005); and the last step follows from (1) of this lemma and (B.2).

Based on the above development on Λ_1 and Λ_2, the result follows.

(5). Write

$$\sup_{F \in \mathsf{D}_F}\left|\frac{1}{NT}\sum_{i=1}^N \phi_i[\Delta_m]' M_F \phi_i[\Delta_m]\right| \le \left|\frac{1}{NT}\sum_{i=1}^N \phi_i[\Delta_m]' \phi_i[\Delta_m]\right| = O_P\left(m^{-\mu}\right),$$

where the last equality follows from the development of Dong and Linton (2018) using Assumption 2.1.

(6). Write

$$\sup_{F \in \mathsf{D}_F}\left|\frac{1}{NT}\sum_{i=1}^N \phi_i[\Delta_m]' M_F F_0 \gamma_{0i}\right| = \sup_{F \in \mathsf{D}_F}\left|\frac{1}{NT}\mathrm{tr}\left(M_F F_0 \Gamma_0' \Delta'\right)\right|$$

$$\le \frac{r}{NT}\sup_{F \in \mathsf{D}_F}\|M_F\|_{\mathrm{sp}}\|F_0\|_{\mathrm{sp}}\|\Gamma_0\|_{\mathrm{sp}}\|\Delta\|_{\mathrm{sp}} = O_P(m^{-\frac{\mu}{2}}),$$

where $\Delta = (\phi_1(\Delta_m), \ldots, \phi_N(\Delta_m))$; and the second equality follows from $\frac{1}{NT}\|\Delta\|^2 = O_P\left(m^{-\mu}\right)$ as in (5) of this lemma.

(7). Write

$$\sup_{\|C_\beta\|\leq M,\, F\in \mathsf{D}_F} \left| \frac{1}{NT} \sum_{i=1}^{N} \phi_i \left[\Delta_m\right]' M_F \left\{ \phi_i \left[\beta_m\right] - \phi_i \left[\beta_{0,m}\right] \right\} \right|$$

$$\leq \left\{ \frac{1}{NT} \sum_{i=1}^{N} \|\phi_i \left[\Delta_m\right]\|^2 \right\}^{1/2} \cdot \sup_{\|C_\beta\|\leq M} \left\{ \frac{1}{NT} \sum_{i=1}^{N} \|\phi_i \left[\beta_m\right] - \phi_i \left[\beta_{0,m}\right]\|^2 \right\}^{1/2} = O_P(m^{-\frac{\mu}{2}}),$$

where the last equality follows from (5) of this lemma. The proof is now complete. ∎

Proof of Lemma A.3:

(1). For notational simplicity, let $\Delta\phi_i[\beta_m] = \phi_i[\beta_{0,m}] - \phi_i[\beta_m]$. Let $\xi_F = \text{vec}(M_F F_0)$, $A_{1F} = \frac{1}{NT} \sum_{i=1}^{N} \mathcal{Z}_i' M_F \mathcal{Z}_i$, $A_2 = \frac{1}{NT}(\Gamma_0' \Gamma_0) \otimes I_T$, and $A_{3F} = \frac{1}{NT} \sum_{i=1}^{N} \gamma_{0i} \otimes (M_F \mathcal{Z}_i)$, where $\mathcal{Z}_i$ has been defined in Assumption 2. By the definition of (3.3) and Lemma A.2, we have

$$0 \geq \frac{1}{NT} Q_\lambda(\widehat{C}_\beta, \widehat{F}) - \frac{1}{NT} Q_\lambda(C_{\beta_0}, F_0)$$

$$= \frac{1}{NT} \sum_{i=1}^{N} \left(\Delta\phi_i[\widehat{\beta}_m] + F_0 \gamma_{0i} \right)' M_{\widehat{F}} \left(\Delta\phi_i[\widehat{\beta}_m] + F_0 \gamma_{0i} \right) + \frac{1}{NT} \sum_{i=1}^{N} \mathcal{E}_i' M_{\widehat{F}} \mathcal{E}_i$$

$$+ \frac{2}{NT} \sum_{i=1}^{N} \left(\Delta\phi_i[\widehat{\beta}_m] + F_0 \gamma(v_i) \right)' M_{\widehat{F}} \mathcal{E}_i + \sum_{j=1}^{p} \frac{\lambda_j}{NT} \|\widehat{C}_{\beta,j}\|$$

$$- \frac{1}{NT} \sum_{i=1}^{N} (\phi_i[\Delta_m] + \mathcal{E}_i)' M_{F_0} (\phi_i[\Delta_m] + \mathcal{E}_i) - \sum_{j=1}^{p^*} \frac{\lambda_j}{NT} \|C_{\beta_0,j}\|$$

$$= \frac{1}{NT} \sum_{i=1}^{N} \left(\Delta\phi_i[\widehat{\beta}_m] + F_0 \gamma_{0i} \right)' M_{\widehat{F}} \left(\Delta\phi_i[\widehat{\beta}_m] + F_0 \gamma_{0i} \right)$$

$$+ \sum_{j=1}^{p} \frac{\lambda_j}{NT} \|\widehat{C}_{\beta,j}\| - \sum_{j=1}^{p^*} \frac{\lambda_j}{NT} \|C_{\beta_0,j}\| + O_P \left(\frac{1}{\sqrt[4]{\xi_{NT}}} + m^{-\frac{\mu}{2}} \right)$$

$$= \text{vec}(C_{\beta_0} - \widehat{C}_\beta)' \frac{1}{NT} \sum_{i=1}^{N} \mathcal{Z}_i' M_{\widehat{F}} \mathcal{Z}_i \, \text{vec}(C_{\beta_0} - \widehat{C}_\beta) + \frac{1}{NT} \text{tr} \left(M_{\widehat{F}} F_0 \Gamma_0' \Gamma_0 F_0' M_{\widehat{F}} \right)$$

$$+ 2 \, \text{vec}(C_{\beta_0} - \widehat{C}_\beta)' \frac{1}{NT} \sum_{i=1}^{N} \mathcal{Z}_i' M_{\widehat{F}} F_0 \gamma_{0i} + \sum_{j=1}^{p} \frac{\lambda_j}{NT} \|\widehat{C}_{\beta,j}\| - \sum_{j=1}^{p^*} \frac{\lambda_j}{NT} \|C_{\beta_0,j}\|$$

$$+ O_P \left(\frac{1}{\sqrt[4]{\xi_{NT}}} + m^{-\frac{\mu}{2}} \right),$$

where the second equality follows from Lemma A.2. Thus, we can further write

$$\sum_{j=1}^{p^*} \frac{\lambda_j}{NT} \|C_{\beta_0,j}\| \geq \text{vec}(C_{\beta_0} - \widehat{C}_\beta)' \frac{1}{NT} \sum_{i=1}^{N} \mathcal{Z}_i' M_{\widehat{F}} \mathcal{Z}_i \, \text{vec}(C_{\beta_0} - \widehat{C}_\beta) + \frac{1}{NT} \text{tr} \left(M_{\widehat{F}} F_0 \Gamma_0' \Gamma_0 F_0' M_{\widehat{F}} \right)$$

$$+ 2 \, \text{vec}(C_{\beta_0} - \widehat{C}_\beta)' \frac{1}{NT} \sum_{i=1}^{N} \mathcal{Z}_i' M_{\widehat{F}} F_0 \gamma_{0i} + O_P \left(\frac{1}{\sqrt[4]{\xi_{NT}}} + m^{-\frac{\mu}{2}} \right)$$

$$\geq \text{vec}(C_{\beta_0} - \widehat{C}_\beta)' \left(A_{1\widehat{F}} - A_{3\widehat{F}}' A_2^{-1} A_{3\widehat{F}} \right) \text{vec}(C_{\beta_0} - \widehat{C}_\beta)$$

4

$$+[\xi'_{\widehat{F}} + \text{vec}(C_{\beta_0} - \widehat{C}_\beta)' A'_{3\widehat{F}} A_2^{-1}] A_2 [\xi_{\widehat{F}} + A_2^{-1} A_{3\widehat{F}} \text{vec}(C_{\beta_0} - \widehat{C}_\beta)]$$

$$+ O_P \left(\frac{1}{\sqrt[4]{\xi_{NT}}} + m^{-\frac{\mu}{2}} \right)$$

$$\geq O_P(1)\|C_{\beta_0} - \widehat{C}_\beta\|^2 + O_P \left(\frac{1}{\sqrt[4]{\xi_{NT}}} + m^{-\frac{\mu}{2}} \right). \tag{B.3}$$

Till now, we can conclude that

$$\|C_{\beta_0} - \widehat{C}_\beta\|^2 = O_P \left(\frac{1}{\sqrt[4]{\xi_{NT}}} + m^{-\frac{\mu}{2}} + \frac{\lambda^*_{\max}}{NT} \right) = o_P(1), \tag{B.4}$$

where the second equality follows from Assumption 3.

(2). By (B.3) and (B.4), we can further obtain that

$$o_p(1) \geq \frac{1}{NT} \text{tr} \left[\left(F_0' M_{\widehat{F}} F_0 \right) \left(\Gamma_0' \Gamma_0 \right) \right] + o_P(1),$$

so $\frac{1}{NT} \text{tr} \left[\left(F_0' M_{\widehat{F}} F_0 \right) \left(\Gamma_0' \Gamma_0 \right) \right] = o_P(1)$. As in Bai (2009, p. 1265), we can further conclude that $\frac{1}{T} \text{tr} \left(F_0' M_{\widehat{F}} F_0 \right) = o_P(1)$, $\left\| P_{\widehat{F}} - P_{F_0} \right\| = o_P(1)$, and $\frac{1}{T} \widehat{F}' F_0$ is invertible with probability approaching one. Thus, the second result of this lemma follows.

(3). For the results result of this lemma, note that minimizing (3.2) with respect to F is equivalent to minimizing $\sum_{i=1}^N (Y_i - \phi_i[\beta_m])' M_F (Y_i - \phi_i[\beta_m])$ without involving the penalty term. Thus, following the same arguments as in Bai (2009, p. 1236), the estimate $\widehat{F}$ of (3.3) is obtained by

$$\frac{1}{NT} \sum_{i=1}^N \left(Y_i - \phi_i[\widehat{\beta}_m] \right) \left(Y_i - \phi_i[\widehat{\beta}_m] \right)' \widehat{F} = \widehat{F} V_{NT}, \tag{B.5}$$

where V_{NT} is a diagonal matrix with the diagonal being the r largest eigenvalues of

$$\frac{1}{NT} \sum_{i=1}^N \left(Y_i - \phi_i[\widehat{\beta}_m] \right) \left(Y_i - \phi_i[\widehat{\beta}_m] \right)'$$

arranged in descending order.

We now consider V_{NT} and write

$$\widehat{F} V_{NT} = \left[\frac{1}{NT} \sum_{i=1}^N \left(Y_i - \phi_i[\widehat{\beta}_m] \right) \left(Y_i - \phi_i[\widehat{\beta}_m] \right)' \right] \widehat{F}$$

$$= \left[\frac{1}{NT} \sum_{i=1}^N \left(\phi_i[\beta_0] + F_0 \gamma_{0i} + \mathcal{E}_i - \phi_i[\widehat{\beta}_m] \right) \left(\phi_i[\beta_0] + F_0 \gamma_{0i} + \mathcal{E}_i - \phi_i[\widehat{\beta}_m] \right)' \right] \widehat{F}$$

$$= \frac{1}{NT} \sum_{i=1}^N \left(\phi_i[\beta_0] - \phi_i[\widehat{\beta}_m] \right) \left(\phi_i[\beta_0] - \phi_i[\widehat{\beta}_m] \right)' \widehat{F}$$

$$+ \frac{1}{NT} \sum_{i=1}^N \left(\phi_i[\beta_0] - \phi_i[\widehat{\beta}_m] \right) (F_0 \gamma_{0i})' \widehat{F} + \frac{1}{NT} \sum_{i=1}^N (F_0 \gamma_{0i}) \left(\phi_i[\beta_0] - \phi_i[\widehat{\beta}_m] \right)' \widehat{F}$$

5

$$+\frac{1}{NT}\sum_{i=1}^{N}\left(\phi_i[\beta_0]-\phi_i[\widehat{\beta}_m]\right)\mathcal{E}_i'\widehat{F}+\frac{1}{NT}\sum_{i=1}^{N}\mathcal{E}_i\left(\phi_i[\beta_0]-\phi_i[\widehat{\beta}_m]\right)'\widehat{F}$$

$$+\frac{1}{NT}\sum_{i=1}^{N}\mathcal{E}_i\mathcal{E}_i'\widehat{F}+\frac{1}{NT}\sum_{i=1}^{N}F_0\gamma_{0i}\mathcal{E}_i'\widehat{F}+\frac{1}{NT}\sum_{i=1}^{N}\mathcal{E}_i\gamma_{0i}'F_0'\widehat{F}+\frac{1}{NT}\sum_{i=1}^{N}F_0\gamma_{0i}\gamma_{0i}'F_0'\widehat{F}$$

$$:=I_{1NT}(\widehat{\beta}_m,\widehat{F})+\cdots+I_{5NT}(\widehat{\beta}_m,\widehat{F})+I_{6NT}(\widehat{F})+\cdots+I_{9NT}(\widehat{F}),$$

where the definitions of $I_{1NT}(\beta,F)$ to $I_{5NT}(\beta,F)$ and $I_{6NT}(F)$ to $I_{9NT}(F)$ should be obvious.

Note that $I_{9NT}(\widehat{F})=F_0(\Gamma_0'\Gamma_0/N)(F_0'\widehat{F}/T)$. Thus, we can write

$$\widehat{F}V_{NT}-F_0(\Gamma_0'\Gamma_0/N)(F_0'\widehat{F}/T)$$
$$=I_{1NT}(\widehat{\beta}_m,\widehat{F})+\cdots+I_{5NT}(\widehat{\beta}_m,\widehat{F})+I_{6NT}(\widehat{F})+\cdots+I_{8NT}(\widehat{F}). \tag{B.6}$$

Right multiplying each side of (B.6) by $(F_0'\widehat{F}/T)^{-1}(\Gamma_0'\Gamma_0/N)^{-1}$, we obtain

$$\widehat{F}V_{NT}(F_0'\widehat{F}/T)^{-1}(\Gamma_0'\Gamma_0/N)^{-1}-F_0$$
$$=\left[I_{1NT}(\widehat{\beta}_m,\widehat{F})+\cdots+I_{8NT}(\widehat{F})\right](F_0'\widehat{F}/T)^{-1}(\Gamma_0'\Gamma_0/N)^{-1}. \tag{B.7}$$

Below, we examine each term on the right hand side of (B.7) and show that V_{NT} is non-singular. Write

$$\frac{1}{\sqrt{T}}\left\|\widehat{F}V_{NT}(F_0'\widehat{F}/T)^{-1}(\Gamma_0'\Gamma_0/N)^{-1}-F_0\right\|$$
$$\leq\frac{1}{\sqrt{T}}\left[\|I_{1NT}(\widehat{\beta}_m,\widehat{F})\|+\cdots+\|I_{8NT}(\widehat{F})\|\right]\cdot\|(F_0'\widehat{F}/T)^{-1}(\Gamma_0'\Gamma_0/N)^{-1}\|. \tag{B.8}$$

We already know $(F_0'\widehat{F}/T)^{-1}=O_P(1)$ by the proofs of the second result of this lemma and $(\Gamma_0'\Gamma_0/N)^{-1}=O_P(1)$, so focus on $\frac{1}{\sqrt{T}}\|I_{jNT}(\widehat{\beta}_m,\widehat{F})\|$ with $j=1,2,\ldots,5$ and $\frac{1}{\sqrt{T}}\|I_{jNT}(\widehat{F})\|$ with $j=6,7,8$.

For $I_{1NT}(\widehat{\beta}_m,\widehat{F})$, we have

$$\frac{1}{\sqrt{T}}\left\|I_{1NT}(\widehat{\beta}_m,\widehat{F})\right\|\leq\frac{\sqrt{r}}{NT}\sum_{i=1}^{N}\left\|\phi_i[\beta_0]-\phi_i[\widehat{\beta}_m]\right\|^2$$

$$\leq\frac{\sqrt{r}}{NT}\sum_{i=1}^{N}\left\|\phi_i[\beta_{0,m}]-\phi_i[\widehat{\beta}_m]\right\|^2+\frac{\sqrt{r}}{NT}\sum_{i=1}^{N}\|\Delta_m\|^2$$

$$=\text{vec}(C_{\beta_0}-\widehat{C}_{\beta})'\frac{\sqrt{r}}{NT}\sum_{i=1}^{N}\mathcal{Z}_i'\mathcal{Z}_i\,\text{vec}(C_{\beta_0}-\widehat{C}_{\beta})+O_P(m^{-\mu})$$

$$=O_P(\|C_{\beta_0}-\widehat{C}_{\beta}\|^2)+O_P(m^{-\mu})=O_P(\|\widehat{\beta}_m-\beta_0\|_{L^2}^2),$$

where the first and second equalities follow from Assumption 2.1.

For $I_{2NT}(\widehat{\beta}_m,\widehat{F})$, write

$$\frac{1}{\sqrt{T}}\left\|I_{2NT}(\widehat{\beta}_m,\widehat{F})\right\|\leq\frac{\sqrt{r}}{NT}\sum_{i=1}^{N}\left\|\left(\phi_i[\beta_0]-\phi_i[\widehat{\beta}_m]\right)(F_0\gamma_{0i})'\right\|$$

6

$$\leq \sqrt{r} \Big\{ \frac{1}{NT} \sum_{i=1}^{N} \Big\| \phi_i[\beta_0] - \phi_i[\widehat{\beta}_m] \Big\|^2 \Big\}^{1/2} \Big\{ \frac{1}{NT} \sum_{i=1}^{N} \| F_0 \gamma_{0i} \|^2 \Big\}^{1/2}$$

$$= O_P(\| \widehat{\beta}_m - \beta_0 \|_{L^2}), \tag{B.9}$$

where the second inequality follows from Cauchy-Schwarz inequality; and the last line follows from the same arguments given for $I_{1NT}(\widehat{\beta}_m, \widehat{F})$ and the fact that $\frac{1}{NT} \sum_{i=1}^{N} \| F_0 \gamma_{0i} \|^2 = O_P(1)$.

Similar to (B.9), we have $\frac{1}{\sqrt{T}} \Big\| I_{jNT}(\widehat{\beta}_m, \widehat{F}) \Big\| = O_P(\| \widehat{\beta}_m - \beta_0 \|_{L^2})$ for $j = 3, 4, 5$. By (1) of Lemma A.2 and $\frac{1}{\sqrt{T}} \| \widehat{F} \| = O(1)$, we also obtain $\frac{1}{\sqrt{T}} \| I_{6NT}(\widehat{F}) \| = O_P\left(\frac{1}{\sqrt{N}} \right) + O_P\left(\frac{1}{\sqrt{T}} \right)$.

For $I_{7NT}(\widehat{F})$ and $I_{8NT}(\widehat{F})$, write

$$E \Big\| \frac{1}{NT} \sum_{i=1}^{N} F_0 \gamma_{0i} \mathcal{E}_i' \Big\|^2 = \sum_{t=1}^{T} \sum_{s=1}^{T} \frac{1}{N^2 T^2} \sum_{i=1}^{N} \sum_{j=1}^{N} E[f_{0t}' \gamma_{0i} \varepsilon_{is} f_{0t}' \gamma_{0j} \varepsilon_{js}]$$

$$= \sum_{t=1}^{T} \sum_{s=1}^{T} \frac{1}{N^2 T^2} \sum_{i=1}^{N} \sum_{j=1}^{N} E[f_{0t}' \gamma_{0i} f_{0t}' \gamma_{0j}] E[\varepsilon_{is} \varepsilon_{js}]$$

$$\leq O(1) \sum_{t=1}^{T} \sum_{s=1}^{T} \frac{1}{N^2 T^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \big\{ E \| f_{0t} \|^4 E \| \gamma_{0i} \|^4 E \| f_{0t} \|^4 E \| \gamma_{0j} \|^4 \big\}^{1/4} |\sigma_{ij}|$$

$$\leq O(1) \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} |\sigma_{ij}| = O\left(\frac{1}{N} \right),$$

where the first inequality follows from Cauchy-Schwarz inequality and Assumption 1. We then can conclude that $\frac{1}{\sqrt{T}} \| I_{7NT}(\widehat{F}) \| = \frac{1}{\sqrt{T}} \| I_{8NT}(\widehat{F}) \| = O_P\left(\frac{1}{\sqrt{N}} \right)$.

Based on the above analysis and by left multiplying (B.6) by $\widehat{F}'/T$, we obtain

$$V_{NT} - (\widehat{F}' F_0 / T)(\Gamma_0' \Gamma_0 / N)(F_0' \widehat{F} / T) = \frac{1}{T} \widehat{F}' \Big[I_{1NT}(\widehat{\beta}_m, \widehat{F}) + \cdots + I_{8NT}(\widehat{F}) \Big] = o_P(1).$$

Thus, $V_{NT} = (\widehat{F}' F_0 / T)(\Gamma_0' \Gamma_0 / N)(F_0' \widehat{F} / T) + o_P(1)$. When proving the second result of this lemma, we have shown that $F_0' \widehat{F} / T$ is non-singular with probability approaching one, which implies that V_{NT} is invertible with probability approaching one. We now left multiply (B.6) by F_0'/T to obtain

$$(F_0' \widehat{F} / T) V_{NT} = (F_0' F_0 / T)(\Gamma_0' \Gamma_0 / N)(F_0' \widehat{F} / T) + o_P(1)$$

based on the above analysis. It shows that the columns of $F_0' \widehat{F} / T$ are the (non-normalized) eigenvectors of the matrix $(F_0' F_0 / T)(\Gamma_0' \Gamma_0 / N)$, and V_{NT} consists of the eigenvalues of the same matrix (in the limit). Thus, the first result of this lemma follows.

(4). According to the above analysis, (B.7) can be summarized by

$$\frac{1}{\sqrt{T}} \| \widehat{F} \Pi_{NT}^{-1} - F_0 \| = O_P(\| \widehat{\beta}_m - \beta_0 \|_{L^2}) + O_P\left(\frac{1}{\sqrt{N}} \right) + O_P\left(\frac{1}{\sqrt{T}} \right).$$

(5). According to (B.7),

$$\frac{1}{T}F_0'(\widehat{F} - F_0\Pi_{NT}) = \frac{1}{T}F_0'\left[I_{1NT}(\widehat{\beta}_m, \widehat{F}) + \cdots + I_{8NT}(\widehat{F})\right]V_{NT}^{-1}.$$

Note that $V_{NT}^{-1} = O_P(1)$, so we focus on $\frac{1}{T}F_0'\left[I_{1NT}(\widehat{\beta}_m, \widehat{F}) + \cdots + I_{8NT}(\widehat{F})\right]$ below. By the proof given for the first result of this lemma, it is easy to show that

$$\left\|\frac{1}{T}F_0'\left[I_{1NT}(\widehat{\beta}_m, \widehat{F}) + \cdots + I_{5NT}(\widehat{\beta}_m, \widehat{F})\right]\right\| = O_P(\|\widehat{\beta}_m - \beta_0\|_{L^2}).$$

We now consider $\left\|\frac{1}{T}F_0'I_{6NT}(\widehat{F})\right\|$. Firstly, note that it is easy to show $\frac{1}{NT}\sum_{i=1}^{N}\|F_0'\mathcal{E}_i\|^2 = O_P(1)$. Secondly, for $\frac{1}{NT}\sum_{i=1}^{N}\left\|\mathcal{E}_i'\widehat{F}\right\|^2$, we have

$$\frac{1}{NT}\sum_{i=1}^{N}\left\|\mathcal{E}_i'\widehat{F}\right\|^2 \leq \frac{2}{NT}\sum_{i=1}^{N}\left\|\mathcal{E}_i'F_0\Pi_{NT}\right\|^2 + \frac{2}{NT}\sum_{i=1}^{N}\left\|\mathcal{E}_i'\left(\widehat{F} - F_0\Pi_{NT}\right)\right\|^2$$

$$= \frac{2}{NT}\sum_{i=1}^{N}\left\|\mathcal{E}_i'F_0\Pi_{NT}\right\|^2 + \frac{2}{NT}\sum_{i=1}^{N}\mathrm{tr}\left\{\mathcal{E}_i'\left(\widehat{F} - F_0\Pi_{NT}\right)\left(\widehat{F} - F_0\Pi_{NT}\right)'\mathcal{E}_i\right\}$$

$$= \frac{2}{NT}\sum_{i=1}^{N}\left\|\mathcal{E}_i'F_0\Pi_{NT}\right\|^2 + \frac{2}{NT}\mathrm{tr}\left\{\left(\widehat{F} - F_0\Pi_{NT}\right)\left(\widehat{F} - F_0\Pi_{NT}\right)'\mathcal{E}'\mathcal{E}\right\}$$

$$\leq O_P(1) + O(1)\frac{1}{N}\left\|\mathcal{E}'\mathcal{E}\right\|\frac{1}{T}\left\|\widehat{F} - F_0\Pi_{NT}\right\|^2,$$

where $\mathcal{E}$ has been defined in Lemma A.2. In connection with (1) of Lemma A.2 and (2) of this lemma, it gives that

$$\left\|\frac{1}{T}F_0'I_{6NT}(\widehat{F})\right\| \leq \frac{1}{T}\left(\frac{1}{NT}\sum_{i=1}^{N}\|F_0'\mathcal{E}_i\|^2\right)^{1/2}\left(\frac{1}{NT}\sum_{i=1}^{N}\left\|\mathcal{E}_i'\widehat{F}\right\|^2\right)^{1/2}$$

$$= O_P(1)\frac{1}{T} + O_P\left(\frac{1}{\sqrt{T}}\right)\left\{\frac{1}{NT}\left\|\mathcal{E}'\mathcal{E}\right\|\frac{1}{T}\left\|\widehat{F} - F_0\Pi_{NT}\right\|^2\right\}^{1/2}$$

$$= O_P(1)\frac{1}{T} + O_P\left(\frac{1}{\sqrt{T}}\right)O_P\left(\frac{1}{\sqrt[4]{N}} + \frac{1}{\sqrt[4]{T}}\right)O_P\left(\|\widehat{\beta}_m - \beta_0\|_{L^2} + \frac{1}{\sqrt{N}} + \frac{1}{\sqrt{T}}\right)$$

$$= O_P(1)\left\{\frac{1}{T} + \frac{\|\widehat{\beta}_m - \beta_0\|_{L^2}}{\sqrt{T}\sqrt[4]{N}} + \frac{1}{\sqrt{T}\sqrt[4]{N^3}}\right\}$$

$$\leq o_P(1)\|\widehat{\beta}_m - \beta_0\|_{L^2} + O_P(1)\frac{1}{T} + O_P(1)\frac{1}{\sqrt{T}\sqrt[4]{N^3}},$$

where the second equality follows from (1) of Lemma A.2 and the second result of this lemma.

For $\left\|\frac{1}{T}F_0'I_{7NT}(\widehat{F})\right\|$, we have

$$\left\|\frac{1}{T}F_0'I_{7NT}(\widehat{F})\right\| \leq \left\|\frac{1}{T}F_0'F_0\right\|\cdot\left\|\frac{1}{NT}\sum_{i=1}^{N}\gamma_{0i}\mathcal{E}_i'(\widehat{F} - F_0\Pi_{NT})\right\|$$

$$+ \left\|\frac{1}{T}F_0'F_0\right\|\cdot\left\|\frac{1}{NT}\sum_{i=1}^{N}\gamma_{0i}\mathcal{E}_i'F_0\Pi_{NT}\right\|$$

$$\leq \left\|\frac{1}{T}F_0'F_0\right\|\cdot\left\|\frac{1}{N\sqrt{T}}\sum_{i=1}^{N}\gamma_{0i}\mathcal{E}_i'\right\|\cdot\frac{1}{\sqrt{T}}\|\widehat{F} - F_0\Pi_{NT}\|$$

8

$$+\left\|\frac{1}{T}F_0'F_0\right\| \cdot \left\|\frac{1}{NT}\sum_{i=1}^{N}\gamma_{0i}\mathcal{E}_i'F_0\right\| \cdot \|\Pi_{NT}\|.$$

By Assumption 1.2, $\left\|\frac{1}{T}F_0'F_0\right\| = O_P(1)$. By the first two results of this lemma, we have $\|\Pi_{NT}\| = O_P(1)$ and $\frac{1}{\sqrt{T}}\|\widehat{F} - F_0\Pi_{NT}\| = O_P(\|\widehat{\beta}_m - \beta_0\|_{L^2}) + O_P\left(\frac{1}{\sqrt{N}}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$. Therefore, we focus on $\left\|\frac{1}{N\sqrt{T}}\sum_{i=1}^{N}\gamma_{0i}\mathcal{E}_i'\right\|$ and $\left\|\frac{1}{NT}\sum_{i=1}^{N}\gamma_{0i}\mathcal{E}_i'F_0\right\|$ below. Write

$$E\left\|\frac{1}{N\sqrt{T}}\sum_{i=1}^{N}\gamma_{0i}\mathcal{E}_i'\right\|^2 = \frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}E[\gamma_{0i}'\gamma_{0j}]E[\varepsilon_{it}\varepsilon_{jt}]$$

$$\leq O(1)\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}|E[\varepsilon_{it}\varepsilon_{jt}]| = O(1)\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}|\sigma_{ij}| = O\left(\frac{1}{N}\right) \tag{B.10}$$

and using Assumption 1, it is easy to show that

$$E\left\|\frac{1}{NT}\sum_{i=1}^{N}\gamma_{0i}\mathcal{E}_i'F_0\right\|^2 = O\left(\frac{1}{NT}\right), \tag{B.11}$$

which immediately yields

$$\left\|\frac{1}{T}F_0'I_{7NT}(\widehat{F})\right\| = O_P\left(\|\widehat{\beta}_m - \beta_0\|_{L^2} \cdot \frac{1}{\sqrt{N}}\right) + O_P\left(\frac{1}{N}\right) + O_P\left(\frac{1}{\sqrt{NT}}\right)$$

$$\leq O_P\left(\|\widehat{\beta}_m - \beta_0\|_{L^2}^2\right) + O_P\left(\frac{1}{N}\right) + O_P\left(\frac{1}{T}\right).$$

Similarly, $\left\|\frac{1}{T}F_0'I_{8NT}(\widehat{F})\right\| = O_P\left(\|\widehat{\beta}_m - \beta_0\|_{L^2}^2\right) + O_P\left(\frac{1}{N}\right) + O_P\left(\frac{1}{T}\right)$.

Based on the above analysis, we have

$$\left\|\frac{1}{T}F_0'(\widehat{F} - F_0\Pi_{NT})\right\| = O_P(\|\widehat{\beta}_m - \beta_0\|_{L^2}) + O_P\left(\frac{1}{N}\right) + O_P\left(\frac{1}{T}\right), \tag{B.12}$$

which further indicates

$$\left\|\frac{1}{T}\widehat{F}'(\widehat{F} - F_0\Pi_{NT})\right\| = \left\|\frac{1}{T}(\widehat{F} - F_0\Pi_{NT} + F_0\Pi_{NT})'(\widehat{F} - F_0\Pi_{NT})\right\|$$

$$\leq \left\|\frac{1}{T}(\widehat{F} - F_0\Pi_{NT})'(\widehat{F} - F_0\Pi_{NT})\right\| + \|\Pi_{NT}\| \cdot \left\|\frac{1}{T}F_0'(\widehat{F} - F_0\Pi_{NT})\right\|$$

$$= O_P(\|\widehat{\beta}_m - \beta_0\|_{L^2}) + O_P\left(\frac{1}{N}\right) + O_P\left(\frac{1}{T}\right). \tag{B.13}$$

(6). Note (B.12) and (B.13) can be respectively expressed as

$$\frac{1}{T}F_0'\widehat{F} - \frac{1}{T}F_0'F_0\Pi_{NT} = O_P(\|\widehat{\beta}_m - \beta_0\|_{L^2}) + O_P\left(\frac{1}{N}\right) + O_P\left(\frac{1}{T}\right)$$

and

$$I_r - \frac{1}{T}\widehat{F}'F_0\Pi_{NT} = O_P(\|\widehat{\beta}_m - \beta_0\|_{L^2}) + O_P\left(\frac{1}{N}\right) + O_P\left(\frac{1}{T}\right),$$

9

which further give

$$\frac{1}{T}\Pi'_{NT}F'_0\widehat{F} - \frac{1}{T}\Pi'_{NT}F'_0F_0\Pi_{NT} = O_P(\|\widehat{\beta}_m - \beta_0\|_{L^2}) + O_P\left(\frac{1}{N}\right) + O_P\left(\frac{1}{T}\right)$$

and

$$I_r - \frac{1}{T}\Pi'_{NT}F'_0\widehat{F} = O_P(\|\widehat{\beta}_m - \beta_0\|_{L^2}) + O_P\left(\frac{1}{N}\right) + O_P\left(\frac{1}{T}\right).$$

Summing up the above two equations yields

$$I_r - \frac{1}{T}\Pi'_{NT}F'_0F_0\Pi_{NT} = O_P(\|\widehat{\beta}_m - \beta_0\|_{L^2}) + O_P\left(\frac{1}{N}\right) + O_P\left(\frac{1}{T}\right). \tag{B.14}$$

Note that it is easy to show that

$$\left\|P_{\widehat{F}} - P_{F_0}\right\|^2 = \operatorname{tr}\left[(P_{\widehat{F}} - P_{F_0})^2\right] = \operatorname{tr}\left[P_{\widehat{F}} - P_{\widehat{F}}P_{F_0} - P_{F_0}P_{\widehat{F}} + P_{F_0}\right]$$
$$= \operatorname{tr}\left[I_r\right] - 2\operatorname{tr}\left[P_{\widehat{F}}P_{F_0}\right] + \operatorname{tr}\left[I_r\right] = 2\operatorname{tr}\left[I_r - \widehat{F}'P_{F_0}\widehat{F}/T\right]$$

and, when proving this lemma, we have shown that

$$\frac{F'_0\widehat{F}}{T} = \frac{F'_0F_0}{T}\Pi_{NT} + O_P(\|\widehat{\beta}_m - \beta_0\|_{L^2}) + O_P\left(\frac{1}{N}\right) + O_P\left(\frac{1}{T}\right).$$

Therefore, we can write

$$\widehat{F}'P_{F_0}\widehat{F}/T = \Pi'_{NT}\left(\frac{F'_0F_0}{T}\right)\Pi_{NT} + O_P(\|\widehat{\beta}_m - \beta_0\|_{L^2}) + O_P\left(\frac{1}{N}\right) + O_P\left(\frac{1}{T}\right).$$

In connection with (B.14), we then obtain that

$$\widehat{F}'P_{F_0}\widehat{F}/T = I_r + O_P(\|\widehat{\beta}_m - \beta_0\|_{L^2}) + O_P\left(\frac{1}{N}\right) + O_P\left(\frac{1}{T}\right). \tag{B.15}$$

Then the proof of the last result of this lemma is completed. ∎

Proof of Lemma A.4:

For simplicity, we show that $\Pr(\|\widehat{C}_{\beta,p}\| = 0) \to 1$ only. The proofs for $\|\widehat{C}_{\beta,j}\|$ with $j = p^* + 1, \ldots, p - 1$ are the same. By (B.4) and Assumption 3, we can conclude that

$$\|C_{\beta_0} - \widehat{C}_\beta\| = O_P\left(\frac{1}{\sqrt[8]{\xi_{NT}}}\right). \tag{B.16}$$

If $\|\widehat{C}_{\beta,p}\| \neq 0$, the following equation must hold:

$$0 = \frac{\partial}{\partial C_{\beta,p}}Q_\lambda(C_\beta, F)\Big|_{(C,F)=(\widehat{C}_\beta,\widehat{F})} = -2B_1 + B_2, \tag{B.17}$$

where $B_1 = \sum_{i=1}^N \mathbf{Z}'_{ip}M_{\widehat{F}}(Y_i - \phi_i[\widehat{\beta}_m])$, $\mathbf{Z}_{ip} = (x_{i1,p}H_m(z_{i1}), \ldots, x_{iT,p}H_m(z_{iT}))'$ and $B_2 = \frac{\lambda_p}{\|\widehat{C}_{\beta,p}\|}\widehat{C}'_{\beta,p}$. For B_1, write

10

$$\frac{1}{NT}B_1 = \frac{1}{NT}\sum_{i=1}^{N}\mathsf{Z}_{ip}'M_{\widehat{F}}(\phi_i[\beta_0] - \phi_i[\widehat{\beta}_m] + F_0\gamma_i + \mathcal{E}_i)$$

In view of (B.16) and the development of Lemma A.3, it is easy to know $\frac{\sqrt[8]{\xi_{NT}}}{\sqrt{m}} \cdot \frac{1}{NT}B_1 = O_P(1)$. On the other hand, $\left\| \frac{\sqrt[8]{\xi_{NT}}}{\sqrt{m}} \cdot \frac{1}{NT}B_2 \right\| \geq \frac{\sqrt[8]{\xi_{NT}}\lambda_{\min}^{\dagger}}{\sqrt{m}NT} \geq \kappa_1$ by Assumption 3. Therefore, $\Pr(\|B_1\| < \|B_2\|) \to 1$, which implies that, with a probability tending to 1, (B.17) does not hold. The above analysis implies that $\widehat{C}_{\beta,p}$ must be located at a place where the objective function $Q_\lambda(C_\beta, F)$ is not differentiable with respect to $C_{\beta,p}$. Since $Q_\lambda(C_\beta, F)$ is not differentiable with respect to $C_{\beta,p}$ only at the origin, we immediately obtain that $\Pr(\|\widehat{C}_{\beta,p}\| = 0) \to 1$. Similarly, we can show $\Pr(\|\widehat{C}_{\beta,j}\| = 0) \to 1$ with $j = p^* + 1, \ldots, p - 1$. The proof is complete. ∎

Proof of Lemma A.5:

Note that (B.16) only gives a slow rate. Below, we aim to improve this rate. Having proved $\Pr(\|\widehat{C}_{\beta}^{\dagger}\| = 0) \to 1$, we delete the corresponding rows of $\widehat{C}_\beta$ and $x_{it,j}$ for $j = p^* + 1, \ldots, p$ from the objective function. Thus, following the same arguments as in Bai (2009, p. 1236), the estimator $\widehat{C}_{\beta}^*$ given by (3.3) can be written as

$$\text{vec}(\widehat{C}_{\beta}^*) = \left(\sum_{i=1}^{N} \mathcal{Z}_i^{*\prime}M_{\widehat{F}}\mathcal{Z}_i^* + \frac{D_{m,p^*}}{2} \right)^{-1} \sum_{i=1}^{N} \mathcal{Z}_i^{*\prime}M_{\widehat{F}}Y_i,$$

where $\mathcal{Z}_i^* = (\mathcal{Z}_{i1}^*, \ldots, \mathcal{Z}_{iT}^*)'$ and $D_{m,p^*} = I_m \otimes \text{diag}\left\{ \frac{\lambda_1}{\|\widehat{C}_{\beta,1}\|}, \ldots, \frac{\lambda_{p^*}}{\|\widehat{C}_{\beta,p^*}\|} \right\}$. Correspondingly, we denote that

$$\text{vec}(\widehat{C}_{\beta}^{\sharp}) = \left(\sum_{i=1}^{N} \mathcal{Z}_i^{*\prime}M_{\widehat{F}}\mathcal{Z}_i^* \right)^{-1} \sum_{i=1}^{N} \mathcal{Z}_i^{*\prime}M_{\widehat{F}}Y_i.$$

Thus, we can write

$$\widehat{C}_{\beta}^* - C_{\beta_0}^* = (\widehat{C}_{\beta}^* - \widehat{C}_{\beta}^{\sharp}) + (\widehat{C}_{\beta}^{\sharp} - C_{\beta_0}^*). \tag{B.18}$$

Below, we investigate each term on the right hand side of (B.18).

Firstly, consider $\widehat{C}_{\beta}^* - \widehat{C}_{\beta}^{\sharp}$, and write

$$\text{vec}(\widehat{C}_{\beta}^*) - \text{vec}(\widehat{C}_{\beta}^{\sharp}) = \left\{ \left(\sum_{i=1}^{N} \mathcal{Z}_i^{*\prime}M_{\widehat{F}}\mathcal{Z}_i^* + \frac{D_{m,p^*}}{2} \right)^{-1} - \left(\sum_{i=1}^{N} \mathcal{Z}_i^{*\prime}M_{\widehat{F}}\mathcal{Z}_i^* \right)^{-1} \right\} \sum_{i=1}^{N} \mathcal{Z}_i^{*\prime}M_{\widehat{F}}Y_i.$$

By Lemma A.1 and Assumption 2, we just need to consider the next term in order to get the difference between $\widehat{C}_{\beta}^*$ and $\widehat{C}_{\beta}^{\sharp}$.

$$\left\| \frac{1}{NT}\sum_{i=1}^{N} \mathcal{Z}_i^{*\prime}M_{\widehat{F}}\mathcal{Z}_i^* + \frac{D_{m,p^*}}{2NT} - \frac{1}{NT}\sum_{i=1}^{N} \mathcal{Z}_i^{*\prime}M_{\widehat{F}}\mathcal{Z}_i^* \right\|$$

$$= \left\| \frac{D_{m,p^*}}{2NT} \right\| = O\left(\frac{\sqrt{m}\lambda_{\max}^*}{NT} \right). \tag{B.19}$$

Moreover, it is easy to know $\left\| \frac{1}{NT} \sum_{i=1}^{N} \mathcal{Z}_i^{*\prime} M_{\widehat{F}} Y_i \right\| = O_P(\sqrt{m})$, which in connection with (B.19) indicates $\|\widehat{C}_\beta^* - \widehat{C}_\beta^\sharp\| = O_P\left(\frac{m\lambda_{\max}^*}{NT} \right)$.

We now focus on $\widehat{C}_\beta^\sharp - C_{\beta_0}^*$, and write

$$
\begin{aligned}
\mathrm{vec}(\widehat{C}_\beta^\sharp) - \mathrm{vec}(C_{\beta_0}^*) &= \left[\sum_{i=1}^{N} \mathcal{Z}_i^{*\prime} M_{\widehat{F}} \mathcal{Z}_i^* \right]^{-1} \sum_{i=1}^{N} \mathcal{Z}_i^{*\prime} M_{\widehat{F}} \mathcal{E}_i \\
&\quad + \left[\sum_{i=1}^{N} \mathcal{Z}_i^{*\prime} M_{\widehat{F}} \mathcal{Z}_i^* \right]^{-1} \sum_{i=1}^{N} \mathcal{Z}_i^{*\prime} M_{\widehat{F}} F_0 \gamma_{0i} \\
&\quad + \left[\sum_{i=1}^{N} \mathcal{Z}_i^{*\prime} M_{\widehat{F}} \mathcal{Z}_i^* \right]^{-1} \sum_{i=1}^{N} \mathcal{Z}_i^{*\prime} M_{\widehat{F}} \phi_i^* [\Delta_m^*] \\
&:= \Lambda_1 + \Lambda_2 + \Lambda_3,
\end{aligned}
$$

where the definitions of Λ_1-Λ_3 should be obvious. Note

$$\frac{1}{NT} \sum_{i=1}^{N} \mathcal{Z}_i^{*\prime} M_{\widehat{F}} \mathcal{Z}_i^* = \frac{1}{NT} \sum_{i=1}^{N} \mathcal{Z}_i^{*\prime} M_{F_0} \mathcal{Z}_i^* \cdot (1 + o_P(1)) = \Sigma_{\mathcal{Z},f}^* \cdot (1 + o_P(1)),$$

where $\Sigma_{\mathcal{Z},f}^* = E[\mathcal{Z}_{11}^* \mathcal{Z}_{11}^{*\prime}] - E[\mathcal{Z}_{11}^* f_{01}']\Sigma_f^{-1} E[f_{01}\mathcal{Z}_{11}^{*\prime}]$. Similar to (A.5) of Su and Jin (2012), we obtain $\|\Lambda_3\| = O_P\left(m^{-\frac{\mu}{2}} \right)$. In the following, we focus on studying Λ_2 at first, and then turn to Λ_1.

In the rest proofs of this lemma, we always let $\Xi_{NT} = (F_0'\widehat{F}/T)^{-1}(\Gamma_0'\Gamma_0/N)^{-1}$ for simplicity, and we have shown $\|\Xi_{NT}\| = O_P(1)$ in the proof of Lemma A.3. Recall that we have denoted Π_{NT} and V_{NT} in Lemma A.3, so $\Pi_{NT}^{-1} = V_{NT}\Xi_{NT}$. Then we start our investigation on Λ_2, and write

$$
\begin{aligned}
\frac{1}{NT} \sum_{i=1}^{N} \mathcal{Z}_i^{*\prime} M_{\widehat{F}} F_0 \gamma_{0i} &= \frac{1}{NT} \sum_{i=1}^{N} \mathcal{Z}_i^{*\prime} M_{\widehat{F}} \left(\widehat{F}\Pi_{NT}^{-1} - F_0 \right) \gamma_{0i} \\
&= \frac{1}{NT} \sum_{i=1}^{N} \mathcal{Z}_i^{*\prime} M_{\widehat{F}} \left[I_{1NT}(\widehat{\beta}_m^*, \widehat{F}) + \cdots + I_{8NT}(\widehat{F}) \right] \Xi_{NT} \gamma_{0i} \\
&:= J_{1NT} + \cdots + J_{8NT},
\end{aligned}
$$

where the second equality follows from (B.7); $I_{1NT}(\beta, F)$ to $I_{8NT}(F)$ have been defined in the proof of Lemma A.3 but excluding $x_{it,j}$ for $j = p^* + 1, \ldots, p$; and the definitions of J_{1NT} to J_{8NT} should be obvious. In view of the decomposition of J_{2NT} below, it is easy to know that $\|J_{1NT}\| = o_P(\|\widehat{C}_\beta^* - C_{\beta_0}^*\|)$. Thus, we start from J_{2NT} and write

$$J_{2NT} = \frac{1}{NT} \sum_{i=1}^{N} \mathcal{Z}_i^{*\prime} M_{\widehat{F}} I_{2NT}(\widehat{\beta}_m^*, \widehat{F}) \Xi_{NT} \gamma_{0i}$$

$$
= \frac{1}{NT} \sum_{i=1}^{N} \mathcal{Z}_i^{*\prime} M_{\widehat{F}} \frac{1}{NT} \sum_{j=1}^{N} \left(\phi_j^*[\beta_{0,m}^*] - \phi_j^*[\widehat{\beta}_m^*] \right) (F_0 \gamma_{0j})' \widehat{F} \Xi_{NT} \gamma_{0i}
$$

$$
+ \frac{1}{NT} \sum_{i=1}^{N} \mathcal{Z}_i^{*\prime} M_{\widehat{F}} \frac{1}{NT} \sum_{j=1}^{N} \phi_j^*[\Delta_m^*] (F_0 \gamma_{0j})' \widehat{F} \Xi_{NT} \gamma_{0i}
$$

$$
= \frac{1}{N^2 T} \sum_{i=1}^{N} \sum_{j=1}^{N} \mathcal{Z}_i^{*\prime} M_{\widehat{F}} \mathcal{Z}_j^* \gamma_{0j}' \left(\frac{F_0' \widehat{F}}{T} \right) \left(\frac{F_0' \widehat{F}}{T} \right)^{-1} \left(\frac{\Gamma_0' \Gamma_0}{N} \right)^{-1} \gamma_{0i} \left[\mathrm{vec}(\widehat{C}_\beta^*) - \mathrm{vec}(C_{\beta_0}^*) \right]
$$

$$
+ \frac{1}{NT} \sum_{i=1}^{N} \mathcal{Z}_i^{*\prime} M_{\widehat{F}} \frac{1}{NT} \sum_{j=1}^{N} \phi_j^*[\Delta_m^*] (F_0 \gamma_{0j})' \widehat{F} \Xi_{NT} \gamma_{0i}
$$

$$
= \frac{1}{N^2 T} \sum_{i=1}^{N} \sum_{j=1}^{N} \mathcal{Z}_i^{*\prime} M_{\widehat{F}} \mathcal{Z}_j^* \gamma_{0j}' \left(\frac{\Gamma_0' \Gamma_0}{N} \right)^{-1} \gamma_{0i} \left[\mathrm{vec}(\widehat{C}_\beta^*) - \mathrm{vec}(C_{\beta_0}^*) \right]
$$

$$
+ \frac{1}{NT} \sum_{i=1}^{N} \mathcal{Z}_i^{*\prime} M_{\widehat{F}} \frac{1}{NT} \sum_{j=1}^{N} \phi_j^*[\Delta_m^*] (F_0 \gamma_{0j})' \widehat{F} \Xi_{NT} \gamma_{0i}
$$

$$
:= J_{2NT,1} + J_{2NT,2}.
$$

By a derivation similar (A.5) of Su and Jin (2012), we know $\left\| \left[\sum_{i=1}^{N} \mathcal{Z}_i^{*\prime} M_{\widehat{F}} \mathcal{Z}_i^* \right]^{-1} NT J_{2NT,2} \right\| = O_P \left(m^{-\frac{\mu}{2}} \right)$, so negligible. We will further study $J_{2NT,1}$ later.

For J_{3NT}, write

$$
J_{3NT} = \frac{1}{NT} \sum_{i=1}^{N} \mathcal{Z}_i^{*\prime} M_{\widehat{F}} I_{3NT}(\widehat{\beta}_m^*, \widehat{F}) \Xi_{NT} \gamma_{0i}
$$

$$
= \frac{1}{NT} \sum_{i=1}^{N} \mathcal{Z}_i^{*\prime} M_{\widehat{F}} \frac{1}{NT} \sum_{j=1}^{N} F_0 \gamma_{0j} \left(\phi_j^*[\beta_0^*] - \phi_j^*[\widehat{\beta}_m^*] \right)' \widehat{F} \Xi_{NT} \gamma_{0i}
$$

$$
= \frac{1}{NT} \sum_{i=1}^{N} \mathcal{Z}_i^{*\prime} M_{\widehat{F}} (\widehat{F} \Pi_{NT}^{-1} - F_0) \frac{1}{NT} \sum_{j=1}^{N} \gamma_{0j} \left(\phi_j^*[\beta_0^*] - \phi_j^*[\widehat{\beta}_m] \right)' \widehat{F} \Xi_{NT} \gamma_{0i}
$$

$$
:= \frac{1}{NT} \sum_{i=1}^{N} \mathcal{Z}_i^{*\prime} M_{\widehat{F}} J_{3NT,i},
$$

where $J_{3NT,i} = (\widehat{F} \Pi_{NT}^{-1} - F_0) \frac{1}{NT} \sum_{j=1}^{N} \gamma_{0j} (\phi_j^*[\beta_0^*] - \phi_j^*[\widehat{\beta}_m^*])' \widehat{F} \Xi_{NT} \gamma_{0i}$. Below, we are going to show that

$$
\left\| \left[\sum_{i=1}^{N} \mathcal{Z}_i^{*\prime} M_{\widehat{F}} \mathcal{Z}_i^* \right]^{-1} NT J_{3NT} \right\| = o_P(\|\widehat{C}_\beta - C_{\beta_0}\|) \tag{B.20}
$$

By the procedure similar to (A.5) of Su and Jin (2012), we just need to focus on $\frac{1}{NT} \sum_{i=1}^{N} \|J_{3NT,i}\|^2$.

$$
\frac{1}{NT} \sum_{i=1}^{N} \|J_{3NT,i}\|^2 \leq \frac{1}{NT} \sum_{i=1}^{N} \|\widehat{F} \Pi_{NT}^{-1} - F_0\|^2 \left\| \frac{1}{NT} \sum_{j=1}^{N} \gamma_{0j} (\phi_j^*[\beta_0^*] - \phi_j^*[\widehat{\beta}_m^*])' \right\|^2 \|\widehat{F} \Xi_{NT} \gamma_{0i}\|^2
$$

$$
\leq O_P(1) \frac{1}{T} \|\widehat{F} \Pi_{NT}^{-1} - F_0\|^2 \left(\frac{1}{N\sqrt{T}} \sum_{j=1}^{N} \|\phi_j^*[\beta_0^*] - \phi_j^*[\widehat{\beta}_m^*]\| \right)^2
$$

13

$$= O_P(1) \frac{1}{T} \|\widehat{F} \Pi_{NT}^{-1} - F_0\|^2 \left(\frac{1}{N} \sum_{j=1}^{N} \left\{ \frac{1}{T} \|\phi_j^*[\beta_0^*] - \phi_j^*[\widehat{\beta}_m^*]\|^2 \right\}^{1/2} \right)^2$$

$$= o_P(\|\widehat{C}_\beta - C_{\beta_0}\|^2),$$

where the second inequality follows from $\Xi_{NT} = O_P(1)$ and $\frac{1}{\sqrt{T}} \|\widehat{F}\| = O(1)$; and the last equality follows from $\frac{1}{\sqrt{T}} \|\widehat{F} \Pi_{NT}^{-1} - F_0\| = o_P(1)$. Thus, we can conclude that (B.20) holds.

For J_{4NT}, write

$$J_{4NT} = \frac{1}{NT} \sum_{i=1}^{N} \mathcal{Z}_i^{*\prime} M_{\widehat{F}} I_{4NT}(\widehat{\beta}_m^*, \widehat{F}) \Xi_{NT} \gamma_{0i}$$

$$\leq \frac{1}{N^2 T^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \mathcal{Z}_i^{*\prime} M_{\widehat{F}} \left(\phi_j^*[\beta_{0,m}^*] - \phi_j^*[\widehat{\beta}_m^*] \right) \mathcal{E}_j' F_0 \Pi_{NT} \Xi_{NT} \gamma_{0i}$$

$$+ \frac{1}{N^2 T^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \mathcal{Z}_i^{*\prime} M_{\widehat{F}} \phi_j^*[\Delta_m^*] \mathcal{E}_j' F_0 \Pi_{NT} \Xi_{NT} \gamma_{0i}$$

$$+ \frac{1}{N^2 T^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \mathcal{Z}_i^{*\prime} M_{\widehat{F}} \left(\phi_j^*[\beta_0^*] - \phi_j^*[\widehat{\beta}_m^*] \right) \left[\mathcal{E}_j'(\widehat{F} - F_0 \Pi_{NT}) \right] \Xi_{NT} \gamma_{0i}$$

$$:= J_{4NT,1} + J_{4NT,2} + J_{4NT,3}.$$

For $J_{4NT,1}$, write

$$\|J_{4NT,1}\| = \left\| \frac{1}{N^2 T^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \mathcal{Z}_i^{*\prime} M_{\widehat{F}} \left(\phi_j^*[\beta_{0,m}^*] - \phi_j^*[\widehat{\beta}_m^*] \right) \mathcal{E}_j' F_0 \Pi_{NT} \Xi_{NT} \gamma_{0i} \right\|$$

$$= \left\| \frac{1}{N^2 T^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \mathcal{Z}_i^{*\prime} M_{\widehat{F}} \mathcal{Z}_j^* \operatorname{vec}(\widehat{C}_\beta^* - C_{\beta_0}^*) \mathcal{E}_j' F_0 \Pi_{NT} \Xi_{NT} \gamma_{0i} \right\|$$

$$\leq \left\| \frac{1}{N^2 T^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \mathcal{Z}_i^{*\prime} M_{\widehat{F}} \mathcal{Z}_j^* \mathcal{E}_j' F_0 \Pi_{NT} \Xi_{NT} \gamma_{0i} \right\| \cdot \| \operatorname{vec}(\widehat{C}_\beta^* - C_{\beta_0}^*) \|$$

$$\leq O_P(1) \frac{1}{NT} \sum_{i=1}^{N} \| \mathcal{Z}_i^{*\prime} M_{\widehat{F}} \| \|\gamma_{0i}\| \cdot \frac{1}{N} \sum_{j=1}^{N} \| \mathcal{Z}_j^* \| \frac{1}{T} \|\mathcal{E}_j' F_0\| \cdot \| \operatorname{vec}(\widehat{C}_\beta^* - C_{\beta_0}^*) \|$$

$$\leq \frac{1}{T} O_P(\sqrt{mT}) \cdot O_P(\sqrt{mT}) \cdot O_P(T^{-1/2}) \cdot \| \operatorname{vec}(\widehat{C}_\beta - C_{\beta_0}) \|$$

$$= o_P(\|\widehat{C}_\beta^* - C_{\beta_0}^*\|),$$

where the last line follows from $\frac{m^2}{T} \to 0$. Thus, $\|J_{4NT,1}\|$ is negligible. Similarly, we can show both $\|J_{4NT,2}\|$ and $\|J_{4NT,3}\|$ are negligible by taking $\frac{1}{T} \|\phi_j^*[\Delta_m^*]\|^2 = O(m^{-\mu})$ and $\frac{1}{\sqrt{T}} \|\widehat{F} \Pi_{NT}^{-1} - F_0\| = o_P(1)$ into account, respectively. Analogous to the derivations of J_{3NT} and J_{4NT}, we can obtain that $\|J_{5NT}\|$ is negligible.

Below, we take a careful look at J_{6NT}. According to Assumption 1, let $\Omega_e = E[\mathcal{E}_i \mathcal{E}_i']$, which is a deterministic matrix uniformly in i. Thus, write

14

$$J_{6NT} = \frac{1}{NT} \sum_{i=1}^{N} \mathcal{Z}_i^{*\prime} M_{\widehat{F}} \frac{1}{NT} \sum_{j=1}^{N} \mathcal{E}_j \mathcal{E}_j^{\prime} \widehat{F} \Xi_{NT} \gamma_{0i}$$

$$= \frac{1}{NT^2} \sum_{i=1}^{N} \mathcal{Z}_i^{*\prime} M_{\widehat{F}} \Omega_e \widehat{F} \Xi_{NT} \gamma_{0i}$$

$$+ \frac{1}{NT} \sum_{i=1}^{N} \mathcal{Z}_i^{*\prime} M_{\widehat{F}} \frac{1}{NT} \sum_{j=1}^{N} \left(\mathcal{E}_j \mathcal{E}_j^{\prime} - \Omega_e \right) \widehat{F} \Xi_{NT} \gamma_{0i}$$

$$:= J_{6NT,1} + J_{6NT,2}.$$

We focus on $J_{6NT,2}$ at first.

$$J_{6NT,2} = \frac{1}{N^2 T^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \mathcal{Z}_i^{*\prime} \left(\mathcal{E}_j \mathcal{E}_j^{\prime} - \Omega_e \right) \widehat{F} \Xi_{NT} \gamma_{0i}$$

$$+ \frac{1}{N^2 T^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \mathcal{Z}_i^{*\prime} P_{\widehat{F}} \left(\mathcal{E}_j \mathcal{E}_j^{\prime} - \Omega_e \right) \widehat{F} \Xi_{NT} \gamma_{0i}$$

$$:= J_{6NT,21} + J_{6NT,22}.$$

Further decompose $J_{6NT,21}$ as

$$J_{6NT,21} = \frac{1}{N^2 T^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \mathcal{Z}_i^{*\prime} \left(\mathcal{E}_j \mathcal{E}_j^{\prime} - \Omega_e \right) F_0 \Pi_{NT} \Xi_{NT} \gamma_{0i}$$

$$+ \frac{1}{N^2 T^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \mathcal{Z}_i^{*\prime} \left(\mathcal{E}_j \mathcal{E}_j^{\prime} - \Omega_e \right) (\widehat{F} - F_0 \Pi_{NT}) \Xi_{NT} \gamma_{0i}$$

$$:= J_{6NT,211} + J_{6NT,212}.$$

Then by a development similar to Jiang et al. (2017, pp. 30-31), we obtain that $\|J_{6NT,21}\| = o_P \left(\sqrt{\frac{m}{NT}} \right)$. Similarly, $\|J_{6NT,22}\| = o_P \left(\sqrt{\frac{m}{NT}} \right)$. Therefore, we obtain $\|J_{6NT,2}\| = o_P \left(\sqrt{\frac{m}{NT}} \right)$.

We will consider $J_{6NT,1}$ together with $J_{2NT,1}$ and J_{8NT} later on. Then we only have one term J_{7NT} left to consider.

$$J_{7NT} = \frac{1}{NT} \sum_{i=1}^{N} \mathcal{Z}_i^{*\prime} M_{\widehat{F}} \frac{1}{NT} \sum_{j=1}^{N} F_0 \gamma_{0j} \mathcal{E}_j^{\prime} \widehat{F} \Xi_{NT} \gamma_{0i}$$

$$= \frac{1}{NT} \sum_{i=1}^{N} \mathcal{Z}_i^{*\prime} M_{\widehat{F}} (F_0 - \widehat{F} \Pi_{NT}^{-1}) \frac{1}{NT} \sum_{j=1}^{N} \gamma_{0j} \mathcal{E}_j^{\prime} \widehat{F} \Xi_{NT} \gamma_{0i}.$$

Notice that

$$\frac{1}{NT} \sum_{j=1}^{N} \gamma_{0j} \mathcal{E}_j^{\prime} \widehat{F} = \frac{1}{NT} \sum_{j=1}^{N} \gamma_{0j} \mathcal{E}_j^{\prime} F_0 + \frac{1}{NT} \sum_{j=1}^{N} \gamma_{0j} \mathcal{E}_j^{\prime} (F_0 - \widehat{F} \Pi_{NT}^{-1})$$

$$= O_P \left(\frac{1}{\sqrt{NT}} \right) + \left\| \frac{1}{N\sqrt{T}} \sum_{j=1}^{N} \gamma_{0j} \mathcal{E}_j^{\prime} \right\| \frac{1}{\sqrt{T}} \| F_0 - \widehat{F} \Pi_{NT}^{-1} \|$$

$$= O_P\left(\frac{1}{\sqrt{NT}}\right) + O_P\left(\frac{1}{\sqrt{N}}\right)\frac{1}{\sqrt{T}}\|F_0 - \widehat{F}\Pi_{NT}^{-1}\|,$$

where the second equality follows from (B.11); and the third equality follows from (B.10). Following the arguments given for $J6$ of Bai (2009, pp. 1271-1272), it is easy to show that $\|J_{7NT}\| = o_P\left(\sqrt{\frac{m}{NT}}\right) + o_P(\|\widehat{C}_\beta - C_{\beta_0}\|)$.

Based on the above analyses and Assumption 4, we have

$$\text{vec}(\widehat{C}_\beta^\sharp) - \text{vec}(C_{\beta_0}^*) - \Sigma_{\mathcal{Z},f}^{*-1}J_{2NT,1}\cdot(1+o_P(1))$$

$$= \Sigma_{\mathcal{Z},f}^{*-1}\left\{\frac{1}{NT}\sum_{i=1}^N \mathcal{Z}_i^{*\prime}M_{\widehat{F}}\mathcal{E}_i + J_{6NT,1} + J_{8NT}\right\}\cdot(1+o_P(1))$$

$$= \Sigma_{\mathcal{Z},f}^{*-1}\cdot\frac{1}{NT}\sum_{i=1}^N\left\{\mathcal{Z}_i^{*\prime}M_{\widehat{F}} + \frac{1}{N}\sum_{j=1}^N \mathcal{Z}_j^{*\prime}M_{\widehat{F}}\gamma_{0j}'(\Gamma_0'\Gamma_0/N)^{-1}\gamma_{0i}\right\}\mathcal{E}_i\cdot(1+o_P(1))$$

$$+\Sigma_{\mathcal{Z},f}^{*-1}\cdot J_{6NT,1}\cdot(1+o_P(1)).$$

Further organise the above equation, we have

$$\text{vec}(\widehat{C}_\beta^\sharp) - \text{vec}(C_{\beta_0}^*) = A_{1NT}^{-1}\Sigma_{\mathcal{Z},f}^{*-1}\cdot\frac{1}{NT}\sum_{i=1}^N\left\{\mathcal{Z}_i^{*\prime}M_{\widehat{F}} + A_{3,i}\right\}\mathcal{E}_i\cdot(1+o_P(1))$$

$$+A_{1NT}^{-1}\Sigma_{\mathcal{Z},f}^{*-1}\cdot J_{6NT,1}\cdot(1+o_P(1)),$$

where

$$A_{1NT} = I_{mp^*} - \Sigma_{\mathcal{Z},f}^{*-1}A_{2NT}\cdot(1+o_P(1)),$$

$$A_{2NT} = \frac{1}{N^2T}\sum_{i=1}^N\sum_{j=1}^N \mathcal{Z}_i^{*\prime}M_{\widehat{F}}\mathcal{Z}_j^*\gamma_{0j}'\left(\frac{\Gamma_0'\Gamma_0}{N}\right)^{-1}\gamma_{0i},$$

$$A_{3,i} = \frac{1}{N}\sum_{j=1}^N \mathcal{Z}_j^{*\prime}M_{\widehat{F}}\gamma_{0j}'(\Gamma_0'\Gamma_0/N)^{-1}\gamma_{0i}. \tag{B.21}$$

Note that

$$\sqrt{\frac{NT}{m}}J_{6NT,1} = \frac{1}{(mN)^{\frac{1}{2}}T^{\frac{3}{2}}}\sum_{i=1}^N \mathcal{Z}_i^{*\prime}M_{\widehat{F}}\Omega_e\widehat{F}\Xi_{NT}\gamma_{0i}$$

$$= \frac{\sqrt{N}}{\sqrt{mT}}\cdot\frac{1}{NT}\sum_{i=1}^N \mathcal{Z}_i^{*\prime}M_{\widehat{F}}\Omega_e\widehat{F}\Xi_{NT}\gamma_0(v_i) = O_P\left(\sqrt{\frac{N}{T}}\right) = O_P(1). \tag{B.22}$$

where the last equality follows from Assumption 3. Thus, we obtain $\|J_{6NT,1}\| = O_P\left(\sqrt{\frac{m}{NT}}\right)$. Moreover, it is easy to show $\frac{1}{NT}\sum_{i=1}^N\left\{\mathcal{Z}_i^{*\prime}M_{\widehat{F}} + A_{3,i}\right\}\mathcal{E}_i = O_P\left(\sqrt{\frac{m}{NT}}\right)$. Based on the above development, the proof is complete. ∎

Note that under the HD setting, the elements of $\beta_m(z)$ belonging to $L^2(V_z)$ indicates that $\|C_\beta\| \le a_0\sqrt{p}$ with a_0 being a large constant. We will be repeatedly using this fact below.

Proof of Lemma A.6:

(1). Write

$$\frac{1}{NT} \sum_{i=1}^{N} (\phi_i[\beta_{0,m}] - \phi_i[\beta_m])' M_F \mathcal{E}_i$$

$$= \frac{1}{NT} \sum_{i=1}^{N} (\phi_i[\beta_{0,m}] - \phi_i[\beta_m])' \mathcal{E}_i + \frac{1}{NT} \sum_{i=1}^{N} (\phi_i[\beta_{0,m}] - \phi_i[\beta_m])' P_F \mathcal{E}_i$$

$$:= \Lambda_1 + \Lambda_2.$$

For Λ_1, write

$$\sup_{\|C_\beta\| \leq a_0\sqrt{p}} \left| \frac{1}{NT} \sum_{i=1}^{N} (\phi_i[\beta_{0,m}] - \phi_i[\beta_m])' \mathcal{E}_i \right| \leq \left\| \frac{1}{NT} \sum_{i=1}^{N} \mathcal{Z}_i' \mathcal{E}_i \right\| \cdot \sup_{\|C_\beta\| \leq a_0\sqrt{p}} \|\mathrm{vec}(C_{\beta_0} - C_\beta)\|$$

$$= O\left(\sqrt{\frac{mp}{NT}} \right) \cdot \sup_{\|C_\beta\| \leq a_0\sqrt{p}} \|C_{\beta_0} - C_\beta\| = O_P\left(\sqrt{\frac{mp^2}{NT}} \right)$$

where the first equality follows from some standard analysis on the term $\frac{1}{NT} \sum_{i=1}^{N} \mathcal{Z}_i' \mathcal{E}_i$ using Assumption 1.1; and the last equality follows from $\|C_\beta\| \leq a_0\sqrt{p}$.

In order to consider Λ_2, let $\Delta b = (\phi_1[\beta_{0,m}] - \phi_1[\beta_m], \ldots, \phi_N[\beta_{0,m}] - \phi_N[\beta_m])$, and note that

$$\sup_{\|C_\beta\| \leq a_0\sqrt{p}} \frac{1}{NT} \|\Delta b\|^2 = \sup_{\|C_\beta\| \leq a_0\sqrt{p}} \frac{1}{N} \sum_{i=1}^{N} (C_\beta - C_{\beta_0})' \mathcal{Z}_i' \mathcal{Z}_i (C_\beta - C_{\beta_0})$$

$$\leq O_P(1) \sup_{\|C_\beta\| \leq a_0\sqrt{p}} \|C_\beta - C_{\beta_0}\|^2 = O_P(p), \tag{B.23}$$

where the inequality follows from Assumption 2.1.

Then we are able to write

$$\sup_{\|C_\beta\| \leq a_0\sqrt{p}, \, F \in \mathsf{D}_F} \left| \frac{1}{NT} \sum_{i=1}^{N} (\phi_i[\beta_{0,m}] - \phi_i[\beta_m])' P_F \mathcal{E}_i \right|$$

$$= \sup_{\|C_\beta\| \leq a_0\sqrt{p}, \, F \in \mathsf{D}_F} \left| \frac{1}{NT} \mathrm{tr}\left(P_F \mathcal{E}' \Delta b' \right) \right| \leq \frac{r}{NT} \sup_{\|C_\beta\| \leq a_0\sqrt{p}, \, F \in \mathsf{D}_F} \|P_F \mathcal{E}' \Delta b'\|_{\mathrm{sp}}$$

$$\leq \sup_{\|C_\beta\| \leq a_0\sqrt{p}, \, F \in \mathsf{D}_F} \frac{r}{NT} \|P_F\|_{\mathrm{sp}} \|\mathcal{E}\|_{\mathrm{sp}} \|\Delta b\|_{\mathrm{sp}} = O_P\left(\sqrt{\frac{p\, \xi_{NT}}{NT}} \right),$$

where the last equality follows from Assumption 5.1 and (B.23).

Based on the above development, the result follows.

(2). Write

$$\sup_{F \in \mathsf{D}_F} \left| \frac{1}{NT} \sum_{i=1}^{N} \phi_i[\Delta_m]' M_F \phi_i[\Delta_m] \right| \leq \left| \frac{1}{NT} \sum_{i=1}^{N} \phi_i[\Delta_m]' \phi_i[\Delta_m] \right|$$

$$= \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \|x_{it}^*\|^2 \|\Delta_m^*(z_{it})\|^2 = O_P\left(p^* m^{-\mu} \right), \tag{B.24}$$

17

where the last equality follows from Assumption 2.1 and $E\|x_{it}^*\|^2 = O(p^*)$.

(3). Write

$$
\sup_{F \in \mathsf{D}_F} \left| \frac{1}{NT} \sum_{i=1}^{N} \phi_i\left[\Delta_m\right]' M_F F_0 \gamma_{0i} \right| = \sup_{F \in \mathsf{D}_F} \left| \frac{1}{NT} \mathrm{tr}\left(M_F F_0 \Gamma_0' \Delta' \right) \right|
$$

$$
\leq \frac{r}{NT} \sup_{F \in \mathsf{D}_F} \|M_F\|_{\mathrm{sp}} \|F_0\|_{\mathrm{sp}} \|\Gamma_0\|_{\mathrm{sp}} \|\Delta\|_{\mathrm{sp}} = O_P(\sqrt{p^*} m^{-\frac{\mu}{2}}),
$$

where $\Delta = (\phi_1(\Delta_m), \ldots, \phi_N(\Delta_m))$; and the second equality follows from that $\frac{1}{NT}\|\Delta\|^2 = O_P\left(p^* m^{-\mu}\right)$ as in (2) of this lemma.

(4). Similar to the proof for (3) of this lemma, the result follows.

(5). Write

$$
\sup_{\|C_\beta\| \leq a_0 \sqrt{p},\, F \in \mathsf{D}_F} \left| \frac{1}{NT} \sum_{i=1}^{N} \phi_i\left[\Delta_m\right]' M_F \left\{ \phi_i\left[\beta_m\right] - \phi_i\left[\beta_{0,m}\right] \right\} \right|
$$

$$
\leq \left\{ \frac{1}{NT} \sum_{i=1}^{N} \|\phi_i\left[\Delta_m\right]\|^2 \right\}^{1/2} \cdot \sup_{\|C_\beta\| \leq a_0 \sqrt{p}} \left\{ \frac{1}{NT} \sum_{i=1}^{N} \|\phi_i\left[\beta_m\right] - \phi_i\left[\beta_{0,m}\right]\|^2 \right\}^{1/2}
$$

$$
= O_P(\sqrt{p^*} m^{-\frac{\mu}{2}}) \cdot O_P(\sqrt{p}),
$$

where the equality follows from (B.23) and (B.24). Then the proof is complete. ∎

Proof of Lemma A.7:

(1). Still, let $\Delta\phi_i[\beta_m] = \phi_i[\beta_{0,m}] - \phi_i[\beta_m]$, $\xi_F = \mathrm{vec}\left(M_F F_0\right)$, $A_{1F} = \frac{1}{NT} \sum_{i=1}^{N} \mathcal{Z}_i' M_F \mathcal{Z}_i$, $A_2 = \frac{1}{NT}\left(\Gamma_0' \Gamma_0\right) \otimes I_T$, and $A_{3F} = \frac{1}{NT} \sum_{i=1}^{N} \gamma_{0i} \otimes \left(M_F \mathcal{Z}_i\right)$. By the definition of (3.3) and Lemma A.6, we have

$$
0 \geq \frac{1}{NT} Q_\lambda(\widehat{C}_\beta, \widehat{F}) - \frac{1}{NT} Q_\lambda(C_{\beta_0}, F_0)
$$

$$
= \frac{1}{NT} \sum_{i=1}^{N} \left(\Delta\phi_i[\widehat{\beta}_m] + F_0 \gamma_{0i} \right)' M_{\widehat{F}} \left(\Delta\phi_i[\widehat{\beta}_m] + F_0 \gamma_{0i} \right) + \frac{1}{NT} \sum_{i=1}^{N} \mathcal{E}_i' M_{\widehat{F}} \mathcal{E}_i
$$

$$
+ \frac{2}{NT} \sum_{i=1}^{N} \left(\Delta\phi_i[\widehat{\beta}_m] + F_0 \gamma(v_i) \right)' M_{\widehat{F}} \mathcal{E}_i + \sum_{j=1}^{p} \frac{\lambda_j}{NT} \|\widehat{C}_{\beta,j}\|
$$

$$
- \frac{1}{NT} \sum_{i=1}^{N} \left(\phi_i[\Delta_m] + \mathcal{E}_i \right)' M_{F_0} \left(\phi_i[\Delta_m] + \mathcal{E}_i \right) - \sum_{j=1}^{p^*} \frac{\lambda_j}{NT} \|C_{\beta_0,j}\|
$$

$$
= \frac{1}{NT} \sum_{i=1}^{N} \left(\Delta\phi_i[\widehat{\beta}_m] + F_0 \gamma_{0i} \right)' M_{\widehat{F}} \left(\Delta\phi_i[\widehat{\beta}_m] + F_0 \gamma_{0i} \right)
$$

$$
+ \sum_{j=1}^{p} \frac{\lambda_j}{NT} \|\widehat{C}_{\beta,j}\| - \sum_{j=1}^{p^*} \frac{\lambda_j}{NT} \|C_{\beta_0,j}\| + O_P\left(\frac{1}{\sqrt[4]{\xi_{NT}}} + \sqrt{\frac{p(\xi_{NT} + mp)}{NT}} + \sqrt{p\, p^*} m^{-\frac{\mu}{2}} \right)
$$

$$
= \mathrm{vec}(C_{\beta_0} - \widehat{C}_\beta)' \frac{1}{NT} \sum_{i=1}^{N} \mathcal{Z}_i' M_{\widehat{F}} \mathcal{Z}_i \,\mathrm{vec}(C_{\beta_0} - \widehat{C}_\beta) + \frac{1}{NT} \mathrm{tr}\left(M_{\widehat{F}} F_0 \Gamma_0' \Gamma_0 F_0' M_{\widehat{F}} \right)
$$

18

$$+2\operatorname{vec}(C_{\beta_0} - \widehat{C}_\beta)'\frac{1}{NT}\sum_{i=1}^{N}\mathcal{Z}_i'M_{\widehat{F}}F_0\gamma_{0i} + \sum_{j=1}^{p}\frac{\lambda_j}{NT}\|\widehat{C}_{\beta,j}\| - \sum_{j=1}^{p^*}\frac{\lambda_j}{NT}\|C_{\beta_0,j}\|$$

$$+O_P\left(\frac{1}{\sqrt[4]{\xi_{NT}}} + \sqrt{\frac{p(\xi_{NT}+mp)}{NT}} + \sqrt{p\,p^*}m^{-\frac{\mu}{2}}\right),$$

where the second equality follows from (1) of Lemma A.2, and Lemma A.6. Thus, we can further write

$$\sum_{j=1}^{p^*}\frac{\lambda_j}{NT}\|C_{\beta_0,j}\| \geq \operatorname{vec}(C_{\beta_0} - \widehat{C}_\beta)'\frac{1}{NT}\sum_{i=1}^{N}\mathcal{Z}_i'M_{\widehat{F}}\mathcal{Z}_i\operatorname{vec}(C_{\beta_0} - \widehat{C}_\beta) + \frac{1}{NT}\operatorname{tr}\left(M_{\widehat{F}}F_0\Gamma_0'\Gamma_0F_0'M_{\widehat{F}}\right)$$

$$+2\operatorname{vec}(C_{\beta_0} - \widehat{C}_\beta)'\frac{1}{NT}\sum_{i=1}^{N}\mathcal{Z}_i'M_{\widehat{F}}F_0\gamma_{0i}$$

$$+O_P\left(\frac{1}{\sqrt[4]{\xi_{NT}}} + \sqrt{\frac{p(\xi_{NT}+mp)}{NT}} + \sqrt{p\,p^*}m^{-\frac{\mu}{2}}\right)$$

$$\geq \operatorname{vec}(C_{\beta_0} - \widehat{C}_\beta)'\left(A_{1\widehat{F}} - A_{3\widehat{F}}'A_2^{-1}A_{3\widehat{F}}\right)\operatorname{vec}(C_{\beta_0} - \widehat{C}_\beta)$$

$$+[\xi_{\widehat{F}}' + \operatorname{vec}(C_{\beta_0} - \widehat{C}_\beta)'A_{3\widehat{F}}'A_2^{-1}]A_2[\xi_{\widehat{F}} + A_2^{-1}A_{3\widehat{F}}\operatorname{vec}(C_{\beta_0} - \widehat{C}_\beta)]$$

$$+O_P\left(\frac{1}{\sqrt[4]{\xi_{NT}}} + \sqrt{\frac{p(\xi_{NT}+mp)}{NT}} + \sqrt{p\,p^*}m^{-\frac{\mu}{2}}\right)$$

$$\geq O_P(1)\|C_{\beta_0} - \widehat{C}_\beta\|^2 + O_P\left(\frac{1}{\sqrt[4]{\xi_{NT}}} + \sqrt{\frac{p(\xi_{NT}+mp)}{NT}} + \sqrt{p\,p^*}m^{-\frac{\mu}{2}}\right). \quad \text{(B.25)}$$

Till now, we can conclude that

$$\|C_{\beta_0} - \widehat{C}_\beta\|^2 = O_P\left(\frac{1}{\sqrt[4]{\xi_{NT}}} + \sqrt{\frac{p(\xi_{NT}+mp)}{NT}} + \sqrt{p\,p^*}m^{-\frac{\mu}{2}} + \frac{p^*\lambda_{\max}^*}{NT}\right) = o_P(1), \quad \text{(B.26)}$$

where the second equality follows from Assumption 5.2.

(2). By (B.25) and (B.26), we can further write that

$$o_p(1) \geq \frac{1}{NT}\operatorname{tr}\left[\left(F_0'M_{\widehat{F}}F_0\right)\left(\Gamma_0'\Gamma_0\right)\right] + o_P(1),$$

so $\frac{1}{NT}\operatorname{tr}\left[\left(F_0'M_{\widehat{F}}F_0\right)\left(\Gamma_0'\Gamma_0\right)\right] = o_P(1)$. As in Bai (2009, p. 1265), we can further conclude that $\frac{1}{T}\operatorname{tr}\left(F_0'M_{\widehat{F}}F_0\right) = o_P(1)$, $\left\|P_{\widehat{F}} - P_{F_0}\right\| = o_P(1)$, and $\frac{1}{T}\widehat{F}'F_0$ is invertible with probability approaching one. Thus, the second result of this theorem follows. ∎

References

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

Bernstein, D. S. (2005), *Matrix Mathematics: Theory, Facts, and Formulas*, Princeton University Press.

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

Jiang, B., Yang, Y., Gao, J. and Hsiao, C. (2017), Recursive estimation in large panel data models: Theory and practice. Working paper available at https://ssrn.com/abstract=2915749.

Su, L. and Jin, S. (2012), 'Sieve estimation of panel data models with cross section dependence', *Journal of Econometrics* **169**(1), 34–47.