# Semiparametric Conditional Factor Models: Estimation and Inference*

Qihui Chen† CUHK-Shenzhen

Nikolai Roussanov‡ The Wharton School

Xiaoliang Wang§ UPenn

February 10, 2022

**Abstract**

This paper introduces a simple and tractable sieve estimation of semiparametric conditional factor models with latent factors. We establish large-$N$-asymptotic properties of the estimators and the tests without requiring large $T$. We also develop a simple bootstrap procedure for conducting inference about the conditional pricing errors as well as the shapes of the factor loadings functions. These results enable us to estimate conditional factor structure of a large set of individual assets by utilizing arbitrary nonlinear functions of a number of characteristics without the need to pre-specify the factors, while allowing us to disentangle the characteristics' role in capturing factor betas from alphas (i.e., undiversifiable risk from mispricing). We apply these methods to the cross-section of individual U.S. stock returns and find strong evidence of large nonzero pricing errors that combine to produce arbitrage portfolios with Sharpe ratios above 3.

KEYWORDS: Characteristics, managed portfolios, factor models, PCA, Sieve estimation, conditional moments, nonparametric estimation, strong approximation

---

*We thank Frank Schorfheide and Liangjun Su for helpful discussions and comments.

†School of Management and Economics and Shenzhen Finance Institute, The Chinese University of Hong Kong, Shenzhen (CUHK-Shenzhen); qihuichen@cuhk.edu.cn

‡The Wharton School, University of Pennsylvania and NBER; nroussan@wharton.upenn.edu

§Department of Economics, University of Pennsylvania; xlwang@sas.upenn.edu

# 1 Introduction

We consider the following semiparametric factor model

$$y_{it} = \alpha(z_{it}) + \beta(z_{it})' f_t + \varepsilon_{it}, i = 1, \ldots, N, t = 1, \ldots, T, \tag{1}$$

where $f_t$ is a $K \times 1$ vector of unobserved factors, $\beta(\cdot)$ is a $K \times 1$ vector of unknown factor loading functions, $\alpha(\cdot)$ is an unknown intercept function, $\varepsilon_{it}$ is the idiosyncratic component that cannot be explained by the common component, and $y_{it}$ and $z_{it}$—an $M \times 1$ vector of covariates—are observed. Our main focus is on cross-sectional asset pricing, where $y_{it}$ are asset return realizations while $z_{it}$ are pre-specified asset characteristics (i.e. they are known at the beginning of period $t$).[1] In this case (1) describes a *conditional* factor model, in the sense that it captures time-variation in asset return exposures to the common factors (i.e., $\beta(z_{it})$) as well as the pricing errors (i.e., $\alpha(z_{it})$), which are both functions of characteristics (i.e., $z_{it}$). As emphasized by (Cochrane, 2011), this model is central to empirical asset pricing, since it potentially allows for distinguishing between "risk" and "mispricing" explanations of the role of characteristics in predicting asset returns.[2] Pooling the information in a multitude of stock characteristics and summarizing the common variation using a small number of factors would amount to "taming the zoo" of factors that proliferate in empirical asset pricing. The challenge to doing so is threefold: first, the identities of the common factors $f_t$ are unknown since the factors are latent; second, the functional forms of the alpha and beta functions are also generally unknown; finally, the cross-sectional dimension $N$ is typically much larger than the time-series length $T$, which renders standard tools of factor analysis inapplicable, especially when conditional covariances are time-varying.

We introduce a simple and tractable estimation method to recover both the latent factors and the functional parameters of the model, as well as develop formal inference procedures. First, we develop an easy-to-compute estimator for $\alpha(\cdot)$, $\beta(\cdot)$ and $f_t$ based on a sieve approximation to the nonparametric functions $\alpha(\cdot)$ and $\beta(\cdot)$. The estimators can be easily obtained by first running the regression of $y_{it}$ on sieves of $z_{it}$ for each $t$ and then applying principle component analysis (PCA) to the estimated coefficient matrix. Throughout the paper, we refer to the two-step procedure as *the regressed-PCA*. The first step of our procedure is a cross-sectional regression (Fama and MacBeth, 1973). Thus, in asset pricing settings the regressed-PCA boils down to applying PCA to a relatively small set of characteristic-managed portfolios constructed via the Fama-MacBeth regressions. Second, we establish large sample properties of the estimators including

[1] Other potential applications include modelling the implied volatility of options (Park et al., 2009) and describing consumer demand system (Lewbel, 1991), among others.

[2] While useful, it might not be sufficient to resolve the debate, since distinguishing between the different explanations requires understanding the economic nature of the latent factors - e.g., see (Kozak et al., 2018) .

consistency, rate of convergence, and asymptotic normality under mild conditions. In particular, we establish a strong approximation for the distribution of the estimator of the large dimensional coefficient matrix in the sieve approximation of $\alpha(\cdot)$ and $\beta(\cdot)$. These asymptotic results have several attractive properties: (i) they do not require large $T$; (ii) they allow $z_{it}$ to vary over time in a potentially non-stationary manner; (iii) they are applicable to unbalanced panels (which is useful since individual securities have varying life spans). Third, we provide two consistent estimators for the number of factors $K$, which are also easy to compute. This enables us to conduct regressed-PCA without specifying the number of factors *a priori*.

In asset pricing, testing the restriction that $\alpha(\cdot)$ is equal to zero for a given set of factors $f_t$ is central for evaluating and comparing factor models. We show that linear specifications of $\alpha(\cdot)$ and $\beta(\cdot)$ that are widely used in existing literature may adversely influence estimation of $f_t$ when the true underlying functional relationships are nonlinear. Therefore, along with the flexible nonparametric estimators we provide a specification test for the shape of $\alpha(\cdot)$ and $\beta(\cdot)$ functions. We develop a simple bootstrap inference procedure for testing significance of pricing error $\alpha(\cdot)$ as well as for linearity of $\alpha(\cdot)$ and $\beta(\cdot)$. First, we propose a weighted bootstrap procedure to approximate the distribution of the estimator of the large-dimensional coefficient matrix in the sieve approximation of $\alpha(\cdot)$ and $\beta(\cdot)$ as well as construct a Wald-type test for examining the significance of $\alpha(\cdot)$. The main challenge to developing a valid bootstrap is that the asymptotic distribution usually involves a rotational transformation matrix, which could be different under the bootstrap distribution, invalidating the procedure. In order to solve this problem we enforce the same factor estimator in the bootstrap samples as in the actual data. Second, we develop an LR-type test for examining the linearity of $\alpha(\cdot)$ and $\beta(\cdot)$. Specifically, we construct the test statistic by comparing estimators under the null hypothesis and the alternative hypothesis. The novelty of our construction is that we use the unrestricted factor estimator from the alternative to obtain the estimators of $\alpha(\cdot)$ and $\beta(\cdot)$ under the null. This ensures the same rotational transformation matrix under the null and the alternative and thus the consistency of our test. Both of these tests also enjoy the aforementioned attractive features of our estimators: our Monte Carlo simulations show that the finite sample performance of our estimators and tests is satisfactory and encouraging for large $N$, even when $T$ is small.

We apply our new methodology to analyze the cross section of individual stock returns in the US market. We use the same data set as in Kelly et al. (2019), which is the closest study to ours in terms of its empirical aims, although both our econometric approach and empirical findings are quite different. First, in contrast to Kelly et al. (2019, 2020) we allow for $\alpha(\cdot)$ and $\beta(\cdot)$ functions to be non-linear. In fact, we are able to test – and reject – the validity of the linear specifications. Second, we are able to conduct rolling small sub-sample analyses to accommodate changing factor dynamics as our methods do not require large sample length $T$. Third, we are able to consistently estimate the number of latent factors. Our empirical findings reveal that only one latent factor is detected by the formal tests when we consider linear dependence of alpha and beta functions on characteristics, and two factors when we allow for nonlinearity - this is also in contrast to Kelly et al. (2019), who advocate a five-factor model. Still, our tests reject the risk-based model, since the pricing errors associated with many characteristics are statistically different from zero. Their economic magnitudes are also substantial, as we are able to construct pure-alpha arbitrage portfolios with annualized Sharpe ratios typically above 3. These Sharpe ratios tend to rise with the number of factors (we consider up to ten), indicating that adding factors does not improve the asset pricing properties of the model, even though it might help capture more time-series variation in returns. This result provides strong empirical evidence that the characteristics contain information about both risk exposures and mispricing. In addition, the nonlinear models often produce more reasonable estimated relationships between the risk exposures and characteristics than the linear model. For instance, the estimates from our nonlinear models show that firms with higher book-to-market ratios bear more systematic risk and hence have higher expected returns, whereas the estimates from the linear model often give the opposite result. The importance of nonlinearity is highlighted by several empirical studies (Connor et al., 2012; Kirby, 2020), and has also been addressed by machine learning methods in recent studies (Gu et al., 2021; Chen et al., 2020). By emphasizing the importance of nonlinearity our method also relates to the method of characteristic-sorted portfolios.

Our paper relates to several strands of literature. Several studies estimate models similar to (1) under the assumption that $z_{it}$ are time-invariant, at least over subsamples. Connor and Linton (2007) develop a two-step kernel estimation procedure for the case with $\alpha(\cdot) = 0$. Connor et al. (2012) propose an alternating least squares procedure based on kernel smoothing for a given consistent initial estimator. Fan et al. (2016a) consider a sieve estimation which facilitates global inference, and propose a projected-PCA approach for the case with $\alpha(\cdot) = 0$. Kim et al. (2020) extend the projected-PCA to allow for nonzero $\alpha(\cdot)$, and use it to construct an arbitrage portfolio. We contribute to this literature by introducing a robust sieve estimation to allow for $z_{it}$ to vary over time and developing global inference for $\alpha(\cdot)$ and $\beta(\cdot)$. Despite some similarities, our regressed-PCA is genetically different from the projected-PCA. The regression in the former serves to extract $z_{it}$ from the common component for a consistent estimation, whereas the projection in the latter serves to remove the noise part of the factor loadings for a more efficient estimation. Therefore, the projected-PCA may fail to obtain consistent estimators when $z_{it}$ is time-varying. In contrast, our regressed-PCA is consistent even when $z_{it}$ is nonstationary.

Our study also contributes to the literature on time-varying factor models. Motta et al. (2011) and Su and Wang (2017) consider the time-varying factor model with factor loadings being smooth functions of $t/T$ and propose local versions of PCA based on kernel smoothing.[3] Pelger and Xiong (2021) assume that factor loadings are smooth functions of state variables and study a similar estimation procedure. Gagliardini and Ma (2019) study a time-varying factor model with no arbitrage and extract local factors from conditional variance matrices. However, none of them are directly suitable for testing asset pricing models, since they all impose $\alpha(\cdot) = 0$. Many empirical findings suggest that characteristics contain information about both pricing errors and risk exposures, which can be distinguished in our approach. There are numerous studies of conditional models with observed factors. For example, Gagliardini et al. (2016) specify factor loadings as linear functions of both time-varying characteristics and state variables in a model imposing $\alpha(\cdot) = 0$. We may refer to Gagliardini et al. (2020) for a comprehensive review.

The literature on the cross section of asset returns is vast; here we focus on multi-factor models motivated by the arbitrage pricing theory of Ross (1976) and its generalizations (Chamberlain and Rothschild, 1982; Connor and Korajczyk, 1986, 1988; Reisman, 1992). Empirical analysis that exploits the ability of stock characteristics to predict asset returns typically follows either the portfolio-sorting approach (Fama and French, 1993; Daniel and Titman, 1997; Fama and French, 2015) or the characteristic-based approach (Rosenberg and McKibben, 1973; Jacobs and Levy, 1988; Lewellen, 2015; Green et al., 2017; Freyberger et al., 2020; Kirby, 2020; Giglio and Xiu, 2019). The central issue with both of these approaches is that they are unable to distinguish between the two roles played by characteristics: capturing time-varying risk exposures and representing mispricing. We complement the literature by introducing a semiparametric time-varying characteristic-based factor model that provides a simple, tractable and robust method for estimation and inference. Briefly, the new methodology enables us to estimate conditional (dynamic) behavior of a large set of individual assets from a number of characteristics exhibiting nonlinearity without the need to pre-specify factors, while allowing us to disentangle the risk and mispricing explanations, as least from the standpoint of arbitrage-based models.

The remainder of the paper is organized as follows. Section 2 presents three relevant application examples. Section 3 introduces the estimation method—the regressed-PCA. Section 4 establishes large sample properties of the estimators, including consistency, rate of convergence, and asymptotic distribution. Section 5 introduces a weighted bootstrap and develops two tests. Section 6 provides two consistent estimators of the number of factors. Section 7 presents simulation studies. Section 8 applies our new methodology to analyze the cross section of individual stock returns in the US market. Section 9

[3]There is a large literature on conditional models that considers time-varying factor loadings that are functions of aggregate variables rather than firm-specific characteristics, e.g. Ferson and Harvey (1999) use a linear formulation while Roussanov (2014) considers nonparametric kernel-based specifications.

briefly concludes. Proofs are collected in the appendices.

## 2 Application Examples

**Example 2.1** (Asset return)**.** Connor and Linton (2007), Connor et al. (2012), Kelly et al. (2019), and Kim et al. (2020) use the model (1) to study the cross section of asset returns. Here, $y_{it}$ is the excess return of asset $i$ (e.g., stock $i$) in time period $t$, and $z_{it}$ is the vector of asset characteristics (e.g., book-to-market ratio and market capitalization). The model assumes that the pricing error (i.e., $\alpha(z_{it})$) and the risk exposures to factors (i.e., $\beta(z_{it})$) are associated with characteristics (i.e., $z_{it}$), unifying the characteristic-based model (Rosenberg and McKibben, 1973; Daniel and Titman, 1997) and the risk-based model (Fama and French, 1993, 2015). This modeling not only provides a way to disentangle alpha and (multi-factor) betas, but also allows us to estimate a model for a large set of individual stocks. Another advantage of the model is that one does not need to rely on ex ante knowledge of the factors, allowing them to be latent. The first paper above develops a two-step kernel estimation procedure when $z_{it}$ is not time-varying and $\alpha(\cdot) = 0$. The second paper proposes an alternating least squares procedure based on kernel smoothing for a given consistent initial estimator. The third paper allows for time-varying $z_{it}$ and considers the least squares estimation, assuming that $\alpha(\cdot)$ and $\beta(\cdot)$ are linear. The authors use the model to describe the realized return variation (i.e., systematic risks) and the cross-sectional differences in average returns (i.e., risk compensation). The fourth paper assumes time-invariant $z_{it}$ and extends Fan et al. (2016a)'s projected-PCA to allow for nonzero $\alpha(\cdot)$. The authors use the extended method to construct arbitrage portfolios.

**Example 2.2** (Implied volatility)**.** Fengler et al. (2007) and Park et al. (2009) use the model (1) to describe the dynamics of the implied volatility from traded options. Here, $y_{it}$ is the log-implied volatility where $t$ is an index of day and $i$ is an intra-day numbering of the option traded on day $t$, and $z_{it}$ is the two-dimensional vector of moneyness and time-to-maturity. Their main interest is the estimation of $f_t$ and its dynamics. The first paper considers a kernel estimation, and the second paper considers the least squares estimation based on sieve approximations. It is noted that the $i$th traded option on day $t$ and the $i$th traded option on day $s$ are different options for $t \neq s$ (abuse of notation), so the data does not necessarily exhibit a panel data structure. However, our results are also applicable for such data structure, as long as the imposed assumptions are satisfied.

**Example 2.3** (Consumer demand)**.** Lewbel (1991) uses the model (1) to describe a consumer's demand system. Here, $y_{it}$ represents the budget share of good $t$ for household $i$, and $z_{it} = z_i$ represents household $i$'s total expenditure. The main interest is the number of factors $K$, which is referred to as the rank of the demand system. Estimation of the rank of the demand system is important, because it provides evidence on consistency of consumer behaviors with utility maximization and has implications for welfare comparisons and aggregation across goods and across consumers.

# 3 Estimation Method

Before we proceed we need introduce some notation that is used throughout the paper. For a symmetric matrix $A$, we denote its $k$th largest eigenvalue by $\lambda_k(A)$, and its smallest and largest eigenvalues by $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$. For a matrix $A$, we denote its operator norm by $\|A\|_2$, its Frobenius norm by $\|A\|_F$, and its vectorization by $\text{vec}(A)$. The Euclidian norm of a column vector $x$ is denoted $\|x\|$. For matrices $A$ and $B$, we use $A \otimes B$ to denote their Kronecker product.

We begin by illustrating the idea behind our regressed-PCA method by assuming that $\alpha(\cdot)$ is null and $\beta(\cdot)$ is linear: $\alpha(\cdot) = 0$ and $\beta(z_{it}) = \Gamma' z_{it}$ for some $M \times K$ matrix $\Gamma$. Let $Y_t \equiv (y_{1t}, \dots, y_{Nt})'$, $Z_t \equiv (z_{1t}, \dots, z_{Nt})'$, and $\varepsilon_t \equiv (\varepsilon_{1t}, \dots, \varepsilon_{Nt})'$. Then we may write (1) in a matrix form

$$Y_t = Z_t \Gamma f_t + \varepsilon_t. \tag{2}$$

The main challenge of applying the standard PCA for estimating $\Gamma$ and $f_t$ is the presence of $Z_t$ in the first term on the right-hand side of (2). To circumvent it, we regress (2) on $Z_t$. Thus, we obtain

$$(Z_t' Z_t)^{-1} Z_t' Y_t = \Gamma f_t + (Z_t' Z_t)^{-1} Z_t' \varepsilon_t. \tag{3}$$

Heuristically, the variation of the common component $Z_t \Gamma f_t$ over $t$ has two sources: $Z_t$ and $f_t$, and regressing $Y_t$ on $Z_t$ can easily isolate the two sources or extract $Z_t$ from the common component. Given the factor structure on the right-hand side of (3), we may apply the standard PCA to the regressed data—$\{(Z_t' Z_t)^{-1} Z_t' Y_t\}_{t \le T}$—to obtain estimators of $\Gamma$ and $f_t$. We call the two-step procedure *the regressed-PCA*.

We next consider the general case with nonzero $\alpha(\cdot)$ and address how to estimate $\alpha(\cdot)$ and $\beta(\cdot) = (\beta_1(\cdot), \dots, \beta_K(\cdot))'$ nonparametrically. To estimate $\alpha(\cdot)$ and $\beta_k(\cdot)$ without the curse of dimensionality when $z_{it}$ is multivariate, we assume $\alpha(\cdot)$ and $\beta_k(\cdot)$ to be separable. Specifically, we assume that there are $\{\alpha_m(\cdot)\}_{m \le M}$ and $\{\beta_{km}(\cdot)\}_{m \le M}$ such that

$$\alpha(z_{it}) = \sum_{m=1}^{M} \alpha_m(z_{it,m}) \text{ and } \beta_k(z_{it}) = \sum_{m=1}^{M} \beta_{km}(z_{it,m}), \tag{4}$$

where $z_{it,m}$ is the $m$th entry of $z_{it}$. To estimate $\alpha_m(\cdot)$ and $\beta_{km}(\cdot)$, we adopt the sieve method. Let $\{\phi_j(\cdot)\}_{j \ge 1}$ be a set of basis functions (e.g., B-spline, Fourier series, polynomials), which spans a dense linear space of the functional space for $\alpha_m(\cdot)$ and $\beta_{km}(\cdot)$. Then we may write

$$\alpha_m(z_{it,m}) = \sum_{j=1}^{J} a_{m,j}\phi_j(z_{it,m}) + r_{m,J}(z_{it,m}), \tag{5}$$

$$\beta_{km}(z_{it,m}) = \sum_{j=1}^{J} b_{km,j}\phi_j(z_{it,m}) + \delta_{km,J}(z_{it,m}). \tag{6}$$

Here, $\{a_{m,j}\}_{j\le J}$ and $\{b_{km,j}\}_{j\le J}$ are the sieve coefficients; $r_{m,J}(\cdot)$ and $\delta_{km,J}(\cdot)$ are "remaining functions" representing the approximation errors; $J$ denotes the sieve size. The basic assumption for the sieve method is that $\sup_z |r_{m,J}(z)| \to 0$ and $\sup_z |\delta_{km,J}(z)| \to 0$ as $J \to \infty$. Let $\bar{\phi}(z_{it,m}) \equiv (\phi_1(z_{it,m}), \ldots, \phi_J(z_{it,m}))'$, $\phi(z_{it}) \equiv (\bar{\phi}(z_{it,1})', \ldots, \bar{\phi}(z_{it,M})')'$, $a \equiv (a_{1,1}, \ldots, a_{1,J}, \ldots, a_{M,1}, \ldots, a_{M,J})'$ which is a $JM \times 1$ vector of the sieve coefficients, $b_k \equiv (b_{k1,1}, \ldots, b_{k1,J}, \ldots, b_{kM,1}, \ldots, b_{kM,J})'$, and $B \equiv (b_1, \cdots, b_K)$ which is a $JM \times K$ matrix of the sieve coefficients. Let $r(z_{it}) \equiv \sum_{m=1}^{M} r_{m,J}(z_{it,m})$ and $\delta(z_{it}) \equiv (\sum_{m=1}^{M} \delta_{1m,J}(z_{it,m}), \ldots, \sum_{m=1}^{M} \delta_{Km,J}(z_{it,m}))'$. Then

$$\alpha(z_{it}) = a'\phi(z_{it}) + r(z_{it}) \text{ and } \beta(z_{it}) = B'\phi(z_{it}) + \delta(z_{it}). \tag{7}$$

Thus, $\alpha(z_{it})$ and $\beta(z_{it})$ can be well approximated by $a'\phi(z_{it})$ and $B'\phi(z_{it})$ under the basic sieve assumption, and estimating $\alpha(\cdot)$ and $\beta(\cdot)$ reduces to estimating $a$ and $B$.

We now introduce the estimation of $a$, $B$ and $f_t$ based on the above sieve approximation in (7) by adapting the regressed-PCA. Let $\Phi(Z_t) \equiv (\phi(z_{1t}), \ldots, \phi(z_{Nt}))'$, $R(Z_t) \equiv (r(z_{1t}), \ldots, r(z_{Nt}))'$ and $\Delta(Z_t) \equiv (\delta(z_{1t}), \ldots, \delta(z_{Nt}))'$. Using the sieve approximation in (7), we may write (1) in a matrix form

$$Y_t = \Phi(Z_t)a + \Phi(Z_t)Bf_t + R(Z_t) + \Delta(Z_t)f_t + \varepsilon_t. \tag{8}$$

Under the basic sieve assumption, the term "$R(Z_t) + \Delta(Z_t)f_t$" is negligible, so the main challenge for applying the standard PCA to estimate $a$, $B$ and $f_t$ is the presence of $\Phi(Z_t)$ in the first two terms on the right-hand side of (8). To solve the challenge, we use the regressed-PCA. Specifically, we may regress (8) on $\Phi(Z_t)$ to obtain

$$\tilde{Y}_t = a + Bf_t + (\Phi(Z_t)'\Phi(Z_t))^{-1}\Phi(Z_t)'(R(Z_t) + \Delta(Z_t)f_t + \varepsilon_t), \tag{9}$$

where $\tilde{Y}_t = (\Phi(Z_t)'\Phi(Z_t))^{-1}\Phi(Z_t)'Y_t$. Thus, we may estimate $a$, $B$ and $f_t$ as follows. First, we may remove $a$ by subtracting $\bar{\tilde{Y}} = \sum_{t=1}^{T} \tilde{Y}_t/T$ from $\tilde{Y}_t$ and estimate $B$ by applying the standard PCA to $\{\tilde{Y}_t - \bar{\tilde{Y}}\}_{t\le T}$. Second, we may impose $a'B = 0$ and estimate $a$ from $\bar{\tilde{Y}}$, since $a + B\bar{f}$ is approximated by $\bar{\tilde{Y}}$ where $\bar{f} = \sum_{t=1}^{T} f_t/T$.

The estimators of $a$, $B$, $\alpha(\cdot)$, $\beta(\cdot)$ and $F = (f_1, \ldots, f_T)'$ are defined as follows. Denote the estimators by $\hat{a}$, $\hat{B}$, $\hat{\alpha}(\cdot)$, $\hat{\beta}(\cdot)$ and $\hat{F}$. Let $M_T \equiv I_T - 1_T 1_T'/T$, where $1_T$ denotes a $T \times 1$ vector of ones. We use the following normalization: $B'B = I_K$ and $F'M_T F/T$ being diagonal with diagonal entries in descending order. Let $\tilde{Y} \equiv (\tilde{Y}_1, \ldots, \tilde{Y}_T)$. Then the columns of $\hat{B}$ are the eigenvectors corresponding to the first $K$ largest eigenvalues of the $JM \times JM$ matrix $\tilde{Y} M_T \tilde{Y}'/T$, $\hat{a} = (I_{JM} - \hat{B}\hat{B}')\bar{\tilde{Y}}$, and

$$\hat{\alpha}(z) = \hat{a}'\phi(z), \hat{\beta}(z) = \hat{B}'\phi(z) \text{ and } \hat{F} = (\hat{f}_1, \ldots, \hat{f}_T)' = \tilde{Y}'\hat{B}. \tag{10}$$

The intuitions for $\hat{a}$ and $\hat{F}$ are as follows. First, since $a + B\bar{f}$ is approximated by $\bar{\tilde{Y}}$ and $a'B = 0$, $B'B\bar{f}$ is approximated by $B'\bar{\tilde{Y}}$. Thus, we may estimate $\bar{f}$ by $(B'B)^{-1}B'\bar{\tilde{Y}}$, and $a$ by $[I_{JM} - B(B'B)^{-1}B']\bar{\tilde{Y}}$. Second, from (9), we may estimate $f_t$ by $(B'B)^{-1}B'(\tilde{Y}_t - a) = (B'B)^{-1}B'\tilde{Y}_t$ since $a'B = 0$. Here, we assume that $K$—the number of factors—is known, and conduct asymptotic analysis and develop inference method in Sections 4 and 5. In Section 6, we develop two consistent estimators of $K$, so all the results carry over to the unknown $K$ case using a conditioning argument.

**Remark 3.1.** Our estimation procedure is applicable for unbalanced panels. The key step is to obtain $\tilde{Y}_t$. To this end, we may write $\tilde{Y}_t = [\sum_{i=1}^N \phi(z_{it})\phi(z_{it})']^{-1} \sum_{i=1}^N \phi(z_{it}) y_{it}$. In the presence of unbalanced panels, we may obtain $\tilde{Y}_t$ by taking the two sums over $i$'s, for which both $z_{it}$ and $y_{it}$ are observed in time period $t$. This is equivalent to replacing missing data with zeros and proceeding as balanced panels. The asymptotic results established in the following sections continue to hold as $\min_{t \le T} N_t \to \infty$, where $N_t$ is the sample size in time period $t$.

**Remark 3.2.** The approximated model in (8) can be alternatively viewed as a panel data model with time-varying slope coefficients $a + Bf_t$, which exhibit a factor structure. The regressed-PCA first estimates the time-varying slope coefficients by period-by-period cross-sectional regressions, and then exploits the factor structure by using PCA. The period-by-period cross-sectional regressions are known as Fama-MacBeth regressions (Fama and MacBeth, 1973). In asset pricing applications, $\tilde{Y}_t$ can be interpreted as the time $t$ realization of returns on a set of $JM$ managed portfolios. Thus, the regressed-PCA boils down to applying PCA to a set of characteristic-managed portfolios constructed via the Fama-MacBeth regressions. The $\ell$th entry of $\tilde{Y}_t$ is a weighted average of asset returns with weights determined by the $\ell$th row of $(\Phi(Z_t)'\Phi(Z_t))^{-1}\Phi(Z_t)'$, which is a standardized version of $\Phi(Z_t)$. If $\Phi(Z_t)'\Phi(Z_t)$ is diagonal, the portfolios are normalized by the second moment of $\Phi(Z_t)$. If the polynomial basis functions are used, $\Phi(Z_t)$ consists of powers of $Z_t$. This allows us to investigate nonlinearity of characteristics in pricing errors and risk exposures.

To end this section, we compare our regressed-PCA with existing methods. First, the regressed-PCA is different from projected-PCA of Fan et al. (2016a). The projected-PCA applies the standard PCA to the projected data—$\{\Phi(Z_t)(\Phi(Z_t)'\Phi(Z_t))^{-1}\Phi(Z_t)'Y_t\}_{t \le T}$. The regression in the regressed-PCA is designed to extract $Z_t$ from the common component for a consistent estimation, whereas the projection in the projected-PCA is designed to remove the noise part of the factor loadings for a more efficient estimation. Therefore, the projected-PCA may fail to provide consistent estimators when $Z_t$ is time-varying. Indeed, as discussed in Appendix G of Fan et al. (2016b), one may need to impose certain smoothness conditions of $Z_t$ over $t$ to ensure the consistency of the projected-PCA; the regressed-PCA does not require such conditions (see below). In view of this, our regressed-PCA is more robust. Second, the regressed-PCA is superior to the least squares estimation (Park et al., 2009) that is the core of the instrumented PCA (IPCA) of Kelly et al. (2019) in terms of computation and asymptotic properties. The least squares estimation may not be explicitly solved because the problem is nonconvex, so Park et al. (2009) develop a Newton-Raphson algorithm to numerically find the estimators, and Kelly et al. (2019) propose to use the alternating least squares procedure. However, both numerical methods may require a good choice of initial values[4], and their asymptotic properties have not been well understood. Our estimators can always be explicitly solved for, and their computation is easy since it involves only regression and PCA.

# 4 Asymptotic Analysis

In this section, we conduct asymptotic analysis for our estimators. Specifically, we establish consistency, rate of convergence, and asymptotic distribution.

## 4.1 Consistency

To establish consistency, we impose the following assumptions.

**Assumption 4.1** (Basis functions). *(i) There are positive constants $c_{\min}$ and $c_{\max}$ such that: with probability approaching one (as $N \to \infty$),*

$$c_{\min} < \min_{t \leq T} \lambda_{\min}(\hat{Q}_t) \leq \max_{t \leq T} \lambda_{\max}(\hat{Q}_t) < c_{\max},$$

*where $\hat{Q}_t = \Phi(Z_t)'\Phi(Z_t)/N$; (ii)* $\max_{m \leq M, j \leq J, i \leq N, t \leq T} E[\phi_j^2(z_{it,m})] < \infty$.

Note that $\hat{Q}_t = \sum_{i=1}^N \phi(z_{it})\phi(z_{it})'/N$ is an $JM \times JM$ matrix with $JM$ much smaller than $N$. Thus, Assumption 4.1(i) can follow from the law of large numbers for finite $T$ and its uniform variant for $T \to \infty$; see Proposition I.1 for its justification when $\{z_{it}\}_{i \leq N, t \leq T}$ are independent across $i$. The assumption can be easily verified for commonly used basis functions such as B-spline, Fourier series, and polynomials. Since $Z_t$

[4] Kelly et al. (2019) do not provide proofs for the convergence of the alternating least squares procedure; as noted by (Park et al., 2009), "it is not guaranteed to converge to a solution of the original problem."

is allowed to change over $t$, we need the well-conditionedness of $\hat{Q}_t$ for all $t$. We allow $Z_t$ to be nonstationary over $t$. When $Z_t$ is not changing over $t$, Assumption 4.1 reduces to Assumptions 3.3 of Fan et al. (2016a).

**Assumption 4.2** (Loading functions and factors)**.** *There are positive constants* $d_{\min}$ *and* $d_{\max}$ *such that: (i)* $d_{\min} < \lambda_{\min}(B'B) \leq \lambda_{\max}(B'B) < d_{\max}$*; (ii)* $\max_{t \leq T} \|f_t\| < d_{\max}$*; (iii)* $\lambda_{\min}(F'M_T F/T) > d_{\min}$*; (iv)* $\max_{m \leq M} \sup_z |r_{m,J}(z)| = O(J^{-\kappa})$ *and* $\max_{k \leq K, m \leq M} \sup_z |\delta_{km,J}(z)| = O(J^{-\kappa})$ *for some constant* $\kappa > 1/2$*.*

Assumption 4.2(i) is similar to the *pervasive* condition on the factor loadings in Stock and Watson (2002). Similar assumptions also are imposed in Assumption B of Bai (2003) and Assumption 4.1(ii) of Fan et al. (2016a). For simplicity of proof, we assume that $f_t$ is nonrandom. All the results continue to hold, when $\{f_t\}_{t \leq T}$ are random and independent of $\{Z_t, \varepsilon_t\}_{t \leq T}$. Since the dimension of $B$ is $JM \times K$, Assumption 4.2(i) requires $JM \geq K$. Since the rank of $M_T$ is $T-1$, Assumption 4.2(iii) requires $T \geq K+1$, which implies $T \geq 2$. These two requirements are reasonable, since we assume $K$ to be fixed throughout the paper. Assumption 4.2(iv) is standard in the sieve literature. It can be easily satisfied by using B-spline or polynomials basis functions under certain smoothness of $\alpha(\cdot)$ and $\beta(\cdot)$; see, for example, Lorentz (1986) and Chen (2007).

**Assumption 4.3** (Data generating process)**.** *(i)* $\{\varepsilon_t\}_{t \leq T}$ *is independent of* $\{Z_t\}_{t \leq T}$*; (ii)* $E[\varepsilon_{it}] = 0$ *for all* $i \leq N$ *and* $t \leq T$*; (iii) there is* $0 < C_1 < \infty$ *such that*

$$\max_{i \leq N, t \leq T} \sum_{j=1}^{N} |E[\varepsilon_{it}\varepsilon_{jt}]| < C_1 \text{ and } \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} |E[\varepsilon_{it}\varepsilon_{js}]| < C_1.$$

Assumption 4.3 is standard in the literature. In particular, Assumption 4.3(iii) requires $\{\varepsilon_{it}\}_{i \leq N, t \leq T}$ to be weakly dependent over both $i$ and $t$, and is commonly imposed for high-dimensional factor analysis; see, for example, Stock and Watson (2002), Bai (2003), and Fan et al. (2016a). When $Z_t$ is not changing over $t$, Assumption 4.3 reduces to Assumptions 3.4 (i) and (iii) of Fan et al. (2016a).

**Assumption 4.4** (Intercept function)**.** $a'B = 0$ *and* $\|a\| < C_0$ *for some* $0 < C_0 < \infty$*.*

Assumption 4.4 is imposed for the identification of $\alpha(\cdot)$. Similar assumption is imposed in Connor et al. (2012) and Assumption 3.1(i) of Kim et al. (2020).

To proceed, let $H \equiv (F'M_T\hat{F})(\hat{F}'M_T\hat{F})^{-1}$, which is a rotational transformation matrix that is needed to define the convergence limits of $\hat{B}$ and $\hat{F}$. The first result is established as follows.

**Theorem 4.1.** *Suppose Assumptions 4.1-4.4 hold. Let* $\hat{a}$*,* $\hat{B}, \hat{F}$*,* $\hat{\alpha}(\cdot)$ *and* $\hat{\beta}(\cdot)$ *be given in* (10)*. Assume (i)* $N \to \infty$*; (ii)* $T \to \infty$ *or* $T \geq K+1$ *is finite; (iii)* $J \to \infty$ *with*

$J = o(\sqrt{N})$. *Then*

$$\|\hat{a} - a\|^2 = O_p\left(\frac{1}{J^{2\kappa}} + \frac{J^2}{N^2} + \frac{J}{NT}\right),$$

$$\|\hat{B} - BH\|_F^2 = O_p\left(\frac{1}{J^{2\kappa}} + \frac{J^2}{N^2} + \frac{J}{NT}\right),$$

$$\frac{1}{T}\|\hat{F} - F(H')^{-1}\|_F^2 = O_p\left(\frac{1}{J^{2\kappa}} + \frac{J}{N}\right),$$

$$\sup_z |\hat{\alpha}(z) - \alpha(z)|^2 = O_p\left(\frac{1}{J^{2\kappa-1}} + \frac{J^3}{N^2} + \frac{J^2}{NT}\right) \max_{j\le J} \sup_z |\phi_j(z)|^2,$$

$$\sup_z \|\hat{\beta}(z) - H'\beta(z)\|^2 = O_p\left(\frac{1}{J^{2\kappa-1}} + \frac{J^3}{N^2} + \frac{J^2}{NT}\right) \max_{j\le J} \sup_z |\phi_j(z)|^2.$$

Theorem 4.1 implies that $a$ and $\alpha(\cdot)$ can be consistently estimated by $\hat{a}$ and $\hat{\alpha}(\cdot)$, and $B$, $F$ and $\beta(\cdot)$ can be consistently estimated by $\hat{B}$, $\hat{F}$ and $\hat{\beta}(\cdot)$ up to a rotational transformation. We have two interesting findings. First, the consistency does not require $T \to \infty$, as similar to Fan et al. (2016a). To quickly see this, let us return to the illustrative model (2) and assume that $T$ is finite, $F'M_TF$ is diagonal, and $\Gamma'\Gamma = I_K$. Let $\Theta$ and $\hat{\Theta}$ be $M \times T$ matrices with $\Theta = \Gamma F' M_T$ and $\hat{\Theta} = ((Z_1'Z_1)^{-1}Z_1'Y_1, \ldots, (Z_T'Z_T)^{-1}Z_T'Y_T)M_T$. Then $\sqrt{N}(\hat{\Theta} - \Theta) = O_p(1)$ by the central limit theorem. Since the columns of $\Gamma$ and $\hat{B}$ are eigenvectors of $\Theta\Theta'$ and $\hat{\Theta}\hat{\Theta}'$, the consistency of $\hat{B}$ thus can be easily understood from the matrix perturbation theorem; see, for example, Yu et al. (2014).

Second, the consistency of $\hat{F}$ requires $J \to \infty$, as different from Fan et al. (2016a). This is because a large sieve approximation error of $\alpha(\cdot)$ and $\beta(\cdot)$ may cause inconsistent estimation of $F$. To quickly see this, let us look at the following simple linear models

$$Y_t = W_t\Pi + Z_t\Gamma f_t + \varepsilon_t, \tag{11}$$

$$Y_t = (Z_t\Gamma + W_t\Pi) f_t + \varepsilon_t, \tag{12}$$

where $Z_t$ and $W_t$ are $N \times 1$ vectors, $f_t$ is a scalar factor, and $\varepsilon_t$ is independent of $Z_t$ and $W_t$. Let us further assume $\Pi = \Gamma$ and $W_t = Z_t g_t + v_t$, where $g_t$ is a scalar coefficient, and $v_t$ is independent of $Z_t$. Then (11) and (12) can be rewritten as

$$Y_t = Z_t\Gamma f_t^\star + \varepsilon_t^\star, \tag{13}$$

$$Y_t = Z_t\Gamma f_t^{\star\star} + \varepsilon_t^{\star\star}, \tag{14}$$

where $f_t^\star = f_t + g_t$, $\varepsilon_t^\star = v_t\Gamma + \varepsilon_t$, $f_t^{\star\star} = f_t(1 + g_t)$, and $\varepsilon_t^{\star\star} = v_t\Gamma f_t + \varepsilon_t$. Thus, if only $Z_t$ is used in (11) (the sieve approximation error of $\alpha(\cdot)$ is large), then $\hat{F}$ can consistently estimate $F^\star = (f_1^\star, \ldots, f_T^\star)'$ up to a scalar; if only $Z_t$ is used (12) (namely, the sieve approximation error of $\beta(\cdot)$ is large), then $\hat{F}$ can consistently estimate $F^{\star\star} = (f_1^{\star\star}, \ldots, f_T^{\star\star})'$ up to a scalar. In both cases, $\hat{F}$ fails to consistently estimate the space spanned by $F$, unless $g_t$ is proportional to $f_t$ in the former case and is not changing over $t$ in the latter case. See Appendix G for a formal analysis for the general model under $\alpha(\cdot) = 0$. The result also implies that misspefication of $\alpha(\cdot)$ and $\beta(\cdot)$ may cause inconsistent estimation of $F$. This motivates us to develop a specification test for $\alpha(\cdot)$ and $\beta(\cdot)$; see Section 5.2.

## 4.2 Rate of Convergence

Theorem 4.1 also gives a preliminary convergence rate of $\hat{a}$, $\hat{B}$, $\hat{F}$, $\hat{\alpha}(\cdot)$ and $\hat{\beta}(\cdot)$. However, the rate is not optimal, which may create challenge for deriving the asymptotic distribution. To improve the rate, we impose the following assumption.

**Assumption 4.5** (Rate of convergence)**.** *(i)* $\max_{m\leq M, j\leq J, i\leq N, t\leq T} E[\phi_j^4(z_{it,m})] < \infty$*; (ii)* $0 < \min_{i\leq N,t\leq T} \lambda_{\min}(Q_{it}) \leq \max_{i\leq N,t\leq T} \lambda_{\max}(Q_{it}) < \infty$*, where* $Q_{it} = E[\phi(z_{it})\phi(z_{it})']$*; (iii)* $\{z_{it}\}_{i\leq N,t\leq T}$ *are independent across* $i \leq N$*; (iv) there is* $0 < C_2 < \infty$ *such that*

$$\max_{t\leq T} \frac{1}{N^2} \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{\ell=1}^{N} |E[\varepsilon_{it}\varepsilon_{jt}\varepsilon_{kt}\varepsilon_{\ell t}]| < C_2$$

*and*

$$\frac{1}{N^2 T}\sum_{t=1}^{T}\left(\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N} |E[\varepsilon_{it}\varepsilon_{js}]|\right)^2 < C_2.$$

Assumption 4.5 is also standard in the literature, though not required by Fan et al. (2016a). Assumption 4.5(i) strengthens Assumption 4.1(ii). Assumption 4.5(ii) requires that the second moment matrix $E[\phi(z_{it})\phi(z_{it})']$ is bounded and nonsingular for all $i$ and $t$, which is standard in the sieve literature; see, for example, Newey (1997) and Huang (1998). Assumption 4.5(iii) is also standard in the sieve literature, which is used to justify the asymptotic convergence of $\hat{Q}_t$. Assumption 4.5(iv) allows for weak dependence of $\{\varepsilon_{it}\}_{i\leq N,t\leq T}$ over both $i$ and $t$, which is standard in the literature as Assumption 4.3(iii). The second condition is similar to the second condition in Assumption 4.3(iii); both are satisfied if $\max_{t\leq T} \sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N} |E[\varepsilon_{it}\varepsilon_{js}]/N$ is bounded.

To proceed, let $\xi_J \equiv \sup_z \|\bar{\phi}(z)\|$, which plays an important role in justifying the asymptotic convergence of $\hat{Q}_t$. In particular, $\xi_J = O(\sqrt{J})$ for B-spline and Fourier series, and $\xi_J = O(J)$ for polynomials; see Section 3 of Belloni et al. (2015) for more discussions on $\xi_J$. The second result is established as follows.

**Theorem 4.2.** *Suppose Assumptions 4.1-4.5 hold. Let* $\hat{a}$*,* $\hat{B}, \hat{F}$*,* $\hat{\alpha}(\cdot)$ *and* $\hat{\beta}(\cdot)$ *be given in* (10)*. Assume (i)* $N \to \infty$*; (ii)* $T \to \infty$ *or* $T \geq K+1$ *is finite; (iii)* $J \to \infty$ *with*

$J^2\xi_J^2 \log J = o(N)$. *Then*

$$\|\hat{a} - a\|^2 = O_p\left(\frac{1}{J^{2\kappa}} + \frac{J}{N^2} + \frac{J}{NT}\right),$$
$$\|\hat{B} - BH\|_F^2 = O_p\left(\frac{1}{J^{2\kappa}} + \frac{J}{N^2} + \frac{J}{NT}\right),$$
$$\frac{1}{T}\|\hat{F} - F(H')^{-1}\|_F^2 = O_p\left(\frac{1}{J^{2\kappa}} + \frac{1}{N}\right),$$
$$\sup_z |\hat{\alpha}(z) - \alpha(z)|^2 = O_p\left(\frac{1}{J^{2\kappa-1}} + \frac{J^2}{N^2} + \frac{J^2}{NT}\right) \max_{j\le J} \sup_z |\phi_j(z)|^2,$$
$$\sup_z \|\hat{\beta}(z) - H'\beta(z)\|^2 = O_p\left(\frac{1}{J^{2\kappa-1}} + \frac{J^2}{N^2} + \frac{J^2}{NT}\right) \max_{j\le J} \sup_z |\phi_j(z)|^2.$$

Theorem 4.2 shows that the rates of $\hat{B}$, $\hat{F}$ and $\hat{\beta}(\cdot)$ are equal to the rates in Fan et al. (2016a). The requirement $J^2\xi_J^2 \log J = o(N)$ is standard in the sieve literature; see, for example, Belloni et al. (2015). See Appendix H for a special case analysis without Assumption 4.5 and the requirement. It is worthwhile to discuss the rate of $\hat{F}$. Assume $J^{-2\kappa}N = O(1)$, which can be satisfied for sufficiently large $\kappa$ under the restriction $J^2\xi_J^2 \log J = o(N)$. Then $\hat{F}$ attains the optimal rate $1/N$, which is the fastest rate that one can obtain when $\alpha(\cdot)$ and $\beta(\cdot)$ were known. This implies that the nonparametric specification of $\alpha(\cdot)$ and $\beta(\cdot)$ does not deteriorate the optimal rate for estimating $F$ as long as $\alpha(\cdot)$ and $\beta(\cdot)$ are sufficiently smooth (i.e., $\kappa$ is sufficiently large), or the linear specification of $\alpha(\cdot)$ and $\beta(\cdot)$ does not necessarily improve the estimation of $F$. This implication is important in developing specification test for $\alpha(\cdot)$ and $\beta(\cdot)$ in Section 5.2.

**Remark 4.1.** Our proofs for the consistency and rate of convergence are not trivial relative to Fan et al. (2016a) for several reasons. First, the regressed-PCA is different from the projected-PCA, as mentioned in Section 3. Second, we allow for nonzero $\alpha(\cdot)$, and need to additionally study the properties of $\hat{\alpha}(\cdot)$. Third, the presence of $z_{it}$ varying over $t$ makes the proofs challenging. In particular, the proof for the rate of convergence relies on LLNs for matrices derived from the Khinchin inequality.

### 4.3 Asymptotic Distribution

We focus on deriving the asymptotic distributions of $\hat{a}$ and $\hat{B}$, since our main concern is the inference on $\alpha(\cdot)$ and $\beta(\cdot)$. To this end, we impose the following assumption.

**Assumption 4.6** (Asymptotic distribution)**.** *(i) The eigenvalues of* $(F'M_TF/T)B'B$ *are distinct; (ii)* $\{\varepsilon_{it}\}_{i\le N, t\le T}$ *are independent across* $i \le N$*; (iii) there is* $0 < C_3 < \infty$ *such that*

$$\max_{i\le N} \frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T \sum_{u=1}^T \sum_{v=1}^T |E[\varepsilon_{it}\varepsilon_{is}\varepsilon_{iu}\varepsilon_{iv}]| < C_3.$$

The distinct eigenvalue condition in Assumption 4.6(i) is necessary to establish the asymptotic normality, as known in the literature; see, for example, Bai (2003) and Chen and Fang (2019). Assumption 4.6(ii) imposes independence of $\{\varepsilon_{it}\}_{i\leq N,t\leq T}$ across $i$ for simplicity.[5] Assumption 4.6(iii) allows for weak dependence of $\{\varepsilon_{it}\}_{i\leq N,t\leq T}$ over $t$, which is standard in the literature as Assumptions 4.3(iii) and 4.5(iv).

To proceed, let $\Omega \equiv \sum_{i=1}^N \sum_{t=1}^T \sum_{s=1}^T f_t^\dagger f_s^{\dagger\prime} \otimes Q_t^{-1} E[\phi(z_{it})\phi(z_{is})']Q_s^{-1} E[\varepsilon_{it}\varepsilon_{is}]/NT$, where $f_t^\dagger = (1, (f_t - \bar{f})')'$ and $Q_t = E[\hat{Q}_t] = \sum_{i=1}^N Q_{it}/N$. It is a variance-covariance matrix, which will appear in the asymptotic distributions of $\hat{a}$ and $\hat{B}$. The third result is established as follows.

**Theorem 4.3.** *Suppose Assumptions 4.1-4.6 hold. Let $\hat{a}$ and $\hat{B}$ be given in* (10). *Assume (i) $N \to \infty$; (ii) $T \to \infty$ or $T \geq K+1$ is finite; (iii) $J \to \infty$ with $J^2\xi_J^2 \log J = o(N)$. Then there exists a $JM \times (K+1)$ random matrix $\mathbb{N}$ with* $\mathrm{vec}(\mathbb{N}) \sim N(0,\Omega)$ *such that*

$$\|\sqrt{NT}(\hat{a}-a) - \mathbb{G}_a\| = O_p\left(\frac{\sqrt{NT}}{J^\kappa} + \frac{\sqrt{TJ}}{\sqrt{N}} + \frac{J^{5/6}}{N^{1/6}} + \frac{\sqrt{J\xi_J}\log^{1/4} J}{N^{1/4}}\right)$$

*and*

$$\|\sqrt{NT}(\hat{B}-BH) - \mathbb{G}_B\|_F = O_p\left(\frac{\sqrt{NT}}{J^\kappa} + \frac{\sqrt{TJ}}{\sqrt{N}} + \frac{J^{5/6}}{N^{1/6}} + \frac{\sqrt{J\xi_J}\log^{1/4} J}{N^{1/4}}\right),$$

*where $\mathbb{G}_a = (I_{JM} - B\mathcal{H}\mathcal{H}'B')(\mathbb{N}_1 - \mathbb{G}_B\mathcal{H}^{-1}\bar{f}) - B\mathcal{H}\mathbb{G}_B' a$ and $\mathbb{G}_B = \mathbb{N}_2 B'B\mathcal{M}$, $\mathcal{H}$ and $\mathcal{M}$ are nonrandom matrices given in Lemma C.3, and $\mathbb{N}_1$ and $\mathbb{N}_2$ are the first column and the last $K$ columns of $\mathbb{N}$.*

Theorem 4.3 establishes a strong approximation: $(\sqrt{NT}(\hat{a}-a), \sqrt{NT}(\hat{B}-BH))$ can be well approximated by a normal random matrix $(\mathbb{G}_a, \mathbb{G}_B)$, in the sense that their difference converges in probability to zero when $T = o(N)$, $J = o(\min\{N^{1/5}, N/T\})$, and $NTJ^{-2\kappa} = o(1)$. Therefore, $(\sqrt{NT}(\hat{a}-a), \sqrt{NT}(\hat{B}-BH))$ behaves like a normal random matrix. Here, $\sqrt{NT}(\hat{a}-a)$ and $\sqrt{NT}(\hat{B}-BH)$ exhibit growing dimensions, so the classical central limit theorem does not apply. Instead, we use the Yurinskiis coupling (which is collected in Lemma C.5 for ease of reference) to establish the strong approximation. We stress that the strong approximation allows for weak dependence of $\{\varepsilon_{it}\}_{i\leq N,t\leq T}$ over $t$. Similar results are not available in Fan et al. (2016a).

[5] This assumption allows us to use the Yurinskiis coupling. In fact, we may relax this assumption and alternatively use Li and Liao (2019)'s coupling, so that the dependence across $i$ can be allowed. However, it is challenging to develop an inference procedure allowing the dependence over both $i$ and $t$. Therefore, we stick with this assumption.

# 5 Bootstrap Inference

In this section, we develop a weighted bootstrap to estimate the distribution of $(\mathbb{G}_a, \mathbb{G}_B)$ in Theorem 4.3, and a specification test for linearity of $\alpha(\cdot)$ and $\beta(\cdot)$.

## 5.1 Weighted Bootstrap

It seems straightforward to estimate the distribution of $(\mathbb{G}_a, \mathbb{G}_B)$ by estimating its unknown components. However, it may be challenging to estimate $\Omega$, especially since we allow for weak dependence of $\{\varepsilon_{it}\}_{i\leq N, t\leq T}$ over $t$. To circumvent the challenge, we develop a weighted bootstrap, which additionally may have computational advantage. To this end, we impose the following assumption.

**Assumption 5.1** (Bootstrap). *(i) $\{w_i\}_{i\leq N}$ is a sequence of independently and identically (i.i.d.) positive random variables with $E[w_i] = 1$ and $var(w_i) = \omega_0 > 0$, and is independent of $\{Z_t, \varepsilon_t\}_{t\leq T}$; (ii) there are positive constants $e_{\min}$ and $e_{\max}$ such that: with probability approaching one (as $N \to \infty$),*

$$e_{\min} < \min_{t\leq T} \lambda_{\min}(\hat{Q}_t^*) \leq \max_{t\leq T} \lambda_{\max}(\hat{Q}_t^*) < e_{\max},$$

*where $\hat{Q}_t^* = \Phi(Z_t)^{*\prime}\Phi(Z_t)/N$ and $\Phi(Z_t)^* = (\phi(z_{1t})w_1, \ldots, \phi(z_{Nt})w_N)'$; (iii) $\lambda_{\min}(\Omega) > 0$.*

Assumption 5.1(i) defines the bootstrap weight $w_i$ for each $i$. Specifically, we assign $w_i$ to all observations over $t$ for each $i$ to maintain the dependence of $\{\varepsilon_{it}\}_{i\leq N,t\leq T}$ over $t$. Note that $\hat{Q}_t^* = \sum_{i=1}^N \phi(z_{it})\phi(z_{it})'w_i/N$ is a $JM \times JM$ matrix with $JM$ much smaller than $N$. Thus, Assumption 5.1(ii) can follow from the law of large numbers for finite $T$ and its uniform variant for $T \to \infty$, similar to Assumption 4.1(i). Assumption 5.1(iii) requires nonsingularity of the variance-covariance matrix $\Omega$.

To define the bootstrap estimators of $a$ and $B$, let $\tilde{Y}_t^* \equiv (\Phi(Z_t)^{*\prime}\Phi(Z_t))^{-1}\Phi(Z_t)^{*\prime}Y_t$, $\tilde{Y}^* \equiv (\tilde{Y}_1^*, \ldots, \tilde{Y}_T^*)$, and $\bar{\tilde{Y}}^* \equiv \sum_{t=1}^T \tilde{Y}_t^*/T$. The bootstrap estimators are given by

$$\hat{B}^* = \tilde{Y}^* M_T \hat{F}(\hat{F}' M_T \hat{F})^{-1} \text{ and } \hat{a}^* = (I_{JM} - \hat{B}^*(\hat{B}^{*\prime}\hat{B}^*)^{-1}\hat{B}^{*\prime})\bar{\tilde{Y}}^*, \tag{15}$$

which mimic $\hat{B}$ and $\hat{a}$ following the formulas $\hat{B} = \tilde{Y} M_T \hat{F}(\hat{F}' M_T \hat{F})^{-1}$ and $\hat{a} = (I_{JM} - \hat{B}\hat{B}')\bar{\tilde{Y}} = (I_{JM} - \hat{B}(\hat{B}'\hat{B})^{-1}\hat{B}')\bar{\tilde{Y}}$. We propose to estimate the distribution of $(\mathbb{G}_a, \mathbb{G}_B)$ by the distribution of $(\sqrt{NT/\omega_0}\ (\hat{a}^* - \hat{a}), \sqrt{NT/\omega_0}(\hat{B}^* - \hat{B}))$ conditional on the data. The validity of the bootstrap for $\hat{B}$ can be quickly seen when $T = 2$ and $K = 1$.[6]

[6]In this case, $\hat{B} = (\tilde{Y}_1 - \tilde{Y}_2)/\|\tilde{Y}_1 - \tilde{Y}_2\|$, $BH = B(f_1 - f_2)/\|\tilde{Y}_1 - \tilde{Y}_2\|$ and $\hat{B}^* = (\tilde{Y}_1^* - \tilde{Y}_2^*)/\|\tilde{Y}_1 - \tilde{Y}_2\|$. Thus, the distribution of $\sqrt{NT}(\hat{B} - BH) = \sqrt{NT}(\tilde{Y}_1 - \tilde{Y}_2 - B(f_1 - f_2))/\|\tilde{Y}_1 - \tilde{Y}_2\|$ can be estimated by the distribution of $\sqrt{NT/\omega_0}(\tilde{Y}_1^* - \tilde{Y}_2^* - (\tilde{Y}_1 - \tilde{Y}_2))/\|\tilde{Y}_1 - \tilde{Y}_2\| = \sqrt{NT/\omega_0}(\hat{B}^* - \hat{B})$ conditional on the data by the weighted bootstrap in Belloni et al. (2015).

**Remark 5.1.** The bootstrap can be easily adapted for unbalanced panels. The key step is to obtain $\tilde{Y}_t^*$. To this end, we may write $\tilde{Y}_t^* = [\sum_{i=1}^N \phi(z_{it})\phi(z_{it})' w_i]^{-1} \sum_{i=1}^N \phi(z_{it}) y_{it} w_i$. In the presence of unbalanced panels, we may obtain $\tilde{Y}_t^*$ by taking the two sums over $i$'s, for which both $z_{it}$ and $y_{it}$ are observed in time period $t$. This is equivalent to replacing missing data with zeros and proceeding as balanced panels. Prior to this, we need to generate $\{w_i\}_{i \le N_{\max}}$ once, where $N_{\max}$ is the number of all observation unit $i$'s. The asymptotic results established below continue to hold as $\min_{t \le T} N_t \to \infty$, where $N_t$ is the sample size in time period $t$.

To proceed, let $p^*$ denote the probability measure with respect to $\{w_i\}_{i \le N}$ conditional on $\{Y_t, Z_t\}_{t \le T}$. The fourth result is established as follows.

**Theorem 5.1.** *Suppose Assumptions 4.1-4.6 and 5.1 hold. Let $\hat{a}$, $\hat{B}$, $\hat{a}^*$ and $\hat{B}^*$ be given in* (10) *and* (15). *Assume (i) $N \to \infty$; (ii) $T \to \infty$ or $T \ge K+1$ is finite; (iii) $J \to \infty$ with $J^2 \xi_J^2 \log J = o(N)$. Then there exists a $JM \times (K+1)$ random matrix $\mathbb{N}^*$ with $\mathrm{vec}(\mathbb{N}^*) \sim N(0, \Omega)$ conditional on $\{Y_t, Z_t\}_{t \le T}$ such that*

$$\|\sqrt{NT/\omega_0}(\hat{a}^* - \hat{a}) - \mathbb{G}_a^*\| = O_{p^*}\left(\frac{\sqrt{NT}}{J^\kappa} + \frac{\sqrt{TJ}}{\sqrt{N}} + \frac{J^{5/6}}{N^{1/6}} + \frac{\sqrt{J\xi_J}\log^{1/4} J}{N^{1/4}}\right)$$

*and*

$$\|\sqrt{NT/\omega_0}(\hat{B}^* - \hat{B}) - \mathbb{G}_B^*\|_F = O_{p^*}\left(\frac{\sqrt{NT}}{J^\kappa} + \frac{\sqrt{TJ}}{\sqrt{N}} + \frac{J^{5/6}}{N^{1/6}} + \frac{\sqrt{J\xi_J}\log^{1/4} J}{N^{1/4}}\right),$$

*where $\mathbb{G}_a^* = (I_{JM} - B\mathcal{H}\mathcal{H}'B')(\mathbb{N}_1^* - \mathbb{G}_B^* \mathcal{H}^{-1}\bar{f}) - B\mathcal{H}\mathbb{G}_B^{*\prime} a$ and $\mathbb{G}_B^* = \mathbb{N}_2^* B' B \mathcal{M}$, $\mathcal{H}$ and $\mathcal{M}$ are nonrandom matrices given in Lemma C.3, $\mathbb{N}_1^*$ and $\mathbb{N}_2^*$ are the first column and the last $K$ columns of $\mathbb{N}^*$.*

Theorem 5.1 implies that the distribution of $(\mathbb{G}_a, \mathbb{G}_B)$, which is equal to the distribution of $(\mathbb{G}_a^*, \mathbb{G}_B^*)$, can be well approximated by the distribution of $(\sqrt{NT/\omega_0}(\hat{a}^* - \hat{a}), \sqrt{NT/\omega_0}(\hat{B}^* - \hat{B}))$ conditional on the data, when $T = o(N)$, $J = o(\min\{N^{1/5}, N/T\})$ and $NTJ^{-2\kappa} = o(1)$. We reiterate that the result is appealing in the sense that it allows for the same weak dependence of $\{\varepsilon_{it}\}_{i \le N, t \le T}$ over $t$ as Theorem 4.3.

**Remark 5.2.** An alternative bootstrap estimator for $B$ is given by $\hat{B}^{**}$, whose columns are the eigenvectors corresponding to the first $K$ largest eigenvalues of $\tilde{Y}^* M_T \tilde{Y}^{*\prime}/T$. We notice that $\sqrt{NT/\omega_0}(\hat{B}^{**} - \hat{B})$ conditional on the data may fail to estimate the distribution of $\mathbb{G}_B$. The key part of the proof is to show that $\sqrt{NT}(\hat{B}^* - BH)$ and $\sqrt{NT}(\hat{B} - BH)$ share a similar asymptotic expansion. Specifically, we show

$$\left\|\sqrt{NT}(\hat{B} - BH) - \frac{1}{\sqrt{NT}} \sum_{t=1}^T Q_t^{-1} \Phi(Z_t)' \varepsilon_t (f_t - \bar{f})' B' B \mathcal{M}\right\|_F = O_p(\delta_{NT})$$

and

$$\left\| \sqrt{NT}(\hat{B}^* - BH) - \frac{1}{\sqrt{NT}} \sum_{t=1}^{T} Q_t^{-1} \Phi(Z_t)^{*\prime} \varepsilon_t (f_t - \bar{f})' B' B \mathcal{M} \right\|_F = O_p(\delta_{NT}),$$

where $\delta_{NT} = \sqrt{NT} J^{-\kappa} + \sqrt{TJ/N} + \sqrt{J\xi_J}(\log J/N)^{1/4}$. Let $\hat{F}^* \equiv \tilde{Y}^{*\prime}\hat{B}^{**}$ and $H^* \equiv (F' M_T \hat{F}^*)(\hat{F}^{*\prime} M_T \hat{F}^*)^{-1}$. Similarly, we can show

$$\left\| \sqrt{NT}(\hat{B}^{**} - BH^*) - \frac{1}{\sqrt{NT}} \sum_{t=1}^{T} Q_t^{-1} \Phi(Z_t)^{*\prime} \varepsilon_t (f_t - \bar{f})' B' B \mathcal{M} \right\|_F = O_p(\delta_{NT}).$$

Thus, $\sqrt{NT/\omega_0}(\hat{B}^{**} - \hat{B})$ conditional on the data may fail to estimate the distribution of $\mathbb{G}_B$, since $\sqrt{NT/\omega_0}(H^* - H)$ is not asymptotically negligible due to the relatively slow convergence rate of $\hat{F}$ and $\hat{F}^*$. Since $\hat{B}^{**} = \tilde{Y}^* M_T \hat{F}^* (\hat{F}^{*\prime} M_T \hat{F}^*)^{-1}$, it is important to use $\hat{F}$ rather than $\hat{F}^*$ in (15) to ensure that $\hat{B}^*$ and $\hat{B}$ share a common rotational transformation matrix and are centered around the same quantity $BH$. Therefore, it is important to use $\hat{F}$ rather than $\hat{F}^*$ in (15) to ensure the validity of the bootstrap.

**Significance test for $\alpha(\cdot)$ and $\beta(\cdot)$.** We can immediately use Theorems 4.3 and 5.1 to test whether $\phi_j(z_{it,m})$'s are jointly significant in $\beta(z_{it})$ for some given $j$'s and $m$'s, which is equivalent to whether certain rows of $BH$ are jointly zero. To see this, let us consider testing the null that the first row of $BH$ is zero. Let $\hat{b}_1'$ and $\hat{b}_1^{*\prime}$ be the first row of $\hat{B}$ and $\hat{B}^*$. The distribution of $NT\hat{b}_1'\hat{b}_1$ under the null can be estimated by the distribution of $NT(\hat{b}_1^* - \hat{b}_1)'(\hat{b}_1^* - \hat{b}_1)/\omega_0$ conditional on the data. Thus, we may construct the test as follows: reject the null if $NT\hat{b}_1'\hat{b}_1$ is greater than the $1-\alpha$ quantile of $NT(\hat{b}_1^* - \hat{b}_1)'(\hat{b}_1^* - \hat{b}_1)/\omega_0$ conditional on the data for $0 < \alpha < 1$. Note that we are not able to do significance test for each entry of $\beta(z_{it})$ due to the lack of identification, and we cannot use Theorems 4.3 and 5.1 to test whether $\beta(\cdot) = 0$ due to the full rank requirement in Assumption 4.2(i). Similarly, we may test whether $\phi_j(z_{it,m})$'s are significant in $\alpha(z_{it})$ for some given $j$'s and $m$'s. In addition, we may test whether $\alpha(\cdot) = 0$ by comparing $NT\hat{a}'\hat{a}$ with the $1-\alpha$ quantile of $NT(\hat{a}^* - \hat{a})'(\hat{a}^* - \hat{a})/\omega_0$ conditional on the data for $0 < \alpha < 1$.

## 5.2 Specification Test

To test for linearity of $\alpha(\cdot)$ and $\beta(\cdot)$, we develop a test by comparing their estimators under the null and the alternative. Specifically, we consider the following the hypothesis:

$$\begin{aligned} &\mathrm{H}_0 : \alpha(z_{it}) = \gamma' z_{it} \text{ and } \beta(z_{it}) = \Gamma' z_{it} \text{ for some } \gamma, \Gamma \text{ and all } i \le N, t \le T \text{ v.s.} \\ &\mathrm{H}_1 : \inf_{i \le N, t \le T} \inf_{\pi} E[|\alpha(z_{it}) - \pi' z_{it}|^2] > 0 \text{ or } \inf_{i \le N, t \le T} \inf_{\Pi} E[\|\beta(z_{it}) - \Pi' z_{it}\|^2] > 0. \end{aligned} \tag{16}$$

Estimators of $\alpha(\cdot)$ and $\beta(\cdot)$ under $\mathrm{H}_1$ are already given by $\hat{\alpha}(\cdot)$ and $\hat{\beta}(\cdot)$ in (10). Let $\vec{Y}_t \equiv (Z_t'Z_t)^{-1}Z_t'Y_t$, $\vec{Y} \equiv (\vec{Y}_1, \ldots, \vec{Y}_T)$, and $\bar{\vec{Y}} \equiv \sum_{t=1}^T \vec{Y}_t/T$. Estimators of $\alpha(z_{it})$ and $\beta(z_{it})$ under $\mathrm{H}_0$ are given by $\hat{\gamma}'z_{it}$ and $\hat{\Gamma}'z_{it}$, where $\hat{\Gamma} = \vec{Y}M_T\hat{F}(\hat{F}'M_T\hat{F})^{-1}$ and $\hat{\gamma} = \bar{\vec{Y}} - \hat{\Gamma}\hat{B}'\bar{\vec{Y}}$. Three remarks for $\hat{\gamma}$ and $\hat{\Gamma}$ are as follows. First, we use the unrestricted estimator $\hat{f}_t$ rather than a restricted estimator of $f_t$ by imposing $\mathrm{H}_0$ to ensure that $\hat{\Gamma}'z_{it}$ and $\hat{\beta}(z_{it})$ share a common rotational transformation matrix, which is important in justifying the validity of the test. Second, in $\hat{\gamma}$ we use $\hat{B}'\bar{\vec{Y}} = \sum_{t=1}^T \hat{f}_t/T$, which is an unrestricted estimator of $\bar{f}$, rather than the restricted estimator $(\hat{\Gamma}'\hat{\Gamma})^{-1}\hat{\Gamma}'\bar{\vec{Y}}$ under $\mathrm{H}_0$ to avoid the full rank requirement of $\Gamma$. Third, we note that using $\hat{f}_t$ does not cause efficiency loss in estimating $\Gamma$ and $\gamma$ , since $\hat{f}_t$ has attained the optimal rate as discussed after Theorem 4.2. Our test statistic is given by

$$\mathcal{S} = \frac{1}{J}\sum_{i=1}^N\sum_{t=1}^T |\hat{\gamma}'z_{it} - \hat{\alpha}(z_{it})|^2 + \frac{1}{J}\sum_{i=1}^N\sum_{t=1}^T \|\hat{\Gamma}'z_{it} - \hat{\beta}(z_{it})\|^2. \tag{17}$$

To obtain critical values, we use the bootstrap method. Let $\vec{Y}_t^* \equiv (Z_t^{*\prime}Z_t)^{-1}Z_t^{*\prime}Y_t$, $\vec{Y}^* \equiv (\vec{Y}_1^*, \ldots, \vec{Y}_T^*)$, and $\bar{\vec{Y}}^* \equiv \sum_{t=1}^T \vec{Y}_t^*/T$, where $Z_t^* = (z_{1t}w_1, \ldots, z_{Nt}w_N)'$. It is shown in Appendix E that under $\mathrm{H}_0$, $\mathcal{S} = \sum_{i=1}^N\sum_{t=1}^T |(\hat{\gamma} - \gamma)'z_{it} - (\hat{a} - a)'\phi(z_{it})|^2/J + \sum_{i=1}^N\sum_{t=1}^T \|(\hat{\Gamma} - \Gamma H)'z_{it} - (\hat{B} - BH)'\phi(z_{it})\|^2/J + o_p(J^{-1/2})$. Given this, we can estimate the null distribution of $\mathcal{S}$ by the distribution of

$$\begin{aligned}\mathcal{S}^* = &\frac{1}{J\omega_0}\sum_{i=1}^N\sum_{t=1}^T |(\hat{\gamma}^* - \hat{\gamma})'z_{it} - (\hat{a}^* - \hat{a})'\phi(z_{it})|^2 \\ &+ \frac{1}{J\omega_0}\sum_{i=1}^N\sum_{t=1}^T \|(\hat{\Gamma}^* - \hat{\Gamma})'z_{it} - (\hat{B}^* - \hat{B})'\phi(z_{it})\|^2\end{aligned} \tag{18}$$

conditional on the data, where $\hat{\Gamma}^* = \vec{Y}^*M_T\hat{F}(\hat{F}'M_T\hat{F})^{-1}$ and $\hat{\gamma}^* = \bar{\vec{Y}}^* - \hat{\Gamma}^*(\hat{B}^{*\prime}\hat{B}^*)^{-1}\hat{B}^{*\prime}\bar{\vec{Y}}^*$. For $0 < \alpha < 1$, let $c_{1-\alpha}$ be the $1 - \alpha$ quantile of $\mathcal{S}^*$ conditional on the data. Thus, we construct the test as follows: reject $\mathrm{H}_0$ if $\mathcal{S} > c_{1-\alpha}$.

To establish the validity of our test, we impose the following assumption.

**Assumption 5.2** (Specification test). *(i) There are positive constants $g_{\min}$ and $g_{\max}$ such that: with probability approaching one (as $N \to \infty$),*

$$g_{\min} < \min_{t\leq T}\lambda_{\min}(Z_t'Z_t/N) \leq \max_{t\leq T}\lambda_{\max}(Z_t'Z_t/N) < g_{\max},$$

*(ii)* $\max_{i\leq N, t\leq T} E[\|z_{it}\|^4] < \infty$*; (iii)* $\min_{i\leq N,t\leq T}\lambda_{\min}(E[z_{it}z_{it}']) > 0$*; (iv) with probability approaching one (as $N \to \infty$),*

$$g_{\min} < \min_{t\leq T}\lambda_{\min}(Z_t^{*\prime}Z_t/N) \leq \max_{t\leq T}\lambda_{\max}(Z_t^{*\prime}Z_t/N) < g_{\max};$$

*(v)* $\sup_z |\alpha(z)| < \infty$ *and* $\sup_z \|\beta(z)\| < \infty$.

Assumptions 5.2(i)-(iv) are analogous to Assumptions 4.1(i), 4.5(i), (ii) and 5.1(ii), respectively. When $z_{it}$ is included as a part of $\phi(z_{it})$, which is true in the case of polynomial basis functions, the formers are implied by the latter ones. In this case, Assumptions 5.2(i)-(iv) are redundant.

The following theorem, as the fifth result of the paper, demonstrates that our test has size control under $\mathrm{H}_0$ and is consistent under $\mathrm{H}_1$.

**Theorem 5.2.** *Suppose Assumptions 4.1-4.6, 5.1 and 5.2 hold. Let $\mathcal{S}$ be given in* (17), *and $c_{1-\alpha}$ be given after* (18) *for $0 < \alpha < 1$. Assume (i) $N \to \infty$; (ii) $T \to \infty$ or $T \geq K+1$ is finite; (iii) $J \to \infty$ with $J^2 \xi_J^2 \log J = o(N)$. In addition, assume $T = o(N)$, $J = o(\min\{N^{1/5}, N/T\})$ and $NTJ^{-2\kappa} = o(1)$. Then under* $\mathrm{H}_0$,

$$P(\mathcal{S} > c_{1-\alpha}) \to \alpha.$$

*Furthermore, under* $\mathrm{H}_1$,

$$P(\mathcal{S} > c_{1-\alpha}) \to 1.$$

The validity of the test does not require $T \to \infty$, as all above results do. It also holds when $T \to \infty$ but at a slower rate than $N$, which is usually true in asset pricing.

## 6 Determining the Number of Factors

We now address the problem of estimating the number of factors $K$. To solve the problem, we develop two estimators: one by maximizing the ratio of two adjacent eigenvalues by extending Ahn and Horenstein (2013) and Fan et al. (2016a) and another by counting the number of "large" eigenvalues similar to Bai and Ng (2002). To define the estimators, let $\lambda_k(\tilde{Y} M_T \tilde{Y}'/T)$ denote the $k$th largest eigenvalue of the $JM \times JM$ matrix $\tilde{Y} M_T \tilde{Y}'/T$. The first estimator of $K$ is given by

$$\hat{K} = \underset{1 \leq k \leq JM/2}{\arg\max} \frac{\lambda_k(\tilde{Y} M_T \tilde{Y}'/T)}{\lambda_{k+1}(\tilde{Y} M_T \tilde{Y}'/T)}. \tag{19}$$

Here, $\hat{K}$ is constrained to between 1 and $JM/2$. This is not restrictive, since we assume that $K \geq 1$ is fixed and $J \to \infty$ throughout the paper. The second estimator of $K$ is given by

$$\tilde{K} = \#\{1 \leq k \leq JM : \lambda_k(\tilde{Y} M_T \tilde{Y}'/T) \geq \lambda_{NT}\}, \tag{20}$$

where $\#A$ denotes the cardinality of $A$ and $0 < \lambda_{NT} \to 0$ is a tuning parameter.

Establishing the consistency of $\tilde{K}$ is straightforward. However, to establish the consistency of $\hat{K}$, we impose the following assumption.

**Assumption 6.1** (Determination of $K$). *(i)* $0 < \min_{t\leq T} \lambda_{\min}(E[\varepsilon_t \varepsilon_t'])\leq \max_{t\leq T} \lambda_{\max}$ $(E[\varepsilon_t \varepsilon_t']) < \infty$*; ii) there is* $0 < C_4 < \infty$ *such that*

$$\frac{1}{N^2T + T^2N} \sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{\ell=1}^{N} |cov(\varepsilon_{it}\varepsilon_{jt}, \varepsilon_{ks}\varepsilon_{\ell s})| < C_4.$$

Assumption 6.1(i) requires that the covariance matrix $E[\varepsilon_t\varepsilon_t']$ is bounded and nonsingular for all $t$. In particular, $\max_{t\leq T} \lambda_{\max}(E[\varepsilon_t\varepsilon_t']) < \infty$ allows for weak dependence of $\{\varepsilon_{it}\}_{i\leq N, t\leq T}$ across $i$. When $\{\varepsilon_{it}\}_{i\leq N,t\leq T}$ are independent across $i$, the condition is satisfied when $\min_{i\leq N,t\leq T} E[\varepsilon_{it}^2] > 0$ and $\max_{i\leq N,t\leq T} E[\varepsilon_{it}^2] < \infty$. Assumption 6.1(ii) allows for weak dependence of $\{\varepsilon_{it}\}_{i\leq N,t\leq T}$ over both $i$ and $t$, which is standard in the literature as Assumptions 4.3(iii), 4.5(iv) and 4.6(iii); see Proposition I.2 for its justification when $\{\varepsilon_{it}\}_{i\leq N,t\leq T}$ are independent across $i$. It is noted that Assumption 6.1 is different from Assumption 6.1 in Fan et al. (2016a) (which is required for the number of factor estimator); Assumption 6.1 appears less complicated.

**Theorem 6.1.** *(A) Suppose Assumptions 4.1-4.3, 4.5(i) and 6.1 hold. Let* $\hat{K}$ *be given in* (19)*. Assume (i)* $N \to \infty$*; (ii)* $T \to \infty$*; (iii)* $J \to \infty$ *with* $J = o(\min\{\sqrt{N}, \sqrt{T}\})$ *and* $NJ^{-2\kappa} = o(1)$*. Then*

$$P(\hat{K} = K) \to 1.$$

*(B) Suppose Assumptions 4.1-4.3 hold. Let* $\tilde{K}$ *be given in* (20)*. Assume (i)* $N \to \infty$*; (ii)* $T \to \infty$ *or* $T \geq K + 1$ *is finite; (iii)* $J \to \infty$ *with* $J = o(\sqrt{N})$*; (iv)* $0 < \lambda_{NT} \to 0$ *and* $\lambda_{NT} \min\{N/J, J^{2\kappa}\} \to \infty$*. Then*

$$P(\tilde{K} = K) \to 1.$$

This is the last result of the paper. Theorem 6.1 demonstrates that $\hat{K}$ and $\tilde{K}$ are consistent estimators of $K$. The consistency of $\hat{K}$ requires $T \to \infty$, as similar to Fan et al. (2016a). While the consistency of $\tilde{K}$ does not require $T \to \infty$ and Assumption 6.1, it relies on the choice of $\lambda_{NT}$. In practice, $\hat{K}$ is recommended when $T$ is large, and $\tilde{K}$ is recommended when $T$ is small.

**Remark 6.1.** To sum up, all the estimators except $\hat{K}$ and the inference procedures do not require $T \to \infty$, allow $z_{it}$ to vary over $t$ even in a nonstationary pattern, and are applicable for unbalanced panels. These attractive features shall make our estimators and inference procedures appealing to researchers in asset pricing, because panels of asset returns and characteristics are usually unbalanced, many assets have small $T$, and characteristics of assets are often time-varying. Furthermore, the small $T$ properties of our estimators and tests allow us to conduct rolling small sub-sample analyses to accommodate changing factor dynamics.

# 7 Simulation Studies

In this section, we conduct small-scale Monte Carlo simulations to examine the finite sample performance of our estimators and bootstrap inference methods.

We consider the following data generating process. For $\theta \geq 0$ and $\delta \geq 0$, we assume

$$\alpha(z_{it}) = \theta z_{it,1} + \delta z_{it,1}^2 \text{ and } \beta(z_{it}) = (z_{it,2} + \delta z_{it,2}^2, 2z_{it,3} + 2\delta z_{it,3}^2)', \tag{21}$$

so $K = 2$ and $M = 3$. Here, $\alpha(\cdot) = 0$ when $\theta = \delta = 0$, and both $\alpha(z_{it})$ and $\beta(z_{it})$ are nonlinear functions of $z_{it}$ when $\delta > 0$. Let

$$z_{it,1} = \sigma_t * u_{it,1}, z_{it,2} = 0.3z_{i(t-1),2} + u_{it,2} \text{ and } z_{it,3} = u_{it,3}, \tag{22}$$

where $u_{it} = (u_{it,1}, u_{it,2}, u_{it,3})'$ are i.i.d. $N(0, I_3)$ across both $i$ and $t$, $\sigma_t$'s are i.i.d. $U(1, 2)$ over $t$, and $z_{i0,2}$'s are i.i.d. $N(0, 1)$. Here, all the entries of $z_{it}$ are varying over $t$ but in different ways. Let $f_t = 0.3f_{t-1} + \eta_t$, where $\eta_t$'s are i.i.d. $N(0, I_K)$ and $f_0 \sim N(0, I_K/0.91)$. For $0 \leq \rho < 1$,

$$\varepsilon_t = \rho\varepsilon_{t-1} + e_t, \tag{23}$$

where $e_t$'s are i.i.d. $N(0, I_N)$ and $\varepsilon_0 \sim N(0, I_N/(1 - \rho^2))$. Note that $\rho$ is a measure of weak dependence of $\varepsilon_{it}$ over $t$. Here, $u_{it}$'s, $\sigma_t$'s, $z_{i0}$'s $\eta_t$'s, $f_0$, $e_t$'s and $\varepsilon_0$ are mutually independent. We generate $y_{it}$ according to the model (1).

To implement the regressed-PCA, we choose $\phi(z_{it}) = (z_{it,1}, z_{it,1}^2, z_{it,2}, z_{it,2}^2, z_{it,3}, z_{it,3}^2)'$, so $J = 2$ and the sieve approximation error is zero. We let $\lambda_{NT} = 1/\log(N)$ in the implementation of $\tilde{K}$. To implement the weighted bootstrap, we let $w_i$'s be i.i.d. random variables with the standard exponential distribution. First, we investigate the performance of $\hat{a}$, $\hat{B}$, $\hat{F}$, $\hat{K}$ and $\tilde{K}$ under different $(N, T)$'s. We run simulations for combinations of $\theta = 0, 0.1, 0.2, \ldots, 1$, $\delta = 0, 0.1, 0.2, \ldots, 0.5$ and $\rho = 0, 0.3, 0.7$. Here we report the results for $\theta = 1$, $\delta = 0.5$ and $\rho = 0, 0.3, 0.7$, while the results for other values of $\theta$ and $\delta$ are similar and available upon request. Specifically, we report the mean square errors of $\hat{a}$, $\hat{B}$ and $\hat{F}$ in Table 1, and the correct rates of $\hat{K}$ and $\tilde{K}$ in Table 2. Second, we investigate the performance of testing $\alpha(\cdot) = 0$ and linearity of $\alpha(\cdot)$ and $\beta(\cdot)$. To test $\alpha(\cdot) = 0$, we fix $\delta = 0$. Then $\alpha(\cdot) = 0$ if and only if $\theta = 0$. We report the rejection rates for $\theta = 0, 0.01, 0.02, \ldots, 0.1$ under $\rho = 0.3$, while the results under $\rho = 0$ and 0.7

are similar and available upon request. To test linearity of $\alpha(\cdot)$ and $\beta(\cdot)$, we fix $\theta = 1$. Then $\alpha(\cdot)$ and $\beta(\cdot)$ are linear if and only if $\delta = 0$. We report the rejection rates for $\delta = 0, 0.01, 0.02, \ldots, 0.1$ under $\rho = 0.3$, while the results under $\rho = 0$ and $0.7$ are similar and available upon request. The number of simulation replications is set to $1,000$ and the number of bootstrap draws is set to 499 for each replication.

The main findings are summarized as follows. First, as shown in Table 1, the mean square errors of $\hat{a}$, $\hat{B}$ and $\hat{F}$ decrease as $N$ increases, even for $T = 10$. This implies that the estimators are consistent as $N \to \infty$ even for small $T$. While increasing $T$ further reduces the mean square errors of $\hat{a}$ and $\hat{B}$, it does not reduce the mean square error of $\hat{F}$. Both findings are true regardless of whether $\rho = 0, 0.3$ or $0.7$, so the results allow for weak dependence of $\varepsilon_{it}$ over $t$. They are consistent with Theorems 4.1 and 4.2. Second, as shown in Table 2, $\hat{K}$ and $\tilde{K}$ can correctly estimate $K$ in all cases except for some cases when both $N$ and $T$ are small. This is consistent with Theorem 6.1. Third, both tests perform well. As shown in Table 3, the rejection rate of the first test may overreject $\alpha(\cdot) = 0$ a little bit for $\theta = 0$ when $N = 50$, but can quickly approach the significance level 5% when $N$ increases, even for $T = 10$. This implies that the test has size control as $N \to \infty$ even for small $T$. As $\theta$ increases, the rejection rate approaches one, even for $T = 10$. This implies that the test is consistent as $N \to \infty$ even for small $T$. We find that increasing $T$ may improve the power of the test (when $\theta > 0$, the rejection rate increases as $T$ increases for all $N$), but meanwhile it may hurt the size of the test (for example, when $\theta = 0$, the rejection rate increases as $T$ increases from 10 to 100 for $N = 200$.) This can be explained by the requirement $T = o(N)$ in Theorem 5.2 or underlying in Theorem 5.1. As shown in Table 4, the second test has a similar performance; the details are omitted. The findings of the second test are consistent with Theorem 5.2. To sum up, the performance of our estimators and bootstrap inference methods is encouraging for large $N$, even when $T$ is small.

# 8 Empirical Applications

A central question in empirical asset pricing is why different assets earn different average returns. While asset pricing theory attributes cross-sectional differences in asset returns to risk exposures, there is substantial evidence suggesting a role for mispricing captured by dependence of returns on asset characteristics, which suggests potential market inefficiency. Much of the debate centers around multi-factor models that aim to link average returns to factor loadings following Fama and French (1993), who pursue a portfolio-sorting approach to constructing asset pricing factors. Since their seminal paper, hundreds of factors have been proposed, collectively dubbed a "*factor zoo*" by Cochrane (2011) and further discussed by Harvey et al. (2016). While some of the factor models have an explicit justification based on economic theory, many implicitly

Table 1: Mean square errors of $\hat{a}$, $\hat{B}$ and $\hat{F}$ when $\theta = 1$ and $\delta = 0.5^{\dagger}$

| (N,T) | $\rho = 0$ | | | $\rho = 0.3$ | | | $\rho = 0.7$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\hat{a}$ | $\hat{B}$ | $\hat{F}$ | $\hat{a}$ | $\hat{B}$ | $\hat{F}$ | $\hat{a}$ | $\hat{B}$ | $\hat{F}$ |
| $(50, 10)$ | 0.0077 | 0.0154 | 0.0394 | 0.0088 | 0.0170 | 0.0435 | 0.0171 | 0.0295 | 0.0799 |
| $(100, 10)$ | 0.0034 | 0.0064 | 0.0168 | 0.0039 | 0.0071 | 0.0186 | 0.0075 | 0.0127 | 0.0336 |
| $(200, 10)$ | 0.0016 | 0.0030 | 0.0079 | 0.0018 | 0.0034 | 0.0087 | 0.0033 | 0.0058 | 0.0155 |
| $(500, 10)$ | 0.0006 | 0.0012 | 0.0030 | 0.0007 | 0.0013 | 0.0033 | 0.0013 | 0.0022 | 0.0060 |
| $(50, 50)$ | 0.0012 | 0.0022 | 0.0423 | 0.0014 | 0.0025 | 0.0466 | 0.0028 | 0.0049 | 0.0842 |
| $(100, 50)$ | 0.0005 | 0.0009 | 0.0184 | 0.0006 | 0.0010 | 0.0203 | 0.0012 | 0.0019 | 0.0365 |
| $(200, 50)$ | 0.0002 | 0.0004 | 0.0086 | 0.0003 | 0.0004 | 0.0095 | 0.0006 | 0.0008 | 0.0170 |
| $(500, 50)$ | 0.0000 | 0.0001 | 0.0033 | 0.0001 | 0.0002 | 0.0037 | 0.0002 | 0.0003 | 0.0065 |
| $(50, 100)$ | 0.0005 | 0.0010 | 0.0431 | 0.0006 | 0.0011 | 0.0473 | 0.0013 | 0.0024 | 0.0850 |
| $(100, 100)$ | 0.0002 | 0.0004 | 0.0187 | 0.0003 | 0.0004 | 0.0206 | 0.0006 | 0.0008 | 0.0370 |
| $(200, 100)$ | 0.0001 | 0.0002 | 0.0087 | 0.0001 | 0.0002 | 0.0096 | 0.0003 | 0.0003 | 0.0172 |
| $(500, 100)$ | 0.0000 | 0.0001 | 0.0034 | 0.0000 | 0.0001 | 0.0037 | 0.0001 | 0.0001 | 0.0066 |

$^{\dagger}$ The mean square errors of $\hat{a}$ , $\hat{B}$ and $\hat{F}$ are given by $\sum_{\ell=1}^{1000} \|\hat{a}^{(\ell)} - a\|^2/1000$, $\sum_{\ell=1}^{1000} \|\hat{B}^{(\ell)} - BH^{(\ell)}\|_F^2/1000$ and $\sum_{\ell=1}^{1000} \|\hat{F}^{(\ell)} - F(H^{(\ell)\prime})^{-1}\|_F^2/1000T$, where $\hat{a}^{(\ell)}$, $\hat{B}^{(\ell)}$ and $\hat{F}^{(\ell)}$ are estimators in the $\ell$th simulation replication, and $H^{(\ell)} \equiv (F'M_T\hat{F}^{(\ell)})(\hat{F}^{(\ell)\prime}M_T\hat{F}^{(\ell)})^{-1}$ is a rotational transformation matrix.

Table 2: Correct rates of $\hat{K}$ and $\tilde{K}$ when $\theta = 1$ and $\delta = 0.5$

| (N,T) | $\rho = 0$ | | $\rho = 0.3$ | | $\rho = 0.7$ | |
|---|---|---|---|---|---|---|
| | $\hat{K}$ | $\tilde{K}$ | $\hat{K}$ | $\tilde{K}$ | $\hat{K}$ | $\tilde{K}$ |
| $(50, 10)$ | 0.999 | 1.000 | 0.999 | 1.000 | 0.994 | 1.000 |
| $(100, 10)$ | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 1.000 |
| $(200, 10)$ | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $(500, 10)$ | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $(50, 50)$ | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $(100, 50)$ | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $(200, 50)$ | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $(500, 50)$ | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $(50, 100)$ | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $(100, 100)$ | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $(200, 100)$ | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $(500, 100)$ | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

Table 3: Rejection rates of testing $\alpha(\cdot) = 0$ when $\delta = 0$ and $\rho = 0.3^{\dagger}$

| (N,T) | $\theta$ | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0 | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 | 0.1 |
| $(50, 10)$ | 0.089 | 0.096 | 0.117 | 0.150 | 0.186 | 0.222 | 0.283 | 0.349 | 0.435 | 0.512 | 0.593 |
| $(100, 10)$ | 0.096 | 0.113 | 0.133 | 0.184 | 0.274 | 0.383 | 0.502 | 0.616 | 0.727 | 0.827 | 0.904 |
| $(200, 10)$ | 0.057 | 0.080 | 0.162 | 0.270 | 0.442 | 0.628 | 0.790 | 0.901 | 0.970 | 0.990 | 0.999 |
| $(500, 10)$ | 0.048 | 0.099 | 0.297 | 0.573 | 0.822 | 0.951 | 0.994 | 1.000 | 1.000 | 1.000 | 1.000 |
| $(50, 50)$ | 0.094 | 0.129 | 0.232 | 0.415 | 0.615 | 0.784 | 0.915 | 0.978 | 0.997 | 0.998 | 1.000 |
| $(100, 50)$ | 0.085 | 0.165 | 0.391 | 0.691 | 0.913 | 0.989 | 0.998 | 1.000 | 1.000 | 1.000 | 1.000 |
| $(200, 50)$ | 0.073 | 0.235 | 0.643 | 0.941 | 0.996 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $(500, 50)$ | 0.052 | 0.451 | 0.960 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $(50, 100)$ | 0.089 | 0.151 | 0.360 | 0.693 | 0.901 | 0.985 | 0.999 | 1.000 | 1.000 | 1.000 | 1.000 |
| $(100, 100)$ | 0.076 | 0.256 | 0.685 | 0.956 | 0.997 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $(200, 100)$ | 0.073 | 0.381 | 0.925 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $(500, 100)$ | 0.059 | 0.737 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

$^{\dagger}$ The significance level $\alpha = 5\%$.

Table 4: Rejection rates of testing linearity of $\alpha(\cdot)$ and $\beta(\cdot)$ when $\theta = 1$ and $\rho = 0.3^{\dagger}$

| (N,T) | $\delta$ | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0 | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 | 0.1 |
| $(50, 10)$ | 0.086 | 0.097 | 0.158 | 0.288 | 0.464 | 0.641 | 0.801 | 0.910 | 0.963 | 0.990 | 0.998 |
| $(100, 10)$ | 0.080 | 0.130 | 0.309 | 0.565 | 0.839 | 0.962 | 0.993 | 1.000 | 1.000 | 1.000 | 1.000 |
| $(200, 10)$ | 0.058 | 0.181 | 0.555 | 0.932 | 0.995 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $(500, 10)$ | 0.038 | 0.397 | 0.963 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $(50, 50)$ | 0.093 | 0.248 | 0.669 | 0.965 | 0.999 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $(100, 50)$ | 0.100 | 0.443 | 0.966 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $(200, 50)$ | 0.070 | 0.771 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $(500, 50)$ | 0.047 | 0.998 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $(50, 100)$ | 0.096 | 0.459 | 0.971 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $(100, 100)$ | 0.085 | 0.846 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $(200, 100)$ | 0.066 | 0.994 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $(500, 100)$ | 0.057 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

$^{\dagger}$ The significance level $\alpha = 5\%$.

rely on the idea that factors capture common variation in portfolio returns, thus appealing to arbitrage pricing theory and its extensions (Ross, 1976; Chamberlain and Rothschild, 1982; Connor and Korajczyk, 1986, 1988). Since implementing the latter requires knowledge of the conditional covariance matrix of returns, which is infeasible to estimate when $N$ is larger than $T$, most studies rely on stock characteristics to proxy for (imperfectly measured) factor exposures. However, this makes distinguishing between the two types of explanations virtually impossible, as exemplified by the "characteristics versus covariances" debate (Daniel and Titman, 1997). Our method is perfectly suited for resolving this debate, since it allows characteristics to simultaneously appear in both pricing errors and conditional covariances with unobserved common factors, which they also help recover.

We consider the following semiparametric characteristic-based factor model

$$r_{it} = \alpha(z_{i,t-1}) + \beta(z_{i,t-1})' f_t + \varepsilon_{it}, i = 1, \ldots, N, t = 1, \ldots, T, \tag{24}$$

where $r_{it}$ is the excess return of asset $i$ (e.g., stock $i$) in time period $t$, $z_{i,t-1}$ is a vector of characteristics in time period $t-1$, $f_t$ is a $K \times 1$ vector of unobserved latent factors, the pricing error (i.e., $\alpha(z_{i,t-1})$) and the risk exposures to factors (i.e., $\beta(z_{i,t-1})$) are nonparametric functions of characteristics (i.e., $z_{i,t-1}$). The model falls into the general framework of model (1), where we need to interpret $z_{it}$ as characteristics in time period $t-1$. This model provides a unified approach for studying the cross section of asset returns that nests the characteristic-based model and the risk-based model. The modelling of the pricing error and the risk exposures not only provides a way to disentangle the alpha versus beta explanations, but also allows us to estimate a model for a large set of individual stocks. In addition, we do not need to rely on ex ante knowledge to pre-specify the latent factors. Distinct from the models in Connor and Linton (2007), Connor et al. (2012), Kelly et al. (2019), and Kim et al. (2020), we allow for time-varying characteristics, nonzero pricing error, nonlinearity of $\alpha(\cdot)$ and $\beta(\cdot)$, and unknown number of factors. These are not intended to complicate the analysis, but are crucial. For example, as illustrated in Section 4.1, failure to take into account the time-varying features of characteristics or linear specifications of $\alpha(\cdot)$ and $\beta(\cdot)$ may result in misleading estimation of factors and the number of factors.

## 8.1 Data and Methodology

We use the same dataset used in Kelly et al. (2019), which is originally from Freyberger et al. (2020). The data set contains monthly returns of $12,813$ individual stocks and 36 associated characteristics with sample periods from July, 1962 to May, 2014. The data is in the form of an unbalanced panel, for which our methods are applicable. See the above two papers for the detailed descriptions of the data. For ease of comparison, we also use the same 36 characteristics as those authors. By following the same procedure in Kelly et al. (2019), we transform the values of the characteristics to relative ranking values with range $[-0.5, 0.5]$. This can make the contributions of individual characteristic in pricing error and risk exposures comparable, and can further avoid the distorting effects from the outliers. To satisfy the large $N$ requirement, we select the sample period with at least $1,000$ individual stocks that have observations on both returns and the 36 characteristics, which is different from the case with at least 100 individual stocks in Kelly et al. (2019). This yields a sample from September, 1968 to May, 2014.

To estimate the model, we implement the regressed-PCA by choosing the basis functions $\phi(z_{it})$ as $(1, z_{it}')'$ and linear B-splines of $z_{it}$. Using $(1, z_{it}')'$ leads to linear specifications of $\alpha(\cdot)$ and $\beta(\cdot)$, while using linear B-splines of $z_{it}$ leads to nonlinear specifications of $\alpha(\cdot)$ and $\beta(\cdot)$, where $\alpha(\cdot)$ and $\beta(\cdot)$ are continuous piecewise linear functions.[7] To estimate the number of factors $K$, we use $\hat{K}$ in (19). To implement the weighted bootstrap, we let $w_i$'s be i.i.d. random variables with the standard exponential distribution. To implement the tests of $\alpha(\cdot) = 0$ and linearity of $\alpha(\cdot)$ and $\beta(\cdot)$, we set the number of bootstrap draws to 499.

### 8.2 Empirical Results

In order to evaluate the performance of the regressed-PCA, we compute several measures of fit. First, we calculate Fama-MacBeth cross sectional regression $R^2_{\tilde{Y}}$, which captures the variation in individual stock returns explained by "managed portfolios" constructed from the sieve functions of characteristics. Next, we report the panel regression $R^2_K$ which captures the variations of these managed portfolios explained by different sets of extracted factors. Then, we consider the following three types of $R^2$ measures that directly speak to the ability of the factor models to explain the cross-section of individual stock returns. The first one is total $R^2$ as used in Kelly et al. (2019).The second one measures the cross-sectional average of time series $R^2$ across all stocks, which reflects the ability of the extracted factors to capture common variation in asset returns. The third measures the time series average of cross-sectional goodness of fit measures. As such, it corresponds to the $R^2$ of the Fama-MacBeth cross-sectional regression, and is the one of interest for evaluating the model's ability to explain the cross-section of average returns. Fama-MacBeth regression slopes can be interpreted as returns on pure-play characteristic portfolios (corresponding to $\alpha(\cdot)$) and factor-mimicking portfolios (for $\beta(\cdot)$) - i.e. portfolios that have unit loading on one characteristic/factor and zero on all the others). Thus, the Fama-MacBeth $R^2$ reflects how much ex post variation in returns these portfolios can explain, as pointed out by Fama (1976) and emphasized by

[7]The one dimensional linear B-spline $\{\psi_j(z)\}_{j=1}^J$ is defined on a set of consecutive equidistant knots: $\{z_1, ..., z_{J+1}\}$. For $j < J$, $\psi_j(z) = (z - z_j)/(z_{j+1} - z_j)$ on $(z_j, z_{j+1}]$, $\psi_j(z) = (z_{j+2} - z)/(z_{j+2} - z_{j+1})$ on $(z_{j+1}, z_{j+2}]$, and 0 elsewhere. For $j = J$, $\psi_j(z) = (z - z_j)/(z_{j+1} - z_j)$ on $(z_j, z_{j+1}]$ and 0 elsewhere.

Lewellen (2015).

$$R^2 = 1 - \frac{\sum_{i,t}[r_{it} - \hat{\alpha}(z_{i,t-1}) - \hat{\beta}(z_{i,t-1})'\hat{f}_t]^2}{\sum_{i,t} r_{i,t}^2}, \tag{25}$$

$$R_{T,N}^2 = 1 - \frac{1}{N}\sum_i \frac{\sum_t[r_{it} - \hat{\alpha}(z_{i,t-1}) - \hat{\beta}(z_{i,t-1})'\hat{f}_t]^2}{\sum_t r_{i,t}^2}, \tag{26}$$

$$R_{N,T}^2 = 1 - \frac{1}{T}\sum_t \frac{\sum_i[r_{it} - \hat{\alpha}(z_{i,t-1}) - \hat{\beta}(z_{i,t-1})'\hat{f}_t]^2}{\sum_i r_{i,t}^2}. \tag{27}$$

Second, we consider a version of these goodness-of-fit measures that zero in on the role of factors in explaining the time-series as well as the cross-section of stock returns, by excluding the conditional intercepts:

$$R_f^2 = 1 - \frac{\sum_{i,t}[r_{it} - \hat{\beta}(z_{i,t-1})'\hat{f}_t]^2}{\sum_{i,t} r_{i,t}^2}, \tag{28}$$

$$R_{f,T,N}^2 = 1 - \frac{1}{N}\sum_i \frac{\sum_t[r_{it} - \hat{\beta}(z_{i,t-1})'\hat{f}_t]^2}{\sum_t r_{i,t}^2}, \tag{29}$$

$$R_{f,N,T}^2 = 1 - \frac{1}{T}\sum_t \frac{\sum_i[r_{it} - \hat{\beta}(z_{i,t-1})'\hat{f}_t]^2}{\sum_i r_{i,t}^2}. \tag{30}$$

Third, we assess the out-of-sample prediction. For $t \geq 120$, we use the data through $t-1$ to implement the regressed-PCA and obtain estimators, say $\hat{\alpha}_{t-1}(\cdot)$, $\hat{\beta}_{t-1}(\cdot)$, $\hat{F}_{t-1}' \equiv (\hat{f}_1^{(t-1)}, \ldots, \hat{f}_{t-1}^{(t-1)})$; and then compute the out-of-sample prediction of $r_{it}$ as $\hat{\alpha}_{t-1}(z_{i,t-1}) + \hat{\beta}_{t-1}(z_{i,t-1})'\hat{\lambda}_t$, where $\hat{\lambda}_t = \sum_{s\leq t-1} \hat{f}_s^{(t-1)}/(t-1)$, that is, the average of factor estimators through $t-1$. We can define three types of out-of-sample predictive $R^2$'s analogously by replacing $\hat{\alpha}(\cdot)$, $\hat{\beta}(\cdot)$ and $\hat{f}_t$ with $\hat{\alpha}_{t-1}(\cdot)$, $\hat{\beta}_{t-1}(\cdot)$ and $\hat{\lambda}_t$,

$$R_O^2 = 1 - \frac{\sum_{i,t\geq 120}[r_{it} - \hat{\alpha}_{t-1}(z_{i,t-1}) - \hat{\beta}_{t-1}(z_{i,t-1})'\hat{\lambda}_t]^2}{\sum_{i,t\geq 120} r_{i,t}^2}, \tag{31}$$

$$R_{T,N,O}^2 = 1 - \frac{1}{N}\sum_i \frac{\sum_{t\geq 120}[r_{it} - \hat{\alpha}_{t-1}(z_{i,t-1}) - \hat{\beta}_{t-1}(z_{i,t-1})'\hat{\lambda}_t]^2}{\sum_{t\geq 120} r_{i,t}^2}, \tag{32}$$

$$R_{N,T,O}^2 = 1 - \frac{1}{T-120}\sum_{t\geq 120} \frac{\sum_i[r_{it} - \hat{\alpha}_{t-1}(z_{i,t-1}) - \hat{\beta}_{t-1}(z_{i,t-1})'\hat{\lambda}_t]^2}{\sum_i r_{i,t}^2}. \tag{33}$$

Finally, we construct an arbitrage portfolio based on a pure-alpha strategy and evaluate its performance. By (9) and Theorem 4.1, it is easy to see that $\tilde{Y}_t'\hat{a} \xrightarrow{p} \|a\|^2$ for each $t$ as $N \to \infty$. This allows us to construct an arbitrage portfolio based on an estimate of $a$. For $t \geq 120$, we use the data through $t-1$ to implement the regressed-PCA and obtain an estimator of $a$, say $\hat{a}_{t-1}$; and then compute the portfolio weights by $\omega_t = \Phi(Z_{t-1})(\Phi(Z_{t-1})'\Phi(Z_{t-1}))^{-1}\hat{a}_{t-1}$ and the excess return of the portfolio by $R_t'\omega_t$, where $R_t = (r_{1t}, \ldots, r_{Nt})'$. We evaluate the annualized Sharpe ratio of this portfolio.

Table 5: Results under linear specifications of $\alpha(\cdot)$ and $\beta(\cdot)$ with 36 characteristics†

| Unrestricted ($\alpha(\cdot) \neq 0$) | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $K$ | $R_K^2$ | $R^2$ | $R^2_{T,N}$ | $R^2_{N,T}$ | $R^2_f$ | $R^2_{f,T,N}$ | $R^2_{f,N,T}$ | $R^2_O$ | $R^2_{T,N,O}$ | $R^2_{N,T,O}$ | Mean | Std | SR |
| 1* | 26.55 | 2.54 | 1.37 | 0.36 | 2.07 | 0.59 | 0.11 | 0.54 | 0.64 | 0.21 | 1.72 | 0.54 | 3.18 |
| 2 | 36.42 | 4.52 | 2.43 | 1.76 | 4.08 | 1.75 | 1.37 | 0.54 | 0.64 | 0.21 | 1.74 | 0.52 | 3.36 |
| 3 | 45.03 | 5.70 | 3.70 | 2.70 | 5.24 | 2.95 | 2.31 | 0.54 | 0.64 | 0.21 | 1.77 | 0.50 | 3.56 |
| 4 | 52.55 | 11.69 | 8.55 | 9.27 | 11.28 | 7.92 | 8.69 | 0.54 | 0.64 | 0.21 | 1.77 | 0.47 | 3.74 |
| 5 | 58.65 | 11.90 | 8.73 | 9.48 | 11.49 | 7.99 | 8.90 | 0.54 | 0.64 | 0.21 | 1.70 | 0.44 | 3.84 |
| 6 | 64.20 | 13.90 | 10.30 | 11.80 | 13.53 | 9.79 | 11.24 | 0.54 | 0.64 | 0.21 | 1.68 | 0.44 | 3.78 |
| 7 | 69.15 | 15.59 | 12.23 | 13.76 | 15.23 | 11.71 | 13.23 | 0.54 | 0.64 | 0.21 | 1.63 | 0.44 | 3.73 |
| 8 | 72.84 | 15.93 | 12.59 | 13.98 | 15.56 | 12.00 | 13.44 | 0.54 | 0.64 | 0.21 | 1.61 | 0.42 | 3.79 |
| 9 | 76.26 | 16.08 | 12.67 | 14.19 | 15.72 | 12.15 | 13.64 | 0.54 | 0.64 | 0.21 | 1.61 | 0.42 | 3.80 |
| 10 | 79.15 | 16.23 | 12.82 | 14.35 | 15.87 | 12.34 | 13.80 | 0.54 | 0.64 | 0.21 | 1.60 | 0.42 | 3.82 |
| Restricted ($\alpha(\cdot) = 0$) | | | | | | | | | | | | | |
| $K$ | $R_K^2$ | $R^2$ | $R^2_{T,N}$ | $R^2_{N,T}$ | $R^2_f$ | $R^2_{f,T,N}$ | $R^2_{f,N,T}$ | $R^2_O$ | $R^2_{T,N,O}$ | $R^2_{N,T,O}$ | Mean | Std | SR |
| 1* | 26.62 | NA | NA | NA | 2.14 | 0.58 | 0.06 | 0.20 | 0.09 | 0.07 | NA | NA | NA |
| 2 | 36.48 | NA | NA | NA | 4.18 | 1.72 | 1.37 | 0.28 | 0.34 | 0.02 | NA | NA | NA |
| 3 | 45.10 | NA | NA | NA | 5.32 | 2.98 | 2.30 | 0.26 | 0.31 | 0.01 | NA | NA | NA |
| 4 | 52.62 | NA | NA | NA | 11.45 | 8.03 | 8.86 | 0.31 | 0.39 | -0.01 | NA | NA | NA |
| 5 | 58.72 | NA | NA | NA | 11.69 | 8.18 | 9.10 | 0.36 | 0.47 | -0.04 | NA | NA | NA |
| 6 | 64.28 | NA | NA | NA | 13.85 | 10.06 | 11.58 | 0.38 | 0.47 | -0.11 | NA | NA | NA |
| 7 | 69.26 | NA | NA | NA | 15.20 | 11.71 | 13.17 | 0.40 | 0.50 | -0.13 | NA | NA | NA |
| 8 | 72.98 | NA | NA | NA | 15.53 | 11.99 | 13.44 | 0.41 | 0.53 | -0.13 | NA | NA | NA |
| 9 | 76.40 | NA | NA | NA | 15.73 | 12.15 | 13.68 | 0.40 | 0.53 | -0.08 | NA | NA | NA |
| 10 | 79.29 | NA | NA | NA | 15.90 | 12.37 | 13.85 | 0.41 | 0.51 | -0.06 | NA | NA | NA |

† $K$: the number of factor specified (* denotes the estimated one by our methods); Fama-MacBeth cross sectional regression $R^2$: $R^2_{\bar{Y}} = 20.89\%$; $R^2_K$ meassures the variations of managed portfolios captured by different numbers of factors from PCA; $R^2$, $R^2_{T,N}$, $R^2_{N,T}$: various in-sample $R^2$'s (%), see (25)-(27); $R^2_f$, $R^2_{f,T,N}$, $R^2_{f,N,T}$: various in-sample $R^2$'s (%) without $\alpha$, see (28)-(30); $R^2_O$, $R^2_{T,N,O}$, $R^2_{N,T,O}$: various out-sample predictive $R^2$'s (%), see (31)-(33); Mean: out-of-sample annualized means of the pure-alpha arbitrage strategy(%); Std: out-of-sample annualized standard deviations of the pure-alpha arbitrage strategy(%); SR: out-of-sample annualized Sharpe ratios of the pure-alpha arbitrage strategy.

Table 6: Results under continuous piecewise linear specifications of $\alpha(\cdot)$ and $\beta(\cdot)$ with 18 characteristics and one internal knot[†]

| Unrestricted ($\alpha(\cdot) \neq 0$) | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $K$ | $R^2_K$ | $R^2$ | $R^2_{T,N}$ | $R^2_{N,T}$ | $R^2_f$ | $R^2_{f,T,N}$ | $R^2_{f,N,T}$ | $R^2_O$ | $R^2_{T,N,O}$ | $R^2_{N,T,O}$ | Mean | Std | SR |
| 1 | 41.61 | 5.94 | 3.47 | 3.60 | 5.52 | 2.99 | 3.11 | 0.59 | 0.64 | 0.28 | 2.46 | 0.69 | 3.54 |
| 2* | 59.05 | 9.56 | 6.17 | 6.91 | 9.18 | 5.67 | 6.33 | 0.59 | 0.64 | 0.28 | 2.39 | 0.57 | 4.22 |
| 3 | 64.47 | 10.42 | 6.78 | 7.96 | 10.03 | 6.27 | 7.38 | 0.59 | 0.64 | 0.28 | 2.36 | 0.57 | 4.17 |
| 4 | 68.99 | 13.83 | 10.26 | 11.52 | 13.40 | 9.80 | 10.90 | 0.59 | 0.64 | 0.28 | 2.19 | 0.53 | 4.12 |
| 5 | 72.33 | 14.32 | 10.73 | 11.98 | 13.91 | 10.29 | 11.38 | 0.59 | 0.64 | 0.28 | 2.19 | 0.51 | 4.26 |
| 6 | 75.35 | 14.71 | 10.97 | 12.40 | 14.29 | 10.55 | 11.86 | 0.59 | 0.64 | 0.28 | 1.95 | 0.49 | 3.96 |
| 7 | 77.63 | 15.28 | 11.78 | 12.99 | 14.84 | 11.27 | 12.42 | 0.59 | 0.64 | 0.28 | 1.90 | 0.48 | 3.93 |
| 8 | 80.83 | 15.44 | 11.98 | 13.16 | 15.10 | 11.59 | 12.73 | 0.59 | 0.64 | 0.28 | 1.73 | 0.47 | 3.66 |
| 9 | 82.88 | 15.84 | 12.33 | 13.49 | 15.48 | 11.87 | 13.05 | 0.59 | 0.64 | 0.28 | 1.31 | 0.40 | 3.26 |
| 10 | 85.61 | 16.39 | 12.89 | 13.93 | 15.71 | 11.80 | 13.14 | 0.59 | 0.64 | 0.28 | 0.88 | 0.28 | 3.14 |
| Restricted ($\alpha(\cdot) = 0$) | | | | | | | | | | | | | |
| $K$ | $R^2_K$ | $R^2$ | $R^2_{T,N}$ | $R^2_{N,T}$ | $R^2_f$ | $R^2_{f,T,N}$ | $R^2_{f,N,T}$ | $R^2_O$ | $R^2_{T,N,O}$ | $R^2_{N,T,O}$ | Mean | Std | SR |
| 1 | 41.75 | NA | NA | NA | 5.61 | 3.00 | 3.14 | 0.30 | 0.34 | -0.12 | NA | NA | NA |
| 2* | 59.20 | NA | NA | NA | 9.14 | 5.56 | 6.26 | 0.34 | 0.30 | -0.38 | NA | NA | NA |
| 3 | 65.00 | NA | NA | NA | 9.80 | 6.24 | 7.12 | 0.60 | 0.76 | 0.29 | NA | NA | NA |
| 4 | 70.17 | NA | NA | NA | 10.79 | 7.23 | 8.37 | 0.60 | 0.80 | 0.19 | NA | NA | NA |
| 5 | 74.44 | NA | NA | NA | 14.28 | 10.57 | 11.98 | 0.52 | 0.66 | 0.29 | NA | NA | NA |
| 6 | 77.39 | NA | NA | NA | 14.58 | 10.88 | 12.18 | 0.52 | 0.63 | 0.22 | NA | NA | NA |
| 7 | 80.12 | NA | NA | NA | 14.91 | 11.07 | 12.61 | 0.53 | 0.58 | 0.22 | NA | NA | NA |
| 8 | 82.36 | NA | NA | NA | 15.43 | 11.93 | 13.17 | 0.54 | 0.57 | 0.27 | NA | NA | NA |
| 9 | 84.34 | NA | NA | NA | 15.80 | 12.28 | 13.45 | 0.53 | 0.54 | 0.27 | NA | NA | NA |
| 10 | 86.23 | NA | NA | NA | 15.94 | 12.37 | 13.59 | 0.53 | 0.54 | 0.27 | NA | NA | NA |

[†] $K$: the number of factor specified (* denotes the estimated one by our methods); Fama-MacBeth cross sectional regression $R^2$: $R^2_{\tilde{Y}} = 21.11\%$; $R^2_K$ meassures the variations of managed portfolios captured by different numbers of factors from PCA; $R^2$, $R^2_{T,N}$, $R^2_{N,T}$: various in-sample $R^2$'s (%), see (25)-(27); $R^2_f$, $R^2_{f,T,N}$, $R^2_{f,N,T}$: various in-sample $R^2$'s (%) without $\alpha$, see (28)-(30); $R^2_O$, $R^2_{T,N,O}$, $R^2_{N,T,O}$: various out-sample predictive $R^2$'s (%), see (31)-(33); Mean: out-of-sample annualized means of the pure-alpha arbitrage strategy(%); Std: out-of-sample annualized standard deviations of the pure-alpha arbitrage strategy(%); SR: out-of-sample annualized Sharpe ratios of the pure-alpha arbitrage strategy.

Table 7: Results under continuous piecewise linear specifications of $\alpha(\cdot)$ and $\beta(\cdot)$ with 12 characteristics and two internal knots[†]

| Unrestricted ($\alpha(\cdot) \neq 0$) | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $K$ | $R_K^2$ | $R^2$ | $R_{T,N}^2$ | $R_{N,T}^2$ | $R_f^2$ | $R_{f,T,N}^2$ | $R_{f,N,T}^2$ | $R_O^2$ | $R_{T,N,O}^2$ | $R_{N,T,O}^2$ | Mean | Std | SR |
| 1 | 42.78 | 5.57 | 2.98 | 3.32 | 5.19 | 2.54 | 2.83 | 0.57 | 0.57 | 0.27 | 3.29 | 0.99 | 3.33 |
| 2* | 61.36 | 9.56 | 5.97 | 6.87 | 9.18 | 5.51 | 6.26 | 0.57 | 0.57 | 0.27 | 3.01 | 0.80 | 3.78 |
| 3 | 67.77 | 10.59 | 6.65 | 7.88 | 10.20 | 6.15 | 7.29 | 0.57 | 0.57 | 0.27 | 2.94 | 0.80 | 3.69 |
| 4 | 72.86 | 13.62 | 10.09 | 11.35 | 13.17 | 9.64 | 10.67 | 0.57 | 0.57 | 0.27 | 2.97 | 0.78 | 3.81 |
| 5 | 76.92 | 14.14 | 10.43 | 12.01 | 13.73 | 10.01 | 11.48 | 0.57 | 0.57 | 0.27 | 2.98 | 0.76 | 3.91 |
| 6 | 80.63 | 14.94 | 11.45 | 12.75 | 14.42 | 10.51 | 12.05 | 0.57 | 0.57 | 0.27 | 1.51 | 0.41 | 3.73 |
| 7 | 84.29 | 15.17 | 11.59 | 12.94 | 14.76 | 10.77 | 12.44 | 0.57 | 0.57 | 0.27 | 1.01 | 0.33 | 3.09 |
| 8 | 87.42 | 15.45 | 11.87 | 13.23 | 15.26 | 11.47 | 12.98 | 0.57 | 0.57 | 0.27 | 0.73 | 0.22 | 3.36 |
| 9 | 89.11 | 16.33 | 12.68 | 13.94 | 16.16 | 12.31 | 13.72 | 0.57 | 0.57 | 0.27 | 0.71 | 0.20 | 3.62 |
| 10 | 90.72 | 16.54 | 12.91 | 14.17 | 16.38 | 12.54 | 13.95 | 0.57 | 0.57 | 0.27 | 0.69 | 0.18 | 3.90 |
| Restricted ($\alpha(\cdot) = 0$) | | | | | | | | | | | | | |
| $K$ | $R_K^2$ | $R^2$ | $R_{T,N}^2$ | $R_{N,T}^2$ | $R_f^2$ | $R_{f,T,N}^2$ | $R_{f,N,T}^2$ | $R_O^2$ | $R_{T,N,O}^2$ | $R_{N,T,O}^2$ | Mean | Std | SR |
| 1 | 42.95 | NA | NA | NA | 5.34 | 2.59 | 2.90 | 0.32 | 0.34 | -0.10 | NA | NA | NA |
| 2 | 61.58 | NA | NA | NA | 9.12 | 5.45 | 6.15 | 0.33 | 0.21 | -0.56 | NA | NA | NA |
| 3 | 68.02 | NA | NA | NA | 10.15 | 6.08 | 7.25 | 0.62 | 0.68 | 0.16 | NA | NA | NA |
| 4 | 74.08 | NA | NA | NA | 10.77 | 6.99 | 7.98 | 0.57 | 0.65 | 0.24 | NA | NA | NA |
| 5 | 78.98 | NA | NA | NA | 14.15 | 10.49 | 11.93 | 0.55 | 0.57 | 0.23 | NA | NA | NA |
| 6 | 82.66 | NA | NA | NA | 14.43 | 10.68 | 12.32 | 0.56 | 0.53 | 0.23 | NA | NA | NA |
| 7 | 85.44 | NA | NA | NA | 14.93 | 11.31 | 12.76 | 0.55 | 0.55 | 0.25 | NA | NA | NA |
| 8 | 87.85 | NA | NA | NA | 15.37 | 11.78 | 13.13 | 0.56 | 0.54 | 0.27 | NA | NA | NA |
| 9 | 89.53 | NA | NA | NA | 16.28 | 12.57 | 13.85 | 0.56 | 0.52 | 0.27 | NA | NA | NA |
| 10 | 91.13 | NA | NA | NA | 16.49 | 12.78 | 14.08 | 0.57 | 0.55 | 0.27 | NA | NA | NA |

[†] $K$: the number of factor specified (* denotes the estimated one by our methods); Fama-MacBeth cross sectional regression $R^2$: $R_{\tilde{Y}}^2 = 20.72\%$; $R_K^2$ meassures the variations of managed portfolios captured by different numbers of factors from PCA; $R^2$, $R_{T,N}^2$, $R_{N,T}^2$: various in-sample $R^2$'s (%), see (25)-(27); $R_f^2$, $R_{f,T,N}^2$, $R_{f,N,T}^2$: various in-sample $R^2$'s (%) without $\alpha$, see (28)-(30); $R_O^2$, $R_{T,N,O}^2$, $R_{N,T,O}^2$: various out-sample predictive $R^2$'s (%), see (31)-(33); Mean: out-of-sample annualized means of the pure-alpha arbitrage strategy(%); Std: out-of-sample annualized standard deviations of the pure-alpha arbitrage strategy(%); SR: out-of-sample annualized Sharpe ratios of the pure-alpha arbitrage strategy.

We report the results for three specifications of $\alpha(\cdot)$ and $\beta(\cdot)$ by fixing the dimension of sieve series for both linear and nonlinear cases. In Table 5, we consider linear specifications by letting $\phi(z_{it}) = (1, z_{it}')'$. In Table 6, we consider continuous piecewise linear specifications with 18 characteristics with one internal knot by letting $\phi(z_{it})$ be linear B-splines of $z_{it}$, where we split $[-0.5, 0.5]$ into two equal length intervals. We also refer it to as nonlinear specificathis ions with 18 characteristics. In Table 7, we further consider continuous piecewise linear specifications with 12 characteristics with two internal knots by letting $\phi(z_{it})$ be linear B-splines of $z_{it}$, where we split $[-0.5, 0.5]$ into three equal length intervals. We also refer it to as nonlinear specifications with 12 characteristics. For the selection of 12 or 18 characteristics, we choose the significant characteristics based on both (Kelly et al., 2019) and our estimation results from linear model, where the list of variables can be found in Tables 10 and 11, separately. For each case, we report the results for imposing $\alpha(\cdot) = 0$ and not imposing $\alpha(\cdot) = 0$, respectively. In addition to estimating the number of factors, we report the results for $K = 1, \ldots, 10$.

The main findings are summarized as follows. First, formal tests select one factor in the linear cases and two factors in the nonlinear cases, which is in contrast to the arguments of Kelly et al. (2019) that five factors are needed. Second, the out-of-sample $R_O^2$ based on our estimated one or two factor model with nonzero $\alpha(\cdot)$ is 0.54 in the linear specification, 0.59 and 0.57 in the two nonlinear specifications, all of which are comparable to 0.60 in Kelly et al. (2019)'s linear specifications with five factors. We notice that the total in-sample $R^2$'s from this estimated single factor models is smaller than Kelly et al. (2019)'s. This is not surprising, since the objective of their IPCA estimation is (essentially) maximizing total (in-sample) $R^2$. The nonlinear specifications with nonzero $\alpha(\cdot)$ give the higher $R^2_{N,T,O}$ than the linear case and nonlinear cases with zero $\alpha(\cdot)$. Third, by increasing the number of factors, we can impove the in-sample fit, since all three in-sample $R^2$'s increase with $K$. However, increasing the number of factors does not necessarily improve the out-of-sample fit of the model, since factor betas simply soaks up the variation that is otherwise captured by alpha.[8] In contrast, when alpha is restricted to be zero, increasing the number of factors does improve the out-of sample fit for the cross-section of expected returns, $R^2_{N,T,O}$, while both the total and the time-series out of sample fit measures, $R^2_O$ and $R^2_{O,T,N}$, are hump-shaped in the number of factors, peaking around three or four factors, depending on specification (naturally, these measures are small since they are not meant to capture month-to-month variation in returns by design). Fourth, compared to the linear model, nonlinear models improve in-sample fitting $R^2$ significantly and out-of-sample prediction $R^2$ slightly, meanwhile, the estimation results from nonlinear models are quite close. Both the better performance and robustness show the advantage of of nonlinear model estimation based on linear B-

[8]Formally, this is because $\hat{\alpha}(z_{i,t-1}) + \hat{\beta}(z_{i,t-1})'\hat{F}1_T/T = \phi(z_{i,t-1})'(\hat{a} - \hat{B}\hat{F}1_T/T) = \phi(z_{i,t-1})'\bar{\tilde{Y}}$, which does not depend on $K$, where $\bar{\tilde{Y}}$ is the average of the coefficient estimates from the first-step Fama-MacBetch regressions (also see (9)).

splines. Finally, we provide more associated empirical results in the Online Appendix.

We further use our tests to examine whether factor models explain the cross-section of average stock returns (i.e. $\alpha(\cdot) = 0$) as well as whether $\alpha(\cdot)$ and $\beta(\cdot)$ functions are linear in characteristics. First, we find strong evidence to reject the null hypothesis of $\alpha(\cdot) = 0$ in our estimated factor models whether we consider the one- or two-factor models that are selected by our formal tests, or indeed any number of factors between one and ten (we do not report the p-values here to save space, suffice it to say that in all cases the pricing errors are significant at 1% level). This finding is contrast to the finding in Kelly et al. (2019); they find that increasing the number of factors can turn rejection to failure to reject, settling on a five-factor model. The difference stems from the nature of factors that we extract: our factors are designed to capture common time-series variation of stock returns, in the spirit of the APT, while the IPCA procedure of Kelly et al. (2019) is designed to fit the cross-section of stock returns as well as their common time-variation, potentially giving up on the former to maximize the latter. Indeed, our factors do a better job of capturing common time variation in stock returns, as exhibited both by the $R^2$ measures above and, more importantly, by the Sharpe ratios of the arbitrage portfolios that exploit the non-zero alphas. In particular, we find a high annualized Sharpe ratio for the pure-alpha strategy in *all* of the cases that we consider. The Sharpe ratio increases from 3.18 to 3.82 as we increase $K$ from 1 to 10 in the linear specification, and are in the same range (sometimes exceeding 4) in the nonlinear specifications that utilize B-splines while reducing the number of characteristics used (when we use fewer characteristics, the Sharpe ratio tends to fall with the number of factors in some of the specifications). Importantly, since alphas always decline when additional factors are introduced, the rising Sharpe ratios' as the number of factors grows is clear evidence of the important role of the factors in hedging out common variation in stock returns, which reduces the volatility of the arbitrage portfolio at a rate that exceeds the decline in alpha.

Before proceeding to the detailed investigation of characteristics and nonlinearity, we need to pin down the sign of the single extracted factor. Under the normalization: $B'B = I_K$ and $F'M_TF/T$ being diagonal with diagonal entries in descending order, the sign of the single extracted factor is undetermined. To pin down the sign, we let the sample means of the extracted factors to be positive such that the unconditional risk premium is positive. Further, to interpret the latent factors, we also report the correlation matrix among the extracted factors and six constructed factors in Table 8. The extracted factor from the linear specification is almost uncorrelated with the market excess return factor, and the second extracted factor from the nonlinear specifications. The extracted factors from the nonlinear specification have much higher correlations with the market excess return, SMB, and RMW. It is noteworthy to point out that the fourth factor in all three models is highly correlated with MKT and SMB, although

Table 8: Factors correlation†

| | MKT | SMB | HML | MOM | RMW | CMA |
|---|---|---|---|---|---|---|
| Linear specifications with 36 characteristics | | | | | | |
| **Factor1** | 0.02 | 0.11 | 0.10 | -0.36 | -0.11 | 0.02 |
| Factor 2 | 0.26 | 0.26 | 0.05 | -0.16 | -0.20 | -0.02 |
| Factor 3 | -0.26 | -0.22 | -0.05 | 0.18 | 0.11 | 0.07 |
| Factor 4 | 0.58 | 0.46 | -0.33 | -0.06 | -0.30 | -0.30 |
| Factor 5 | 0.10 | 0.05 | -0.15 | 0.05 | 0.01 | -0.16 |
| Factor 6 | 0.32 | 0.24 | -0.13 | -0.04 | -0.20 | -0.13 |
| Factor 7 | 0.30 | 0.19 | 0.02 | -0.16 | -0.09 | -0.10 |
| Factor 8 | -0.04 | -0.05 | 0.21 | -0.41 | 0.14 | 0.16 |
| Factor 9 | 0.02 | 0.03 | -0.08 | 0.29 | -0.06 | -0.03 |
| Factor 10 | -0.03 | 0.01 | 0.13 | 0.01 | 0.09 | 0.08 |
| Nonlinear specifications with 18 characteristics | | | | | | |
| **Factor1** | 0.24 | 0.37 | 0.07 | -0.37 | -0.30 | -0.02 |
| **Factor2** | 0.41 | 0.31 | -0.37 | -0.09 | -0.30 | -0.29 |
| Factor 3 | -0.22 | -0.11 | 0.45 | -0.50 | 0.23 | 0.33 |
| Factor 4 | 0.48 | 0.21 | -0.04 | -0.34 | -0.15 | -0.18 |
| Factor 5 | -0.12 | -0.01 | -0.25 | 0.14 | -0.06 | -0.10 |
| Factor 6 | 0.07 | -0.21 | -0.04 | 0.09 | 0.08 | -0.08 |
| Factor 7 | -0.15 | -0.22 | -0.09 | -0.12 | -0.06 | -0.03 |
| Factor 8 | -0.01 | -0.08 | 0.14 | 0.05 | 0.09 | 0.05 |
| Factor 9 | -0.17 | 0.03 | 0.08 | 0.10 | 0.10 | 0.09 |
| Factor 10 | -0.21 | -0.04 | 0.11 | 0.07 | 0.06 | 0.06 |
| Nonlinear specifications with 12 characteristics | | | | | | |
| **Factor1** | 0.22 | 0.36 | 0.05 | -0.32 | -0.32 | -0.01 |
| **Factor2** | 0.39 | 0.31 | -0.27 | -0.34 | -0.26 | -0.26 |
| Factor 3 | -0.19 | -0.19 | 0.49 | -0.57 | 0.25 | 0.30 |
| Factor 4 | 0.50 | 0.21 | -0.12 | -0.13 | -0.16 | -0.24 |
| Factor 5 | -0.04 | -0.15 | -0.11 | 0.12 | -0.02 | -0.07 |
| Factor 6 | -0.20 | -0.20 | -0.14 | 0.08 | 0.06 | -0.07 |
| Factor 7 | -0.05 | -0.04 | 0.16 | 0.00 | 0.12 | 0.04 |
| Factor 8 | -0.04 | 0.19 | 0.28 | 0.18 | 0.06 | 0.22 |
| Factor 9 | 0.27 | 0.06 | 0.18 | 0.11 | -0.14 | 0.16 |
| Factor 10 | 0.17 | -0.27 | -0.22 | -0.06 | 0.21 | -0.20 |

† MKT: market excess return; SMB: "small minus big" factor; HML: "high minus low" factor; MOM: "momentum" factor; RMW: "robust minus weak" factor; CMA: "conservative minus aggressive" factor. The highlighted ones are selected factors in the models.

is not a common factor that is selected by the formal tests. Finally, we find that the correlation of extracted factors with MOM is negative for the first two to four factors, shifting to positive for the third factor in the linear case and the fifth (and some of the higher-order) factors in the nonlinear cases.[9]

Table 9: Characteristic significance under linear specifications of $\alpha(\cdot)$ and $\beta(\cdot)^{\dagger}$

| Char | Alpha | Beta |
|---|---|---|
| a2me | 0.0034 | −0.2356*** |
| assets | −0.0048*** | 0.5638*** |
| ato | −0.0023* | 0.0398*** |
| bm | 0.0045** | −0.0885*** |
| beta | −0.0056*** | 0.0829*** |
| bidask | 0.0004 | 0.0336*** |
| c | 0.0017*** | 0.0079 |
| cto | −0.0013 | −0.0426* |
| d2a | 0.0047*** | 0.0221*** |
| dpi2a | −0.0017** | −0.0109 |
| e2p | −0.0034*** | 0.0067 |
| fc2y | 0.0000 | 0.0372*** |
| free_cy | 0.0019*** | −0.0133* |
| idiovol | −0.0075*** | −0.0077 |
| invest | −0.0019** | 0.0003 |
| lev | −0.0012 | −0.0200** |
| mktcap | −0.0022*** | −0.7576*** |
| turn | 0.0075*** | 0.0532*** |
| noa | −0.0036*** | 0.0217** |
| oa | −0.0007 | −0.0083 |
| ol | 0.0041* | 0.0466* |
| pcm | 0.0064*** | −0.0320** |
| pm | 0.0020 | 0.0319** |
| prof | 0.0008 | −0.0341** |
| q | −0.0027 | −0.0552** |
| w52h | 0.0014* | −0.0374*** |
| rna | 0.0008 | −0.0074 |
| roa | 0.0015 | −0.0370** |
| roe | 0.0041*** | 0.0295** |
| mom | 0.0078*** | −0.0202* |
| intmom | 0.0010 | 0.0072 |
| strev | −0.0272*** | −0.0512*** |
| ltrev | −0.0031*** | −0.0629*** |
| s2p | 0.0044** | −0.0151 |
| sga2m | −0.0008 | 0.0291 |
| suv | 0.0143*** | 0.0147*** |
| Constant | 0.0060*** | 0.0829*** |

$^{\dagger}$ Char = characteristic; ***: $p-$value $<$ 1%; **: $p-$value $<$ 5%; *: $p-$value $<$ 10%.

[9] We also find that the single factor from the linear model and the first extracted factors from the nonlinear cases are highly correlated, and the second extracted factors from the nonlinear cases are also highly correlated, which shows the robustness of our model estimation.

Table 10: Characteristics significance under nonlinear specifications of $\alpha(\cdot)$ and $\beta(\cdot)$ with 18 characteristics and one internal knot[†]

| Char | Alpha | | Beta1 | | Beta2 | |
|---|---|---|---|---|---|---|
| | $B_1$ | $B_2$ | $B_1$ | $B_2$ | $B_1$ | $B_2$ |
| assets | 0.0013** | −0.0043*** | 0.0975*** | 0.2703*** | −0.2603*** | −0.5175*** |
| ato | −0.0002 | −0.0005 | 0.0193*** | 0.0168** | −0.0058 | −0.0356*** |
| bm | 0.0069*** | 0.0098*** | −0.0606*** | −0.1212*** | 0.0416*** | 0.0761*** |
| beta | −0.0009** | −0.0085*** | 0.0466*** | 0.1293*** | 0.0562*** | 0.1740*** |
| d2a | 0.0037*** | 0.0051*** | 0.0197*** | 0.0175*** | −0.0175** | −0.0375*** |
| idiovol | 0.0030*** | −0.0085*** | −0.0133*** | 0.0186*** | 0.0003 | −0.0078 |
| invest | 0.0000 | −0.0033*** | −0.0316*** | −0.0189*** | −0.0060 | 0.0197** |
| mktcap | −0.0088*** | −0.0023*** | −0.2815*** | −0.5338*** | 0.2094*** | 0.4082*** |
| turn | 0.0072*** | 0.0064*** | 0.0282*** | 0.0328*** | 0.0343*** | 0.1099*** |
| noa | −0.0012** | −0.0056*** | −0.0030 | −0.0019 | −0.0183** | −0.0238*** |
| pcm | 0.0010* | 0.0044*** | −0.0047 | −0.0061 | −0.0294*** | −0.0551*** |
| prof | 0.0022*** | 0.0027*** | −0.0257*** | −0.0479*** | 0.0452*** | 0.1033*** |
| w52h | −0.0049*** | 0.0018** | −0.1087*** | −0.0823*** | −0.1042*** | −0.0984*** |
| roe | 0.0058*** | 0.0061*** | −0.0427*** | −0.0228*** | −0.0739*** | −0.0801*** |
| mom | 0.0063*** | 0.0090*** | −0.0835*** | −0.0799*** | −0.0804*** | −0.1060*** |
| strev | −0.0140*** | −0.0255*** | −0.0458*** | −0.0993*** | −0.0785*** | −0.1252*** |
| ltrev | −0.0018*** | −0.0028*** | −0.0974*** | −0.1033*** | −0.0744*** | −0.0706*** |
| suv | 0.0042*** | 0.0143*** | −0.0048 | 0.0089** | −0.0029 | −0.0016 |
| Constant | −0.0039*** | | 0.6643*** | | 0.5443*** | |

[†] Char = characteristic; ***: $p-$value $<$ 1%; **: $p-$value $<$ 5%; *: $p-$value $<$ 10%. $B_i$ is the estimated coefficients corresponding to the $i$-th sieve basis function.

Table 11: Characteristics significance under nonlinear specifications of $\alpha(\cdot)$ and $\beta(\cdot)$ with 12 characteristics and two internal knots†

| Char | Alpha | | | Beta1 | | | Beta2 | | |
|---|---|---|---|---|---|---|---|---|---|
| | $B_1$ | $B_2$ | $B_3$ | $B_1$ | $B_2$ | $B_3$ | $B_1$ | $B_2$ | $B_3$ |
| assets | 0.0002 | 0.0011** | −0.0059*** | 0.0782*** | 0.0780*** | 0.2297*** | −0.2479*** | −0.2104*** | −0.4264*** |
| bm | 0.0079*** | 0.0066*** | 0.0096*** | −0.0731*** | −0.0671*** | −0.1236*** | 0.0675*** | 0.0507*** | 0.1005*** |
| beta | −0.0024*** | −0.0017*** | −0.0102*** | 0.0614*** | 0.0404*** | 0.1378*** | 0.0669*** | 0.0403*** | 0.1941*** |
| idiovol | 0.0018*** | 0.0021*** | −0.0095*** | 0.0004 | −0.0201*** | 0.0381*** | −0.0064 | 0.0110* | −0.0175* |
| invest | 0.0001 | 0.0012** | −0.0063*** | −0.0369*** | −0.0363*** | −0.0219*** | −0.0392*** | −0.0401*** | −0.0161** |
| mktcap | −0.0113*** | −0.0063*** | −0.0005 | −0.2845*** | −0.2625*** | −0.4995*** | 0.2184*** | 0.2057*** | 0.3388*** |
| turn | 0.0068*** | 0.0066*** | 0.0058*** | 0.0310*** | 0.0288*** | 0.0392*** | 0.0621*** | 0.0578*** | 0.1398*** |
| prof | 0.0048*** | 0.0038*** | 0.0052*** | −0.0318*** | −0.0198*** | −0.0458*** | 0.0307*** | 0.0254*** | 0.0784*** |
| mom | 0.0077*** | 0.0070*** | 0.0129*** | −0.1411*** | −0.1096*** | −0.1179*** | −0.2299*** | −0.1928*** | −0.2641*** |
| strev | −0.0165*** | −0.0110*** | −0.0251*** | −0.0788*** | −0.0522*** | −0.1243*** | −0.1241*** | −0.0980*** | −0.1697*** |
| ltrev | −0.0003 | 0.0004 | −0.0008 | −0.1164*** | −0.0933*** | −0.1132*** | −0.0848*** | −0.0782*** | −0.0740*** |
| suv | 0.0051*** | 0.0039*** | 0.0153*** | −0.0031 | −0.0030 | 0.0109*** | −0.0060 | −0.0013 | −0.0043 |
| Constant | −0.0029*** | | | 0.6074*** | | | 0.4194*** | | |

† Char = characteristic; ***: $p-$value $< 1\%$; **: $p-$value $< 5\%$; *: $p-$value $< 10\%$. $B_i$ is the estimated coefficients corresponding to the $i$-th sieve basis function.

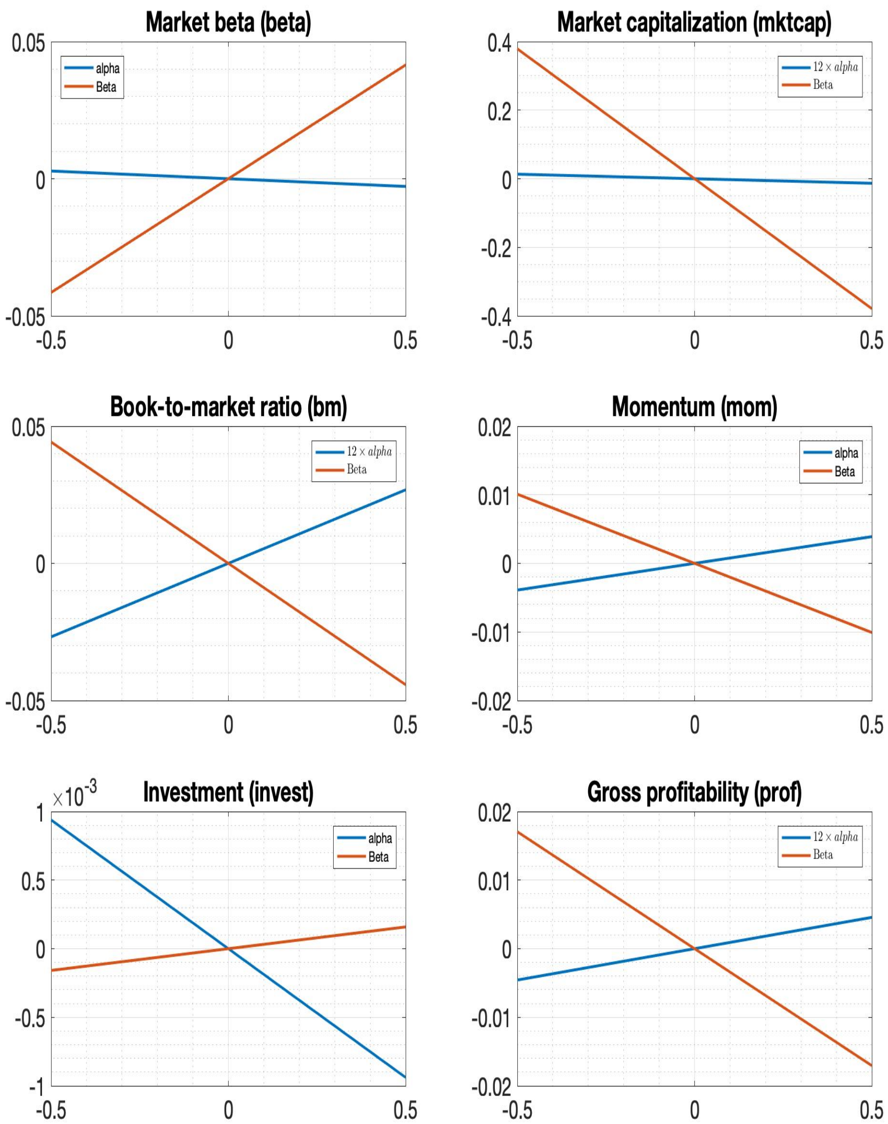
To further investigate the nonlinearity and contributions of each individual characteristic to pricing errors and risk exposures, we report coefficient estimates for three different specifications in Tables 9-11. We examine the significance of each individual term in $\phi(z_{it})$ by the weighted bootstrap in Section 5.1. In contrast to the finding in Kelly et al. (2019) of 13 out of 36 characteristics being significant in driving risk exposures, we find 26 to be significant in the linear specification. We find 22 characteristics to have a significant effect on alpha in the linear specification. The findings indicate that most of the characteristics contain relevant information about both pricing errors and risk exposures (rather just one or the other). In the nonlinear cases, we find that almost all the sieve coefficients are significant, which indicates that it is necessary to introduce the nonlinear terms.

Empirical studies show that stocks with smaller market capitalization, higher book-to-market ratio (Fama and French, 1993), or better past performance (Jegadeesh and Titman, 1993) tend to have higher returns, often referred to as "size", "value", and "momentum" anomalies in equity market. The presumed "rational" explanation for these anomalies is that smaller or value firms or firms with better past performance have larger exposures to priced systematic risky factors. In order to test this hypothesis, we report the plots of the pricing error and the risk exposure versus six important characteristics. Figure 1 reports the results for the linear specification, and we find a downward sloping curve for book-to-market ratio, which rejects a (conditional) one-factor-model explanation of value. Figures 2-3 report the results for the nonlinear specifications, while we find the associated nonlinear and upward sloping exposure to the second extracted factor, which is more consistent with the risk-based views. Similarly, we find opposite curve slopes from the linear and nonlinear specifications for investment and profitability, where the results from the nonlinear specifications are more consistent with the findings in Fama and French (2015): the firms with low investment and high profitability bear the larger risks. Overall, more estimates from the nonlinear specifications are consistent with the traditional view than those from the linear specification.

Finally, we check the estimated contribution of each individual characteristic in the pricing error with different numbers of factors. In Figure 4, we report the estimates and their associated 95% confidence intervals for the coefficients in the linear specification. We find that increasing the number of factors does not affect the estimates and confidence intervals significantly. This implies that the estimation of *alpha* is not sensitive to the number of factors and is also contrast to the estimation procedure in (Kelly et al., 2019).

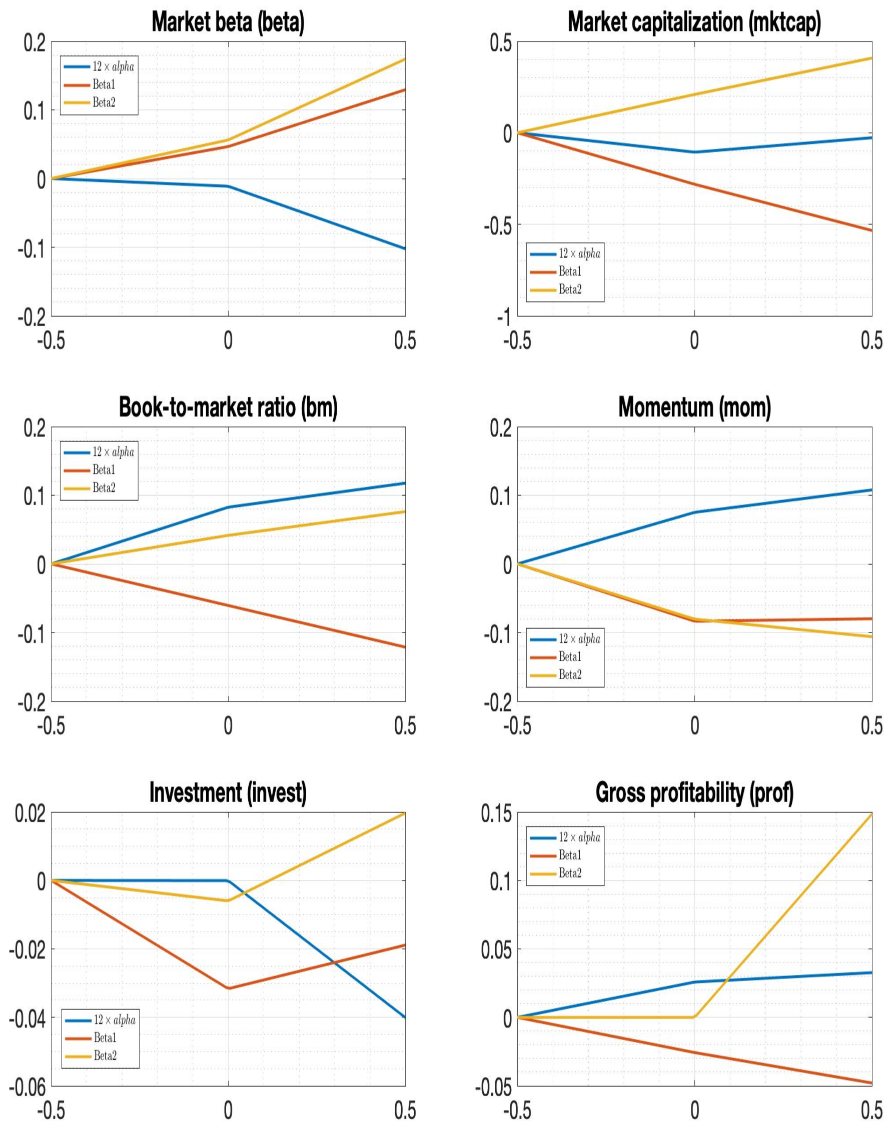
To conclude, we not only find strong evidence of nonlinearity but also identify the source of the nonlinearity in both pricing errors and risk exposures.

Figure 1: Characteristics-alpha and characteristics-beta plots under linear specifications of $\alpha(\cdot)$ and $\beta(\cdot)$ with 36 characteristics



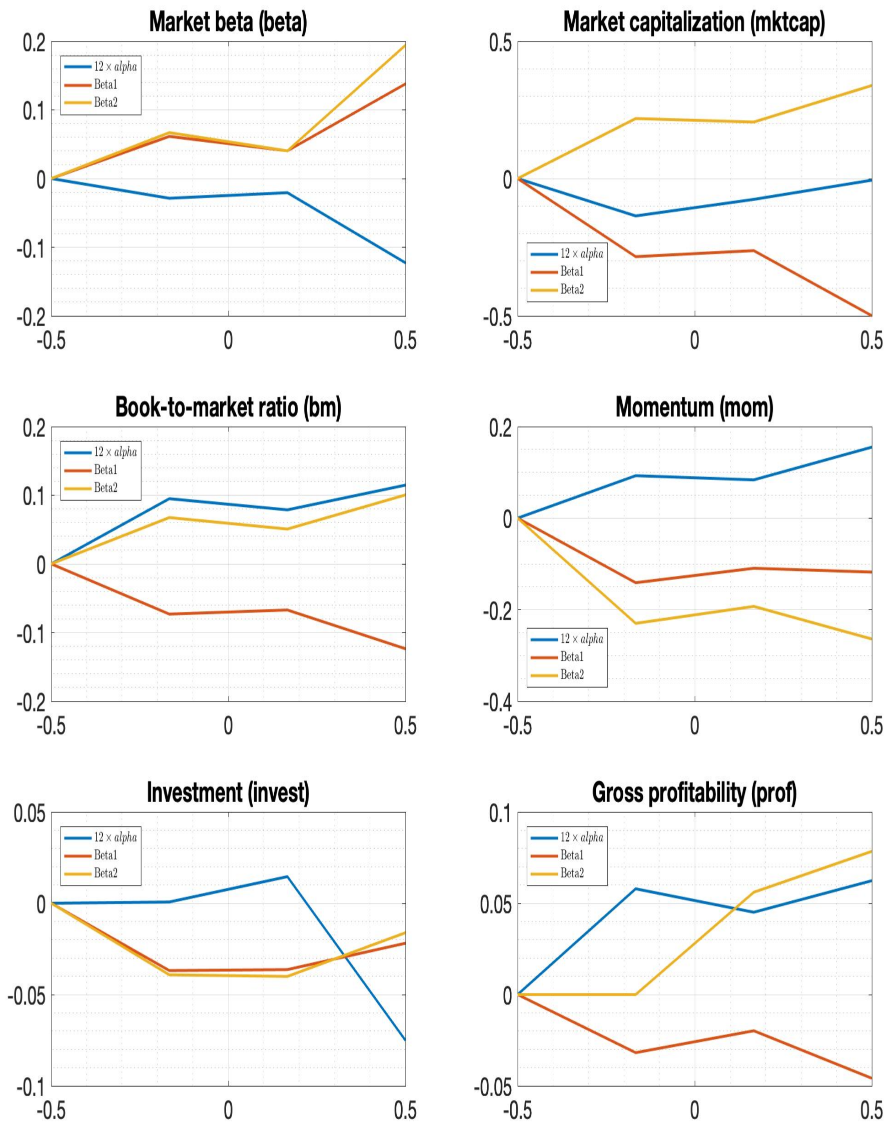

Notes: the six important characteristics are market capitalization (mktcap), market beta (beta), book-to-market ratio (bm), momentum (mom), investment (invest), and gross profitaility (prof). To make the magnitude comparable, the annualized values are reported for some alphas.

Figure 2: Characteristics-alpha and characteristics-beta plots under continuous piece-wise linear specifications of $\alpha(\cdot)$ and $\beta(\cdot)$ with 18 characteristics and one internal knot



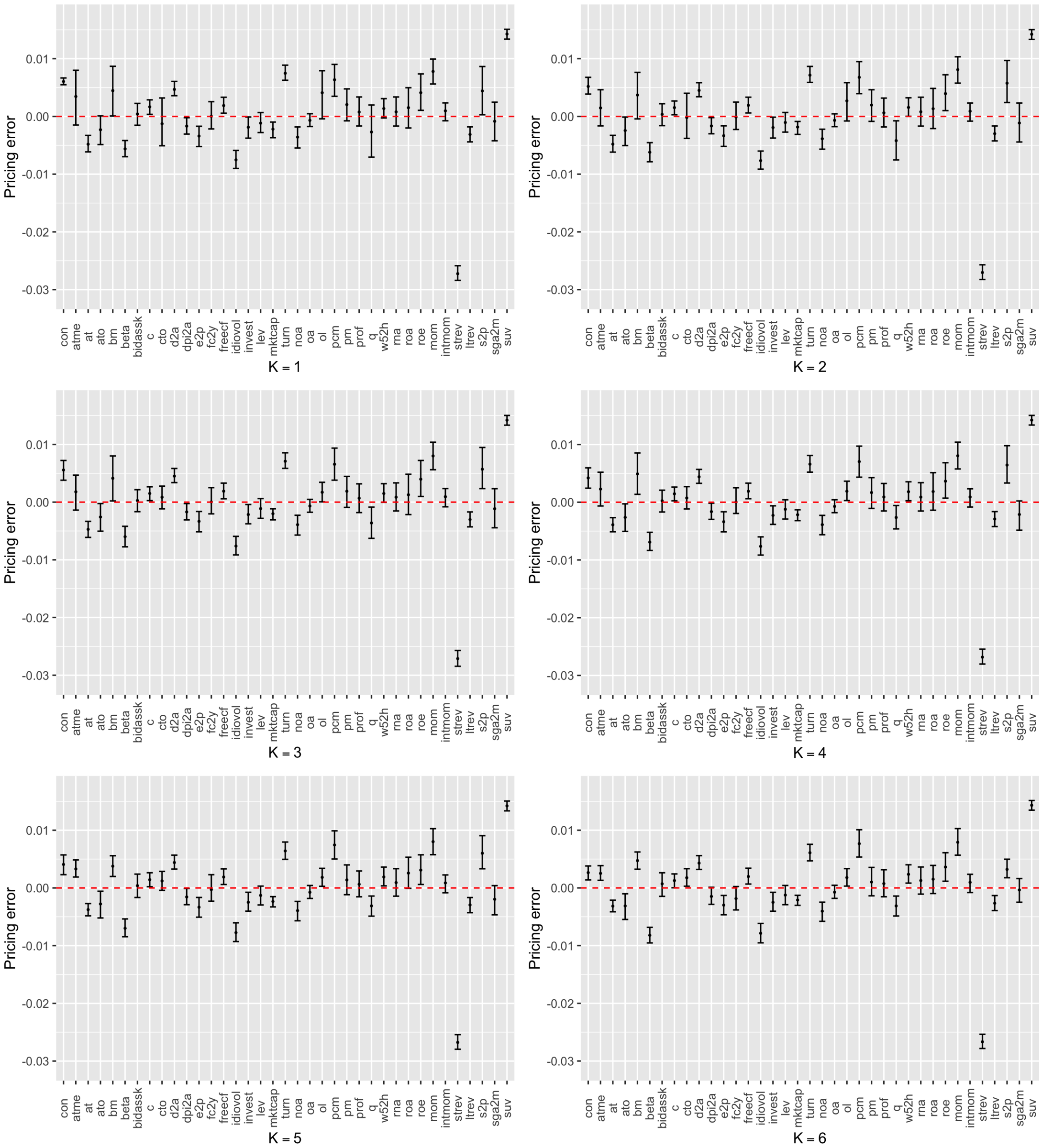

Notes: the six important characteristics are market capitalization (mktcap), market beta (beta), book-to-market ratio (bm), momentum (mom), investment (invest), and gross profitaility (prof). To make the magnitude comparable, the annualized alpha is reported.

Figure 3: Characteristics-alpha and characteristics-beta plots under continuous piecewise linear specifications of $\alpha(\cdot)$ and $\beta(\cdot)$ with 12 characteristics and two internal knots




Notes: the six important characteristics are market capitalization (mktcap), market beta (beta), book-to-market ratio (bm), momentum (mom), investment (invest), and gross profitaility (prof). To make the magnitude comparable, the annualized alpha is reported.

Figure 4: Estimates and 95% confidence intervals of coefficients in alpha under linear specifications of $\alpha(\cdot)$ and $\beta(\cdot)$



# 9 Conclusion

In this paper, we developed a simple and tractable semiparametric sieve estimation for conditional factor models with time-varying covariances and latent factors, and a weighted bootstrap for conducting inference on the intercept and factor loading functions. We established large sample properties of the estimators and validity of the tests for large $N$, even when $T$ is small. These results enable us to estimate conditional (dynamic) behavior of a large set of individual assets from a number of characteristics exhibiting nonlinearity without the need to pre-specify factors, while allowing us to disentangle the alpha from betas. We applied these methods to explain the cross-sectional differences of individual stock returns in the US market. We found strong evidence of conditional factor structure as well as nonlinearity in conditional alpha and beta functions. Importantly, although only one or two factors are selected by the formal tests, even when a large number of common factors is considered, conditional pricing errors remain large, resulting in arbitrage portfolios with high Sharpe ratios (typically above 3).

# Appendix A - Proof of Theorem 4.1

## A.1 Proof of Theorem 4.1

Proof of Theorem 4.1: Let us begin by defining some notation. Let $\tilde{A}_t \equiv (\Phi(Z_t)'\Phi(Z_t))^{-1}\Phi(Z_t)'A_t$ for $A_t = \Delta_t$ and $\varepsilon_t$, where $\Delta_t = R(Z_t) + \Delta(Z_t)f_t$. Let $\tilde{\Delta} \equiv (\tilde{\Delta}_1, \ldots, \tilde{\Delta}_T)$ and $\tilde{E} \equiv (\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_T)$. Then (9) can be written as

$$\tilde{Y} = a1_T' + BF' + \tilde{\Delta} + \tilde{E}, \tag{A.1}$$

where $1_T$ denote a $T \times 1$ vector of ones. Recall $M_T = I_T - 1_T1_T'/T$. Post-multiplying (A.1) by $M_T$ to remove $a$, we thus obtain

$$\tilde{Y}M_T = B(M_TF)' + \tilde{\Delta}M_T + \tilde{E}M_T. \tag{A.2}$$

Let $V$ be a $K\times K$ diagonal matrix of the first $K$ largest eigenvalues of $\tilde{Y}M_T\tilde{Y}'/T$. By the definitions of $\hat{B}$ and $\hat{F}$, $(\tilde{Y}M_T\tilde{Y}'/T)\hat{B} = \hat{B}V$ and $M_T\hat{F} = M_T\tilde{Y}'\hat{B}$. Thus, $\hat{F}'M_T\hat{F}/T = \hat{B}'(\tilde{Y}M_T\tilde{Y}'/T)\hat{B} = V$ and $H = (F'M_T\hat{F})(\hat{F}'M_T\hat{F})^{-1} = (F'M_T\tilde{Y}'\hat{B}/T)V^{-1}$. We may substitute (A.2) to $(\tilde{Y}M_T\tilde{Y}'/T)\hat{B} = \hat{B}V$ to obtain

$$\hat{B} - BH = [(\tilde{\Delta} + \tilde{E})M_T\tilde{Y}'/T]\hat{B}V^{-1} = \sum_{j=1}^{6} D_j\hat{B}V^{-1}, \tag{A.3}$$

where $D_1 = \tilde{\Delta}M_TFB'/T$, $D_2 = \tilde{\Delta}M_T\tilde{\Delta}'/T$, $D_3 = D_6' = \tilde{\Delta}M_T\tilde{E}'/T$, $D_4 = \tilde{E}M_TFB'/T$ and $D_5 = \tilde{E}M_T\tilde{E}'/T$. By the Cauchy-Schwartz inequality and the fact that $\|C+D\|_F \leq \|C\|_F + \|D\|_F$ and $\|CD\|_F \leq \|C\|_2\|D\|_F$, (A.3) implies

$$\|\hat{B} - BH\|_F^2 \leq 6\|\hat{B}\|_2^2\|V^{-1}\|_2^2 \left(\sum_{j=1}^{6} \|D_j\|_F^2\right) = O_p\left(\frac{1}{J^{2\kappa}} + \frac{J^2}{N^2} + \frac{J}{NT}\right), \tag{A.4}$$

where the equality follows by Lemmas A.1 and A.2(i) and the fact that $\|D_3\|_F = \|D_6\|_F$. Since $\hat{B}'\hat{B} = I_K$, $\|I_{JM} - \hat{B}\hat{B}'\|_2 = 1$ and $(I_{JM} - \hat{B}\hat{B}')\hat{B} = 0$. By the definition of $\hat{a}$,

$$\begin{aligned}\hat{a} - a = &-\hat{B}(\hat{B} - BH)'a + (I_{JM} - \hat{B}\hat{B}')(BH - \hat{B})H^{-1}\bar{f} \\ &+ (I_{JM} - \hat{B}\hat{B}')\tilde{\Delta}1_T/T + (I_{JM} - \hat{B}\hat{B}')\tilde{E}1_T/T.\end{aligned} \tag{A.5}$$

where $H^{-1}$ is well defined with probability approaching one by (A.4) and Lemma A.2(ii), and we have used $a'B = 0$ and $(I_{JM} - \hat{B}\hat{B}')\hat{B} = 0$. By the Cauchy-Schwartz inequality and the fact that $\|x + y\| \leq \|x\| + \|y\|$ and $\|Ax\| \leq \|A\|_2\|x\|$, (A.5) implies

$$|\hat{a} - a|^2 \leq 4\Bigg(\|\hat{B} - BH\|_F^2\|a\|^2 + \|BH - \hat{B}\|_F^2\|H^{-1}\|_2^2 \max_{t\leq T}\|f_t\|^2$$

$$+\frac{1}{T}\|\tilde{\Delta}\|_F^2+\frac{1}{T^2}\|\tilde{E}1_T\|^2\Big)=O_p\left(\frac{1}{J^{2\kappa}}+\frac{J^2}{N^2}+\frac{J}{NT}\right), \tag{A.6}$$

where the equality follows by (A.4), Assumptions 4.2(ii) and 4.4, and Lemmas A.2(ii), A.3(i) and A.4(ii). Noting $\hat{B}'\hat{B}=I_K$, we may substitute (A.1) to $\hat{F}=\tilde{Y}'\hat{B}$ to obtain

$$\hat{F}-F(H')^{-1}=1_Ta'(\hat{B}-BH)+F(H')^{-1}(BH-\hat{B})'\hat{B}+\tilde{\Delta}'\hat{B}+\tilde{E}'\hat{B}. \tag{A.7}$$

where $(H')^{-1}$ is well defined with probability approaching one by (A.4) and Lemma A.2(ii), and we have used $a'B=0$. By the Cauchy-Schwartz inequality and the fact that $\|C+D\|_F\le\|C\|_F+\|D\|_F$ and $\|CD\|_F\le\|C\|_2\|D\|_F$, (A.7) implies

$$\begin{aligned}\frac{1}{T}\|\hat{F}-F(H')^{-1}\|_F^2 &\le \frac{4}{T}\left(\|F\|_2^2\|H^{-1}\|_2^2\|BH-\hat{B}\|_F^2+\|\tilde{\Delta}\|_F^2+\|\tilde{E}\|_F^2\right)\|\hat{B}\|_2^2\\ &\quad+\frac{4}{T}\|1_T\|^2\|BH-\hat{B}\|_F^2\|a\|^2=O_p\left(\frac{1}{J^{2\kappa}}+\frac{J}{N}\right),\end{aligned} \tag{A.8}$$

where the equality follows from (A.4), Assumptions 4.2(ii) and 4.4, and Lemmas A.2(ii), A.3(i) and (ii) by noting that $J=o(\sqrt{N})$. Since $\hat{\beta}(z)=\hat{B}'\phi(z)$ and $\beta(z)=B'\phi(z)+\delta(z)$,

$$\hat{\beta}(z)-H'\beta(z)=\hat{B}'\phi(z)-(BH)'\phi(z)+H'\delta(z). \tag{A.9}$$

By the Cauchy-Schwartz inequality and the fact that $\|x+y\|\le\|x\|+\|y\|$, $\|Ax\|\le\|A\|_2\|x\|$ and $\|A\|_2\le\|A\|_F$, (A.9) implies

$$\begin{aligned}\sup_z\|\hat{\beta}(z)-H'\beta(z)\|^2 &\le 2\|\hat{B}-BH\|_F^2\sup_z\|\phi(z)\|^2+2\|H\|_2^2\sup_z\|\delta(z)\|^2\\ &=O_p\left(\frac{1}{J^{2\kappa-1}}+\frac{J^3}{N^2}+\frac{J^2}{NT}\right)\max_{j\le J}\sup_z|\phi_j(z)|^2,\end{aligned} \tag{A.10}$$

where the equality follows from (A.4) and Lemma A.2(i) by noting that $\sup_z\|\phi(z)\|^2\le JM\max_{j\le J}\sup_z|\phi_j(z)|^2$ and $\sup_z\|\delta(z)\|^2\le KM^2\max_{k\le K,m\le M}\sup_z|\delta_{km,J}(z)|^2=O(J^{-2\kappa})$ due to Assumption 4.2(iv). The proof of the second last result is similar. This completes the proof of the theorem. ■

## A.2 Technical Lemmas

**Lemma A.1.** *Let $D_1,D_2,D_3,D_4,D_5$ be given in the proof of Theorem 4.1.*
*(i) Under Assumptions 4.1(i), 4.2(i), (ii) and (iv), $\|D_1\|_F^2=O_p(J^{-2\kappa})$.*
*(ii) Under Assumptions 4.1(i), 4.2(ii) and (iv), $\|D_2\|_F^2=O_p(J^{-4\kappa})$.*
*(iii) Under Assumptions 4.1, 4.2(ii), (iv) and 4.3, $\|D_3\|_F^2=O_p(J^{-2\kappa}J/N)$.*
*(iv) Under Assumptions 4.1, 4.2(i), (ii) and 4.3, $\|D_4\|_F^2=O_p(J/NT)$.*
*(v) Under Assumptions 4.1 and 4.3, $\|D_5\|_F^2=O_p(J^2/N^2)$.*

Proof: (i) Since $\|M_T\|_2=1$, $\|D_1\|_F\le\|B\|_2\|F\|_2\|\tilde{\Delta}\|_F/T$. The result then immediately follows from Assumptions 4.2(i), (ii) and Lemma A.3(i).

(ii) Since $\|M_T\|_2 = 1$, $\|D_2\|_F \leq \|\tilde{\Delta}\|_F^2/T$. The result then immediately follows from Lemma A.3(i).

(iii) Since $\|M_T\|_2 = 1$, $\|D_3\|_F \leq \|\tilde{\Delta}\|_F\|\tilde{E}\|_F/T$. The result then immediately follows from Lemma A.3(i) and (ii).

(iv) Since $\|D_4\|_F \leq \|B\|_2\|\tilde{E}M_T F\|_F/T$, the result then immediately follows from Assumption 4.2(i) and Lemma A.3(iii).

(v) Since $\|M_T\|_2 = 1$, $\|D_5\|_F \leq \|\tilde{E}\|_F^2/T$. The result then immediately follows from Lemma A.3(ii). ■

**Lemma A.2.** *Suppose Assumptions 4.1-4.3 hold. Let $V$ be given in the proof of Theorem 4.1. Assume (i) $N \to \infty$; (ii) $T \to \infty$ or $T \geq K+1$ is finite; (iii) $J \to \infty$ with $J = o(\sqrt{N})$. Then (i) $\|V\|_2 = O_p(1)$, $\|V^{-1}\|_2 = O_p(1)$, and $\|H\|_2 = O_p(1)$; (ii) $\|H^{-1}\|_2 = O_p(1)$, if $\|\hat{B} - BH\|_F = o_p(1)$.*

Proof: (i) Let $D_7 \equiv D_1'$ and $D_8 \equiv D_4'$. Then by (A.2), $\tilde{Y}M_T\tilde{Y}'/T = BF'M_T FB'/T + \sum_{j=1}^{8} D_j$, where $D_1, \ldots, D_6$ are given below (A.2). By the fact that $\|C + D\|_F \leq \|C\|_F + \|D\|_F$,

$$\|\tilde{Y}M_T\tilde{Y}'/T - BF'M_T FB'/T\|_F \leq \sum_{j=1}^{8} \|D_j\|_F = O_p\left(\frac{1}{J^\kappa} + \frac{J}{N} + \frac{\sqrt{J}}{\sqrt{NT}}\right), \quad \text{(A.11)}$$

where the equality follows by Lemma A.1 and the fact that $\|D_6\|_F = \|D_3\|_F$, $\|D_7\|_F = \|D_1\|_F$ and $\|D_8\|_F = \|D_4\|_F$. Let $\mathcal{V}$ be a $K \times K$ diagonal matrix of the eigenvalues of $(F'M_T F/T)B'B$, which are equal to the first $K$ largest eigenvalues of $BF'M_T FB'/T$. By the Weyl's inequality and the fact that $\|A\|_2 \leq \|A\|_F$,

$$\|V - \mathcal{V}\|_2 \leq \|\tilde{Y}M_T\tilde{Y}'/T - BF'M_T FB'/T\|_2 = O_p\left(\frac{1}{J^\kappa} + \frac{J}{N} + \frac{\sqrt{J}}{\sqrt{NT}}\right). \quad \text{(A.12)}$$

Thus, $\|V\|_2 = O_p(1)$ and $\|V^{-1}\|_2 = \lambda_{\min}^{-1}(V) = O_p(1)$ follow from (A.12) and Assumptions 4.2(i)-(iii). Let $H^\diamond \equiv (F'M_T F/T)B'\hat{B}V^{-1}$. Recall that $H = (F'M_T\tilde{Y}'\hat{B}/T)V^{-1}$. Then by the fact that $\|A\|_2 \leq \|A\|_F$ and $\|M_T\|_2 = 1$,

$$\|H - H^\diamond\|_2 \leq \frac{1}{T}(\|F\|_2\|\tilde{\Delta}\|_F + \|\tilde{E}M_T F\|_F)\|\hat{B}\|_2\|V^{-1}\|_2 = O_p\left(\frac{1}{J^\kappa} + \frac{\sqrt{J}}{\sqrt{NT}}\right), \quad \text{(A.13)}$$

where the equality follows from the second result in (i), Assumption 4.2(ii) and Lemmas A.3(i) and (iii). Since $\|H^\diamond\|_2 \leq \|F'M_T F/T\|_2\|B\|_2\|\hat{B}\|_2\|V^{-1}\|_2$, the third result in (i) follows from (A.13), the second result in (i) and Assumptions 4.2(i) and (ii).

(ii) By the fact that $\|C+D\|_F \leq \|C\|_F + \|D\|_F$ and $\|CD\|_F \leq \|C\|_2\|D\|_F$,

$$\|\hat{B}'\hat{B} - H'B'BH\|_F \leq \|\hat{B}\|_2\|\hat{B}-BH\|_F + \|\hat{B}-BH\|_F\|B\|_2\|H\|_2. \tag{A.14}$$

Thus, $I_K - H'B'BH = o_p(1)$ by Assumption 4.2(i) and $\|H\|_2 = O_p(1)$. It then follows that $I_K - \lambda_{\min}(B'B)H'H$ is negative semidefinite with probability approaching one, since $H'B'BH - \lambda_{\min}(B'B)H'H$ is positive semidefinite. So, the eigenvalues of $H'H$ are not smaller than $\lambda_{\min}^{-1}(B'B)$ with probability approaching one. Thus, the result in (ii) follows from Assumption 4.2(i). ■

**Lemma A.3.** *Let $\tilde{\Delta}$ and $\tilde{E}$ be given in the proof of Theorem 4.1.*
*(i) Under Assumptions 4.1(i), 4.2(ii) and (iv), $\|\tilde{\Delta}\|_F^2/T = O_p(J^{-2\kappa})$.*
*(ii) Under Assumptions 4.1 and 4.3, $\|\tilde{E}\|_F^2/T = O_p(J/N)$.*
*(iii) Under Assumptions 4.1, 4.2 (ii) and 4.3, $\|\tilde{E}M_T F\|_F^2/T = O_p(J/N)$.*

Proof: (i) By the fact that $\|Ax\| \leq \|A\|_2\|x\|$ and $\|A\|_2 \leq \|A\|_F$,

$$\begin{aligned}\frac{1}{T}\|\tilde{\Delta}\|_F^2 &= \frac{1}{T}\sum_{t=1}^T \|(\Phi(Z_t)'\Phi(Z_t))^{-1}\Phi(Z_t)'(R(Z_t)+\Delta(Z_t)f_t)\|^2 \\ &\leq 2\max_{t\leq T}\|f_t\|^2\left(\min_{t\leq T}\lambda_{\min}(\hat{Q}_t)\right)^{-1}\frac{1}{NT}\sum_{t=1}^T\|\Delta(Z_t)\|_F^2 \\ &\quad + 2\left(\min_{t\leq T}\lambda_{\min}(\hat{Q}_t)\right)^{-1}\frac{1}{NT}\sum_{t=1}^T\|R(Z_t)\|^2 = O_p\left(\frac{1}{J^{2\kappa}}\right),\end{aligned} \tag{A.15}$$

where the last equality follows from Assumptions 4.1(i) and 4.2(ii) and Lemma A.4(iii).

(ii) By the fact that $\|Ax\| \leq \|A\|_2\|x\|$,

$$\begin{aligned}\frac{1}{T}\|\tilde{E}\|_F^2 &= \frac{1}{T}\sum_{t=1}^T\|(\Phi(Z_t)'\Phi(Z_t))^{-1}\Phi(Z_t)'\varepsilon_t\|^2 \\ &\leq \left(\min_{t\leq T}\lambda_{\min}(\hat{Q}_t)\right)^{-2}\frac{1}{N^2T}\sum_{t=1}^T\|\Phi(Z_t)'\varepsilon_t\|^2 = O_p\left(\frac{J}{N}\right),\end{aligned} \tag{A.16}$$

where the last equality follows from Assumption 4.1(i) and Lemma A.4(i).

(iii) By the fact that $\|C+D\|_F \leq \|C\|_F + \|D\|_F$,

$$\begin{aligned}\frac{1}{T}\|\tilde{E}M_T F\|_F^2 \leq{}& \frac{2}{N^2T}\left\|\sum_{t=1}^T\hat{Q}_t^{-1}\Phi(Z_t)'\varepsilon_t f_t'\right\|_F^2 \\ &+ \frac{2\|\bar{f}\|^2}{N^2T}\left\|\sum_{t=1}^T\hat{Q}_t^{-1}\Phi(Z_t)'\varepsilon_t\right\|^2 = O_p\left(\frac{J}{N}\right),\end{aligned} \tag{A.17}$$

where the equality follows from Assumption 4.2(ii) and Lemma A.4(ii). ■

**Lemma A.4.** *(i) Under Assumptions 4.1(ii) and 4.3,*

$$\sum_{t=1}^{T} \|\Phi(Z_t)'\varepsilon_t\|^2 = O_p(NTJ).$$

*(ii) Under Assumptions 4.1, 4.2(ii) and 4.3,*

$$\left\|\sum_{t=1}^{T} \hat{Q}_t^{-1}\Phi(Z_t)'\varepsilon_t f_t'\right\|_F^2 = O_p(NTJ) \text{ and } \left\|\sum_{t=1}^{T} \hat{Q}_t^{-1}\Phi(Z_t)'\varepsilon_t\right\|^2 = O_p(NTJ).$$

*(iii) Under Assumption 4.2(iv),*

$$\sum_{t=1}^{T} \|\Delta(Z_t)\|_F^2 = O_p(NTJ^{-2\kappa}) \text{ and } \sum_{t=1}^{T} \|R(Z_t)\|^2 = O_p(NTJ^{-2\kappa}).$$

Proof: (i) The result follows by the Markov's inequality, since

$$\begin{aligned}
&E\left[\sum_{t=1}^{T} \|\Phi(Z_t)'\varepsilon_t\|^2\right] = E\left[\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N} \phi(z_{it})'\phi(z_{jt})\varepsilon_{it}\varepsilon_{jt}\right] \\
&= \sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N} E[\phi(z_{it})'\phi(z_{jt})]E[\varepsilon_{it}\varepsilon_{jt}] \\
&\leq \max_{i\leq N, t\leq T} E[\|\phi(z_{it})\|^2] \sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N} |E[\varepsilon_{it}\varepsilon_{jt}]| \\
&\leq TJM \max_{m\leq M, j\leq J, i\leq N, t\leq T} E[\phi_j^2(z_{it,m})] \max_{t\leq T} \sum_{i=1}^{N}\sum_{j=1}^{N} |E[\varepsilon_{it}\varepsilon_{jt}]| = O(NTJ), \qquad \text{(A.18)}
\end{aligned}$$

where the second equality follows by the independence in Assumption 4.3(i), the first inequality is due to the Cauchy Schwartz inequality, the second inequality follows since $\max_{i\leq N, t\leq T} E[\|\phi(z_{it})\|^2] \leq JM \max_{m\leq M, j\leq J, i\leq N, t\leq T} E[\phi_j^2(z_{it,m})]$, and the last equality follows from Assumptions 4.1(ii) and 4.3(iii).

(ii) Let $E_\varepsilon$ denote the expectation with respect to $\{\varepsilon_t\}_{t\leq T}$. Since $\|A\|_F^2 = \mathrm{tr}(AA')$,

$$\begin{aligned}
&E_\varepsilon\left[\left\|\sum_{t=1}^{T} \hat{Q}_t^{-1}\Phi(Z_t)'\varepsilon_t f_t'\right\|_F^2\right] = E_\varepsilon\left[\mathrm{tr}\left(\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N} \hat{Q}_t^{-1}\phi(z_{it})\varepsilon_{it}f_t'f_s\varepsilon_{js}\phi(z_{js})'\hat{Q}_s^{-1}\right)\right] \\
&= \sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N} \phi(z_{it})'\hat{Q}_t^{-1}\hat{Q}_s^{-1}\phi(z_{js})f_t'f_s E[\varepsilon_{it}\varepsilon_{js}] \\
&\leq \max_{t\leq T}\|f_t\|^2 \left(\min_{t\leq T}\lambda_{\min}(\hat{Q}_t)\right)^{-2} \sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N} \|\phi(z_{it})\|\|\phi(z_{js})\||E[\varepsilon_{it}\varepsilon_{js}]|, \qquad \text{(A.19)}
\end{aligned}$$

where the second equality follows by the independence in Assumption 4.3(i) and the fact that both expectation and trace operators are linear, and the inequality follows by the fact that $\|Ax\| \leq \|A\|_2\|x\|$. Moreover,

$$
\begin{aligned}
&E\left[\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\|\phi(z_{it})\|\|\phi(z_{js})\||E[\varepsilon_{it}\varepsilon_{js}]|\right] \\
\leq& \max_{i\leq N,t\leq T} E[\|\phi(z_{it})\|^2]\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}|E[\varepsilon_{it}\varepsilon_{js}]| \\
\leq& JM \max_{m\leq M,j\leq J,i\leq N,t\leq T} E[\phi_j^2(z_{it,m})]\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}|E[\varepsilon_{it}\varepsilon_{js}]|, \qquad \text{(A.20)}
\end{aligned}
$$

where the first inequality is due to the Cauchy-Schwartz inequality, and the second inequality follows since $\max_{i\leq N,t\leq T} E[\|\phi(z_{it})\|^2] \leq JM \max_{m\leq M,j\leq J,i\leq N,t\leq T} E[\phi_j^2(z_{it,m})]$. Combining (A.19) and (A.20) implies that $E_\varepsilon[\|\sum_{t=1}^{T}\hat{Q}_t^{-1}\Phi(Z_t)'\varepsilon_t f_t'\|_F^2] = O_p(NTJ)$ by Assumptions 4.1, 4.2(ii) and 4.3(iii). Thus, the first result of the lemma follows by the Markov's inequality and Lemma A.5. The proof of the second result is similar.

(iii) The first result follows since

$$
\sum_{t=1}^{T}\|\Delta(Z_t)\|_F^2 \leq NTKM^2 \max_{k\leq K,m\leq M}\sup_z |\delta_{km,J}(z)|^2 = O_p(NTJ^{-2\kappa}), \qquad \text{(A.21)}
$$

where the inequality follows since $\max_{i\leq N,t\leq T}\|\delta(z_{it})\|^2 \leq M^2K\sup_{k\leq K,m\leq M}\sup_z |\delta_{km,J}(z)|^2$, and the equality follows from Assumption 4.2(iv). The proof of the second result is similar. ■

**Lemma A.5.** *Let $S_1,\ldots,S_N$ be a sequence of random variables and $\mathcal{D}_1,\ldots,\mathcal{D}_N$ be a sequence of random vectors. Then $S_N = O_p(1)$ if and only if $S_N = O_{p|\mathcal{D}_N}(1)$, where $p$ denotes the underlying probability measure and $p|\mathcal{D}_N$ denotes the probability measure conditional on $\mathcal{D}_N$.*

PROOF: By definition, $S_N = O_p(1)$ means that $P(|S_N| > \ell_N) = o(1)$ for any $\ell_N \to \infty$, while $S_N = O_{p|\mathcal{D}_N}(1)$ means that $P(|S_N| > \ell_N|\mathcal{D}_N) = o_p(1)$ for any $\ell_N \to \infty$. The second follows from the first by the Markov inequality because $E[P(|S_N| > \ell_N|\mathcal{D}_N)] = P(|S_N| > \ell_N) = o(1)$. Since $P(|S_N| > \ell_N|\mathcal{D}_N) \leq 1$ for all $N$, $\{P(|S_N| > \ell_N|\mathcal{D}_N)\}_{N\geq 1}$ are uniformly integrable. The first follows from the second by the fact that convergence in probability implies moments convergence for uniformly integrable sequences. ■

# Appendix B - Proof of Theorem 4.2

## B.1 Proof of Theorem 4.2

Proof of Theorem 4.2: Let us first look at (A.3). To improve the rate of $\hat{B}$ in Theorem 4.1, we cannot use the inequality in (A.4). Instead, we need to treat $D_5\hat{B}$ as a whole to establish its rate. By the Cauchy-Schwartz inequality and the fact that $\|C+D\|_F \leq \|C\|_F + \|D\|_F$ and $\|CD\|_F \leq \|C\|_2\|D\|_F$, (A.3) implies

$$\begin{aligned} \|\hat{B} - BH\|_F^2 &\leq 10\|\hat{B}\|_2^2\|V^{-1}\|_2^2 \left(\sum_{j\neq 5}^{6} \|D_j\|_F^2\right) + 2\|V^{-1}\|_2^2\|D_5\hat{B}\|_F^2, \\ &= O_p\left(\frac{1}{J^{2\kappa}} + \frac{J}{N^2} + \frac{J}{NT}\right), \end{aligned} \tag{B.1}$$

where the equality follows by $J = o(\sqrt{N})$, Lemmas A.1(i)-(iv), A.2(i) and B.1(ii) and the fact that $\|D_6\|_F = \|D_3\|_F$. Given the rate of $\|\hat{B} - BH\|_F^2$ in (B.1), the rate of $|\hat{a} - a|^2$ immediately follows from the same argument in (A.6). We now look at (A.7). To improve the rate of $\hat{F}$ in Theorem 4.1, we cannot use the inequality in (A.8). Instead, we need to plug in the expansion of $\hat{B} - BH$, and treat $a'D_4$, $D_4'\hat{B}$, $D_5\hat{B}$ and $\tilde{E}'\hat{B}$ as a whole to establish their rates. By the fact that $\|C+D\|_F \leq \|C\|_F + \|D\|_F$ and $\|CD\|_F \leq \|C\|_2\|D\|_F$, combining (A.3) and (A.7) implies

$$\begin{aligned} \|\hat{F} - F(H')^{-1}\|_F &= \left(\sum_{j\neq 4,5}^{6} \|D_j\|_F\|\hat{B}\|_2\|a\| + \|a'D_4\|\|\hat{B}\|_2 + \|a\|\|D_5\hat{B}\|_F\right)\|V^{-1}\|_2\|1_T\| \\ &\quad + \left(\sum_{j\neq 4,5}^{6} \|D_j\|_F\|\hat{B}\|_2 + \|D_4'\hat{B}\|_F + \|D_5\hat{B}\|_F\right) \\ &\quad \times \|F\|_2\|H^{-1}\|_2\|V^{-1}\|_2\|\hat{B}\|_2 + \|\tilde{\Delta}\|_F\|\hat{B}\|_2 + \|\tilde{E}'\hat{B}\|_F \\ &= O_p\left(\frac{\sqrt{T}}{J^{\kappa}} + \sqrt{\frac{T}{N}}\right), \end{aligned} \tag{B.2}$$

where the equality follows by $J = o(\sqrt{N})$, Assumptions 4.2(ii) and 4.4, Lemmas A.1(i)-(iii), A.2, A.3(i), B.1 and B.2(i) and the fact that $\|D_6\|_F = \|D_3\|_F$. Thus, the third result of the theorem follows from (B.2). The proofs of the last two results of the theorem are similar to the proofs of the last two results of Theorem 4.1. ■

## B.2 Technical Lemmas

**Lemma B.1.** *Let $D_4$ and $D_5$ be given in the proof of Theorem 4.1. Assume (i) $N \to \infty$; (ii) $T \to \infty$ or $T \geq K+1$ is finite; (iii) $J \to \infty$ with $J^2\xi_J^2 \log J = o(N)$.*

*(i) Under Assumptions 4.1-4.5,* $\|D_4'\hat{B}\|_F^2 = O_p(1/NT)$*.*
*(ii) Under Assumptions 4.1-4.5,* $\|D_5\hat{B}\|_F^2 = O_p(J/N^2)$*.*
*(iii) Under Assumptions 4.1-4.5,* $\|D_4'a\|^2 = O_p(1/NT)$*.*

Proof: (i) Since $\|D_4'\hat{B}\|_F \leq \|B\|_2\|\hat{B}'\tilde{E}M_TF\|_F/T$, the result then immediately follows from Assumption 4.2(i) and Lemma B.2(ii).

(ii) Since $\|M_T\|_2 = 1$, $\|D_5\hat{B}\|_F \leq \|\tilde{E}\|_F\|\hat{B}'\tilde{E}\|_F/T$. The result then immediately follows from Lemmas A.3(ii) and B.2(i).

(iii) Since $\|D_4'a\| \leq \|B\|_2\|a'\tilde{E}M_TF\|/T$, the result then immediately follows from Assumption 4.2(i) and Lemma B.2(iii). ■

**Lemma B.2.** *Let* $\tilde{E}$ *be given in the proof of Theorem 4.1. Assume (i)* $N \to \infty$*; (ii)* $T \to \infty$ *or* $T \geq K+1$ *is finite; (iii)* $J \to \infty$ *with* $J^2\xi_J^2 \log J = o(N)$*.*
*(i) Under Assumptions 4.1-4.5,* $\|\hat{B}'\tilde{E}\|_F^2/T = O_p(1/N)$*.*
*(ii) Under Assumptions 4.1-4.5,* $\|\hat{B}'\tilde{E}M_TF\|_F^2/T = O_p(1/N)$*.*
*(iii) Under Assumptions 4.1-4.5,* $\|a'\tilde{E}M_TF\|^2/T = O_p(1/N)$*.*

Proof: By the fact that $\|C+D\|_F \leq \|C\|_F + \|D\|_F$ and $\|CD\|_F \leq \|C\|_2\|D\|_F$,

$$\begin{aligned}\frac{1}{T}\|\hat{B}'\tilde{E}\|_F^2 &\leq \frac{2}{T}\|\tilde{E}\|_F^2\|\hat{B}-BH\|_F^2 + \frac{2}{T}\|H\|_2^2\|B'\tilde{E}\|_F^2 \\ &= \frac{2}{T}\|\tilde{E}\|_F^2\|\hat{B}-BH\|_F^2 + \frac{2}{N^2T}\|H\|_2^2\left(\sum_{t=1}^T \|B'\hat{Q}_t^{-1}\Phi(Z_t)'\varepsilon_t\|^2\right) \\ &= O_p\left(\frac{J}{N}\left(\frac{1}{J^{2\kappa}} + \frac{J^2}{N^2} + \frac{J}{NT}\right) + \frac{1}{N}\right) = O_p\left(\frac{1}{N}\right),\end{aligned} \tag{B.3}$$

where the second equality follows from $J^2\xi_J^2\log J = o(N)$, Lemmas A.2(i), A.3(ii) and B.3(i) and Theorem 4.1, and the last equality is due to $\kappa > 1/2$ and $J = o(\sqrt{N})$.

(ii) By the fact that $\|C+D\|_F \leq \|C\|_F + \|D\|_F$ and $\|CD\|_F \leq \|C\|_2\|D\|_F$,

$$\begin{aligned}\frac{1}{T}\|\hat{B}'\tilde{E}M_TF\|_F^2 &\leq \frac{2}{T}\|\tilde{E}M_TF\|_F^2\|\hat{B}-BH\|_F^2 + \frac{2}{T}\|H\|_2^2\|B'\tilde{E}M_TF\|_F^2 \\ &\leq \frac{2}{T}\|\tilde{E}M_TF\|_F^2\|\hat{B}-BH\|_F^2 + \frac{4}{N^2T}\|H\|_2^2\left\|\sum_{t=1}^T B'\hat{Q}_t^{-1}\Phi(Z_t)'\varepsilon_t f_t'\right\|_F^2 \\ &\quad + \frac{4\|\bar{f}\|^2}{N^2T}\|H\|_2^2\left\|\sum_{t=1}^T B'\hat{Q}_t^{-1}\Phi(Z_t)'\varepsilon_t\right\|^2 \\ &= O_p\left(\frac{J}{N}\left(\frac{1}{J^{2\kappa}} + \frac{J^2}{N^2} + \frac{J}{NT}\right) + \frac{1}{N}\right) = O_p\left(\frac{1}{N}\right),\end{aligned} \tag{B.4}$$

where the first equality follows from $J^2\xi_J^2\log J = o(N)$, Assumption 4.2(ii), Lemmas A.2(i), A.3(iii) and B.3(ii) and Theorem 4.1, and the last equality is due to $\kappa > 1/2$ and $J = o(\sqrt{N})$.

(iii) By the fact that $\|x+y\| \leq \|x\| + \|y\|$,

$$\frac{1}{T}\|a'\tilde{E}M_TF\|^2 \leq \frac{2}{N^2T}\left\|\sum_{t=1}^T a'\hat{Q}_t^{-1}\Phi(Z_t)'\varepsilon_t f_t'\right\|^2 + \frac{2\|\bar{f}\|^2}{N^2T}\left|\sum_{t=1}^T a'\hat{Q}_t^{-1}\Phi(Z_t)'\varepsilon_t\right|^2$$
$$= O_p\left(\frac{1}{N}\right), \tag{B.5}$$

where the equality follows from $J^2\xi_J^2\log J = o(N)$, Assumption 4.2(ii) and Lemma B.3(ii). ■

**Lemma B.3.** *Assume $J \geq 2$ and $\xi_J^2 \log J = o(N)$.*
*(i) Under Assumptions 4.1(i), 4.2(i), 4.3 and 4.5,*

$$\sum_{t=1}^T \|B'\hat{Q}_t^{-1}\Phi(Z_t)'\varepsilon_t\|^2 = O_p\left(NT\left(1 + \frac{J\xi_J^2\log J}{N}\right)\right).$$

*(ii) Under Assumptions 4.1(i), 4.2(i), (ii), 4.3, 4.4 and 4.5,*

$$\left\|\sum_{t=1}^T B'\hat{Q}_t^{-1}\Phi(Z_t)'\varepsilon_t f_t'\right\|_F^2 = O_p\left(NT\left(1 + \frac{J\xi_J\sqrt{\log J}}{\sqrt{N}}\right)\right),$$
$$\left\|\sum_{t=1}^T B'\hat{Q}_t^{-1}\Phi(Z_t)'\varepsilon_t\right\|^2 = O_p\left(NT\left(1 + \frac{J\xi_J\sqrt{\log J}}{\sqrt{N}}\right)\right),$$
$$\left\|\sum_{t=1}^T a'\hat{Q}_t^{-1}\Phi(Z_t)'\varepsilon_t f_t'\right\|^2 = O_p\left(NT\left(1 + \frac{J\xi_J\sqrt{\log J}}{\sqrt{N}}\right)\right),$$
$$\left|\sum_{t=1}^T a'\hat{Q}_t^{-1}\Phi(Z_t)'\varepsilon_t\right|^2 = O_p\left(NT\left(1 + \frac{J\xi_J\sqrt{\log J}}{\sqrt{N}}\right)\right).$$

Proof: (i) Let $Q_t \equiv E[\hat{Q}_t]$. By the fact that $\|x+y\| \leq \|x\| + \|y\|$,

$$\sum_{t=1}^T \|B'\hat{Q}_t^{-1}\Phi(Z_t)'\varepsilon_t\|^2 \leq 2\sum_{t=1}^T \|B'Q_t^{-1}\Phi(Z_t)'\varepsilon_t\|^2$$
$$+ 2\sum_{t=1}^T \|B'(\hat{Q}_t^{-1} - Q_t^{-1})\Phi(Z_t)'\varepsilon_t\|^2 \equiv 2\mathcal{T}_1 + 2\mathcal{T}_2. \tag{B.6}$$

Therefore, it suffices to show that $\mathcal{T}_1 = O_p(NT)$ and $\mathcal{T}_2 = O_p(TJ\xi_J^2\log J)$. The former holds by the Markov's inequality, since

$$E[\mathcal{T}_1] = E\left[\sum_{t=1}^T\sum_{i=1}^N\sum_{j=1}^N \phi(z_{it})'Q_t^{-1}BB'Q_t^{-1}\phi(z_{jt})\varepsilon_{it}\varepsilon_{jt}\right]$$

$$= \sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N} E[\phi(z_{it})' Q_t^{-1} B B' Q_t^{-1} \phi(z_{jt})] E[\varepsilon_{it}\varepsilon_{jt}]$$

$$\leq T \max_{i\leq N, t\leq T} E[\|B' Q_t^{-1} \phi(z_{it})\|^2] \max_{t\leq T} \sum_{i=1}^{N}\sum_{j=1}^{N} |E[\varepsilon_{it}\varepsilon_{jt}]| = O(NT), \quad \text{(B.7)}$$

where the second equality follows by the independence in Assumption 4.3(i), the inequality is due to the Cauchy-Schwartz inequality, and the last equality follows from Assumption 4.3(iii) and Lemma B.4. The latter also holds, since

$$\mathcal{T}_2 \leq C_{NT} \sum_{t=1}^{T} \|\hat{Q}_t - Q_t\|_2^2 \|\Phi(Z_t)'\varepsilon_t\|^2$$

$$\leq C_{NT} \left(\sum_{t=1}^{T} \|\hat{Q}_t - Q_t\|_2^4\right)^{1/2} \left(\sum_{t=1}^{T} \|\Phi(Z_t)'\varepsilon_t\|^4\right)^{1/2} = O_p(TJ\xi_J^2 \log J), \quad \text{(B.8)}$$

where $C_{NT} = \|B\|_2^2 (\min_{t\leq T} \lambda_{\min}(\hat{Q}_t))^{-2} (\min_{i\leq N, t\leq T} \lambda_{\min}(Q_{it}))^{-2}$, the first inequality follows since $\min_{t\leq T} \lambda_{\min}(Q_t) \geq \min_{i\leq N,t\leq T} \lambda_{\min}(Q_{it})$, the second inequality is due to the Cauchy-Schwartz inequality, and the equality follows from Assumptions 4.1(i), 4.2(i) and 4.5(ii) and Lemmas B.5 and B.6. This completes the proof of (i).

(ii) Let $Q_t \equiv E[\hat{Q}_t]$. By the fact that $\|C + D\|_F \leq \|C\|_F + \|D\|_F$,

$$\left\|\sum_{t=1}^{T} B' \hat{Q}_t^{-1} \Phi(Z_t)'\varepsilon_t f_t'\right\|_F^2 \leq 2 \left\|\sum_{t=1}^{T} B' Q_t^{-1} \Phi(Z_t)'\varepsilon_t f_t'\right\|_F^2$$

$$+ 2 \left\|\sum_{t=1}^{T} B' (\hat{Q}_t^{-1} - Q_t^{-1}) \Phi(Z_t)'\varepsilon_t f_t'\right\|_F^2 \equiv 2\mathcal{T}_1 + 2\mathcal{T}_2. \quad \text{(B.9)}$$

Therefore, it suffices to show that $\mathcal{T}_1 = O_p(NT)$ and $\mathcal{T}_2 = O_p(\sqrt{N}TJ\xi_J\sqrt{\log J})$. Note that $\|A\|_F^2 = \text{tr}(AA')$. The former holds by the Markov's inequality, since

$$E[\mathcal{T}_1] = E\left[\text{tr}\left(\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N} B' Q_t^{-1} \phi(z_{it}) \varepsilon_{it} f_t' f_s \varepsilon_{js} \phi(z_{js})' Q_s^{-1} B\right)\right]$$

$$= \sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N} E\left[\phi(z_{it})' Q_t^{-1} B B' Q_s^{-1} \phi(z_{js})\right] f_t' f_s E[\varepsilon_{it}\varepsilon_{js}]$$

$$\leq C_{NT} \max_{i\leq N, t\leq T} E[\|B' Q_t^{-1} \phi(z_{it})\|^2] \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} |E[\varepsilon_{it}\varepsilon_{js}]| = O(NT), \quad \text{(B.10)}$$

where $C_{NT} = \max_{t\leq T} \|f_t\|^2$, the second equality follows by the independence in Assumption 4.3(i) and the fact that both expectation and trace operators are linear, the inequality is due to the Cauchy-Schwartz inequality, and the last equality follows from

Assumptions 4.2(ii) and 4.3(iii) and Lemma B.4. Let $E_\varepsilon$ denote the expectation with respect to $\{\varepsilon_t\}_{t\leq T}$. For the latter, we have

$$\begin{aligned}
E_\varepsilon[\mathcal{T}_2] &= E_\varepsilon\left[\operatorname{tr}\left(\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N} B'(\hat{Q}_t^{-1}-Q_t^{-1})\phi(z_{it})\varepsilon_{it}f_t'f_s\varepsilon_{js}\phi(z_{js})'(\hat{Q}_s^{-1}-Q_s^{-1})B\right)\right] \\
&= \sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N} \phi(z_{it})'(\hat{Q}_t^{-1}-Q_t^{-1})BB'(\hat{Q}_s^{-1}-Q_s^{-1})\phi(z_{js})f_t'f_sE[\varepsilon_{it}\varepsilon_{js}] \\
&\leq C_{NT}^{*}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N} \|\hat{Q}_t-Q_t\|_2\|\phi(z_{it})\|\|\phi(z_{js})\||E[\varepsilon_{it}\varepsilon_{js}]| \\
&\leq C_{NT}^{**}\left(\sum_{t=1}^{T}\left(\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\|\phi(z_{it})\|\|\phi(z_{js})\||E[\varepsilon_{it}\varepsilon_{js}]|\right)^2\right)^{1/2}, \qquad \text{(B.11)}
\end{aligned}$$

where $C_{NT}^{*} = \|B\|_2^2\max_{t\leq T}\|f_t\|^2[(\min_{t\leq T}\lambda_{\min}(\hat{Q}_t))^{-1} + (\min_{i\leq N,t\leq T}\lambda_{\min}(Q_{it}))^{-1}]$ $(\min_{t\leq T}\lambda_{\min}(\hat{Q}_t))^{-1}(\min_{i\leq N,t\leq T}\lambda_{\min}(Q_{it}))^{-1}$ and $C_{NT}^{**} = C_{NT}^{*}(\sum_{t=1}^{T}\|\hat{Q}_t - Q_t\|_2^2)^{1/2}$, the second equality follows by the independence in Assumption 4.3(i) and the fact that both expectation and trace operators are linear, the first inequality follows since $\min_{t\leq T}\lambda_{\min}(Q_t) \geq \min_{i\leq N,t\leq T}\lambda_{\min}(Q_{it})$, and the last inequality is due to the Cauchy-Schwartz inequality. Moreover, we have

$$\begin{aligned}
&E\left[\sum_{t=1}^{T}\left(\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\|\phi(z_{it})\|\|\phi(z_{js})\||E[\varepsilon_{it}\varepsilon_{js}]|\right)^2\right] \\
&\leq \max_{i\leq N,t\leq T}E[\|\phi(z_{it})\|^4]\sum_{t=1}^{T}\left(\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}|E[\varepsilon_{it}\varepsilon_{js}]|\right)^2 \\
&\leq J^2M^2\max_{m\leq M,j\leq J,i\leq N,t\leq T}E[\phi_j^4(z_{it,m})]\sum_{t=1}^{T}\left(\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}|E[\varepsilon_{it}\varepsilon_{js}]|\right)^2, \qquad \text{(B.12)}
\end{aligned}$$

where the first inequality is due to the Cauchy-Schwartz inequality, the second inequality follows since $\max_{i\leq N,t\leq T}E[\|\phi(z_{it})\|^4] \leq J^2M^2\max_{m\leq M,j\leq J,i\leq N,t\leq T}E[\phi_j^4(z_{it,m})]$. By Assumptions 4.1(i), 4.2(i), (ii) and 4.5(ii) and Lemma B.6, $C_{NT}^{**} = O_p(\sqrt{T}\xi_J\sqrt{\log J}/\sqrt{N})$. Combining this, (B.11) and (B.12) implies that $E_\varepsilon[\mathcal{T}_2] = O_p(\sqrt{N}TJ\xi_J\sqrt{\log J})$ by Assumptions 4.5(i) and (iv). Thus, the latter—$\mathcal{T}_2 = O_p(\sqrt{N}TJ\xi_J\sqrt{\log J})$—holds by the Markov's inequality and Lemma A.5. This completes the proof of the first result of (ii), and the proofs of the other three are similar. ■

**Lemma B.4.** *Suppose Assumptions 4.2(i), 4.4 and 4.5(ii) hold. Let $Q_t \equiv E[\hat{Q}_t]$. Then*

$$\max_{i\leq N,t\leq T}E[\|B'Q_t^{-1}\phi(z_{it})\|^2] < \infty \text{ and } \max_{i\leq N,t\leq T}E[|a'Q_t^{-1}\phi(z_{it})|^2] < \infty.$$

PROOF: Since $\|x\|^2 = \mathrm{tr}(xx')$,

$$\begin{aligned} E[\|B'Q_t^{-1}\phi(z_{it})\|^2] &= E[\mathrm{tr}(B'Q_t^{-1}\phi(z_{it})\phi(z_{it})'Q_t^{-1}B)] = \mathrm{tr}(B'Q_t^{-1}Q_{it}Q_t^{-1}B) \\ &\leq \max_{i\leq N, t\leq T} \lambda_{\max}(Q_{it}) \left( \min_{t\leq T} \lambda_{\min}(Q_t) \right)^{-1} K\|B\|_2^2 \\ &\leq \max_{i\leq N, t\leq T} \lambda_{\max}(Q_{it}) \left( \min_{i\leq N, t\leq T} \lambda_{\min}(Q_{it}) \right)^{-1} K\|B\|_2^2, \end{aligned} \tag{B.13}$$

where the second equality follows by the fact that both expectation and trace operators are linear, the first inequality follows since $\mathrm{tr}(B'B) = \|B\|_F^2 \leq K\|B\|_2^2$, and the second inequality follows since $\min_{t\leq T} \lambda_{\min}(Q_t) \geq \min_{i\leq N, t\leq T} \lambda_{\min}(Q_{it})$. Thus, the first result of the lemma follows from (B.13), Assumptions 4.2(i) and 4.5(ii). The proof of the second result is similar. ■

**Lemma B.5.** *Under Assumptions 4.3(i), 4.5(i) and (iv),*

$$\sum_{t=1}^{T} \|\Phi(Z_t)'\varepsilon_t\|^4 = O_p(N^2TJ^2).$$

PROOF: The result follows by the Markov's inequality, since

$$\begin{aligned} E\left[\sum_{t=1}^{T} \|\Phi(Z_t)'\varepsilon_t\|^4\right] &= E\left[\sum_{t=1}^{T} \left(\sum_{i=1}^{N}\sum_{j=1}^{N} \phi(z_{it})'\phi(z_{jt})\varepsilon_{it}\varepsilon_{jt}\right)^2\right] \\ &= E\left[\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{\ell=1}^{N} \phi(z_{it})'\phi(z_{jt})\phi(z_{kt})'\phi(z_{\ell t})\varepsilon_{it}\varepsilon_{jt}\varepsilon_{kt}\varepsilon_{\ell t}\right] \\ &= \sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{\ell=1}^{N} E[\phi(z_{it})'\phi(z_{jt})\phi(z_{kt})'\phi(z_{\ell t})]E[\varepsilon_{it}\varepsilon_{jt}\varepsilon_{kt}\varepsilon_{\ell t}] \\ &\leq \max_{i\leq N, t\leq T} E[\|\phi(z_{it})\|^4] \sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{\ell=1}^{N} |E[\varepsilon_{it}\varepsilon_{jt}\varepsilon_{kt}\varepsilon_{\ell t}]| \\ &\leq J^2M^2 \max_{m\leq M, j\leq J, i\leq N, t\leq T} E[\phi_j^4(z_{it,m})] \sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{\ell=1}^{N} |E[\varepsilon_{it}\varepsilon_{jt}\varepsilon_{kt}\varepsilon_{\ell t}]| \\ &= O(N^2TJ^2), \end{aligned} \tag{B.14}$$

where the third equality follows by the independence in Assumption 4.3(i), the first inequality is due to the Cauchy Schwartz inequality, the second inequality follows since $\max_{i\leq N, t\leq T} E[\|\phi(z_{it})\|^4] \leq J^2M^2 \max_{m\leq M, j\leq J, i\leq N, t\leq T} E[\phi_j^4(z_{it,m})]$, and the last equality follows from Assumptions 4.5(i) and (iv). ■

**Lemma B.6.** *Suppose Assumptions 4.5(ii) and (iii) hold. Let $Q_t \equiv E[\hat{Q}_t]$. Assume*

$J \geq 2$ *and* $\xi_J^2 \log J = o(N)$. *Then*

$$\sum_{t=1}^{T} \|\hat{Q}_t - Q_t\|_2^2 = O_p\left(\frac{T\xi_J^2 \log J}{N}\right) \text{ and } \sum_{t=1}^{T} \|\hat{Q}_t - Q_t\|_2^4 = O_p\left(\frac{T\xi_J^4 \log^2 J}{N^2}\right).$$

Proof: Recall that $\hat{Q}_t = \sum_{i=1}^{N} \phi(z_{it})\phi(z_{it})'/N$. Let $\eta_1, \ldots, \eta_N$ be an i.i.d. sequence of Rademacher variables. It then follows that

$$\begin{aligned}
\mathcal{D}_t &\equiv E[\|\hat{Q}_t - Q_t\|_2^4] \\
&\leq 16E\left[\left\|\frac{1}{N}\sum_{i=1}^{N} \eta_i \phi(z_{it})\phi(z_{it})'\right\|_2^4\right] \\
&\leq 16C\frac{\log^2 JM}{N^2} \sup_z \|\phi(z)\|^4 E\left[\left\|\frac{1}{N}\sum_{i=1}^{N} \phi(z_{it})\phi(z_{it})'\right\|_2^2\right] \\
&\leq 16M^2C\frac{\xi_J^4 \log^2 JM}{N^2} E[\|\hat{Q}_t\|_2^2], \qquad \text{(B.15)}
\end{aligned}$$

where the first inequality follows from the independence in Assumption 4.5(iii) and the symmetrization lemma (e.g., Lemma 2.3.1 of van der Vaart and Wellner (1996)), the second inequality follows by Lemma B.7 and the fact that $\phi(z_{it})'\phi(z_{it}) \leq \sup_z \|\phi(z)\|^2$, the third inequality follows since $\sup_z \|\phi(z)\|^2 \leq M \sup_z \|\bar{\phi}(z)\|^2 = M\xi_J^2$. Let $A = 16M^2C\xi_J^4 \log^2 JM/N^2$. Combining $E[\|\hat{Q}_t\|_2^2] \leq 2\sqrt{\mathcal{D}_t} + 2\|Q_t\|_2^2$ and (B.15) leads to the inequality: $\mathcal{D}_t \leq 2A(\sqrt{\mathcal{D}_t} + \|Q_t\|_2^2)$. Solving the inequality yields

$$E[\|\hat{Q}_t - Q_t\|_2^4] \leq \left(A + \sqrt{A^2 + 2A\|Q_t\|_2^2}\right)^2. \qquad \text{(B.16)}$$

Thus, by the fact that $\max_{t \leq T} \|Q_t\|_2 \leq \max_{i \leq N, t \leq T} \lambda_{\max}(Q_{it})$ and the Markov's inequality, the second result of the lemma follows from (B.16) and Assumption 4.5(ii). The first result of the lemma follows similarly by noting that $E[\|\hat{Q}_t - Q_t\|_2^2] \leq (E[\|\hat{Q}_t - Q_t\|_2^4])^{1/2}$. This completes the proof of the lemma. ■

**Lemma B.7** (Khinchin inequality). *Let $S_1, \ldots, S_N$ be a sequence of symmetric $k \times k$ matrices and $\eta_1, \ldots, \eta_N$ be an i.i.d. sequence of Rademacher variables. Then for $k \geq 2$,*

$$E_\eta\left[\left\|\frac{1}{N}\sum_{i=1}^{N} \eta_i S_i\right\|_2^4\right] \leq C\frac{\log^2 k}{N^2}\left\|\frac{1}{N}\sum_{i=1}^{N} S_i^2\right\|_2^2$$

*for some positive constant $C$, where $E_\eta$ denotes the expectation with respect to $\{\eta_i\}_{i \leq N}$.*

Proof: This is a modified version of Lemma 6.1 in Belloni et al. (2015). The result is trivial for $2 \leq k \leq e^6$. For $k > e^6$, we have

$$
\begin{aligned}
E_\eta\left[\left\|\frac{1}{N}\sum_{i=1}^N \eta_i S_i\right\|_2^4\right] &\leq E_\eta\left[\left\|\frac{1}{N}\sum_{i=1}^N \eta_i S_i\right\|_{S_{\log k}}^4\right] \\
&\leq \left(E_\eta\left[\left\|\frac{1}{N}\sum_{i=1}^N \eta_i S_i\right\|_{S_{\log k}}^{\log k}\right]\right)^{4/\log k} \\
&\leq C_0^4 \frac{\log^2 k}{N^2}\left\|\left(\frac{1}{N}\sum_{i=1}^N S_i^2\right)^{1/2}\right\|_{S_{\log k}}^4 \\
&\leq C_0^4 e^4 \frac{\log^2 k}{N^2}\left\|\frac{1}{N}\sum_{i=1}^N S_i^2\right\|_2^2,
\end{aligned}
\tag{B.17}
$$

where the first inequality follows by (6.44) in Belloni et al. (2015) and $\|\cdot\|_{S_{\log k}}$ is the Schatten norm, the second inequality follows by the Jensen's inequality, the third inequality follows by (6.45) in Belloni et al. (2015) and $C_0$ is some positive constant, and the fourth inequality follows by (6.44) in Belloni et al. (2015) again. Thus, the result of the lemma follows by setting $C = C_0^4 e^4$. ■

# Appendix C - Proof of Theorem 4.3

## C.1 Proof of Theorem 4.3

Proof of Theorem 4.3: Let us first look at (B.1). The asymptotic distribution can be obtained by choosing large $J$ and assuming $T$ not too large such that the terms with $O_p(J^{-2\kappa})$ and $O_p(J/N^2)$ are negligible relative to the term with $O_p(J/NT)$. Thus, the asymptotic distribution is determined by the term with $O_p(J/NT)$. Specifically, by the fact that $\|C + D\|_F \leq \|C\|_F + \|D\|_F$ and $\|CD\|_F \leq \|C\|_2\|D\|_F$, (A.3) implies

$$
\begin{aligned}
\|\sqrt{NT}(\hat{B} - BH) - \sqrt{NT}D_4\hat{B}V^{-1}\|_F &\leq \sqrt{NT}\|V^{-1}\|_2\|D_5\hat{B}\|_F \\
&+ \sqrt{NT}\|\hat{B}\|_2\|V^{-1}\|_2\sum_{j\neq 4,5}^{6}\|D_j\|_F = O_p\left(\frac{\sqrt{NT}}{J^\kappa} + \frac{\sqrt{TJ}}{\sqrt{N}}\right),
\end{aligned}
\tag{C.1}
$$

where the equality follows by $J = o(\sqrt{N})$, Lemmas A.1(i)-(iii), A.2(i) and B.1(ii) and the fact that $\|D_6\|_F = \|D_3\|_F$. Let $\mathcal{L}_{NT} \equiv \sum_{t=1}^T Q_t^{-1}\Phi(Z_t)'\varepsilon_t(f_t - \bar{f})'/\sqrt{NT}$. Since $J = o(\sqrt{N})$, $J^{(1/2-\kappa)} = o(\sqrt{NT}/J^\kappa)$. By the fact that $\|C + D\|_F \leq \|C\|_F + \|D\|_F$, combining (C.1) and Lemma C.1 implies

$$
\|\sqrt{NT}(\hat{B} - BH) - \mathcal{L}_{NT}B'B\mathcal{M}\|_F = O_p\left(\frac{\sqrt{NT}}{J^\kappa} + \frac{\sqrt{TJ}}{\sqrt{N}} + \frac{\sqrt{J\xi_J}\log^{1/4} J}{N^{1/4}}\right). \tag{C.2}
$$

Note that $\mathbb{N}_2$ is a $JM \times K$ matrix from the last $K$ columns of $\mathbb{N}$. Thus, the second result of the theorem follows from (C.2) and Lemma C.2. We now look at (A.5). By the fact that $\|x+y\| \leq \|x\| + \|y\|$, it implies

$$\|\sqrt{NT}(\hat{a}-a) - (I_{JM} - \hat{B}\hat{B}')[\sqrt{N/T}\tilde{E}1_T - \sqrt{NT}(\hat{B}-BH)H^{-1}\bar{f}] \\ + \hat{B}\sqrt{NT}(\hat{B}-BH)'a\| \leq \|(I_{JM} - \hat{B}\hat{B}')\sqrt{N/T}\tilde{\Delta}1_T\| = O_p\left(\frac{\sqrt{NT}}{J^{\kappa}}\right), \tag{C.3}$$

where the equality follows by Lemma A.3(i). Given the rate of $\|\hat{B} - BH\|_F$ in Theorem 4.2 and the rate of $\|N\tilde{E}1_T\|$ in Lemma A.4(ii), we may replace all $\hat{B}$ except those in $\hat{B} - BH$ with $BH$ to obtain

$$\|\sqrt{NT}(\hat{a}-a) - (I_{JM} - BHH'B')[\sqrt{N/T}\tilde{E}1_T - \sqrt{NT}(\hat{B}-BH)H^{-1}\bar{f}] \\ + BH\sqrt{NT}(\hat{B}-BH)'a\| = O_p\left(\frac{\sqrt{NT}}{J^{\kappa}} + \frac{\sqrt{TJ}}{\sqrt{N}} + \frac{J}{\sqrt{NT}}\right) \tag{C.4}$$

by noting that $J = o(\sqrt{N})$ and $J^{(1/2-\kappa)} = o(\sqrt{NT}/J^{\kappa})$. Similarly, given the rate of $H - \mathcal{H}$ in Lemma C.3, we may replace all $H$ except those in $\hat{B} - BH$ with $\mathcal{H}$ to obtain

$$\|\sqrt{NT}(\hat{a}-a) - (I_{JM} - B'\mathcal{H}\mathcal{H}'B')[\sqrt{N/T}\tilde{E}1_T - \sqrt{NT}(\hat{B}-BH)\mathcal{H}^{-1}\bar{f}] \\ + B\mathcal{H}\sqrt{NT}(\hat{B}-BH)'a\| = O_p\left(\frac{\sqrt{NT}}{J^{\kappa}} + \frac{\sqrt{TJ}}{\sqrt{N}} + \frac{J}{\sqrt{NT}}\right) \tag{C.5}$$

Let $\ell_{NT} \equiv \sum_{t=1}^{T} Q_t^{-1}\Phi(Z_t)'\varepsilon_t/\sqrt{NT}$. Given the rate of $\|\sqrt{N/T}\tilde{E}1_T - \ell_{NT}\|$ in Lemma C.1, we may replace $\sqrt{N/T}\tilde{E}1_T$ with $\ell_{NT}$ to obtain

$$\|\sqrt{NT}(\hat{a}-a) - (I_{JM} - B'\mathcal{H}\mathcal{H}'B')[\ell_{NT} - \sqrt{NT}(\hat{B}-BH)\mathcal{H}^{-1}\bar{f}] \\ + B\mathcal{H}\sqrt{NT}(\hat{B}-BH)'a\| = O_p\left(\frac{\sqrt{NT}}{J^{\kappa}} + \frac{\sqrt{TJ}}{\sqrt{N}} + \frac{\sqrt{J\xi_J}\log^{1/4} J}{N^{1/4}}\right) \tag{C.6}$$

by noting that $J/\sqrt{NT} = o(\sqrt{J\xi_J}\log^{1/4} J/N^{1/4})$. The arguments in (C.4)-(C.6) are similar to those for the first result in Lemma C.1. Note that $\mathbb{N}_1$ is a $JM \times 1$ vector from the first column of $\mathbb{N}$. Thus, the first result of the theorem follows from (C.6), Lemma C.2 and the second result of the theorem. This completes proof of the theorem. ■

## C.2 Technical Lemmas

**Lemma C.1.** *Suppose Assumptions 4.1-4.5, 4.6(i) and (ii) hold. Let $\tilde{E}$, $D_4$ and $V$ be given in the proof of Theorem 4.1, and $\ell_{NT}$ and $\mathcal{L}_{NT}$ be given in the proof of Theorem 4.3. Assume (i) $N \to \infty$; (ii) $T \to \infty$ or $T \geq K+1$ is finite; (iii) $J \to \infty$ with*

$\xi_J^2 \log J = o(N)$. *Then*

$$\|\sqrt{NT}D_4\hat{B}V^{-1} - \mathcal{L}_{NT}B'B\mathcal{M}\|_F = O_p\left(\frac{1}{J^{(\kappa-1/2)}} + \frac{\sqrt{J\xi_J}\log^{1/4} J}{N^{1/4}}\right)$$

*and*

$$\|\sqrt{N/T}\tilde{E}1_T - \ell_{NT}\| = O_p\left(\frac{\sqrt{J\xi_J}\log^{1/4} J}{N^{1/4}}\right),$$

*where $\mathcal{M}$ is a nonrandom matrix given in Lemma C.3.*

Proof: For the first result, we have the following decomposition

$$\begin{aligned}\sqrt{NT}D_4\hat{B}V^{-1} &= \sqrt{N/T}\tilde{E}M_TFB'B\mathcal{M} + \sqrt{N/T}\tilde{E}M_TFB'(\hat{B} - BH)V^{-1} \\ &\quad + \sqrt{N/T}\tilde{E}M_TFB'B(HV^{-1} - \mathcal{M}) \equiv \mathcal{T}_1 + \mathcal{T}_2 + \mathcal{T}_3. \end{aligned} \tag{C.7}$$

Therefore, it suffices to show that $\|\mathcal{T}_1 - \mathcal{L}_{NT}B'B\mathcal{M}\|_F = O_p(\sqrt{J\xi_J}\log^{1/4} J/N^{1/4})$, $\|\mathcal{T}_2\|_F = O_p(J^{(1/2-\kappa)} + J^{3/2}/N + J/\sqrt{NT})$ and $\|\mathcal{T}_3\|_F = O_p(J^{(1/2-\kappa)} + J^{3/2}/N + J/\sqrt{NT})$. The first one holds, since

$$\begin{aligned}\|\mathcal{T}_1 - \mathcal{L}_{NT}B'B\mathcal{M}\|_F &\le \|B\|_2^2\|\mathcal{M}\|_2 \left\|\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}(\hat{Q}_t^{-1} - Q_t^{-1})\Phi(Z_t)'\varepsilon_t f_t'\right\|_F \\ &\quad + \|B\|_2^2\|\mathcal{M}\|_2\|\bar{f}\| \left\|\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}(\hat{Q}_t^{-1} - Q_t^{-1})\Phi(Z_t)'\varepsilon_t\right\| \\ &= O_p\left(\frac{\sqrt{J\xi_J}\log^{1/4} J}{N^{1/4}}\right),\end{aligned} \tag{C.8}$$

where the equality follows from Assumptions 4.2(i) and (ii) and Lemma C.4. The latter two follow by a similar argument. The second result also follows by a similar argument as in (C.8). This completes the proof of the lemma. ■

**Lemma C.2.** *Suppose Assumptions 4.2(ii), 4.3(i), (ii), 4.5(i)-(iii), 4.6(ii) and (iii) hold. Let $\ell_{NT}$ and $\mathcal{L}_{NT}$ be given in the proof of Theorem 4.3. Then there exists a $JM \times (K+1)$ random matrix $\mathbb{N}$ with $\mathrm{vec}(\mathbb{N}) \sim N(0, \Omega)$ such that*

$$\|(\ell_{NT}, \mathcal{L}_{NT}) - \mathbb{N}\|_F = O_p\left(\frac{J^{5/6}}{N^{1/6}}\right).$$

Proof: Let $\zeta_i \equiv \sum_{t=1}^{T} f_t^\dagger \otimes Q_t^{-1}\phi(z_{it})\varepsilon_{it}/\sqrt{NT}$. Then $\mathrm{vec}((\ell_{NT}, \mathcal{L}_{NT})) = \sum_{i=1}^{N}\zeta_i$. Note that $E[\zeta_i] = 0$ by Assumptions 4.3(i) and (ii) and $\zeta_1, \ldots, \zeta_N$ are independent by

Assumptions 4.3(i), 4.5(iii) and 4.6(ii). Moreover,

$$\sum_{i=1}^{N} E[\|\zeta_i\|^3] \leq \sum_{i=1}^{N} (E[\|\zeta_i\|^4])^{3/4} = O\left(\frac{J^{3/2}}{\sqrt{N}}\right), \tag{C.9}$$

where the inequality follows by the Liapounovs inequality, and the equality follows from Assumptions 4.2(ii), 4.5(i), (ii) and 4.6(iii) since

$$\begin{aligned} E[\|\zeta_i\|^4] &= \frac{1}{N^2T^2} E\left[\left(\sum_{t=1}^{T}\sum_{s=1}^{T} \phi(z_{it})' Q_t^{-1} Q_s^{-1} \phi(z_{is}) f_t^{\dagger\prime} f_s^{\dagger} \varepsilon_{it}\varepsilon_{is}\right)^2\right] \\ &\leq C_{NT} \max_{i\leq N, t\leq T} E[\|\phi(z_{it})\|^4] \frac{1}{N^2T^2} \sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{u=1}^{T}\sum_{v=1}^{T} |E[\varepsilon_{it}\varepsilon_{is}\varepsilon_{iu}\varepsilon_{iv}]| \\ &\leq C_{NT} \max_{m\leq M, j\leq J, i\leq N, t\leq T} E[\phi_j^4(z_{it,m})] \frac{J^2M^2}{N^2T^2} \sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{u=1}^{T}\sum_{v=1}^{T} |E[\varepsilon_{it}\varepsilon_{is}\varepsilon_{iu}\varepsilon_{iv}]|, \end{aligned} \tag{C.10}$$

where $C_{NT} = \max_{t\leq T} \|f_t^{\dagger}\|^4 (\min_{i\leq N, t\leq T} \lambda_{\min}(Q_{it}))^{-4}$, the first inequality follows by the independence in Assumption 4.3(i), the Cauchy-Schwartz inequality, and the fact that $\min_{t\leq T} \lambda_{\min}(Q_t) \geq \min_{i\leq N, t\leq T} \lambda_{\min}(Q_{it})$, and the second inequality follows since $\max_{i\leq N, t\leq T} E[\|\phi(z_{it})\|^4] \leq J^2M^2 \max_{m\leq M, j\leq J, i\leq N, t\leq T} E[\phi_j^4(z_{it,m})]$. In addition, $\Omega = E[\text{vec}((\ell_{NT}, \mathcal{L}_{NT}))\text{vec}((\ell_{NT}, \mathcal{L}_{NT}))']$. Thus, Lemma C.5 implies that there exists a $JM \times (K+1)$ random matrix $\mathbb{N}$ with $\text{vec}(\mathbb{N}) \sim N(0, \Omega)$ such that

$$\|(\mathcal{L}_{NT}, \ell_{NT}) - \mathbb{N}\|_F = \|\text{vec}((\mathcal{L}_{NT}, \ell_{NT})) - \text{vec}(\mathbb{N})\| = O_p\left(\frac{J^{5/6}}{N^{1/6}}\right). \tag{C.11}$$

This completes the proof of the Lemma. ■

**Lemma C.3.** *Suppose Assumptions 4.1-4.4 and 4.6(i) hold. Let $V$ be given in the proof of Theorem 4.1. Assume (i) $N \to \infty$; (ii) $T \to \infty$ or $T \geq K+1$ is finite; (iii) $J \to \infty$ with $J = o(\sqrt{N})$. Then*

$$H = \mathcal{H} + O_p\left(\frac{1}{J^{\kappa}} + \frac{J}{N} + \frac{\sqrt{J}}{\sqrt{NT}}\right) \text{ and } HV^{-1} = \mathcal{M} + O_p\left(\frac{1}{J^{\kappa}} + \frac{J}{N} + \frac{\sqrt{J}}{\sqrt{NT}}\right),$$

*where $\mathcal{H} = (F'M_TF/T)^{1/2}\Upsilon\mathcal{V}^{-1/2}$, $\mathcal{M} = \mathcal{H}\mathcal{V}^{-1}$, $\mathcal{V}$ is a diagonal matrix of the eigenvalues of $(F'M_TF/T)^{1/2}\ B'B(F'M_TF/T)^{1/2}$ and $\Upsilon$ is the corresponding eigenvector matrix such that $\Upsilon'\Upsilon = I_K$.*

Proof: By the definition of $\hat{B}$, $(\tilde{Y}M_T\tilde{Y}'/T)\hat{B} = \hat{B}V$. Pre-multiply $(\tilde{Y}M_T\tilde{Y}'/T)\hat{B} = \hat{B}V$ on both sides by $(F'M_TF/T)^{1/2}B'$ to obtain

$$(F'M_TF/T)^{1/2}B'(\tilde{Y}M_T\tilde{Y}'/T)\hat{B} = (F'M_TF/T)^{1/2}B'\hat{B}V. \tag{C.12}$$

To simplify the notation, let $\delta_{NT} \equiv (F'M_TF/T)^{1/2}B'(\tilde{Y}M_T\tilde{Y}'/T - B(F'M_TF/T)B')\hat{B}$ and $R_{NT} \equiv (F'M_TF/T)^{1/2}B'\hat{B}$. Then we can rewrite (C.12) as

$$[(F'M_TF/T)^{1/2}B'B(F'M_TF/T)^{1/2} + \delta_{NT}R_{NT}^{-1}]R_{NT} = R_{NT}V. \tag{C.13}$$

Let $D_{NT}$ be a diagonal matrix consisting the diagonal elements of $R'_{NT}R_{NT}$. Denote $\Upsilon_{NT} \equiv R_{NT}D_{NT}^{-1/2}$, which has a unit length. Then we can further rewrite (C.13) as

$$[(F'M_TF/T)^{1/2}B'B(F'M_TF/T)^{1/2} + \delta_{NT}R_{NT}^{-1}]\Upsilon_{NT} = \Upsilon_{NT}V, \tag{C.14}$$

which implies that $\Upsilon_{NT}$ is the eigenvector matrix of $(F'M_TF/T)^{1/2}B'B(F'M_TF/T)^{1/2}+\delta_{NT}R_{NT}^{-1}$ and $V$ is the diagonal eigenvalue matrix. Since $R_{NT} = (F'M_TF/T)^{1/2}B'BH+o_p(1)$ by simple algebra and Theorem 4.1, $R_{NT}^{-1} = O_p(1)$ by Assumptions 4.2(i)-(iii) and Lemma A.2. This together with (A.11) and Assumptions 4.2(i) and (ii) implies that

$$\delta_{NT}R_{NT}^{-1} = O_p\left(\frac{1}{J^\kappa} + \frac{J}{N} + \frac{\sqrt{J}}{\sqrt{NT}}\right). \tag{C.15}$$

Since the eigenvalues of $(F'M_TF/T)B'B$ are equal to those of $(F'M_TF/T)^{1/2}B'B$ $(F'M_TF/T)^{1/2}$, the eigenvalues of $(F'M_TF/T)^{1/2}B'B(F'M_TF/T)^{1/2}$ are distinct by Assumption 4.6(i). By the eigenvector perturbation theory, there exists a unique eigenvector matrix $\Upsilon$ of $(F'M_TF/T)^{1/2}B'B(F'M_TF/T)^{1/2}$ such that

$$\Upsilon_{NT} = \Upsilon + O_p\left(\frac{1}{J^\kappa} + \frac{J}{N} + \frac{\sqrt{J}}{\sqrt{NT}}\right). \tag{C.16}$$

By (A.11) and simple algebra, $R'_{NT}R_{NT} = \hat{B}'B(F'M_TF/T)B'\hat{B} = \hat{B}'(\tilde{Y}M_T\tilde{Y}'/T)\hat{B} + O_p(J^{-\kappa} + J/N + \sqrt{J}/\sqrt{NT}) = V + O_p(J^{-\kappa} + J/N + \sqrt{J}/\sqrt{NT})$. This implies that

$$D_{NT} = V + O_p\left(\frac{1}{J^\kappa} + \frac{J}{N} + \frac{\sqrt{J}}{\sqrt{NT}}\right). \tag{C.17}$$

Recall that $H^\diamond = (F'M_TF/T)B'\hat{B}V^{-1}$ in the proof of Lemma A.2(i). Thus, by (C.16) and (C.17), we have $H^\diamond = (F'M_TF/T)^{1/2}R_{NT}V^{-1} = (F'M_TF/T)^{1/2}\Upsilon_{NT}D_{NT}^{1/2}V^{-1} = \mathcal{H}+O_p(J^{-\kappa}+J/N+\sqrt{J/NT})$, which together with (A.12) and (A.13) leads to the first result of the lemma. The second result of the lemma follows from (A.12), the first result of the lemma and Lemma A.2(i). ∎

**Lemma C.4.** *Suppose Assumptions 4.1(i), 4.2(ii), 4.3(i), (ii), 4.5 and 4.6(ii) hold. Assume* $J \geq 2$ *and* $\xi_J^2 \log J = o(N)$. *Then*

$$\left\|\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}(\hat{Q}_t^{-1} - Q_t^{-1})\Phi(Z_t)'\varepsilon_t f_t'\right\|_F = O_p\left(\frac{\sqrt{J}\xi_J \log^{1/4} J}{N^{1/4}}\right)$$

*and*

$$\left\| \frac{1}{\sqrt{NT}} \sum_{t=1}^{T} (\hat{Q}_t^{-1} - Q_t^{-1}) \Phi(Z_t)' \varepsilon_t \right\| = O_p\left( \frac{\sqrt{J\xi_J} \log^{1/4} J}{N^{1/4}} \right).$$

Proof: Let $\mathcal{T} \equiv \sum_{t=1}^{T} (\hat{Q}_t^{-1} - Q_t^{-1}) \Phi(Z_t)' \varepsilon_t f_t' / \sqrt{NT}$ and $E_\varepsilon$ denote the expectation with respect to $\{\varepsilon_t\}_{t \leq T}$. Since $\|A\|_F^2 = \mathrm{tr}(AA')$,

$$\begin{aligned}
E_\varepsilon[\|\mathcal{T}\|_F^2] &= \frac{1}{NT} E_\varepsilon \left[ \mathrm{tr} \left( \sum_{t=1}^{T} \sum_{s=1}^{T} (\hat{Q}_t^{-1} - Q_t^{-1}) \Phi(Z_t)' \varepsilon_t f_t' f_s \varepsilon_s' \Phi(Z_s) (\hat{Q}_s^{-1} - Q_s^{-1}) \right) \right] \\
&= \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \phi(z_{it})' (\hat{Q}_t^{-1} - Q_t^{-1}) (\hat{Q}_s^{-1} - Q_s^{-1}) \phi(z_{js}) f_t' f_s E[\varepsilon_{it} \varepsilon_{js}] \\
&= \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \phi(z_{it})' (\hat{Q}_t^{-1} - Q_t^{-1}) (\hat{Q}_s^{-1} - Q_s^{-1}) \phi(z_{is}) f_t' f_s E[\varepsilon_{it} \varepsilon_{is}] \\
&\leq C_{NT}^* \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \|\hat{Q}_t - Q_t\|_2 \|\phi(z_{it})\| \|\phi(z_{is})\| |E[\varepsilon_{it} \varepsilon_{is}]| \\
&\leq C_{NT}^{**} \frac{1}{NT} \left( \sum_{t=1}^{T} \left( \sum_{i=1}^{N} \sum_{s=1}^{T} \|\phi(z_{it})\| \|\phi(z_{is})\| |E[\varepsilon_{it} \varepsilon_{is}]| \right)^2 \right)^{1/2},
\end{aligned} \tag{C.18}$$

where $C_{NT}^* = (\min_{t \leq T} \lambda_{\min}(\hat{Q}_t))^{-1} [(\min_{t \leq T} \lambda_{\min}(\hat{Q}_t))^{-1} + (\min_{i \leq N, t \leq T} \lambda_{\min}(Q_{it}))^{-1}]$ $(\min_{i \leq N, t \leq T} \lambda_{\min}(Q_{it}))^{-1} \max_{t \leq T} \|f_t\|^2$ and $C_{NT}^{**} = C_{NT}^* (\sum_{t=1}^{T} \|\hat{Q}_t - Q_t\|_2^2)^{1/2}$, the second equality follows by the independence in Assumption 4.3(i) and the fact that both expectation and trace operators are linear, the third equality follows by Assumption 4.3(ii) and the independence in Assumption 4.6(ii), the first inequality follows since $\min_{t \leq T} \lambda_{\min}(Q_t) \geq \min_{i \leq N, t \leq T} \lambda_{\min}(Q_{it})$, and the last inequality is due to the Cauchy-Schwartz inequality. Moreover, we have

$$\begin{aligned}
& E \left[ \sum_{t=1}^{T} \left( \sum_{i=1}^{N} \sum_{s=1}^{T} \|\phi(z_{it})\| \|\phi(z_{is})\| |E[\varepsilon_{it} \varepsilon_{is}]| \right)^2 \right] \\
&\leq \max_{i \leq N, t \leq T} E[\|\phi(z_{it})\|^4] \sum_{t=1}^{T} \left( \sum_{i=1}^{N} \sum_{s=1}^{T} |E[\varepsilon_{it} \varepsilon_{is}]| \right)^2 \\
&\leq J^2 M^2 \max_{m \leq M, j \leq J, i \leq N, t \leq T} E[\phi_j^4(z_{it,m})] \sum_{t=1}^{T} \left( \sum_{i=1}^{N} \sum_{s=1}^{T} |E[\varepsilon_{it} \varepsilon_{is}]| \right)^2,
\end{aligned} \tag{C.19}$$

where the first inequality is due to the Cauchy-Schwartz inequality, the second inequality follows since $\max_{i \leq N, t \leq T} E[\|\phi(z_{it})\|^4] \leq J^2 M^2 \max_{m \leq M, j \leq J, i \leq N, t \leq T} E[\phi_j^4(z_{it,m})]$. By Assumptions 4.1(i), 4.2(ii), and 4.5(ii) and Lemma B.6, $C_{NT}^{**} = O_p(\sqrt{T} \xi_J \sqrt{\log J} / \sqrt{N})$. Combining this, (C.18) and (C.19) implies that $E_\varepsilon[\|\mathcal{T}\|_F^2] = O_p(J \xi_J \sqrt{\log J} / \sqrt{N})$ by Assumptions 4.5(i) and (iv). Thus, the first result of the lemma follows by the Markov's inequality and Lemma A.5. The proof of the second result is similar. ■

**Lemma C.5** (Yurinskiis coupling). *Let $\zeta_1, \ldots, \zeta_N$ be independent random $k-$vectors with $E[\zeta_i] = 0$ for each $i$ and $\beta = \sum_{i=1}^N E[\|\zeta_i\|^3]$ finite. Let $S = \sum_{i=1}^N \zeta_i$. For each $\delta > 0$, there exists a random vector $\mathbb{S}$ in the same probability space with $S$ with a $N(0, E[SS'])$ distribution such that*

$$P\{\|S - \mathbb{S}\| > 3\delta\} \leq C_0 D_0 \left(1 + \frac{|\log(1/D_0)|}{k}\right)$$

*for some universal constant $C_0$, where $D_0 = \beta k \delta^{-3}$.*

Proof: This is the Yurinskiis coupling, see Theorem 10 in Pollard (2002). ■

# Appendix D - Proof of Theorem 5.1

## D.1 Proof of Theorem 5.1

Proof of Theorem 5.1: Let us begin by defining some notation. Let $\tilde{A}_t^* \equiv (\Phi(Z_t)^{*\prime}\Phi(Z_t))^{-1}\Phi(Z_t)^{*\prime}A_t$ for $A_t = \Delta_t$ and $\varepsilon_t$, where $\Delta_t = R(Z_t) + \Delta(Z_t)f_t$. Let $\tilde{\Delta}^* \equiv (\tilde{\Delta}_1^*, \ldots, \tilde{\Delta}_T^*)$ and $\tilde{E}^* \equiv (\tilde{\varepsilon}_1^*, \ldots, \tilde{\varepsilon}_T^*)$. Then we have

$$\tilde{Y}^* = a1_T' + BF' + \tilde{\Delta}^* + \tilde{E}^*, \tag{D.1}$$

where $1_T$ denotes a $T \times 1$ vector of ones. Recall $M_T = I_T - 1_T 1_T'/T$. Post-multiplying (D.1) by $M_T$ to remove $a$, we thus obtain

$$\tilde{Y}^* M_T = B(M_T F)' + \tilde{\Delta}^* M_T + \tilde{E}^* M_T. \tag{D.2}$$

Recall that $V$ us a $K \times K$ diagonal matrix of the first $K$ largest eigenvalues of $\tilde{Y} M_T \tilde{Y}'/T$ as defined in the proof of Theorem 4.1, $H = F' M_T \hat{F}(\hat{F}' M_T \hat{F})^{-1}$ and $\hat{F}' M_T \hat{F}/T = V$ as showed in the proof of Theorem 4.1. By the definitions of $\hat{B}^*$, $\hat{B}^* = \tilde{Y}^* M_T \hat{F}(\hat{F}' M_T \hat{F})^{-1}$. We may substitute (D.2) to $\hat{B}^* = \tilde{Y}^* M_T \hat{F}(\hat{F}' M_T \hat{F})^{-1}$ to obtain

$$\hat{B}^* - BH = [(\tilde{\Delta}^* + \tilde{E}^*) M_T \tilde{Y}'/T]\hat{B}V^{-1} = \sum_{j=1}^{6} D_j^* \hat{B} V^{-1}, \tag{D.3}$$

where in the first equality we have used $\hat{F}' M_T \hat{F}/T = V$ and $\hat{F} = \tilde{Y}'\hat{B}$, in the second equality we have substituted (A.2) into the equation, and $D_1^* = \tilde{\Delta}^* M_T F B'/T$, $D_2^* = \tilde{\Delta}^* M_T \tilde{\Delta}'/T$, $D_3^* = \tilde{\Delta}^* M_T \tilde{E}'/T$, $D_4^* = \tilde{E}^* M_T F B'/T$, $D_5^* = \tilde{E}^* M_T \tilde{E}'/T$ and $D_6^* = \tilde{E}^* M_T \tilde{\Delta}'/T$. We can conduct the same exercise as in (C.1) to obtain

$$\|\sqrt{NT}(\hat{B}^* - BH) - \sqrt{NT} D_4^* \hat{B} V^{-1}\|_F \leq \sqrt{NT}\|V^{-1}\|_2 \|D_5^* \hat{B}\|_F$$

$$+\sqrt{NT}\|\hat{B}\|_2\|V^{-1}\|_2\sum_{j\neq 4,5}^{6}\|D_j^*\|_F=O_p\left(\frac{\sqrt{NT}}{J^\kappa}+\frac{\sqrt{TJ}}{\sqrt{N}}\right), \tag{D.4}$$

where the equality follows by $J=o(\sqrt{N})$, Lemmas D.1 and A.2(i). Let $\mathcal{L}_{NT}^{**}\equiv\sum_{t=1}^T Q_t^{-1}\Phi(Z_t)^{*\prime}\varepsilon_t(f_t-\bar{f})'/\sqrt{NT}$. Since $J=o(\sqrt{N})$, $J^{(1/2-\kappa)}=o(\sqrt{NT}/J^\kappa)$. By the fact that $\|C+D\|_F\leq\|C\|_F+\|D\|_F$, combining (D.4) and Lemma D.2 implies

$$\|\sqrt{NT}(\hat{B}^*-BH)-\mathcal{L}_{NT}^{**}B'B\mathcal{M}\|_F=O_p\left(\frac{\sqrt{NT}}{J^\kappa}+\frac{\sqrt{TJ}}{\sqrt{N}}+\frac{\sqrt{J\xi_J}\log^{1/4}J}{N^{1/4}}\right). \tag{D.5}$$

Let $\mathcal{L}_{NT}^{*}\equiv\sum_{t=1}^T Q_t^{-1}[\Phi(Z_t)^*-\Phi(Z_t)]'\varepsilon_t(f_t-\bar{f})'/\sqrt{NT}=\mathcal{L}_{NT}^{**}-\mathcal{L}_{NT}$. Note that $\sqrt{NT}(\hat{B}^*-\hat{B})=\sqrt{NT}(\hat{B}^*-BH)-\sqrt{NT}(\hat{B}-BH)$. By the fact that $\|C+D\|_F\leq\|C\|_F+\|D\|_F$, we now may combine (C.2) and (D.5) to obtain

$$\|\sqrt{NT}(\hat{B}^*-\hat{B})-\mathcal{L}_{NT}^{*}B'B\mathcal{M}\|_F=O_p\left(\frac{\sqrt{NT}}{J^\kappa}+\frac{\sqrt{TJ}}{\sqrt{N}}+\frac{\sqrt{J\xi_J}\log^{1/4}J}{N^{1/4}}\right). \tag{D.6}$$

Note that $\mathbb{N}_2^*$ is a $JM\times K$ matrix from the last $K$ columns of $\mathbb{N}^*$. Thus, the second result of the theorem follows from (D.6) and Lemmas A.5 and D.3. We now show the first result of the theorem. By the definition of $\hat{a}^*$,

$$\begin{aligned}\hat{a}^*-a=&-\hat{B}^*(\hat{B}^{*\prime}\hat{B}^*)^{-1}(\hat{B}^*-BH)'a+(I_{JM}-\hat{B}^*(\hat{B}^{*\prime}\hat{B}^*)^{-1}\hat{B}^{*\prime})(BH-\hat{B}^*)H^{-1}\bar{f}\\&+(I_{JM}-\hat{B}^*(\hat{B}^{*\prime}\hat{B}^*)^{-1}\hat{B}^{*\prime})\tilde{\Delta}^*1_T/T+(I_{JM}-\hat{B}^*(\hat{B}^{*\prime}\hat{B}^*)^{-1}\hat{B}^{*\prime})\tilde{E}^*1_T/T,\end{aligned} \tag{D.7}$$

where $H^{-1}$ is well defined with probability approaching one by (A.4) and Lemma A.2(ii), and we have used $a'B=0$ and $(I_{JM}-\hat{B}^*(\hat{B}^{*\prime}\hat{B}^*)^{-1}\hat{B}^{*\prime})\hat{B}^*=0$. Let $\ell_{NT}^{**}\equiv\sum_{t=1}^T Q_t^{-1}\Phi(Z_t)^{*\prime}\varepsilon_t/\sqrt{NT}$. By a similar argument as in (C.3)-(C.6), we have

$$\begin{aligned}&\|\sqrt{NT}(\hat{a}^*-a)-(I_{JM}-B\mathcal{H}\mathcal{H}'B')[\ell_{NT}^{**}-\sqrt{NT}(\hat{B}^*-BH)\mathcal{H}^{-1}\bar{f}\\&\quad+B\mathcal{H}\sqrt{NT}(\hat{B}^*-BH)'a\|=O_p\left(\frac{\sqrt{NT}}{J^\kappa}+\frac{\sqrt{TJ}}{\sqrt{N}}+\frac{\sqrt{J\xi_J}\log^{1/4}J}{N^{1/4}}\right)\end{aligned} \tag{D.8}$$

by noting that $\mathcal{H}'B'B\mathcal{H}=I_K$. Let $\ell_{NT}^{*}\equiv\sum_{t=1}^T Q_t^{-1}[\Phi(Z_t)^*-\Phi(Z_t)]'\varepsilon_t/\sqrt{NT}=\ell_{NT}^{**}-\ell_{NT}$. Note that $\sqrt{NT}(\hat{a}^*-\hat{a})=\sqrt{NT}(\hat{a}^*-a)-\sqrt{NT}(\hat{a}-a)$ and $\sqrt{NT}(\hat{B}^*-\hat{B})=\sqrt{NT}(\hat{B}^*-BH)-\sqrt{NT}(\hat{B}-BH)$. By the fact that $\|x+y\|\leq\|x\|+\|y\|$, we now may combine (C.6) and (D.8) to obtain

$$\begin{aligned}&\|\sqrt{NT}(\hat{a}^*-\hat{a})-(I_{JM}-B\mathcal{H}\mathcal{H}'B')[\ell_{NT}^{*}-\sqrt{NT}(\hat{B}^*-\hat{B})\mathcal{H}^{-1}\bar{f}\\&\quad+B\mathcal{H}\sqrt{NT}(\hat{B}^*-\hat{B})'a\|=O_p\left(\frac{\sqrt{NT}}{J^\kappa}+\frac{\sqrt{TJ}}{\sqrt{N}}+\frac{\sqrt{J\xi_J}\log^{1/4}J}{N^{1/4}}\right).\end{aligned} \tag{D.9}$$

Note that $\mathbb{N}_1^*$ is a $JM \times 1$ vector from the first column of $\mathbb{N}^*$. Thus, the first result of the theorem follows from (D.9), the second result of the theorem and Lemmas A.5 and D.3. This completes the proof of the theorem. ■

## D.2 Technical Lemmas

**Lemma D.1.** *Let $D_1^*, D_2^*, D_3^*, D_5^*, D_6^*$ be given in the proof of Theorem 5.1.*
*(i) Under Assumptions 4.2(i), (ii), (iv), 5.1(i) and (ii), $\|D_1^*\|_F^2 = O_p(J^{-2\kappa})$.*
*(ii) Under Assumptions 4.1(i), 4.2(ii), (iv), 5.1(i) and (ii), $\|D_2^*\|_F^2 = O_p(J^{-4\kappa})$.*
*(iii) Under Assumptions 4.1, 4.2(ii), (iv), 4.3, 5.1(i) and (ii), $\|D_3^*\|_F^2 = O_p(J^{-2\kappa}J/N)$.*
*(iv) Assume (i) $N \to \infty$; (ii) $T \to \infty$ or $T \geq K+1$ is finite; (iii) $J \to \infty$ with $J^2\xi_J^2 \log J = o(N)$. Under Assumptions 4.1-4.5, 5.1(i) and (ii), $\|D_5^*\hat{B}\|_F^2 = O_p(J/N^2)$.*
*(v) Under Assumptions 4.1, 4.2(ii), (iv), 4.3, 5.1(i) and (ii), $\|D_6^*\|_F^2 = O_p(J^{-2\kappa}J/N)$.*

Proof: (i) Since $\|M_T\|_2 = 1$, $\|D_1^*\|_F \leq \|B\|_2\|F\|_2\|\tilde{\Delta}^*\|_F/T$. The result then immediately follows from Assumptions 4.2(i), (ii) and Lemma D.4(i).

(ii) Since $\|M_T\|_2 = 1$, $\|D_2^*\|_F \leq \|\tilde{\Delta}\|_F\|\tilde{\Delta}^*\|_F/T$. The result then immediately follows from Lemmas A.3(i) and D.4(i).

(iii) Since $\|M_T\|_2 = 1$, $\|D_3^*\|_F \leq \|\tilde{\Delta}^*\|_F\|\tilde{E}\|_F/T$. The result then immediately follows from Lemmas A.3(ii) and D.4(i).

(iv) Since $\|M_T\|_2 = 1$, $\|D_5^*\hat{B}\|_F \leq \|\hat{B}'\tilde{E}\|_F\|\tilde{E}^*\|_F/T$. The result then immediately follows from Lemmas B.2(i) and D.4(ii).

(v) Since $\|M_T\|_2 = 1$, $\|D_6^*\|_F \leq \|\tilde{\Delta}\|_F\|\tilde{E}^*\|_F/T$. The result then immediately follows from Lemmas A.3(i) and D.4(ii). ■

**Lemma D.2.** *Suppose Assumptions 4.1-4.5, 4.6(i), (ii), 5.1(i) and (ii) hold. Let $V$ be given in the proof of Theorem 4.1, and $\tilde{E}^*$, $D_4^*$, $\ell_{NT}^{**}$ and $\mathcal{L}_{NT}^{**}$ be given in the proof of Theorem 5.1. Assume (i) $N \to \infty$; (ii) $T \to \infty$ or $T \geq K+1$ is finite; (iii) $J \to \infty$ with $\xi_J^2 \log J = o(N)$. Then*

$$\|\sqrt{NT}D_4^*\hat{B}V^{-1} - \mathcal{L}_{NT}^{**}B'B\mathcal{M}\|_F = O_p\left(\frac{1}{J^{(\kappa-1/2)}} + \frac{\sqrt{J}\xi_J \log^{1/4} J}{N^{1/4}}\right)$$

*and*

$$\|\sqrt{N/T}\tilde{E}^*1_T - \ell_{NT}^{**}\| = O_p\left(\frac{\sqrt{J}\xi_J \log^{1/4} J}{N^{1/4}}\right),$$

*where $\mathcal{M}$ is a nonrandom matrix given in Lemma C.3.*

Proof: For the first result, we have the following decomposition

$$\sqrt{NT}D_4^*\hat{B}V^{-1} = \sqrt{N/T}\tilde{E}^*M_TFB'B\mathcal{M}_2 + \sqrt{N/T}\tilde{E}^*M_TFB'(\hat{B} - BH)V^{-1}$$

$$+\sqrt{N/T}\tilde{E}^* M_T F B' B(HV^{-1}-\mathcal{M}) \equiv \mathcal{T}_1+\mathcal{T}_2+\mathcal{T}_3. \tag{D.10}$$

Therefore, it suffices to show that $\|\mathcal{T}_1-\mathcal{L}^{**}_{NT}B'B\mathcal{M}_2\|_F = O_p(\sqrt{J\xi_J}\log^{1/4}J/N^{1/4})$, $\|\mathcal{T}_2\|_F = O_p(J^{(1/2-\kappa)}+J^{3/2}/N+J/\sqrt{NT})$ and $\|\mathcal{T}_3\|_F = O_p(J^{(1/2-\kappa)}+J^{3/2}/N+J/\sqrt{NT})$. The first one holds, since

$$\begin{aligned}
\|\mathcal{T}_1-\mathcal{L}^{**}_{NT}B'B\mathcal{M}\|_F &\le \|B\|_2^2\|\mathcal{M}\|_2\left\|\frac{1}{\sqrt{NT}}\sum_{t=1}^T(\hat{Q}_t^{*-1}-Q_t^{-1})\Phi(Z_t)^{*\prime}\varepsilon_t f_t'\right\|_F \\
&\quad+\|B\|_2^2\|\mathcal{M}\|_2\|\bar{f}\|\left\|\frac{1}{\sqrt{NT}}\sum_{t=1}^T(\hat{Q}_t^{*-1}-Q_t^{-1})\Phi(Z_t)^{*\prime}\varepsilon_t\right\| \\
&= O_p\left(\frac{\sqrt{J\xi_J}\log^{1/4}J}{N^{1/4}}\right),
\end{aligned} \tag{D.11}$$

where the equality follows from Assumptions 4.2(i) and (ii) and Lemma D.6. The latter two follow by a similar argument. The second result also follows by a similar argument as in (D.11). This completes the proof of the lemma. ■

**Lemma D.3.** *Suppose Assumptions 4.2(ii), 4.3(i), (ii), 4.5(i)-(iii), 4.6(ii), (iii), 5.1(i) and (iii) hold. Let $\ell^*_{NT}$ and $\mathcal{L}^*_{NT}$ be given in the proof of Theorem 5.1. Assume $J = o(\sqrt{N})$. Then there exists a $JM\times(K+1)$ random matrix $\mathbb{N}^*$ with $\mathrm{vec}(\mathbb{N}^*)\sim N(0,\Omega)$ conditional on $\{Y_t,Z_t\}_{t\le T}$ such that*

$$\|(\ell^*_{NT},\mathcal{L}^*_{NT})-\sqrt{\omega_0}\mathbb{N}^*\|_F = O_p\left(\frac{J^{5/6}}{N^{1/6}}\right).$$

Proof: Let $\zeta_i \equiv (w_i-1)\sum_{t=1}^T f_t^\dagger\otimes Q_t^{-1}\phi(z_{it})\varepsilon_{it}/\sqrt{NT}$. Then $\mathrm{vec}((\ell^*_{NT},\mathcal{L}^*_{NT})) = \sum_{i=1}^N\zeta_i$. Let $E_w$ denote the expectation with respect to $\{w_i\}_{i\le N}$. Then conditional on $\{Y_t,Z_t\}_{t\le T}$, $E_w[\zeta_i]=0$ and $\zeta_1,\ldots,\zeta_N$ are independent by Assumption 5.1(i). To proceed, let $\Omega_{NT}\equiv\sum_{i=1}^N\sum_{t=1}^T\sum_{s=1}^T(f_t^\dagger f_s^{\dagger\prime})\otimes Q_t^{-1}\phi(z_{it})\phi(z_{is})'Q_s^{-1}\varepsilon_{it}\varepsilon_{is}/NT$. Then $E_w[\mathrm{vec}((\ell^*_{NT},\mathcal{L}^*_{NT}))\mathrm{vec}((\ell^*_{NT},\mathcal{L}^*_{NT}))'] = \omega_0\Omega_{NT}$. We now apply Lemma C.5 to the independent random vectors $\zeta_1,\ldots,\zeta_N$ conditional on $\{Y_t,Z_t\}_{t\le T}$. There exists a $JM\times(K+1)$ random matrix $\mathbb{N}^{**}$ with $\mathrm{vec}(\mathbb{N}^{**})\sim N(0,\Omega_{NT})$ conditional on $\{Y_t,Z_t\}_{t\le T}$ such that the following holds:

$$\|\mathrm{vec}((\ell^*_{NT},\mathcal{L}^*_{NT}))-\sqrt{\omega_0}\mathrm{vec}(\mathbb{N}^{**})\| = O_{p^*}\left((J\beta)^{1/3}\right), \tag{D.12}$$

where $\beta=\sum_{i=1}^N E[\|\zeta_i\|^3]$. Next, we calculate $\beta$. To the end, we first calculate

$$E[\|\zeta_i\|^4] = E[(w_1-1)^4]\frac{1}{N^2T^2}E\left[\left(\sum_{t=1}^T\sum_{s=1}^T\phi(z_{it})'Q_t^{-1}Q_s^{-1}\phi(z_{is})f_t^{\dagger\prime}f_s^\dagger\varepsilon_{it}\varepsilon_{is}\right)^2\right]$$

$$\leq C_{NT} \max_{i\leq N,t\leq T} E[\|\phi(z_{it})\|^4] \frac{1}{N^2T^2} \sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{u=1}^{T}\sum_{v=1}^{T} |E[\varepsilon_{it}\varepsilon_{is}\varepsilon_{iu}\varepsilon_{iv}]|$$

$$\leq C_{NT} \max_{m\leq M,j\leq J,i\leq N,t\leq T} E[\phi_j^4(z_{it,m})] \frac{J^2M^2}{N^2T^2} \sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{u=1}^{T}\sum_{v=1}^{T} |E[\varepsilon_{it}\varepsilon_{is}\varepsilon_{iu}\varepsilon_{iv}]|, \quad \text{(D.13)}$$

where $C_{NT} = E[(w_1 - 1)^4] \max_{t\leq T} \|f_t^{\dagger}\|^4 (\min_{i\leq N,t\leq T} \lambda_{\min}(Q_{it}))^{-4}$, the first inequality follows by the independence in Assumption 4.3(i), the Cauchy-Schwartz inequality, and the fact that $\min_{t\leq T} \lambda_{\min}(Q_t) \geq \min_{i\leq N,t\leq T} \lambda_{\min}(Q_{it})$, and the second inequality follows since $\max_{i\leq N,t\leq T} E[\|\phi(z_{it})\|^4] \leq J^2M^2 \max_{m\leq M,j\leq J,i\leq N,t\leq T} E[\phi_j^4(z_{it,m})]$. Thus,

$$\beta = \sum_{i=1}^{N} E[\|\zeta_i\|^3] \leq \sum_{i=1}^{N} (E[\|\zeta_i\|^4])^{3/4} = O\left(\frac{J^{3/2}}{\sqrt{N}}\right), \quad \text{(D.14)}$$

where the inequality follows by the Liapounovs inequality, and the last equality follows from (D.13) and Assumptions 4.2(ii), 4.5(i), (ii), 4.6(iii) and 5.1(i). We now may combine (D.12), (D.14) and Lemma A.5 to obtain

$$\|\text{vec}((\ell_{NT}^*, \mathcal{L}_{NT}^*)) - \sqrt{\omega_0}\text{vec}(\mathbb{N}^{**})\| = O_p\left(\frac{J^{5/6}}{N^{1/6}}\right). \quad \text{(D.15)}$$

By Assumption 5.1(iii) and Lemma D.8, $\Omega_{NT}^{-1/2}$ is well defined with probability approaching one since $J = o(\sqrt{N})$. Define $\mathbb{N}^*$ such that $\text{vec}(\mathbb{N}^*) = \Omega^{1/2}\Omega_{NT}^{-1/2}\text{vec}(\mathbb{N}^{**})$. Then $\text{vec}(\mathbb{N}^*) \sim N(0, \Omega)$ conditional on $\{Y_t, Z_t\}_{t\leq T}$. It follows that

$$\begin{aligned}\|\text{vec}((\ell_{NT}^*, \mathcal{L}_{NT}^*)) - \sqrt{\omega_0}\mathbb{N}^*\|_F &\leq \|\text{vec}((\ell_{NT}^*, \mathcal{L}_{NT}^*)) - \sqrt{\omega_0}\text{vec}(\mathbb{N}^{**})\| \\ &+ \sqrt{\omega_0}\|\text{vec}(\mathbb{N}^*) - \text{vec}(\mathbb{N}^{**})\| = O_p\left(\frac{J^{5/6}}{N^{1/6}} + \frac{J^{3/2}}{\sqrt{N}}\right) = O_p\left(\frac{J^{5/6}}{N^{1/6}}\right), \quad \text{(D.16)}\end{aligned}$$

where the first equality follows by (D.15) and the fact that $\|\text{vec}(\mathbb{N}^*) - \text{vec}(\mathbb{N}^{**})\| \leq \|\Omega_{NT}^{1/2} - \Omega^{1/2}\|_2 \|\Omega_{NT}^{-1/2}\text{vec}(\mathbb{N}^{**})\| = O_p(J^{3/2}/\sqrt{N})$, which is due to Lemma D.8. This completes the proof of the lemma. ■

**Lemma D.4.** *Let $\tilde{\Delta}^*$ and $\tilde{E}^*$ be given in the proof of Theorem 5.1.*
*(i) Under Assumptions 4.2(ii), (iv), 5.1(i) and (ii), $\|\tilde{\Delta}^*\|_F^2/T = O_p(J^{-2\kappa})$.*
*(ii) Under Assumptions 4.1(ii), 4.3, 5.1(i) and (ii), $\|\tilde{E}^*\|_F^2/T = O_p(J/N)$.*

Proof: (i) By the fact that $\|Ax\| \leq \|A\|_2\|x\|$ and $\|A\|_2 \leq \|A\|_F$,

$$\begin{aligned}\frac{1}{T}\|\tilde{\Delta}^*\|_F^2 &= \frac{1}{T}\sum_{t=1}^{T} \|(\Phi(Z_t)^{*\prime}\Phi(Z_t))^{-1}\Phi(Z_t)^{*\prime}(R(Z_t) + \Delta(Z_t)f_t)\|^2 \\ &\leq 2\max_{t\leq T}\|f_t\|^2 \left(\min_{t\leq T}\lambda_{\min}(\hat{Q}_t^*)\right)^{-1} \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N} w_i\|\delta(z_{it})\|^2\end{aligned}$$

$$+ 2\left(\min_{t\leq T}\lambda_{\min}(\hat{Q}_t^*)\right)^{-1}\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}w_i|r(z_{it})|^2 = O_p\left(\frac{1}{J^{2\kappa}}\right), \tag{D.17}$$

where the last equality follows from Assumptions 4.2(ii) and 5.1(ii) and Lemma D.5(ii).

(ii) By the fact that $\|Ax\| \leq \|A\|_2\|x\|$,

$$\begin{aligned}
\frac{1}{T}\|\tilde{E}^*\|_F^2 &= \frac{1}{T}\sum_{t=1}^{T}\|(\Phi(Z_t)^{*\prime}\Phi(Z_t))^{-1}\Phi(Z_t)^{*\prime}\varepsilon_t\|^2 \\
&\leq \left(\min_{t\leq T}\lambda_{\min}(\hat{Q}_t^*)\right)^{-2}\frac{1}{N^2T}\sum_{t=1}^{T}\|\Phi(Z_t)^{*\prime}\varepsilon_t\|^2 = O_p\left(\frac{J}{N}\right), \qquad (D.18)
\end{aligned}$$

where the last equality follows from Assumption 5.1(ii) and Lemma D.5(i). ■

**Lemma D.5.** *(i) Under Assumptions 4.1(ii), 4.3 and 5.1(i),*

$$\sum_{t=1}^{T}\|\Phi(Z_t)^{*\prime}\varepsilon_t\|^2 = O_p(NTJ).$$

*(ii) Under Assumption 4.2(iv) and 5.1(i),*

$$\sum_{t=1}^{T}\sum_{i=1}^{N}w_i\|\delta(z_{it})\|^2 = O_p(NTJ^{-2\kappa}) \text{ and } \sum_{t=1}^{T}\sum_{i=1}^{N}w_i|r(z_{it})|^2 = O_p(NTJ^{-2\kappa}).$$

Proof: (i) The result follows by the Markov's inequality, since

$$\begin{aligned}
&E\left[\sum_{t=1}^{T}\|\Phi(Z_t)^{*\prime}\varepsilon_t\|^2\right] = E\left[\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\phi(z_{it})'\phi(z_{jt})\varepsilon_{it}\varepsilon_{jt}w_iw_j\right] \\
&= \sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}E[\phi(z_{it})'\phi(z_{jt})]E[\varepsilon_{it}\varepsilon_{jt}]E[w_iw_j] \\
&\leq E[w_1^2]\max_{i\leq N,t\leq T}E[\|\phi(z_{it})\|^2]\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}|E[\varepsilon_{it}\varepsilon_{jt}]| \\
&\leq TJME[w_1^2]\max_{m\leq M,j\leq J,i\leq N,t\leq T}E[\phi_j^2(z_{it,m})]\max_{t\leq T}\sum_{i=1}^{N}\sum_{j=1}^{N}|E[\varepsilon_{it}\varepsilon_{jt}]| = O(NTJ), \qquad (D.19)
\end{aligned}$$

where the second equality follows by the independence in Assumptions 4.3(i) and 5.1(i), the first inequality is due to the Cauchy Schwartz inequality, the second inequality follows since $\max_{i\leq N,t\leq T}E[\|\phi(z_{it})\|^2] \leq JM\max_{m\leq M,j\leq J,i\leq N,t\leq T}E[\phi_j^2(z_{it,m})]$, and the last equality follows from Assumptions 4.1(ii), 4.3(iii) and 5.1(i).

(iii) The first result follows since

$$\sum_{t=1}^{T}\sum_{i=1}^{N} w_i\|\delta(z_{it})\|^2 \le TKM^2 \max_{k\le K,m\le M}\sup_z |\delta_{km,J}(z)|^2 \sum_{i=1}^{N} w_i = O_p(NTJ^{-2\kappa}), \quad \text{(D.20)}$$

where the inequality follows since $w_i$'s are positive and $\max_{i\le N,t\le T}\|\delta(z_{it})\|^2 \le M^2K \sup_{k\le K,m\le M}\sup_z |\delta_{km,J}(z)|^2$, and the equality follows by the law of large numbers and Assumptions 4.2(iv) and 5.1(i). The proof of the second result is similar. ■

**Lemma D.6.** *Suppose Assumptions 4.2(ii), 4.3(i), (ii), 4.5, 4.6(ii), 5.1(i) and (ii) hold. Assume $J \ge 2$ and $\xi_J^2 \log J = o(N)$. Then*

$$\left\| \frac{1}{\sqrt{NT}} \sum_{t=1}^{T} (\hat{Q}_t^{*-1} - Q_t^{-1})\Phi(Z_t)^{*\prime}\varepsilon_t f_t' \right\|_F = O_p\left(\frac{\sqrt{J}\xi_J \log^{1/4} J}{N^{1/4}}\right)$$

*and*

$$\left\| \frac{1}{\sqrt{NT}} \sum_{t=1}^{T} (\hat{Q}_t^{*-1} - Q_t^{-1})\Phi(Z_t)^{*\prime}\varepsilon_t \right\| = O_p\left(\frac{\sqrt{J}\xi_J \log^{1/4} J}{N^{1/4}}\right).$$

Proof: Let $\mathcal{T} \equiv \sum_{t=1}^{T}(\hat{Q}_t^{*-1} - Q_t^{-1})\Phi(Z_t)^{*\prime}\varepsilon_t f_t'/\sqrt{NT}$ and $E_\varepsilon$ denote the expectation with respect to $\{\varepsilon_t\}_{t\le T}$. Since $\|A\|_F^2 = \mathrm{tr}(AA')$,

$$\begin{aligned}
E_\varepsilon[\|\mathcal{T}\|_F^2] &= \frac{1}{NT}E_\varepsilon\left[\mathrm{tr}\left(\sum_{t=1}^{T}\sum_{s=1}^{T}(\hat{Q}_t^{*-1} - Q_t^{-1})\Phi(Z_t)^{*\prime}\varepsilon_t f_t' f_s \varepsilon_s' \Phi(Z_s)^*(\hat{Q}_s^{*-1} - Q_s^{-1})\right)\right] \\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} w_i\phi(z_{it})'(\hat{Q}_t^{*-1} - Q_t^{-1})(\hat{Q}_s^{*-1} - Q_s^{-1})\phi(z_{js})w_j f_t' f_s E[\varepsilon_{it}\varepsilon_{js}] \\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} w_i^2\phi(z_{it})'(\hat{Q}_t^{*-1} - Q_t^{-1})(\hat{Q}_s^{*-1} - Q_s^{-1})\phi(z_{is}) f_t' f_s E[\varepsilon_{it}\varepsilon_{is}] \\
&\le C_{NT}^{*}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\|\hat{Q}_t^* - Q_t\|_2 w_i^2\|\phi(z_{it})\|\|\phi(z_{is})\||E[\varepsilon_{it}\varepsilon_{is}]| \\
&\le C_{NT}^{**}\frac{1}{NT}\left(\sum_{t=1}^{T}\left(\sum_{i=1}^{N}\sum_{s=1}^{T} w_i^2\|\phi(z_{it})\|\|\phi(z_{is})\||E[\varepsilon_{it}\varepsilon_{is}]|\right)^2\right)^{1/2}, \qquad \text{(D.21)}
\end{aligned}$$

where $C_{NT}^* = (\min_{t\le T}\lambda_{\min}(\hat{Q}_t^*))^{-1}[(\min_{t\le T}\lambda_{\min}(\hat{Q}_t^*))^{-1} + (\min_{i\le N,t\le T}\lambda_{\min}(Q_{it}))^{-1}]$ $(\min_{i\le N,t\le T}\lambda_{\min}(Q_{it}))^{-1}\max_{t\le T}\|f_t\|^2$ and $C_{NT}^{**} = C_{NT}^*(\sum_{t=1}^{T}\|\hat{Q}_t^* - Q_t\|_2^2)^{1/2}$, the second equality follows by the independence in Assumptions 4.3(i) and 5.1(i) and the fact that both expectation and trace operators are linear, the third equality follows by Assumption 4.3(ii) and the independence in Assumption 4.6(ii), the first inequality follows since $\min_{t\le T}\lambda_{\min}(Q_t) \ge \min_{i\le N,t\le T}\lambda_{\min}(Q_{it})$, and the last inequality is due to the

Cauchy-Schwartz inequality. Moreover, we have

$$
\begin{aligned}
& E\left[\sum_{t=1}^{T}\left(\sum_{i=1}^{N}\sum_{s=1}^{T} w_i^2\|\phi(z_{it})\|\|\phi(z_{is})\||E[\varepsilon_{it}\varepsilon_{is}]|\right)^2\right] \\
&\leq E[w_1^4]\max_{i\leq N,t\leq T} E[\|\phi(z_{it})\|^4]\sum_{t=1}^{T}\left(\sum_{i=1}^{N}\sum_{s=1}^{T}|E[\varepsilon_{it}\varepsilon_{is}]|\right)^2 \\
&\leq J^2M^2E[w_1^4]\max_{m\leq M,j\leq J,i\leq N,t\leq T} E[\phi_j^4(z_{it,m})]\sum_{t=1}^{T}\left(\sum_{i=1}^{N}\sum_{s=1}^{T}|E[\varepsilon_{it}\varepsilon_{is}]|\right)^2, \qquad \text{(D.22)}
\end{aligned}
$$

where the first inequality is by the Cauchy-Schwartz inequality and the independence in Assumption 5.1(i), the second inequality follows since $\max_{i\leq N,t\leq T} E[\|\phi(z_{it})\|^4] \leq J^2M^2\max_{m\leq M,j\leq J,i\leq N,t\leq T} E[\phi_j^4(z_{it,m})]$. By Assumptions 4.2(ii), 4.5(ii) and 5.1(ii) and Lemma D.7, $C_{NT}^{**} = O_p(\sqrt{T}\xi_J\sqrt{\log J}/\sqrt{N})$. Combining this, (D.21) and (D.22) implies that $E_\varepsilon[\|\mathcal{T}\|_F^2] = O_p(J\xi_J\sqrt{\log J}/\sqrt{N})$ by Assumptions 4.5(i), (iv) and 5.1(i). Thus, the result of the lemma follows by the Markov's inequality and Lemma A.5. The proof of the second result is similar. ■

**Lemma D.7.** *Suppose Assumptions 4.5(ii), (iii) and 5.1(i) hold. Assume $J \geq 2$ and $\xi_J^2\log J = o(N)$. Then*

$$\sum_{t=1}^{T}\|\hat{Q}_t^* - Q_t\|_2^2 = O_p\left(\frac{T\xi_J^2\log J}{N}\right).$$

Proof: The proof is similar to the proof of Lemma B.6, thus omitted for brevity. ■

**Lemma D.8.** *Suppose Assumptions 4.2(ii), 4.3(i), (ii), 4.5(i)-(iii), 4.6(ii), (iii) and 5.1(iii) hold. Let $\Omega_{NT}$ be given in the proof of Lemma D.3. Then*

$$\|\Omega_{NT}^{1/2} - \Omega^{1/2}\|_2 = O_p\left(\frac{J}{\sqrt{N}}\right).$$

Proof: We first show $\|\Omega_{NT}-\Omega\|_F^2 = O_p(J^2/N)$. Let $\zeta_i \equiv \sum_{t=1}^{T} f_t\otimes Q_t^{-1}\phi(z_{it})\varepsilon_{it}/\sqrt{NT}$. Then $\Omega_{NT} = \sum_{i=1}^{N}\zeta_i\zeta_i'$ and $\Omega = \sum_{i=1}^{N}E[\zeta_i\zeta_i']$. Since $\|A\|_F^2 = \mathrm{tr}(AA')$,

$$
\begin{aligned}
E[\|\Omega_{NT} - \Omega\|_F^2] &= E\left[\mathrm{tr}\left(\sum_{i=1}^{N}\sum_{j=1}^{N}(\zeta_i\zeta_i' - E[\zeta_i\zeta_i'])(\zeta_j\zeta_j' - E[\zeta_j\zeta_j'])'\right)\right] \\
&= \sum_{i=1}^{N}\left(E[(\zeta_i'\zeta_i)^2] - \|E[\zeta_i\zeta_i']\|_F^2\right) \leq N\max_{i\leq N} E[\|\zeta_i\|^4] = O\left(\frac{J^2}{N}\right), \qquad \text{(D.23)}
\end{aligned}
$$

where the second equality follows since $\zeta_1,\ldots,\zeta_N$ are independent by Assumptions 4.3(i), 4.5(iii) and 4.6(ii) and both expectation and trace operators are linear, the inequality follows by the Cauchy-Schwartz inequality since $\|E[\zeta_i\zeta_i']\|_F^2 \geq 0$, and the last equality follows from (C.10) and Assumptions 4.2(ii), 4.5(i), (ii) and 4.6(iii). Thus, $\|\Omega_{NT}-\Omega\|_F^2 = O_p(J^2/N)$ follows from (D.23) by the Markov's inequality. The result of the lemma follows from Assumption 5.1(iii) and Lemma A.2 of Belloni et al. (2015). ■

# Appendix E - Proof of Theorem 5.2

## E.1 Proof of Theorem 5.2

Proof of Theorem 5.2: To show the first result, let us assume that $\mathrm{H}_0$ is true. Since $\hat{\alpha}(z_{it}) = \hat{a}'\phi(z_{it})$, $\hat{\beta}(z_{it}) = \hat{B}'\phi(z_{it})$, $\alpha(z_{it}) = a'\phi(z_{it}) + r(z_{it}) = \gamma' z_{it}$ and $\beta(z_{it}) = B'\phi(z_{it}) + \delta(z_{it}) = \Gamma' z_{it}$, we have

$$
\begin{aligned}
\mathcal{S} &= \frac{1}{J}\sum_{i=1}^{N}\sum_{t=1}^{T} |(\hat{\gamma}-\gamma)' z_{it} - (\hat{a}-a)'\phi(z_{it}) + r(z_{it})|^2 \\
&\quad + \frac{1}{J}\sum_{i=1}^{N}\sum_{t=1}^{T} \|(\hat{\Gamma}-\Gamma H)' z_{it} - (\hat{B}-BH)'\phi(z_{it}) + H'\delta(z_{it})\|^2 \\
&= \frac{1}{J}\sum_{i=1}^{N}\sum_{t=1}^{T} |(\hat{\gamma}-\gamma)' z_{it} - (\hat{a}-a)'\phi(z_{it})|^2 + \mathcal{S}_1 + 2\mathcal{S}_2 + 2\mathcal{S}_3 \\
&\quad + \frac{1}{J}\sum_{i=1}^{N}\sum_{t=1}^{T} \|(\hat{\Gamma}-\Gamma H)' z_{it} - (\hat{B}-BH)'\phi(z_{it})\|^2 + \mathcal{S}_4 + 2\mathcal{S}_5 + 2\mathcal{S}_6, \qquad \text{(E.1)}
\end{aligned}
$$

where $\mathcal{S}_1 = \sum_{i=1}^{T}\sum_{t=1}^{N} |r(z_{it})|^2/J$, $\mathcal{S}_2 = \sum_{i=1}^{N}\sum_{t=1}^{T} z_{it}'(\hat{\gamma}-\gamma) r(z_{it})/J$, $\mathcal{S}_3 = \sum_{i=1}^{N}\sum_{t=1}^{T} \phi(z_{it})'(\hat{a}-a) r(z_{it})/J$, $\mathcal{S}_4 = \sum_{i=1}^{T}\sum_{t=1}^{N} \|H'\delta(z_{it})\|^2/J$, $\mathcal{S}_5 = \sum_{i=1}^{N}\sum_{t=1}^{T} z_{it}'(\hat{\Gamma}-\Gamma H)H'\delta(z_{it})/J$ and $\mathcal{S}_6 = \sum_{i=1}^{N}\sum_{t=1}^{T} \phi(z_{it})'(\hat{B}-BH)H'\delta(z_{it})/J$. Let $\mathcal{W}_{NT,a} \equiv (\sqrt{NT}(\hat{\gamma}-\gamma)', -\sqrt{NT}(\hat{a}-a)')'$, $\mathcal{W}_{NT,B} \equiv (\sqrt{NT}(\hat{\Gamma}-\Gamma H)', -\sqrt{NT}(\hat{B}-BH)')'$, $\mathcal{W}_{NT} \equiv (\mathcal{W}_{NT,a}, \mathcal{W}_{NT,B})$ and $\hat{\mathcal{Q}} \equiv \sum_{i=1}^{N}\sum_{t=1}^{T} (z_{it}', \phi(z_{it})')'(z_{it}', \phi(z_{it})')/NT$. By Lemma E.1, (E.1) implies

$$
\begin{aligned}
&\mathcal{S} - \frac{1}{J}\mathcal{W}_{NT,a}'\hat{\mathcal{Q}}\mathcal{W}_{NT,a} - \frac{1}{J}\mathrm{tr}(\mathcal{W}_{NT,B}'\hat{\mathcal{Q}}\mathcal{W}_{NT,B}) \\
&= \mathcal{S} - \frac{1}{J}\mathrm{tr}(\mathcal{W}_{NT}'\hat{\mathcal{Q}}\mathcal{W}_{NT}) = O_p\left(\frac{\sqrt{NT}}{J^{\kappa+1/2}}\right). \qquad \text{(E.2)}
\end{aligned}
$$

Let $\mathcal{Q} \equiv E[\hat{\mathcal{Q}}]$, $\mathbb{W}_a \equiv (\mathbb{G}_\gamma', -\mathbb{G}_a')'$, $\mathbb{W}_B \equiv (\mathbb{G}_\Gamma', -\mathbb{G}_B')'$ and $\mathbb{W} \equiv (\mathbb{W}_a, \mathbb{W}_B)$, where $\mathbb{G}_\gamma$ and $\mathbb{G}_\Gamma$ are given in Lemma E.2. By Lemmas E.2 and E.3 and Theorem 4.3, (E.2) implies

$$
\begin{aligned}
&\mathcal{S} - \frac{1}{J}\mathbb{W}_a'\mathcal{Q}\mathbb{W}_a - \frac{1}{J}\mathrm{tr}(\mathbb{W}_B'\mathcal{Q}\mathbb{W}_B) \\
&= \mathcal{S} - \frac{1}{J}\mathrm{tr}(\mathbb{W}'\mathcal{Q}\mathbb{W}) = O_p\left(\frac{\sqrt{NT}}{J^{\kappa+1/2}} + \frac{J^{1/3}}{N^{1/6}} + \frac{\sqrt{\xi_J}\log^{1/4} J}{N^{1/4}} + \sqrt{\frac{T}{N}}\right). \qquad \text{(E.3)}
\end{aligned}
$$

We let $\mathcal{W}^*_{NT,a} \equiv (\sqrt{NT/\omega_0}(\hat{\gamma}^* - \hat{\gamma})', -\sqrt{NT/\omega_0}(\hat{a}^* - \hat{a})')'$, $\mathcal{W}^*_{NT,B} \equiv (\sqrt{NT/\omega_0}(\hat{\Gamma}^* - \hat{\Gamma})', -\sqrt{NT/\omega_0}(\hat{B}^* - \hat{B})')'$, $\mathcal{W}^*_{NT} \equiv (\mathcal{W}^*_{NT,a}, \mathcal{W}^*_{NT,B})$, $\mathbb{W}^*_a \equiv (\mathbb{G}^{*\prime}_\gamma, -\mathbb{G}^{*\prime}_a)'$, $\mathbb{W}^*_B \equiv (\mathbb{G}^{*\prime}_\Gamma, -\mathbb{G}^{*\prime}_B)'$ and $\mathbb{W}^* \equiv (\mathbb{W}^*_a, \mathbb{W}^*_B)$, where $\mathbb{G}^*_\gamma$ and $\mathbb{G}^*_\Gamma$ are given in Lemma E.4. Then (18) can be written as $\mathcal{S}^* = \mathcal{W}^{*\prime}_{NT,a}\hat{\mathcal{Q}}\mathcal{W}^*_{NT,a}/J + \mathrm{tr}(\mathcal{W}^{*\prime}_{NT,B}\hat{\mathcal{Q}}\mathcal{W}^*_{NT,B})/J = \mathrm{tr}(\mathcal{W}^{*\prime}_{NT}\hat{\mathcal{Q}}\mathcal{W}^*_{NT})/J$. By Lemmas A.5, E.3 and E.4 and Theorem 5.1,

$$\begin{aligned}
&\mathcal{S}^* - \frac{1}{J}\mathbb{W}^{*\prime}_a\mathcal{Q}\mathbb{W}^*_a - \frac{1}{J}\mathrm{tr}(\mathbb{W}^{*\prime}_B\mathcal{Q}\mathbb{W}^*_B) \\
&= \mathcal{S}^* - \frac{1}{J}\mathrm{tr}(\mathbb{W}^{*\prime}\mathcal{Q}\mathbb{W}^*) = O_p\left(\frac{\sqrt{NT}}{J^{\kappa+1/2}} + \frac{J^{1/3}}{N^{1/6}} + \frac{\sqrt{\xi_J}\log^{1/4}J}{N^{1/4}} + \sqrt{\frac{T}{N}}\right).
\end{aligned} \tag{E.4}$$

Let $\gamma_{NT} \equiv (\sqrt{NT}J^{-\kappa} + J^{5/6}/N^{1/6} + \sqrt{J\xi_J}\log^{1/4}J/N^{1/4} + \sqrt{TJ/N})^{1/2}$, which is $o(1)$ by the assumption. Let $c_{0,1-\alpha}$ be the $1-\alpha$ quantile of $\mathrm{tr}(\mathbb{W}^{*\prime}\mathcal{Q}\mathbb{W}^*)/J$, which is also the $1-\alpha$ quantile of $\mathrm{tr}(\mathbb{W}'\mathcal{Q}\mathbb{W})/J$. Then in view of (E.4), Lemma A.1 of Belloni et al. (2015) implies that there exists a sequence $\{\nu_{NT}\}$ such that $\nu_{NT} = o(1)$ and

$$P(c_{1-\alpha} < c_{0,1-\alpha-\nu_{NT}} - \gamma_{NT}/\sqrt{J}) = o(1), \tag{E.5}$$

$$P(c_{1-\alpha} > c_{0,1-\alpha+\nu_{NT}} + \gamma_{NT}/\sqrt{J}) = o(1). \tag{E.6}$$

Note that $\mathrm{tr}(\mathbb{W}'\mathcal{Q}\mathbb{W}) = \mathrm{vec}(\mathbb{W})'(I_K \otimes \mathcal{Q})\mathrm{vec}(\mathbb{W})$. Since $\mathcal{Q}$ has rank not smaller than $JM - M$ and the variance of $\mathrm{vec}(\mathbb{G}_B)$ has full rank, $\mathrm{tr}(\mathbb{W}'\mathcal{Q}\mathbb{W})$ is bounded below by a random variable with a chi-squared distribution with degree of freedom $JM - M$ multiplied by a constant, and above by a random variable with a chi-squared distribution with degree of freedom $JM$ multiplied by a constant. Thus, it follows that

$$\begin{aligned}
P(\mathcal{S} \le c_{1-\alpha}) &\le P(\mathrm{tr}(\mathbb{W}'\mathcal{Q}\mathbb{W})/J \le c_{1-\alpha} + \gamma_{NT}/\sqrt{J}) + o(1) \\
&\le P(\mathrm{tr}(\mathbb{W}'\mathcal{Q}\mathbb{W})/J \le c_{0,1-\alpha+\nu_{NT}} + 2\gamma_{NT}/\sqrt{J}) + o(1) \\
&\le P(\mathrm{tr}(\mathbb{W}'\mathcal{Q}\mathbb{W})/\sqrt{J} \le \sqrt{J}c_{0,1-\alpha+\nu_{NT}} + 2\gamma_{NT}) + o(1) \\
&\le P(\mathrm{tr}(\mathbb{W}'\mathcal{Q}\mathbb{W})/\sqrt{J} \le \sqrt{J}c_{0,1-\alpha+\nu_{NT}}) + o(1) \\
&\le 1 - \alpha + \nu_{NT} + o(1) = 1 - \alpha + o(1),
\end{aligned} \tag{E.7}$$

where the first inequality follows since $P(|\mathcal{S} - \mathrm{tr}(\mathbb{G}'Q\mathbb{G})/J| > \gamma_{NT}/\sqrt{J}) = o(1)$ due to (E.3), the second inequality follows from (E.6), and the fourth inequality follows since $\gamma_{NT} = o(1)$ and $\mathrm{tr}(\mathbb{W}'\mathcal{Q}\mathbb{W})$ is bounded by chi-squared random variables. By a similar argument, $P(\mathcal{S} > c_{1-\alpha}) \le 1 - \alpha + o(1)$. Therefore, the first result of the theorem follows. To show the second result, we now assume that $\mathrm{H}_1$ is true. Since $(x+y)^2 \ge x^2/2 - y^2$,

$$\begin{aligned}
\frac{2J}{NT}\mathcal{S} \ge \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \|\hat{\Gamma}'z_{it} - H'\beta(z_{it})\|^2 - \frac{2}{NT}\sum_{i=1}^N\sum_{t=1}^T \|\hat{\beta}(z_{it}) - H'\beta(z_{it})\|^2 \\
+ \frac{2}{NT}\sum_{i=1}^N\sum_{t=1}^T |\hat{\gamma}'z_{it} - \hat{\alpha}(z_{it})|^2 \ge c_0 + o_p(1) \text{ for some } c_0 > 0,
\end{aligned} \tag{E.8}$$

where the second inequality follows from Lemmas E.5 and E.6. Since $(x+y)^2 \leq 2x^2+2y^2$,

$$
\begin{aligned}
\frac{2J}{NT}\mathcal{S}^* &\leq \frac{4}{NT\omega_0}\sum_{i=1}^{N}\sum_{t=1}^{T}|(\hat{\gamma}^* - \hat{\gamma})' z_{it}|^2 + \frac{4}{NT\omega_0}\sum_{i=1}^{N}\sum_{t=1}^{T}|(\hat{a}^* - \hat{a})'\phi(z_{it})|^2 \\
&\quad + \frac{4}{NT\omega_0}\sum_{i=1}^{N}\sum_{t=1}^{T}\|(\hat{\Gamma}^* - \hat{\Gamma})' z_{it}\|^2 + \frac{4}{NT\omega_0}\sum_{i=1}^{N}\sum_{t=1}^{T}\|(\hat{B}^* - \hat{B})'\phi(z_{it})\|^2 \\
&= o_p(1), \qquad \text{(E.9)}
\end{aligned}
$$

where the equality follows from Lemma E.7. In view of (E.9), Lemma A.1 of Belloni et al. (2015) implies that $2c_{1-\alpha}J/(NT) = o_p(1)$. This together with (E.8) thus concludes the second result of the theorem. This completes the proof of the theorem. ■

## E.2 Technical Lemmas

**Lemma E.1.** *Let $\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3, \mathcal{S}_4, \mathcal{S}_5, \mathcal{S}_6$ be given in the proof of Theorem 5.2. Assume (i) $N \to \infty$; (ii) $T \to \infty$ with $T = o(N)$ or $T \geq K+1$ is finite; (iii) $J \to \infty$ with $J^2\xi_J^2 \log J = o(N)$ and $NTJ^{-(2\kappa+1)} = o(1)$. Assume that $\mathrm{H}_0$ is true.*
*(i) Under Assumption 4.2(iv), $\mathcal{S}_1 = O_p(NTJ^{-(2\kappa+1)})$.*
*(ii) Under Assumptions 4.1-4.6, 5.2(i)-(iii), $\mathcal{S}_2 = O_p(\sqrt{NT}J^{-(\kappa+1)})$.*
*(iii) Under Assumptions 4.1-4.5, $\mathcal{S}_3 = O_p(\sqrt{NT}J^{-(\kappa+1/2)})$.*
*(iv) Under Assumptions 4.1-4.3, $\mathcal{S}_4 = O_p(NTJ^{-(2\kappa+1)})$.*
*(v) Under Assumptions 4.1-4.6, 5.2(i)-(iii), $\mathcal{S}_5 = O_p(\sqrt{NT}J^{-(\kappa+1)})$.*
*(vi) Under Assumptions 4.1-4.5, $\mathcal{S}_6 = O_p(\sqrt{NT}J^{-(\kappa+1/2)})$.*

PROOF: (i) The proof is similar to the proof of (iv).

(ii) The proof is similar to the proof of (v).

(iii) The proof is similar to the proof of (vi).

(iv) It follows that

$$
\mathcal{S}_4 \leq \|H\|_2^2 \sum_{i=1}^{T}\sum_{t=1}^{T}\|\delta(z_{it})\|^2/J \leq (NT/J)\|H\|_2^2 M^2 K \sup_{k\leq K, m\leq M}\sup_{z}|\delta_{km,J}(z)|^2, \quad \text{(E.10)}
$$

where the second inequality follows since $\max_{i\leq N, t\leq T}\|\delta(z_{it})\|^2 \leq M^2K\sup_{k\leq K,m\leq M}\sup_z|\delta_{km,J}(z)|^2$. Thus, the result of the lemma follows from (E.10), Assumption 4.2(iv) and Lemma A.2(i).

(v) By Assumption 5.2(ii), $\sum_{i=1}^{N}\sum_{t=1}^{T}\|z_{it}\|^2/NT = O_p(1)$ by the Markov's inequality. It then follows that

$$\frac{1}{J}\sum_{i=1}^{T}\sum_{t=1}^{T}\|(\hat{\Gamma}-\Gamma H)'z_{it}\|^2 \leq \|\hat{\Gamma}-\Gamma H\|_F^2 \frac{1}{J}\sum_{i=1}^{N}\sum_{t=1}^{T}\|z_{it}\|^2$$
$$= O_p\left(\frac{1}{J}+\frac{T}{J^{2\kappa+1}}+\frac{T}{NJ}\right) = O_p\left(\frac{1}{J}\right), \qquad \text{(E.11)}$$

where the first equality follows from Lemma E.2, and the second equality follows since $T = o(N)$, $NTJ^{-(2\kappa+1)} = o(1)$ and $J = o(\sqrt{N})$. By the Cauchy-Schwartz inequality,

$$|\mathcal{S}_5| \leq \mathcal{S}_4^{1/2}\left(\frac{1}{J}\sum_{i=1}^{T}\sum_{t=1}^{T}\|(\hat{\Gamma}-\Gamma H)'z_{it}\|^2\right)^{1/2}. \qquad \text{(E.12)}$$

Thus, the result of the lemma follows from (E.11) and (E.12) and Lemma E.1(iv).

(vi) By the fact that $\|x\|^2 = \mathrm{tr}(xx')$,

$$\frac{1}{J}\sum_{i=1}^{T}\sum_{t=1}^{T}\|(\hat{B}-BH)'\phi(z_{it})\|^2 = \frac{N}{J}\sum_{t=1}^{T}\mathrm{tr}\left((\hat{B}-BH)'\hat{Q}_t(\hat{B}-BH)\right)$$
$$\leq \frac{NT}{J}\max_{t\leq T}\lambda_{\max}(\hat{Q}_t)\|\hat{B}-BH\|_F^2 = O_p\left(\frac{NT}{J^{2\kappa+1}}+\frac{T}{N}+1\right) = O_p(1), \qquad \text{(E.13)}$$

where the second equality follows from Assumption 4.1(i) and Theorem 4.2, and the last equality follows since $T = o(N)$ and $NTJ^{-(2\kappa+1)} = o(1)$. By the Cauchy-Schwartz inequality,

$$|\mathcal{S}_6| \leq \mathcal{S}_4^{1/2}\left(\frac{1}{J}\sum_{i=1}^{T}\sum_{t=1}^{T}\|(\hat{B}-BH)'\phi(z_{it})\|^2\right)^{1/2}. \qquad \text{(E.14)}$$

Thus, the result of the lemma follows from (E.13) and (E.14) and Lemma E.1(iv). ■

**Lemma E.2.** *Suppose Assumptions 4.1-4.6 and 5.2(i)-(iii) hold. Let $\hat{\gamma}$ and $\hat{\Gamma}$ be given in Section 5.2. Assume (i) $N \to \infty$; (ii) $T \to \infty$ or $T \geq K+1$ is finite; (iii) $J \to \infty$ with $J^2\xi_J^2 \log J = o(N)$. Let $\Omega_z \equiv \sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} f_t^\dagger f_s^{\dagger\prime} \otimes Q_{z,t}^{-1}E[z_{it}z_{is}']Q_{z,s}^{-1}E[\varepsilon_{it}\varepsilon_{is}]/NT$, where $Q_{z,t} = \sum_{i=1}^{N}E[z_{it}z_{it}']/N$. Assume that $\mathrm{H}_0$ is true. Then there exists an $M\times(K+1)$ random matrix $\mathbb{N}_z$ with $\mathrm{vec}(\mathbb{N}_z) \sim N(0,\Omega_z)$ such that*

$$\|\sqrt{NT}(\hat{\gamma}-\gamma)-\mathbb{G}_\gamma\| = O_p\left(\frac{\sqrt{NT}}{J^\kappa}+\frac{\sqrt{TJ}}{\sqrt{N}}+\frac{J^{5/6}}{N^{1/6}}+\frac{\sqrt{J\xi_J}\log^{1/4}J}{N^{1/4}}\right)$$

*and*

$$\|\sqrt{NT}(\hat{\Gamma}-\Gamma H)-\mathbb{G}_\Gamma\|_F = O_p\left(\frac{\sqrt{T}}{J^\kappa}+\frac{\sqrt{T}}{\sqrt{N}}+\frac{1}{N^{1/6}}\right),$$

*where* $\mathbb{G}_\gamma = \mathbb{N}_{z,1} - \mathbb{G}_\Gamma \mathcal{H}^{-1}\bar{f} - \Gamma\mathcal{H}\mathcal{H}'B'(\mathbb{N}_1 - \mathbb{G}_B\mathcal{H}^{-1}\bar{f}) - \Gamma\mathcal{H}\mathbb{G}_B'a$, $\mathbb{G}_\Gamma = \mathbb{N}_{z,2}B'B\mathcal{M}$, $\mathcal{H}$, $\mathcal{M}$, $\mathbb{N}_1$ *and* $\mathbb{G}_B$ *are given in Theorem 4.3, and* $\mathbb{N}_{z,1}$ *and* $\mathbb{N}_{z,2}$ *are the first column and the last* $K$ *columns of* $\mathbb{N}_z$*.*

Proof: Let us begin by defining some notation. Let $\vec{\varepsilon}_t \equiv (Z_t'Z_t)^{-1}Z_t'\varepsilon_t$ and $\vec{E} \equiv (\vec{\varepsilon}_1, \dots, \vec{\varepsilon}_T)$. Then (9) under $\mathrm{H}_0$ can be written as

$$\vec{Y} = \gamma 1_T' + \Gamma F' + \vec{E}, \tag{E.15}$$

where $1_T$ denotes a $T \times 1$ vector of ones. Recall $M_T = I_T - 1_T 1_T'/T$. Post-multiplying (E.15) by $M_T$ to remove $\gamma$, we thus obtain

$$\vec{Y}M_T = \Gamma(M_T F)' + \vec{E}M_T. \tag{E.16}$$

Recall that $V$ is a $K \times K$ diagonal matrix of the first $K$ largest eigenvalues of $\tilde{Y}M_T\tilde{Y}'/T$ as defined in the proof of Theorem 4.1, $H = F'M_T\hat{F}(\hat{F}'M_T\hat{F})^{-1}$ and $\hat{F}'M_T\hat{F}/T = V$ as showed in the proof of Theorem 4.1. By the definition of $\hat{\Gamma}$, $\hat{\Gamma} = \vec{Y}M_T\hat{F}(\hat{F}'M_T\hat{F})^{-1}$. We may substitute (E.16) to $\hat{\Gamma} = \vec{Y}M_T\hat{F}(\hat{F}'M_T\hat{F})^{-1}$ to obtain

$$\hat{\Gamma} - \Gamma H = (\vec{E}M_T\tilde{Y}'/T)\hat{B}V^{-1} = \sum_{j=1}^{3}\mathcal{D}_j\hat{B}V^{-1}, \tag{E.17}$$

where in the first equality we have used $\hat{F}'M_T\hat{F}/T = V$ and $\hat{F} = \tilde{Y}'\hat{B}$, in the second equality we have substituted (A.2) into the equation, and $\mathcal{D}_1 = \vec{E}M_TFB'/T$, $\mathcal{D}_2 = \vec{E}M_T\tilde{E}'/T$ and $\mathcal{D}_3 = \vec{E}M_T\tilde{\Delta}'/T$. We can conduct the same exercise as in (C.1) to obtain

$$\begin{aligned}
&\|\sqrt{NT}(\hat{\Gamma} - \Gamma H) - \sqrt{NT}\mathcal{D}_1\hat{B}V^{-1}\|_F \\
&\leq \sqrt{NT}\|V^{-1}\|_2(\|\mathcal{D}_2\hat{B}\|_F + \|\mathcal{D}_3\|_F\|\hat{B}\|_2) = O_p\left(\frac{\sqrt{T}}{J^\kappa} + \frac{\sqrt{T}}{\sqrt{N}}\right),
\end{aligned} \tag{E.18}$$

where the equality follows by Lemmas A.2(i) and E.8. Thus, the second result of the lemma follows from (E.18) and Lemma E.9. We now show the first result of the lemma. By the definition of $\hat{\gamma}$,

$$\begin{aligned}
\hat{\gamma} - \gamma = \vec{E}1_T/T &+ (\Gamma H - \hat{\Gamma})H^{-1}\bar{f} - \hat{\Gamma}(\hat{B} - BH)'a \\
&- \hat{\Gamma}\hat{B}'(BH - \hat{B})H^{-1}\bar{f} - \hat{\Gamma}\hat{B}'\tilde{E}1_T/T - \hat{\Gamma}\hat{B}'\tilde{\Delta}1_T/T,
\end{aligned} \tag{E.19}$$

where $H^{-1}$ is well defined with probability approaching one by (A.4) and Lemma A.2(ii), and we have used $a'B = 0$ and $\hat{B}'\hat{B} = I_K$. By a similar argument as in (C.3)-(C.5),

$$\|\sqrt{NT}(\hat{\gamma} - \gamma) - [\sqrt{N/T}\vec{E}1_T - \sqrt{NT}(\hat{\Gamma} - \Gamma H)\mathcal{H}^{-1}\bar{f}]$$

$$+\Gamma\mathcal{H}\mathcal{H}'B'[\sqrt{N/T}\tilde{E}1_T-\sqrt{NT}(\hat{B}-BH)\mathcal{H}^{-1}\bar{f}]$$
$$+\Gamma\mathcal{H}\sqrt{NT}(\hat{B}-BH)'a\|=O_p\left(\frac{\sqrt{NT}}{J^\kappa}+\frac{\sqrt{TJ}}{\sqrt{N}}+\frac{J}{\sqrt{NT}}\right). \qquad (E.20)$$

Thus, the first result of the lemma follows from (E.20), Lemmas E.9, C.1 and C.2, Theorem 4.3 and the second result of the lemma. ■

**Lemma E.3.** *Suppose Assumptions 4.5(i), (iii) and 5.2(ii) hold. Let* $\hat{\mathcal{Q}}$ *and* $\mathcal{Q}$ *be given in the proof of Theorem 5.2. Then*

$$\|\hat{\mathcal{Q}}-\mathcal{Q}\|_F^2=O_p\left(\frac{J^2}{N}\right).$$

Proof: Let $\hat{\mathcal{Q}}_t\equiv\sum_{i=1}^N(z_{it}',\phi(z_{it})')(z_{it}',\phi(z_{it})')'/N$ and $\mathcal{Q}_t\equiv E[\hat{\mathcal{Q}}_t]$. Then $\hat{\mathcal{Q}}=\sum_{t=1}^T\hat{\mathcal{Q}}_t/T$ and $\mathcal{Q}=\sum_{t=1}^T\mathcal{Q}_t/T$. It follows that $E[\|\hat{\mathcal{Q}}_t-\mathcal{Q}_t\|_F^2]\leq[((J+1)M)^2/N]$ $(\max_{m\leq M,j\leq J,i\leq N,t\leq T}E[\phi_j^4(z_{it,m})]+\max_{i\leq N,t\leq T}E[\|z_{it}\|^4])$ by the independence in Assumption 4.5(iii). By the Cauchy-Schwartz inequality,

$$E[\|\hat{\mathcal{Q}}-\mathcal{Q}\|_F^2]\leq\frac{1}{T}\sum_{t=1}^T E[\|\hat{\mathcal{Q}}_t-\mathcal{Q}_t\|_F^2]=O\left(\frac{J^2}{N}\right), \qquad (E.21)$$

where the equality follows from Assumptions 4.5(i) and 5.2(ii). By the Markov's inequality, the result of the lemma thus follows from (E.21). ■

**Lemma E.4.** *Suppose Assumptions 4.1-4.6, 5.1 and 5.2(ii)-(iv) hold. Let* $\hat{\gamma}$, $\hat{\Gamma}$, $\hat{\gamma}^*$ *and* $\hat{\Gamma}^*$ *be given in Section 5.2. Assume (i)* $N\to\infty$*; (ii)* $T\to\infty$ *or* $T\geq K+1$ *is finite; (iii)* $J\to\infty$ *with* $J^2\xi_J^2\log J=o(N)$*. Assume that* $\mathrm{H}_0$ *is true. Then there exists an* $M\times(K+1)$ *random matrix* $\mathbb{N}_z^*$ *with* $\mathrm{vec}(\mathbb{N}_z^*)\sim N(0,\Omega_z)$ *conditional on* $\{Y_t,Z_t\}_{t\leq T}$ *such that*

$$\|\sqrt{NT/\omega_0}(\hat{\gamma}^*-\hat{\gamma})-\mathbb{G}_\gamma^*\|=O_{p^*}\left(\frac{\sqrt{NT}}{J^\kappa}+\frac{\sqrt{TJ}}{\sqrt{N}}+\frac{J^{5/6}}{N^{1/6}}+\frac{\sqrt{J\xi_J}\log^{1/4}J}{N^{1/4}}\right)$$

*and*

$$\|\sqrt{NT/\omega_0}(\hat{\Gamma}^*-\hat{\Gamma})-\mathbb{G}_\Gamma^*\|_F=O_{p^*}\left(\frac{\sqrt{T}}{J^\kappa}+\frac{\sqrt{T}}{\sqrt{N}}+\frac{1}{N^{1/6}}\right),$$

*where* $\Omega_z$ *is given in Lemma E.2,* $\mathbb{G}_\gamma^*=\mathbb{N}_{z,1}^*-\mathbb{G}_\Gamma^*\mathcal{H}^{-1}\bar{f}-\Gamma\mathcal{H}\mathcal{H}'B'(\mathbb{N}_1^*-\mathbb{G}_B^*\mathcal{H}^{-1}\bar{f})-\Gamma\mathcal{H}\mathbb{G}_B^{*\prime}a$*,* $\mathbb{G}_\Gamma^*=\mathbb{N}_{z,2}^*B'B\mathcal{M}$*,* $\mathcal{H}$*,* $\mathcal{M}$*,* $\mathbb{N}_1^*$ *and* $\mathbb{G}_B^*$ *are given in Theorem 5.1, and* $\mathbb{N}_{z,1}^*$ *and* $\mathbb{N}_{z,2}^*$ *are the first column and the last* $K$ *columns of* $\mathbb{N}_z^*$*.*

Proof: Let us begin by defining some notation. Let $\vec{\varepsilon}_t^*\equiv(Z_t^{*\prime}Z_t)^{-1}Z_t^{*\prime}\varepsilon_t$ and $\vec{E}^*\equiv$

$(\vec{\varepsilon}_1^*, \ldots, \vec{\varepsilon}_T^*)$. Then under $\mathrm{H}_0$, we have

$$\vec{Y}^* = \gamma 1_T' + \Gamma F' + \vec{E}^*. \tag{E.22}$$

where $1_T$ denotes a $T \times 1$ vector of ones. Recall $M_T = I_T - 1_T 1_T'/T$. Post-multiplying (E.22) by $M_T$ to remove $\gamma$, we thus obtain

$$\vec{Y}^* M_T = \Gamma(M_T F)' + \vec{E}^* M_T. \tag{E.23}$$

Recall that $V$ us a $K \times K$ diagonal matrix of the first $K$ largest eigenvalues of $\tilde{Y} M_T \tilde{Y}'/T$ as defined in the proof of Theorem 4.1, $H = F' M_T \hat{F}(\hat{F}' M_T \hat{F})^{-1}$ and $\hat{F}' M_T \hat{F}/T = V$ as showed in the proof of Theorem 4.1. By the definitions of $\hat{\Gamma}^*$, $\hat{\Gamma}^* = \vec{Y}^* M_T \hat{F}(\hat{F}' M_T \hat{F})^{-1}$. We may substitute (E.23) to $\hat{\Gamma}^* = \vec{Y}^* M_T \hat{F}(\hat{F}' M_T \hat{F})^{-1}$ to obtain

$$\hat{\Gamma}^* - \Gamma H = (\vec{E}^* M_T \tilde{Y}'/T)\hat{B} V^{-1} = \sum_{j=1}^{3} \mathcal{D}_j^* \hat{B} V^{-1}, \tag{E.24}$$

where in the first equality we have used $\hat{F}' M_T \hat{F}/T = V$ and $\hat{F} = \tilde{Y}'\hat{B}$, in the second equality follows we have substituted (A.2) into the equation, and $\mathcal{D}_1^* = \vec{E}^* M_T F B'/T$, $\mathcal{D}_2^* = \vec{E}^* M_T \tilde{E}'/T$ and $\mathcal{D}_3^* = \vec{E}^* M_T \tilde{\Delta}'/T$. We can conduct the same exercise as in (C.1) to obtain

$$\begin{aligned}
&\|\sqrt{NT}(\hat{\Gamma}^* - \Gamma H) - \sqrt{NT}\mathcal{D}_1^* \hat{B} V^{-1}\|_F \\
&\leq \sqrt{NT}\|V^{-1}\|_2(\|\mathcal{D}_2^* \hat{B}\|_F + \|\mathcal{D}_3^*\|_F \|\hat{B}\|_2) = O_p\left(\frac{\sqrt{T}}{J^\kappa} + \frac{\sqrt{T}}{\sqrt{N}}\right),
\end{aligned} \tag{E.25}$$

where the equality follows by Lemmas A.2(i) and E.10. By the fact that $\|C + D\|_F \leq \|C\|_F + \|D\|_F$, we may combine (E.18) and (E.25) to obtain

$$\|\sqrt{NT}(\hat{\Gamma}^* - \hat{\Gamma}) - \sqrt{NT}(\mathcal{D}_1^* - \mathcal{D}_1)\hat{B} V^{-1}\|_F = O_p\left(\frac{\sqrt{T}}{J^\kappa} + \frac{\sqrt{T}}{\sqrt{N}}\right). \tag{E.26}$$

Thus, the second result of the lemma follows from (E.26) and Lemmas A.5 and E.11. We now show the first result of the lemma. By the definition of $\hat{\gamma}^*$,

$$\begin{aligned}
\hat{\gamma}^* - \gamma = {} & \vec{E}^* 1_T/T + (\Gamma H - \hat{\Gamma}^*)H^{-1}\bar{f} - \hat{\Gamma}^*(\hat{B}^{*\prime}\hat{B}^*)^{-1}(\hat{B}^* - BH)'a \\
& - \hat{\Gamma}^*(\hat{B}^{*\prime}\hat{B}^*)^{-1}\hat{B}^{*\prime}(BH - \hat{B}^*)H^{-1}\bar{f} - \hat{\Gamma}^*(\hat{B}^{*\prime}\hat{B}^*)^{-1}\hat{B}^{*\prime}\tilde{E}^* 1_T/T \\
& - \hat{\Gamma}^*(\hat{B}^{*\prime}\hat{B}^*)^{-1}\hat{B}^{*\prime}\tilde{\Delta}^* 1_T/T,
\end{aligned} \tag{E.27}$$

where $H^{-1}$ is well defined with probability approaching one by (A.4) and Lemma A.2(ii), and we have used $a'B = 0$ and $(\hat{B}^{*\prime}\hat{B}^*)^{-1}\hat{B}^{*\prime}\hat{B}^* = I_K$. By a similar argument as in

(C.3)-(C.5),

$$\begin{aligned}&\|\sqrt{NT}(\hat{\gamma}^* - \gamma) - [\sqrt{N/T}\vec{E}^* 1_T - \sqrt{NT}(\hat{\Gamma}^* - \Gamma H)\mathcal{H}^{-1}\bar{f}] \\ &\quad + \Gamma\mathcal{H}\mathcal{H}' B'[\sqrt{N/T}\tilde{E}^* 1_T - \sqrt{NT}(\hat{B}^* - BH)\mathcal{H}^{-1}\bar{f}] \\ &\quad + \Gamma\mathcal{H}\sqrt{NT}(\hat{B}^* - BH)' a\| = O_p\left(\frac{\sqrt{NT}}{J^\kappa} + \frac{\sqrt{TJ}}{\sqrt{N}} + \frac{J}{\sqrt{NT}}\right).\end{aligned} \tag{E.28}$$

By the fact that $\|x+y\| \le \|x\| + \|y\|$, we may combine (E.20) and (E.28) to obtain

$$\begin{aligned}&\|\sqrt{NT}(\hat{\gamma}^* - \hat{\gamma}) - [\sqrt{N/T}(\vec{E}^* 1_T - \vec{E} 1_T) - \sqrt{NT}(\hat{\Gamma}^* - \hat{\Gamma})\mathcal{H}^{-1}\bar{f}] \\ &\quad + \Gamma\mathcal{H}\mathcal{H}' B'[\sqrt{N/T}(\tilde{E}^* 1_T - \tilde{E} 1_T) - \sqrt{NT}(\hat{B}^* - \hat{B})\mathcal{H}^{-1}\bar{f}] \\ &\quad + \Gamma\mathcal{H}\sqrt{NT}(\hat{B}^* - \hat{B})' a\| = O_p\left(\frac{\sqrt{NT}}{J^\kappa} + \frac{\sqrt{TJ}}{\sqrt{N}} + \frac{J}{\sqrt{NT}}\right).\end{aligned} \tag{E.29}$$

Thus, the first result of the lemma follows from (E.29), Lemma E.11, C.1, D.2 and D.3, Theorem 5.1 and the second result of the lemma. ■

**Lemma E.5.** *Suppose Assumptions 4.1-4.4 hold. Assume (i) $N \to \infty$; (ii) $T \to \infty$ or $T \ge K+1$ is finite; (iii) $J \to \infty$ with $J = o(\sqrt{N})$. Then*

$$\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \|\hat{\beta}(z_{it}) - H'\beta(z_{it})\|^2 = o_p(1).$$

Proof: Since $\hat{\beta}(z_{it}) = \hat{B}'\phi(z_{it})$ and $\beta(z_{it}) = B'\phi(z_{it}) + \delta(z_{it})$,

$$\frac{1}{J}\sum_{i=1}^N\sum_{t=1}^T \|\hat{\beta}(z_{it}) - H'\beta(z_{it})\|^2 \le \frac{2}{J}\sum_{i=1}^T\sum_{t=1}^T \|(\hat{B} - BH)'\phi(z_{it})\|^2 + 2\mathcal{S}_4, \tag{E.30}$$

where $\mathcal{S}_4 = \sum_{i=1}^T\sum_{t=1}^N \|H'\delta(z_{it})\|^2/J$ as defined in the proof of Theorem 5.2. Note that (E.13) and Lemma E.1(iv) continue to hold under $\mathrm{H}_1$. Thus, the result of the lemma follows from (E.13) and Lemma E.1(iv). ■

**Lemma E.6.** *Suppose Assumptions 4.1-4.4, 4.5(iii), 5.2(i), (ii) and (v) hold. Assume (i) $N \to \infty$; (ii) $T \to \infty$ or $T \ge K+1$ is finite; (iii) $J \to \infty$ with $J = o(\sqrt{N})$. Assume that $\mathrm{H}_1$ is true. Then there exists positive constant $c_0$ such that*

$$\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \|\hat{\Gamma}' z_{it} - H'\beta(z_{it})\|^2 \ge c_0 + o_p(1).$$

Proof: Let us begin by defining some notation. Let $\vec{A}_t \equiv (Z_t' Z_t)^{-1} Z_t' A_t$ for $A_t = Y_t, \Psi_t, \varepsilon_t$, where $\Psi_t = (\alpha(z_{1t}) + \beta(z_{1t})' f_t, \ldots, \alpha(z_{Nt}) + \beta(z_{Nt})' f_t)'$. Let $\vec{Y} \equiv (\vec{Y}_1, \ldots, \vec{Y}_T)$, $\vec{\Psi} \equiv (\vec{\Psi}_1, \ldots, \vec{\Psi}_T)$ and $\vec{E} \equiv (\vec{\varepsilon}_1, \ldots, \vec{\varepsilon}_T)$. Then $\hat{\Gamma} = (\vec{\Psi} + \vec{E}) M_T \hat{F}(\hat{F}' M_T \hat{F})^{-1}$. It is straightforward to show that $\hat{\Gamma} = (\vec{\Psi} M_T F/T)(F' M_T F/T)^{-1} H + o_p(1)$ by Theorem 4.1, and $\|(\vec{\Psi} M_T F/T)(F' M_T F/T)^{-1}\|_F \le C^*$ for some $C^*$ with probability approaching one. This together with Lemma A.2(ii) implies that $P(\|\hat{\Gamma} H^{-1}\|_F > C) = o(1)$. Therefore, under $\mathrm{H}_1$,

$$
\begin{aligned}
\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\|\hat{\Gamma}' z_{it} - H'\beta(z_{it})\|^2 &\ge \lambda_{\min}(H'H)\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\|(\hat{\Gamma} H^{-1})' z_{it} - \beta(z_{it})\|^2 \\
&= \lambda_{\min}(H'H)\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} E\left[\|\beta(z_{it}) - (\hat{\Gamma} H^{-1})' z_{it}\|^2\right] + o_p(1) \\
&\ge \lambda_{\min}(H'H)\inf_{i\le N, t\le T}\inf_{\Pi} E[\|\beta(z_{it}) - \Pi' z_{it}\|^2] + o_p(1) \\
&\ge c_0 + o_p(1) \text{ for some } c_0 > 0,
\end{aligned}
\tag{E.31}
$$

where the equality follows from Lemma E.12 since $P(\|\hat{\Gamma} H^{-1}\|_F > C) = o(1)$, and the last inequality follows by Lemma A.2(ii). ■

**Lemma E.7.** *Suppose Assumptions 4.1-4.4, 4.5(iii), 5.2 hold. Assume (i) $N \to \infty$; (ii) $T \to \infty$ or $T \ge K + 1$ is finite; (iii) $J \to \infty$ with $J = o(\sqrt{N})$. Then*

$$
\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}|(\hat{\gamma}^* - \hat{\gamma})' z_{it}|^2 + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}|(\hat{a}^* - \hat{a})'\phi(z_{it})|^2 = o_p(1)
$$

*and*

$$
\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\|(\hat{\Gamma}^* - \hat{\Gamma})' z_{it}\|^2 + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\|(\hat{B}^* - \hat{B})'\phi(z_{it})\|^2 = o_p(1).
$$

Proof: We prove the second result, and the proof of the first result is similar. Note that (E.13) continue to hold under $\mathrm{H}_1$, so the second term on the left-hand side of the second result is $o_p(1)$. For the first term, we have

$$
\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\|(\hat{\Gamma}^* - \hat{\Gamma})' z_{it}\|^2 \le \|\hat{\Gamma}^* - \hat{\Gamma}\|_F^2 \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\|z_{it}\|^2. \tag{E.32}
$$

Let $\vec{A}_t^* \equiv (Z_t^{*\prime} Z_t)^{-1} Z_t^{*\prime} A_t$ for $A_t = Y_t, \Psi_t, \varepsilon_t$, where $\Psi_t = (\alpha(z_{1t}) + \beta(z_{1t})' f_t, \dots, \alpha(z_{Nt}) + \beta(z_{Nt})' f_t)'$. Let $\vec{Y}^* \equiv (\vec{Y}_1^*, \dots, \vec{Y}_T^*)$, $\vec{\Psi}^* \equiv (\vec{\Psi}_1^*, \dots, \vec{\Psi}_T^*)$ and $\vec{E}^* \equiv (\vec{\varepsilon}_1^*, \dots, \vec{\varepsilon}_T^*)$. Then $\hat{\Gamma}^* = (\vec{\Psi}^* + \vec{E}^*) M_T \hat{F}(\hat{F}' M_T \hat{F})^{-1}$. It is straightforward to show that $\hat{\Gamma}^* = (\vec{\Psi}^* M_T F/T)(F' M_T F/T)^{-1} H + o_p(1)$ by Theorem 4.1. From the proof of Lemma E.6, $\hat{\Gamma} = (\vec{\Psi} M_T F/T)(F' M_T F/T)^{-1} H + o_p(1)$. Moreover, it can be easily shown that $(\vec{\Psi}^* - \vec{\Psi}) M_T F/T = o_p(1)$. Thus,

$$
\hat{\Gamma}^* - \hat{\Gamma} = (\vec{\Psi}^* - \vec{\Psi}) F/T (F'F/T)^{-1} = o_p(1). \tag{E.33}
$$

By Assumption 5.2(ii), $\sum_{i=1}^{N}\sum_{t=1}^{T}\|z_{it}\|^2/NT = O_p(1)$ by the Markov's inequality. This together with (E.32) and (E.33) implies that the first term is also $o_p(1)$. ■

**Lemma E.8.** *Let $\mathcal{D}_2$ and $\mathcal{D}_3$ be given in the proof of Lemma E.2.*
*(i) Assume (i) $N \to \infty$; (ii) $T \to \infty$ or $T \geq K+1$ is finite; (iii) $J \to \infty$ with $J^2\xi_J^2 \log J = o(N)$. Under Assumptions 4.1-4.5, 5.2(i) and (ii), $\|\mathcal{D}_2\hat{B}\|_F^2 = O_p(1/N^2)$.*
*(ii) Under Assumptions 4.1(i), 4.2(ii), (iv), 4.3, 5.2(i) and (ii), $\|\mathcal{D}_3\|_F^2 = O_p(J^{-2\kappa}/N)$.*

PROOF: (i) By Assumptions 4.3, 5.2(i) and (ii), we may follow a similar argument as in the proof of Lemma A.3(ii) to obtain $\|\vec{E}\|_F^2/T = O_p(1/N)$. Since $\|\mathcal{D}_2\hat{B}\|_F \leq \|\hat{B}'\tilde{E}\|_F\|\vec{E}\|_F/T$, the result then follows from Lemmas B.2(i).

(ii) Note that $\|\vec{E}\|_F^2/T = O_p(1/N)$ from the proof of (i). Since $\|\mathcal{D}_3\|_F \leq \|\tilde{\Delta}\|_F\|\vec{E}\|_F/T$, the result then immediately follows from Lemmas A.3(i). ■

**Lemma E.9.** *Suppose Assumptions 4.1-4.3, 4.5(iii), (iv), 4.6, 5.2(i)-(iii) hold. Let $V$ be given in the proof of Theorem 4.1, $\mathcal{D}_1$ and $\vec{E}$ be given in the proof of Lemma E.2. Assume (i) $N \to \infty$; (ii) $T \to \infty$ or $T \geq K+1$ is finite; (iii) $J \to \infty$ with $J = o(\sqrt{N})$. Then there exists an $M \times (K+1)$ random matrix $\mathbb{N}_z$ with $\mathrm{vec}(\mathbb{N}_z) \sim N(0, \Omega_z)$ such that*

$$\|\sqrt{NT}\mathcal{D}_1\hat{B}V^{-1} - \mathbb{G}_\Gamma\|_F = O_p\left(\frac{1}{J^\kappa} + \frac{1}{N^{1/6}}\right)$$

*and*

$$\|\sqrt{N/T}\vec{E}1_T - \mathbb{N}_{z,1}\| = O_p\left(\frac{1}{N^{1/6}}\right),$$

*where $\Omega_z$ is given in Lemma E.2, $\mathbb{G}_\Gamma = \mathbb{N}_{z,2}B'B\mathcal{M}$, $\mathcal{M}$ is a nonrandom matrix given in Lemma C.3, and $\mathbb{N}_{z,1}$ and $\mathbb{N}_{z,2}$ are first column and the last $K$ columns of $\mathbb{N}_z$.*

PROOF: Let $\mathcal{L}_{NT,z} \equiv \sum_{t=1}^T Q_{t,z}^{-1}Z_t'\varepsilon_t(f_t - \bar{f})'/\sqrt{NT}$ and $\ell_{NT,z} \equiv \sum_{t=1}^T Q_{t,z}^{-1}Z_t'\varepsilon_t/\sqrt{NT}$. By a similar argument as in the proof of Lemma C.1,

$$\|\sqrt{NT}\mathcal{D}_1\hat{B}V^{-1} - \mathcal{L}_{NT,z}B'B\mathcal{M}\|_F = O_p\left(\frac{1}{J^\kappa} + \frac{1}{N^{1/4}}\right) \tag{E.34}$$

and

$$\|\sqrt{N/T}\vec{E}1_T - \ell_{NT,z}\| = O_p\left(\frac{1}{N^{1/4}}\right). \tag{E.35}$$

By a similar argument as in the proof of Lemma C.2, there exists an $M \times (K+1)$ random matrix $\mathbb{N}_z$ with $\mathrm{vec}(\mathbb{N}_z) \sim N(0, \Omega_z)$ such that

$$\|(\ell_{NT,z}, \mathcal{L}_{NT,z}) - \mathbb{N}_z\|_F = O_p\left(\frac{1}{N^{1/6}}\right). \tag{E.36}$$

Thus the result of the lemma follows from (E.34)-(E.36). ■

**Lemma E.10.** *Let $\mathcal{D}_2^*$ and $\mathcal{D}_3^*$ be given in the proof of Lemma E.4.*
*(i) Assume (i) $N \to \infty$; (ii) $T \to \infty$ or $T \geq K+1$ is finite; (iii) $J \to \infty$ with $J^2\xi_J^2 \log J = o(N)$. Under Assumptions 4.1-4.5, 5.1(i), 5.2(ii) and (iv), $\|\mathcal{D}_2^*\hat{B}\|_F^2 = O_p(1/N^2)$.*
*(ii) Under Assumptions 4.1(i), 4.2(ii), (iv), 4.3, 5.1(i), 5.2(ii) and (iv), $\|\mathcal{D}_3^*\|_F^2 = O_p(J^{-2\kappa}/N)$.*

PROOF: (i) By Assumptions 4.3, 5.1(i), 5.2 (ii) and (iv), we may follow a similar argument as in the proof of Lemma D.4(ii) to obtain $\|\vec{E}^*\|_F^2/T = O_p(1/N)$. Since $\|\mathcal{D}_2^*\hat{B}\|_F \leq \|\hat{B}'\tilde{E}\|_F\|\vec{E}^*\|_F/T$, the result then follows from Lemmas B.2(i).

(ii) Note that $\|\vec{E}^*\|_F^2/T = O_p(1/N)$ from the proof of (i). Since $\|\mathcal{D}^*{}_3\|_F \leq \|\tilde{\Delta}\|_F\|\vec{E}^*\|_F/T$, the result then immediately follows from Lemmas A.3(i). ■

**Lemma E.11.** *Suppose Assumptions 4.1-4.3, 4.5(iii), (iv), 4.6, 5.1(i) and 5.2(ii)-(iv) hold. Let $V$ be given in the proof of Theorem 4.1, $\mathcal{D}_1$ and $\vec{E}$ be given in the proof of Lemma E.2, and $\mathcal{D}_1^*$ and $\vec{E}^*$ be given in the proof of Lemma E.4. Assume (i) $N \to \infty$; (ii) $T \to \infty$ or $T \geq K+1$ is finite; (iii) $J \to \infty$ with $J = o(\sqrt{N})$. Then there exists an $M \times (K+1)$ random matrix $\mathbb{N}_z^*$ with $\mathrm{vec}(\mathbb{N}_z^*) \sim N(0, \Omega_z)$ conditional on $\{Y_t, Z_t\}_{t\leq T}$ such that*

$$\|\sqrt{NT}(\mathcal{D}_1^* - \mathcal{D}_1)\hat{B}V^{-1} - \sqrt{\omega_0}\mathbb{G}_\Gamma^*\|_F = O_p\left(\frac{1}{J^\kappa} + \frac{1}{N^{1/6}}\right)$$

*and*

$$\|\sqrt{N/T}(\vec{E}^*1_T - \vec{E}1_T) - \sqrt{\omega_0}\mathbb{N}_{z,1}^*\| = O_p\left(\frac{1}{N^{1/6}}\right),$$

*where $\Omega_z$ is given in Lemma E.2, $\mathbb{G}_\Gamma^* = \mathbb{N}_{z,2}^* B'B\mathcal{M}$, $\mathcal{M}$ is a nonrandom matrix given in Lemma C.3, and $\mathbb{N}_{z,1}^*$ and $\mathbb{N}_{z,2}^*$ are first column and the last $K$ columns of $\mathbb{N}_z^*$.*

PROOF: Let $\mathcal{L}_{NT,z}^{**} \equiv \sum_{t=1}^T Q_{t,z}^{-1} Z_t^{*\prime}\varepsilon_t(f_t - \bar{f})'/\sqrt{NT}$ and $\ell_{NT,z}^{**} \equiv \sum_{t=1}^T Q_{t,z}^{-1} Z_t^{*\prime}\varepsilon_t/\sqrt{NT}$. By a similar argument as in the proof of Lemma D.2,

$$\|\sqrt{NT}\mathcal{D}_1^*\hat{B}V^{-1} - \mathcal{L}_{NT,z}^{**}B'B\mathcal{M}\|_F = O_p\left(\frac{1}{J^\kappa} + \frac{1}{N^{1/4}}\right) \tag{E.37}$$

and

$$\|\sqrt{N/T}\vec{E}^*1_T - \ell_{NT,z}^{**}\| = O_p\left(\frac{1}{N^{1/4}}\right). \tag{E.38}$$

Let $\mathcal{L}_{NT,z}^* \equiv \sum_{t=1}^T Q_{t,z}^{-1}(Z_t^* - Z_t)'\varepsilon_t(f_t - \bar{f})'/\sqrt{NT} = \mathcal{L}_{NT,z}^{**} - \mathcal{L}_{NT,z}$ and $\ell_{NT,z}^* \equiv \sum_{t=1}^T Q_{t,z}^{-1}(Z_t^* - Z_t)'\varepsilon_t/\sqrt{NT} = \ell_{NT,z}^{**} - \ell_{NT,z}$. By a similar argument as in the proof of Lemma D.3, there exists an $M \times (K+1)$ random matrix $\mathbb{N}_z^*$ with $\mathrm{vec}(\mathbb{N}_z^*) \sim N(0, \Omega_z)$

conditional on $\{Y_t, Z_t\}_{t\le T}$ such that

$$\|(\ell^*_{NT,z}, \mathcal{L}^*_{NT,z}) - \sqrt{\omega_0}\mathbb{N}^*_z\|_F = O_p\left(\frac{1}{N^{1/6}}\right). \tag{E.39}$$

Thus, the result of the lemma follows from (E.34),(E.35) and (E.37)-(E.39). ■

**Lemma E.12.** *Suppose Assumptions 4.5(iii), 5.2(ii) and (v) hold. For any given positive constant $C$,*

$$\sup_{\|\Pi\|_F\le C}\left|\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\|\beta(z_{it})-\Pi' z_{it}\|^2-\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T E\left[\|\beta(z_{it})-\Pi' z_{it}\|^2\right]\right|=o_p(1).$$

Proof: Let $\mathcal{A}_C \equiv \{\Pi \in \mathbf{R}^{M\times K}, \|\Pi\|_F \le C\}$ for $C > 0$, and $\mathcal{F}_C \equiv \{\zeta(\cdot,\Pi) : \zeta(z_1,\cdots,z_T,\Pi) = \sum_{t=1}^T \|\beta(z_t) - \Pi' z_t\|^2/T \text{ for } \Pi \in \mathcal{A}_C\}$ be a class of functions $\zeta(\cdot,\Pi)$ indexed by $\Pi \in \mathcal{A}_C$. We aim to show $\sup_{\Pi\in\mathcal{A}_C}|\frac{1}{N}\sum_{i=1}^N \zeta(z_{i1},\cdots,z_{iT},\Pi) - \frac{1}{N}\sum_{i=1}^N E[\zeta(z_{i1},\cdots,z_{iT},\Pi)]| = o_p(1)$. It follows that for any $\Pi_1, \Pi_2 \in \mathcal{A}_C$,

$$\begin{aligned}
&|\zeta(z_1,\cdots,z_T,\Pi_1)-\zeta(z_1,\cdots,z_T,\Pi_2)|\\
\le{}& \|\Pi_1-\Pi_2\|_F\frac{1}{T}\sum_{t=1}^T\|z_t\|(\|\beta(z_t)-\Pi_1' z_t\|+\|\beta(z_t)-\Pi_2' z_t\|)\\
\le{}& \|\Pi_1-\Pi_2\|_F\frac{2}{T}\sum_{t=1}^T(\|z_t\|\|\beta(z_t)\|+C\|z_t\|^2)\equiv\|\Pi_1-\Pi_2\|_F G(z_1,\cdots,z_T).
\end{aligned}\tag{E.40}$$

By Assumptions 5.2(ii) and (v), $\max_{i\le N} E[G(z_{i1},\cdots,z_{iT})] < \infty$. This together with (E.40) implies that and $\mathcal{F}_C$ is a class of functions that are Lipschitz in the index $\Pi \in \mathcal{A}_C$ with envelop function $G$. Since $\mathcal{A}_C$ is compact, for every $\epsilon > 0$, the covering number $N(\epsilon, \mathcal{A}_C, \|\cdot\|_F)$ of $\mathcal{A}_C$ with respect to $\|\cdot\|_F$ is bounded. By Theorem 2.7.11 of van der Vaart and Wellner (1996), for every $\epsilon > 0$, the bracketing number $N_{[]}(\epsilon, \mathcal{F}_C, L_1(P))$ of $\mathcal{F}_C$ with respect to $L_1(P)$ is bounded. Thus, the result of the lemma follows by the Glivenko-Cantelli theorem (e.g., Theorem 2.4.1 of van der Vaart and Wellner (1996)). ■

# Appendix F - Proof of Theorem 6.1

## F.1 Proof of Theorem 6.1

Proof of Theorem 6.1: (A) Let $\theta_k \equiv \lambda_k(\tilde{Y}M_T\tilde{Y}'/T)/\lambda_{k+1}(\tilde{Y}M_T\tilde{Y}'/T)$. If $\hat{K} \ne K$, then there exists some $1 \le k \le K-1$ or $K+1 \le k \le JM/2$ such that $\theta_k \ge \theta_K$. Let $\underline{JM/2}$ be the integer part of $JM/2$. Since $\lambda_1(\tilde{Y}M_T\tilde{Y}'/T)/\lambda_K(\tilde{Y}M_T\tilde{Y}'/T) \ge \theta_k$ for all $1 \le k \le K-1$ and $\lambda_{K+1}(\tilde{Y}M_T\tilde{Y}'/T)/\lambda_{\underline{JM/2}}(\tilde{Y}M_T\tilde{Y}'/T) \ge \theta_k$ for all $K+1 \le k \le JM/2$, the event of $\hat{K} \ne K$ implies the event of $\lambda_1(\tilde{Y}M_T\tilde{Y}'/T)/\lambda_K(\tilde{Y}M_T\tilde{Y}'/T) \ge \theta_K$ or the event of $\lambda_{K+1}(\tilde{Y}M_T\tilde{Y}'/T)/\lambda_{\underline{JM/2}}(\tilde{Y}M_T\tilde{Y}'/T) \geq \theta_K$. Thus, we have

$$P(\hat{K} \neq K) \leq P\left(\frac{\lambda_1(\tilde{Y}M_T\tilde{Y}'/T)}{\lambda_K(\tilde{Y}M_T\tilde{Y}'/T)} \geq \theta_K\right) + P\left(\frac{\lambda_{K+1}(\tilde{Y}M_T\tilde{Y}'/T)}{\lambda_{\underline{JM/2}}(\tilde{Y}M_T\tilde{Y}'/T)} \geq \theta_K\right). \quad \text{(F.1)}$$

By Lemmas F.1 and F.2, $\lambda_1(\tilde{Y}M_T\tilde{Y}'/T)/\lambda_K(\tilde{Y}M_T\tilde{Y}'/T) = O_p(1)$, $\theta_K/N = C + o_p(1)$ for some positive constant $C$, and $\lambda_{K+1}(\tilde{Y}M_T\tilde{Y}'/T)/\lambda_{\underline{JM/2}}(\tilde{Y}M_T\tilde{Y}'/T) = O_p(1)$, since $\underline{JM/2} + 1 < JM - K - 1$ for large $J$. Thus, this together with (F.1) implies $P(\hat{K} \neq K) \to 0$.

(B) If $\tilde{K} \neq K$, then $\lambda_{K-1}(\tilde{Y}M_T\tilde{Y}'/T) < \lambda_{NT}$ or $\lambda_{K+1}(\tilde{Y}M_T\tilde{Y}'/T) \geq \lambda_{NT}$. Thus, we have

$$P(\tilde{K} \neq K) \leq P\left(\lambda_{K-1}(\tilde{Y}M_T\tilde{Y}'/T) < \lambda_{NT}\right) + P\left(\lambda_{K+1}(\tilde{Y}M_T\tilde{Y}'/T) \geq \lambda_{NT}\right). \quad \text{(F.2)}$$

By Lemma F.1 and $\lambda_{NT} \to 0$, $P(\lambda_{K-1}(\tilde{Y}M_T\tilde{Y}'/T) < \lambda_{NT}) \to 0$. For a matrix $A$, let $\sigma_k(A)$ denote the $k$th largest singular value of $A$. Noting that $\lambda_k(AA') = \sigma_k^2(A)$, it follows that

$$\begin{aligned}\lambda_{K+1}(\tilde{Y}M_T\tilde{Y}'/T) &= \sigma_{K+1}^2(\tilde{Y}M_T/\sqrt{T}) = |\sigma_{K+1}(\tilde{Y}M_T/\sqrt{T}) - \sigma_{K+1}(BF'M_T/\sqrt{T})|^2 \\ &\leq \frac{1}{T}\|\tilde{Y}M_T - B(M_TF)'\|_F^2 \leq \frac{2}{T}\|\tilde{\Delta}\|_F^2 + \frac{2}{T}\|\tilde{E}\|_F^2 = O_p\left(\frac{1}{J^{2\kappa}} + \frac{J}{N}\right), \quad \text{(F.3)}\end{aligned}$$

where the second equality follows since the rank of $B(M_TF)'$ is not greater than $K$, the first inequality follows by the Weyl's inequality, the second inequality by (A.2) and the Cauchy-Schwartz inequality, and the last equality follows from Lemmas A.3(i) and (ii). Since $\lambda_{NT}\min\{N/J, J^{2\kappa}\} \to \infty$, (F.3) implies that $P(\lambda_{K+1}(\tilde{Y}M_T\tilde{Y}'/T) \geq \lambda_{NT}) \to 0$. This completes the proof of the theorem. ■

## F.2 Technical Lemmas

**Lemma F.1.** *Suppose Assumptions 4.1-4.3 hold. Assume (i) $N \to \infty$; (ii) $T \to \infty$ or $T \geq K+1$ is finite; (iii) $J \to \infty$ with $J = o(\sqrt{N})$. Then there exist positive constants $c_1$ and $c_2$ such that*

$$c_1 + o_p(1) \leq \lambda_K(\tilde{Y}M_T\tilde{Y}'/T) \leq \lambda_1(\tilde{Y}M_T\tilde{Y}'/T) \leq c_2 + o_p(1).$$

PROOF: By (A.12), $\lambda_k(\tilde{Y}M_T\tilde{Y}'/T) = \lambda_k((F'M_TF/T)B'B) + o_p(1)$ for $k = 1, \ldots, K$. Thus, the result of the lemma immediately follows from Assumptions 4.2(i)-(iii). ■

**Lemma F.2.** *Suppose Assumptions 4.1(i), 4.2(ii), (iv), 4.3(i), 4.5(i) and 6.1 hold. Assume (i) $N \to \infty$; (ii) $T \to \infty$; (iii) $J \to \infty$ with $J = o(\min\{\sqrt{N}, \sqrt{T}\})$ and*

$J^{-2\kappa}N = o(1)$. *Then there exist positive constants* $c_3$ *and* $c_4$ *such that*

$$c_3 + o_p(1) \leq N\lambda_{JM-K-1}(\tilde{Y}M_T\tilde{Y}'/T) \leq N\lambda_{K+1}(\tilde{Y}M_T\tilde{Y}'/T) \leq c_4 + o_p(1).$$

Proof: For a matrix $A$, let $\sigma_k(A)$ denote the $k$th largest singular value of $A$. Noting that $\lambda_k(AA') = \sigma_k^2(A)$, it follows that for $k = 1, \ldots, JM - K$,

$$\begin{aligned}
&|\lambda_{K+k}(\tilde{Y}M_T\tilde{Y}') - \lambda_{K+k}((BF' + \tilde{E})M_T(BF' + \tilde{E})')| \\
&\leq |\sigma_{K+k}(\tilde{Y}M_T) - \sigma_{K+k}((BF' + \tilde{E})M_T)|^2 + 2|\sigma_{K+k}(\tilde{Y}M_T) \\
&\quad - \sigma_{K+k}((BF' + \tilde{E})M_T)|\sigma_{K+k}((BF' + \tilde{E})M_T) \\
&\leq \|\tilde{Y}M_T - (BF' + \tilde{E})M_T\|_F^2 + 2\|\tilde{Y}M_T - (BF' + \tilde{E})M_T\|_F \\
&\quad \times \lambda_{K+k}^{1/2}((BF' + \tilde{E})M_T(BF' + \tilde{E})') \\
&\leq \|\tilde{\Delta}\|_F^2 + 2\|\tilde{\Delta}\|_F\lambda_{K+1}^{1/2}((BF' + \tilde{E})M_T(BF' + \tilde{E})'),
\end{aligned} \tag{F.4}$$

where the first inequality is due to the triangle inequality, the second inequality follows by the Weyl's inequality, and the third inequality follows from (A.2) and the fact that $\lambda_{K+k}((BF' + \tilde{E})M_T(BF' + \tilde{E})') \leq \lambda_{K+1}((BF' + \tilde{E})M_T(BF' + \tilde{E})')$ for $k \geq 1$. We next show that the right-hand side of (F.4) is asymptotically negligible and study the behavior of $\lambda_{K+k}((BF' + \tilde{E})M_T(BF' + \tilde{E})')$. Let $\tilde{B} = B + \tilde{E}M_TF(F'M_TF)^{-1}$ and $M_F = I_T - M_TF(F'M_TF)^{-1}(M_TF)'$. We may decompose $(BF' + \tilde{E})M_T(BF' + \tilde{E})'$ by

$$(BF' + \tilde{E})M_T(BF' + \tilde{E})' = \tilde{B}F'M_TF\tilde{B}' + \tilde{E}M_TM_FM_T\tilde{E}'. \tag{F.5}$$

Then, (F.5) implies that for $k = 1, \ldots, JM - K$,

$$\begin{aligned}
\lambda_{K+k}((BF' + \tilde{E})M_T(BF' + \tilde{E})') &\leq \lambda_{K+1}(\tilde{B}F'M_TF\tilde{B}') \\
&\quad + \lambda_k(\tilde{E}M_TM_FM_T\tilde{E}') \leq \lambda_k(\tilde{E}M_T\tilde{E}') \leq \lambda_k(\tilde{E}\tilde{E}'),
\end{aligned} \tag{F.6}$$

where the first inequality follows by Lemma F.3(i), the second inequality follows by Lemma F.3(ii) and the fact that the rank of $\tilde{B}F'M_TF\tilde{B}'$ is not greater than $K$ and $I - M_F$ is positive semi-definite, and the third inequality follows since $I - M_T$ is positive semi-definite. Moreover, (F.5) also implies that for $k = 1, \ldots, JM - 2K - 1$,

$$\begin{aligned}
&\lambda_{K+k}((BF' + \tilde{E})M_T(BF' + \tilde{E})') \geq \lambda_{K+k}(\tilde{E}M_TM_FM_T\tilde{E}') \\
&\quad = \lambda_{K+k}(E\tilde{M}_TM_FM_T\tilde{E}') + \lambda_{K+1}(\tilde{E}M_T(I - M_F)M_T\tilde{E}') \geq \lambda_{2K+k}(\tilde{E}M_T\tilde{E}') \\
&\quad = \lambda_{2K+k}(\tilde{E}M_T\tilde{E}') + \lambda_2(\tilde{E}(I_T - M_T)\tilde{E}') \geq \lambda_{2K+k+1}(\tilde{E}\tilde{E}'),
\end{aligned} \tag{F.7}$$

where the first inequality follows by Lemma F.3(ii), the first equality follows since the rank of $\tilde{E}M_T(I - M_F)M_T\tilde{E}'$ is not greater than $K$, the second inequality follows by Lemma F.3(i), and the second equality and the third inequality follow similarly. Putting

(F.6) and (F.7) together implies that eigenvalues of $(BF'+\tilde{E})M_T(BF'+\tilde{E})'$ are bounded by those of $\tilde{E}\tilde{E}'$. Thus, we may study the behavior of the eigenvalues of $\tilde{E}\tilde{E}'$. Recall that $\mathcal{A}_{NT} = \sum_{t=1}^{T} \hat{Q}_t^{-1}\Phi(Z_t)'E[\varepsilon_t\varepsilon_t']\Phi(Z_t)\hat{Q}_t^{-1}/NT$ in Lemma F.4. By the Weyl's inequality and Lemma F.4,

$$\sup_{k\leq JM} |\lambda_k(N\tilde{E}\tilde{E}'/T) - \lambda_k(\mathcal{A}_{NT})| \leq \|N\tilde{E}\tilde{E}'/T - \mathcal{A}_{NT}\|_F = o_p(1). \tag{F.8}$$

This implies that the eigenvalues of $N\tilde{E}\tilde{E}'/T$ and $\mathcal{A}_{NT}$ are asymptotically equivalent. Then, it follows from (F.6) and (F.8) that

$$\begin{aligned} &\lambda_{K+1}(N(BF'+\tilde{E})M_T(BF'+\tilde{E})'/T) \\ &\qquad \leq \lambda_1(N\tilde{E}\tilde{E}'/T) \leq \lambda_1\left(\mathcal{A}_{NT}\right) + o_p(1) = O_p(1), \end{aligned} \tag{F.9}$$

where the equality follows since $\lambda_1(\mathcal{A}_{NT}) \leq (\min_{t\leq T}\lambda_{\min}(\hat{Q}_t))^{-1}\max_{t\leq T}\lambda_{\max}(E[\varepsilon_t\varepsilon_t'])$ $= O_p(1)$ by Assumptions 4.1(i) and 6.1 (i). Thus, combining (F.4), (F.9) and Lemma A.3(i) yields

$$\sup_{k\leq JM-K} |N\lambda_{K+k}(\tilde{Y}M_T\tilde{Y}'/T) - N\lambda_{K+k}((BF'+\tilde{E})M_T(BF'+\tilde{E})'/T)| = o_p(1). \tag{F.10}$$

This means that $N\lambda_{K+k}(\tilde{Y}M_T\tilde{Y}'/T)$ and $N\lambda_{K+k}((BF'+\tilde{E})M_T(BF'+\tilde{E})'/T)$ are asymptotically equivalent. By the triangle inequality, it follows from (F.6)-(F.8) and (F.10) that

$$\begin{aligned} &\lambda_{JM}(\mathcal{A}_{NT}) + o_p(1) \leq N\lambda_{JM-K-1}(\tilde{Y}M_T\tilde{Y}'/T) \\ &\qquad \leq N\lambda_{K+1}(\tilde{Y}M_T\tilde{Y}'/T) \leq \lambda_1(\mathcal{A}_{NT}) + o_p(1). \end{aligned} \tag{F.11}$$

Noting that $\lambda_1(\mathcal{A}_{NT}) \leq (\min_{t\leq T}\lambda_{\min}(\hat{Q}_t))^{-1}\max_{t\leq T}\lambda_{\max}(E[\varepsilon_t\varepsilon_t'])$ and $\lambda_{JM}(\mathcal{A}_{NT}) \geq (\max_{t\leq T}\lambda_{\max}(\hat{Q}_t))^{-1}\min_{t\leq T}\lambda_{\min}(E[\varepsilon_t\varepsilon_t'])$, the result of the lemma then follows from (F.11), Assumptions 4.1(i) and 6.1(i). ■

**Lemma F.3** (Weyl's inequalities). *Let $C$ and $D$ be $k\times k$ symmetric matrices.*
*(i) For every $i,j\geq 1$ and $i+j-1\leq k$,*

$$\lambda_{i+j-1}(C+D) \leq \lambda_i(C) + \lambda_j(D).$$

*(ii) If $D$ is positive semi-definite, for all $1\leq i\leq k$,*

$$\lambda_i(C+D) \geq \lambda_i(C).$$

Proof: The results can be found in Section III.2 of Bhatia (1997). Also, see the appendices of Ahn and Horenstein (2013) and Fan et al. (2016b). ■

**Lemma F.4.** *Let* $\mathcal{A}_{NT} \equiv \sum_{t=1}^T \hat{Q}_t^{-1}\Phi(Z_t)'E[\varepsilon_t\varepsilon_t']\Phi(Z_t)\hat{Q}_t^{-1}/NT$ *and* $\tilde{E}$ *be given in the proof of Theorem 4.1. Under Assumptions 4.1(i), 4.3(i), 4.5(i) and 6.1(ii),*

$$\|N\tilde{E}\tilde{E}'/T - \mathcal{A}_{NT}\|_F^2 = O_p\left(\frac{J^2}{N} + \frac{J^2}{T}\right).$$

Proof: Let $E_\varepsilon$ denote the expectation with respect to $\{\varepsilon_t\}_{t\le T}$. To simplify the notation, let $\hat{\psi}_{it} \equiv \phi(z_{it})\hat{Q}_t^{-1}$ and $\nu_{ijt} \equiv \varepsilon_{it}\varepsilon_{jt} - E[\varepsilon_{it}\varepsilon_{jt}]$. Since $\|A\|_F^2 = \text{tr}(AA')$,

$$E_\varepsilon[\|\bar{E}\bar{E}'/NT - \mathcal{A}_{NT}\|_F^2] = \frac{1}{N^2T^2}E_\varepsilon\left[\text{tr}\left(\sum_{t=1}^T\sum_{s=1}^T\sum_{i=1}^N\sum_{j=1}^N\sum_{k=1}^N\sum_{\ell=1}^N \hat{\psi}_{it}\hat{\psi}_{jt}'\nu_{ijt}\nu_{k\ell s}\hat{\psi}_{\ell s}\hat{\psi}_{ks}'\right)\right]$$
$$= \frac{1}{N^2T^2}\sum_{t=1}^T\sum_{s=1}^T\sum_{i=1}^N\sum_{j=1}^N\sum_{k=1}^N\sum_{\ell=1}^N \hat{\psi}_{it}'\hat{\psi}_{ks}\hat{\psi}_{jt}'\hat{\psi}_{\ell s} cov(\varepsilon_{it}\varepsilon_{jt}, \varepsilon_{ks}\varepsilon_{\ell s})$$
$$= C_{NT}\frac{1}{N^2T^2}\sum_{t=1}^T\sum_{s=1}^T\sum_{i=1}^N\sum_{j=1}^N\sum_{k=1}^N\sum_{\ell=1}^N \|\phi(z_{it})\|\|\phi(z_{jt})\|\|\phi(z_{ks})\|\|\phi(z_{\ell s})\||cov(\varepsilon_{it}\varepsilon_{jt}, \varepsilon_{ks}\varepsilon_{\ell s})|$$
$$\text{with } C_{NT} = (\min_{t\le T}\lambda_{\min}(\hat{Q}_t))^{-4}, \tag{F.12}$$

where the second equality follows by the independence in Assumption 4.3 (i) and the fact that both expectation and trace operators are linear, and the inequality follows since $\|\hat{\psi}_{it}\| \le (\lambda_{\min}(\hat{Q}_t))^{-1}\|\phi(z_{it})\|$. Moreover,

$$E\left[\sum_{t=1}^T\sum_{s=1}^T\sum_{i=1}^N\sum_{j=1}^N\sum_{k=1}^N\sum_{\ell=1}^N \|\phi(z_{it})\|\|\phi(z_{jt})\|\|\phi(z_{ks})\|\|\phi(z_{\ell s})\||cov(\varepsilon_{it}\varepsilon_{jt}, \varepsilon_{ks}\varepsilon_{\ell s})|\right]$$
$$\le \max_{i\le N, t\le T} E[\|\phi(z_{it})\|^4]\sum_{t=1}^T\sum_{s=1}^T\sum_{i=1}^N\sum_{j=1}^N\sum_{k=1}^N\sum_{\ell=1}^N |cov(\varepsilon_{it}\varepsilon_{jt}, \varepsilon_{ks}\varepsilon_{\ell s})|$$
$$\le J^2M^2 \max_{\ell\le JM, i\le N, t\le T} E[\phi^4(z_{it,m})]\sum_{t=1}^T\sum_{s=1}^T\sum_{i=1}^N\sum_{j=1}^N\sum_{k=1}^N\sum_{\ell=1}^N |cov(\varepsilon_{it}\varepsilon_{jt}, \varepsilon_{ks}\varepsilon_{\ell s})|, \tag{F.13}$$

where the first inequality is due to the Cauchy-Schwartz inequality, and the second inequality follows since $\max_{i\le N, t\le T} E[\|\phi(z_{it})\|^4] = J^2M^2 \max_{\ell\le JM, i\le N, t\le T} E[\phi^4(z_{it,m})]$. Combining (F.12) and (F.13) implies that $E_\varepsilon[\|\bar{E}\bar{E}'/NT - \mathcal{A}_{NT}\|_F^2] = O_p(J^2/N + J^2/T)$ by Assumptions 4.1(i), 4.5(i) and 6.1(ii). Thus, the result of the lemma follows by the Markov's inequality and Lemma A.5. ■

# Appendix G - Result for Fixed Sieve Size

In this section, we establish the result for fixed sieve size under $\alpha(\cdot) = 0$. Without $\alpha(\cdot)$, we only need one step to estimate both $B$ and $F$. Specifically, the columns of $\hat{B}$ are the eigenvectors corresponding to the first $K$ largest eigenvalues of the $JM \times JM$ matrix

$\tilde{Y}\tilde{Y}'/T$ and $\hat{F} = \tilde{Y}'\hat{B}$. We provide the consistency result for the estimators below. To this end, we replace Assumptions 4.2(i), (iii) and (iv) with the following assumptions.

**Assumption G.1** (Fixed sieve size)**.** *There are positive constants $h_{\min}$ and $h_{\max}$ such that: with probability approaching one (as $N \to \infty$),*

$$h_{\min} < \lambda_K(\Sigma_{NT}) \leq \lambda_{\max}(\Sigma_{NT}) < h_{\max},$$

*where $\Sigma_{NT} = \sum_{t=1}^{T}(\Phi(Z_t)'\Phi(Z_t))^{-1}\Phi(Z_t)'\Lambda(Z_t)f_tf_t'\Lambda(Z_t)'\Phi(Z_t)(\Phi(Z_t)'\Phi(Z_t))^{-1}/T$ with $\Lambda(Z_t) = (\beta(z_{1t}), \ldots, \beta(z_{Nt}))'$, and $\lambda_K(\Sigma_{NT})$ denotes the $K$th largest eigenvalue of $\Sigma_{NT}$.*

Assumption G.1 is a high-level assumption, which does not require $J \to \infty$. Given Assumptions 4.1(i) and 4.2(ii), it is weaker than Assumptions 4.2(i), (iii) and (iv) when $J \to \infty$. It is easy to verify that $\|\Sigma_{NT} - BF'FB'/T\|_F = o_p(1)$ when $J \to \infty$, under Assumptions 4.1(i), 4.2(i), (ii) and (iv). Thus, Assumption G.1 is satisfied under Assumptions 4.1(i) and 4.2. Assumption G.1 requires $T, JM \geq K$, which is reasonable since we assume $K$ to be fixed throughout the paper.

**Theorem G.1.** *Suppose $\alpha(\cdot) = 0$, $\max_{i \leq N, t \leq T} E[\|\beta(z_{it})\|^2] < \infty$, and Assumptions 4.1, 4.2(ii), 4.3 and G.1 hold. Let $\hat{B}$ and $\hat{f}_t$ be given in the beginning of Section G. Assume (i) $N \to \infty$; (ii) $T \to \infty$ or $T \geq K$ be finite; (iii) $J \to \infty$ with $J = o(N)$ or $J \geq K/M$ is fixed. Then*

$$\left\|\hat{B} - \frac{1}{T}\sum_{t=1}^{T}(\Phi(Z_t)'\Phi(Z_t))^{-1}\Phi(Z_t)'\Lambda(Z_t)H_t\right\|_F^2 = O_p\left(\frac{J}{N}\right),$$

$$\frac{1}{T}\sum_{t=1}^{T}\|\hat{f}_t - \hat{B}'(\Phi(Z_t)'\Phi(Z_t))^{-1}\Phi(Z_t)'\Lambda(Z_t)f_t\|^2 = O_p\left(\frac{J}{N}\right),$$

*where $H_t = f_tf_t'\Lambda(Z_t)'\Phi(Z_t)(\Phi(Z_t)'\Phi(Z_t))^{-1}\hat{B}(\hat{F}'\hat{F}/T)^{-1}$.*

The first result of Theorem G.1 implies that $\hat{B}$ consistently estimates a weighted average of the regressed coefficient matrices $(\Phi(Z_t)'\Phi(Z_t))^{-1}\Phi(Z_t)'\Lambda(Z_t)$ with the weighting matrices $H_t$. When $J \to \infty$, the result reduces to one similar to the one in Theorem 4.1. When $J$ is fixed, the regressed coefficient matrices can be highly varying over $t$ and $\hat{B}$ only estimates their weighted average. Let $T$ be finite. The second result implies that $\hat{f}_t$ consistently estimates $f_t$ up to the rotational transformation matrix $\hat{B}'(\Phi(Z_t)'\Phi(Z_t))^{-1}\Phi(Z_t)'\Lambda(Z_t)$. However, the rotational transformation matrix may be highly varying over $t$ when $J$ is fixed. It turns out that $\hat{F}$ may not be consistent to $F$ up to a rotational transformation. That is, the space spanned by the columns of $F$ may not be consistently estimated by the space spanned by the columns of $\hat{F}$ due to a large sieve approximation error. The result also implies that misspecification of $\beta(\cdot)$ may cause inconsistent estimation of $F$.

Proof of Theorem G.1: The proof is similar to the proof of Theorem 4.1. Let $\tilde{\Lambda}_t \equiv (\Phi(Z_t)'\Phi(Z_t))^{-1}\Phi(Z_t)'\Lambda_t$, where $\Lambda_t = \Lambda(Z_t)$. Since $\alpha(\cdot) = 0$, (9) can be written as

$$\tilde{Y}_t = \tilde{\Lambda}_t f_t + \tilde{\varepsilon}_t. \tag{G.1}$$

Let $V$ be a $K \times K$ diagonal matrix of the first $K$ largest eigenvalues of $\tilde{Y}\tilde{Y}'/T$. By the definition of $\hat{B}$, $(\tilde{Y}\tilde{Y}'/T)\hat{B} = \hat{B}V$. We have $V = \hat{F}'\hat{F}/T$, so $H_t = f_t f_t' \tilde{\Lambda}_t' \hat{B} V^{-1}$. We may substitute (G.1) to $(\tilde{Y}\tilde{Y}'/T)\hat{B} = \hat{B}V$ to obtain

$$\hat{B} - \frac{1}{T}\sum_{t=1}^{T} \tilde{\Lambda}_t H_t = \sum_{j=1}^{3} A_j \hat{B} V^{-1}, \tag{G.2}$$

where $A_1 = \tilde{E}\tilde{E}'/T$ and $A_2 = A_3' = \sum_{t=1}^{T} \tilde{\Lambda}_t f_t \tilde{\varepsilon}_t'/T$. By the Cauchy-Schwartz inequality and the fact that $\|C + D\|_F \leq \|C\|_F + \|D\|_F$ and $\|CD\|_F \leq \|C\|_2 \|D\|_F$, (G.2) implies

$$\left\| \hat{B} - \frac{1}{T}\sum_{t=1}^{T} \tilde{\Lambda}_t H_t \right\|_F^2 \leq 3\|\hat{B}\|_2^2 \|V^{-1}\|_2^2 \left( \sum_{j=1}^{3} \|A_j\|_F^2 \right) = O_p\left(\frac{J}{N}\right), \tag{G.3}$$

where the equality follows by Lemmas A.3(ii), G.1, G.2, and the fact that $A_1 = D_5$ and $\|A_3\|_F = \|A_2\|_F$. By the definition of $\hat{f}_t$, $\hat{f}_t = \hat{B}'\tilde{Y}_t$. We may substitute (G.1) to $\hat{f}_t = \hat{B}'\tilde{Y}_t$ to obtain

$$\hat{f}_t - \hat{B}'\tilde{\Lambda}_t f_t = \hat{B}'\tilde{\varepsilon}_t. \tag{G.4}$$

By the fact that $\|Ax\| \leq \|A\|_2 \|x\|$, (G.4) implies

$$\frac{1}{T}\sum_{t=1}^{T} \|\hat{f}_t - \hat{B}'\tilde{\Lambda}_t f_t\|^2 \leq \|\hat{B}\|_2^2 \left( \frac{1}{T}\sum_{t=1}^{T} \|\tilde{\varepsilon}_t\|^2 \right) = \frac{1}{T}\|\hat{B}\|_2^2 \|\tilde{E}\|_F^2 = O_p\left(\frac{J}{N}\right), \tag{G.5}$$

where the last equality follows from Lemma A.3(ii) and $\hat{B}'\hat{B} = I_K$. This completes the proof of the theorem. ■

**Lemma G.1.** *Suppose $\alpha(\cdot) = 0$ and $\max_{i \leq N, t \leq T} E[\|\beta(z_{it})\|^2] < \infty$. Let $A_2$ be given in the proof of Theorem G.1. Under Assumptions 4.1, 4.2(ii) and 4.3, $\|A_2\|_F^2 = O_p(J/N)$.*

Proof: The proof is similar to the proof of Lemma A.1(iii). Let $\tilde{C} \equiv (\tilde{\Lambda}_1 f_1, \ldots, \tilde{\Lambda}_T f_T)$, then $A_2 = \tilde{C}\tilde{E}'/T$. Since $\|A_2\|_F \leq \|\tilde{C}\|_F \|\tilde{E}\|_F / T$, it suffices to show $\|\tilde{C}\|_F = O_p(\sqrt{T})$ by Lemma A.3(ii). We next show this result. By the fact that $\|Ax\| \leq \|A\|_2 \|x\|$, $\|CD\|_2 \leq \|C\|_2 \|D\|_2$ and $\|A\|_2 \leq \|A\|_F$,

$$\|\tilde{C}\|_F^2 = \sum_{t=1}^{T} \|\tilde{\Lambda}_t f_t\|^2 \leq \max_{t \leq T} \|f_t\|^2 \left( \min_{t \leq T} \lambda_{\min}(\hat{Q}_t) \right)^{-1} \frac{1}{N} \sum_{t=1}^{T} \|\Lambda(Z_t)\|_F^2 = O_p(T), \tag{G.6}$$

where the last equality follows from Assumptions 4.1(i) and 4.2(ii) by noting that $\sum_{t=1}^{T} \|\Lambda(Z_t)\|_F^2 = O_p(NT)$. Here, $\sum_{t=1}^{T} \|\Lambda(Z_t)\|_F^2 = O_p(NT)$ holds by the Markov's inequality, since $E[\sum_{t=1}^{T} \|\Lambda(Z_t)\|_F^2] = \sum_{i=1}^{N}\sum_{t=1}^{T} E[\|\beta(z_{it})\|^2] = O(NT)$ which is due to $\max_{i\leq N, t\leq T} E[\|\beta(z_{it})\|^2] < \infty$. This completes the proof of the lemma. ■

**Lemma G.2.** *Suppose $\alpha(\cdot) = 0$, $\max_{i\leq N,t\leq T} E[\|\beta(z_{it})\|^2] < \infty$, and Assumptions 4.1, 4.2(ii), 4.3 and G.1 hold. Let $V$ be given in the proof of Theorem G.1. Assume (i) $N \to \infty$; (ii) $T \to \infty$ or $T \geq K$ is finite; (iii) $J \to \infty$ with $J = o(N)$ or $J \geq K/M$ is fixed. Then $\|V\|_2 = O_p(1)$ and $\|V^{-1}\|_2 = O_p(1)$.*

Proof: The proof of is similar to the proof of Lemma A.2(i). It is noted that $\tilde{Y}\tilde{Y}'/T = \Sigma_{NT} + \sum_{j=1}^{3} A_j$ by (G.1), where $A_1, A_2, A_3$ are given below (G.2). By the fact that $\|C + D\|_F \leq \|C\|_F + \|D\|_F$,

$$\|\tilde{Y}\tilde{Y}'/T - \Sigma_{NT}\|_F \leq \sum_{j=1}^{3} \|A_j\|_F = O_p\left(\frac{\sqrt{J}}{\sqrt{N}}\right), \tag{G.7}$$

where the equality follows by Lemmas A.3(ii) and G.1, and the fact that $A_1 = D_5$ and $\|A_3\|_F = \|A_2\|_F$. Let $\vec{V}$ be a $K \times K$ diagonal matrix of the first $K$ largest eigenvalues of $\Sigma_{NT}$. By the Weyl's inequality and the fact that $\|A\|_2 \leq \|A\|_F$,

$$\|V - \vec{V}\|_2 = \|\tilde{Y}\tilde{Y}'/T - \Sigma_{NT}\|_2 = O_p\left(\frac{\sqrt{J}}{\sqrt{N}}\right). \tag{G.8}$$

Thus, $\|V\|_2 = O_p(1)$ and $\|V^{-1}\|_2 = \lambda_{\min}^{-1}(V) = O_p(1)$ follows from (G.8) and Assumption G.1. This completes the proof of the lemma. ■

# Appendix H - Improved Rate for A Special Case

In this section, we establish the improved rate for the special case when $\hat{Q}_t = I_{JM}$ for all $t \leq T$ without Assumption 4.5 and the requirement $J^2\xi_J^2 \log J = o(N)$.

**Theorem H.1.** *Suppose $\hat{Q}_t = I_{JM}$ for all $t \leq T$, $\max_{i\leq N,t\leq T} E[|\alpha(z_{it})|^2] < \infty$, $\max_{i\leq N,t\leq T} E[\|\beta(z_{it})\|^2] < \infty$, and Assumptions 4.1-4.4 hold. Let $\hat{a}$, $\hat{B}, \hat{F}$, $\hat{\alpha}(\cdot)$ and $\hat{\beta}(\cdot)$ be given in (10). Assume (i) $N \to \infty$; (ii) $T \to \infty$ or $T \geq K+1$ is finite; (iii) $J \to \infty$ with $J = o(\sqrt{N})$. Then*

$$\|\hat{a} - a\|^2 = O_p\left(\frac{1}{J^{2\kappa}} + \frac{J}{N^2} + \frac{J}{NT}\right),$$

$$\|\hat{B} - BH\|_F^2 = O_p\left(\frac{1}{J^{2\kappa}} + \frac{J}{N^2} + \frac{J}{NT}\right),$$

$$\frac{1}{T}\|\hat{F} - F(H')^{-1}\|_F^2 = O_p\left(\frac{1}{J^{2\kappa}} + \frac{1}{N}\right),$$

$$\sup_z |\hat{\alpha}(z) - \alpha(z)|^2 = O_p\left(\frac{1}{J^{2\kappa-1}} + \frac{J^2}{N^2} + \frac{J^2}{NT}\right) \max_{j\leq J}\sup_z |\phi_j(z)|^2,$$
$$\sup_z \|\hat{\beta}(z) - H'\beta(z)\|^2 = O_p\left(\frac{1}{J^{2\kappa-1}} + \frac{J^2}{N^2} + \frac{J^2}{NT}\right) \max_{j\leq J}\sup_z |\phi_j(z)|^2.$$

When $Z_t$ is not changing over $t$, $\max_{i\leq N, t\leq T} E[\|\beta(z_{it})\|^2] < \infty$ reduces to Assumption 4.2(ii) of Fan et al. (2016a). Theorem H.1 implies that the rate of $\hat{B}$, $\hat{F}$ and $\hat{\beta}(\cdot)$ is equal to the rate in Fan et al. (2016a), when $\hat{Q}_t = I_{JM}$ for all $t \leq T$. When $Z_t$ is not changing over $t$, we may orthonormalize the basis functions such that $\hat{Q}_t = I_{JM}$ for all $t \leq T$. But the condition fails to hold for general $Z_t$. See Theorem 4.2 for the general case.

Proof of Theorem H.1: Let us first look at (A.3). To improve the rate of $\hat{B}$ in Theorem 4.1, we cannot use the inequality in (A.4). Instead, we need to treat $D_5\hat{B}$ as a whole to establish its rate. By the Cauchy-Schwartz inequality and the fact that $\|C + D\|_F \leq \|C\|_F + \|D\|_F$ and $\|CD\|_F \leq \|C\|_2\|D\|_F$, (A.3) implies

$$\begin{aligned}\|\hat{B} - BH\|_F^2 &\leq 10\|\hat{B}\|_2^2\|V^{-1}\|_2^2\left(\sum_{j\neq 5}^{6}\|D_j\|_F^2\right) + 2\|V^{-1}\|_2^2\|D_5\hat{B}\|_F^2,\\ &= O_p\left(\frac{1}{J^{2\kappa}} + \frac{J}{N^2} + \frac{J}{NT}\right),\end{aligned} \tag{H.1}$$

where the equality follows by $J = o(\sqrt{N})$, Lemmas A.1(i)-(iv), A.2(i) and H.1(ii) and the fact that $\|D_6\|_F = \|D_3\|_F$. Given the rate of $\|\hat{B} - BH\|_F^2$ in (H.1), the rate of $|\hat{a} - a|^2$ immediately follows from the same argument in (A.6). We now look at (A.7). To improve the rate of $\hat{F}$ in Theorem 4.1, we cannot use the inequality in (A.8). Instead, we need to plug in the expansion of $\hat{B} - BH$, and treat $a'D_4$, $D_4'\hat{B}$, $D_5\hat{B}$ and $\tilde{E}'\hat{B}$ as a whole to establish their rate. By the fact that $\|C + D\|_F \leq \|C\|_F + \|D\|_F$ and $\|CD\|_F \leq \|C\|_2\|D\|_F$, combining (A.3) and (A.7) implies

$$\begin{aligned}\|\hat{F} - F(H')^{-1}\|_F = &\left(\sum_{j\neq 4,5}^{6}\|D_j\|_F\|\hat{B}\|_2\|a\| + \|a'D_4\|\|\hat{B}\|_2 + \|a\|\|D_5\hat{B}\|_F\right)\|V^{-1}\|_2\|1_T\|\\ &+ \left(\sum_{j\neq 4,5}^{6}\|D_j\|_F\|\hat{B}\|_2 + \|D_4'\hat{B}\|_F + \|D_5\hat{B}\|_F\right)\\ &\times \|F\|_2\|H^{-1}\|_2\|V^{-1}\|_2\|\hat{B}\|_2 + \|\tilde{\Delta}\|_F\|\hat{B}\|_2 + \|\tilde{E}'\hat{B}\|_F\\ = &O_p\left(\frac{\sqrt{T}}{J^\kappa} + \sqrt{\frac{T}{N}}\right),\end{aligned} \tag{H.2}$$

where the equality follows by $J = o(\sqrt{N})$, Assumption 4.2(ii) and 4.4, Lemmas A.1(i)-(iii), A.2, A.3(i), H.1 and H.2(i) and the fact that $\|D_6\|_F = \|D_3\|_F$. Thus, the third result follows from (H.2). The proofs of the last two results of the theorem are similar to the proofs of the last two results of Theorem 4.1. ■

**Lemma H.1.** *Suppose* $\hat{Q}_t = I_{JM}$ *for all* $t \leq T$, $\max_{i\leq N,t\leq T} E[|\alpha(z_{it})|^2] < \infty$, *and* $\max_{i\leq N,t\leq T} E[\|\beta(z_{it})\|^2] < \infty$. *Let* $D_4$ *and* $D_5$ *be given in the proof of Theorem 4.1. Assume (i)* $N \to \infty$; *(ii)* $T \to \infty$ *or* $T \geq K+1$ *is finite; (iii)* $J \to \infty$ *with* $J = o(\sqrt{N})$.
*(i) Under Assumptions 4.1-4.4,* $\|D_4'\hat{B}\|_F^2 = O_p(1/NT)$.
*(ii) Under Assumptions 4.1-4.4,* $\|D_5\hat{B}\|_F = O_p(J/N^2)$.
*(iii) Under Assumptions 4.1-4.4,* $\|D_4'a\|^2 = O_p(1/NT)$.

Proof: (i) Since $\|D_4'\hat{B}\|_F \leq \|B\|_2\|\hat{B}'\tilde{E}M_TF\|_F/T$, the result then immediately follows from Assumption 4.2(i) and Lemma H.2(ii).

(ii) Since $\|M_T\|_2 = 1$, $\|D_5\hat{B}\|_F \leq \|\tilde{E}\|_F\|\hat{B}'\tilde{E}\|_F/T$. The result then immediately follows from Lemmas A.3(ii) and H.2(i).

(iii) Since $\|D_4'a\| \leq \|B\|_2\|a'\tilde{E}M_TF\|/T$, the result then immediately follows from Assumption 4.2(i) and Lemma H.2(iii). ■

**Lemma H.2.** *Suppose* $\hat{Q}_t = I_{JM}$ *for all* $t \leq T$, $\max_{i\leq N,t\leq T} E[|\alpha(z_{it})|^2] < \infty$, *and* $\max_{i\leq N,t\leq T} E[\|\beta(z_{it})\|^2] < \infty$. *Let* $\tilde{E}$ *be given in the proof of Theorem 4.1. Assume (i)* $N \to \infty$; *(ii)* $T \to \infty$ *or* $T \geq K+1$ *is finite; (iii)* $J \to \infty$ *with* $J = o(\sqrt{N})$.
*(i) Under Assumptions 4.1-4.4,* $\|\hat{B}'\tilde{E}\|_F^2/T = O_p(1/N)$.
*(ii) Under Assumptions 4.1-4.4,* $\|\hat{B}'\tilde{E}M_TF\|_F^2/T = O_p(1/N)$.
*(iii) Under Assumptions 4.1-4.4,* $\|a'\tilde{E}M_TF\|^2/T = O_p(1/N)$.

Proof: By the fact that $\|C+D\|_F \leq \|C\|_F + \|D\|_F$ and $\|CD\|_F \leq \|C\|_2\|D\|_F$,

$$\begin{aligned}
\frac{1}{T}\|\hat{B}'\tilde{E}\|_F^2 &\leq \frac{2}{T}\|\tilde{E}\|_F^2\|\hat{B}-BH\|_F^2 + \frac{2}{T}\|H\|_2^2\|B'\tilde{E}\|_F^2 \\
&= \frac{2}{T}\|\tilde{E}\|_F^2\|\hat{B}-BH\|_F^2 + \frac{2}{N^2T}\|H\|_2^2\left(\sum_{t=1}^{T}\|B'\Phi(Z_t)'\varepsilon_t\|^2\right) \\
&= O_p\left(\frac{J}{N}\left(\frac{1}{J^{2\kappa}} + \frac{J^2}{N^2} + \frac{J}{NT}\right) + \frac{1}{N}\right) = O_p\left(\frac{1}{N}\right),
\end{aligned} \tag{H.3}$$

where the first equality follows from $\hat{Q}_t = I_{JM}$ for all $t \leq T$, the second equality follows from Lemmas A.2(i), A.3(ii) and H.3(i) and Theorem 4.1, and the last equality is due to $\kappa > 1/2$ and $J = o(\sqrt{N})$.

(ii) By the fact that $\|C+D\|_F \leq \|C\|_F + \|D\|_F$ and $\|CD\|_F \leq \|C\|_2\|D\|_F$,

$$\begin{aligned}
\frac{1}{T}\|\hat{B}'\tilde{E}M_TF\|_F^2 &\leq \frac{2}{T}\|\tilde{E}M_TF\|_F^2\|\hat{B}-BH\|_F^2 + \frac{2}{T}\|H\|_2^2\|B'\tilde{E}M_TF\|_F^2 \\
&\leq \frac{2}{T}\|\tilde{E}M_TF\|_F^2\|\hat{B}-BH\|_F^2 + \frac{4}{N^2T}\|H\|_2^2\left\|\sum_{t=1}^{T}B'\Phi(Z_t)'\varepsilon_tf_t'\right\|_F^2
\end{aligned}$$

$$
\begin{aligned}
&+\frac{4\|\bar{f}\|^2}{N^2T}\|H\|_2^2\left\|\sum_{t=1}^{T}B'\Phi(Z_t)'\varepsilon_t\right\|^2 \\
&= O_p\left(\frac{J}{N}\left(\frac{1}{J^{2\kappa}}+\frac{J^2}{N^2}+\frac{J}{NT}\right)+\frac{1}{N}\right)=O_p\left(\frac{1}{N}\right), \qquad \text{(H.4)}
\end{aligned}
$$

where the second inequality follows from $\hat{Q}_t = I_{JM}$ for all $t \leq T$, the first equality follows from Assumption 4.2(ii), Lemmas A.2(i), A.3(iii) and H.3(ii) and Theorem 4.1, and the last equality is due to $\kappa > 1/2$ and $J = o(\sqrt{N})$.

(iii) By the fact that $\|x+y\| \leq \|x\| + \|y\|$,

$$
\begin{aligned}
\frac{1}{T}\|a'\tilde{E}M_TF\|^2 &\leq \frac{2}{N^2T}\left\|\sum_{t=1}^{T}a'\Phi(Z_t)'\varepsilon_t f_t'\right\|^2+\frac{2\|\bar{f}\|^2}{N^2T}\left|\sum_{t=1}^{T}a'\Phi(Z_t)'\varepsilon_t\right|^2 \\
&= O_p\left(\frac{1}{N}\right), \qquad \text{(H.5)}
\end{aligned}
$$

where the inequality follows from $\hat{Q}_t = I_{JM}$ for all $t \leq T$, and the equality follows from Assumption 4.2(ii) and Lemma H.3(ii). ■

**Lemma H.3.** *Suppose* $\max_{i\leq N,t\leq T} E[|\alpha(z_{it})|^2] < \infty$ *and* $\max_{i\leq N,t\leq T} E[\|\beta(z_{it})\|^2] < \infty$. *(i) Under Assumptions 4.2(iv) and 4.3,*

$$
\sum_{t=1}^{T}\|B'\Phi(Z_t)'\varepsilon_t\|^2 = O_p(NT(1+J^{-2\kappa})).
$$

*(ii) Under Assumptions 4.2(ii), (iv) and 4.3,*

$$
\begin{aligned}
\left\|\sum_{t=1}^{T}B'\Phi(Z_t)'\varepsilon_t f_t'\right\|_F^2 &= O_p(NT(1+J^{-2\kappa})), \\
\left\|\sum_{t=1}^{T}B'\Phi(Z_t)'\varepsilon_t\right\|^2 &= O_p(NT(1+J^{-2\kappa})), \\
\left\|\sum_{t=1}^{T}a'\Phi(Z_t)'\varepsilon_t f_t'\right\|^2 &= O_p(NT(1+J^{-2\kappa})), \\
\left|\sum_{t=1}^{T}a'\Phi(Z_t)'\varepsilon_t\right|^2 &= O_p(NT(1+J^{-2\kappa})).
\end{aligned}
$$

PROOF: (i) Let $\Lambda(Z_t) \equiv (\beta(z_{1t}), \ldots, \beta(z_{Nt}))'$, then $\Phi(Z_t)B = \Lambda(Z_t) - \Delta(Z_t)$. By the fact that $\|x+y\| \leq \|x\| + \|y\|$,

$$
\sum_{t=1}^{T}\|B'\Phi(Z_t)'\varepsilon_t\|^2 \leq 2\sum_{t=1}^{T}\|\Lambda(Z_t)'\varepsilon_t\|^2 + 2\sum_{t=1}^{T}\|\Delta(Z_t)'\varepsilon_t\|^2 \equiv 2\mathcal{T}_1 + 2\mathcal{T}_2. \qquad \text{(H.6)}
$$

Therefore, it suffices to show that $\mathcal{T}_1 = O_p(NT)$ and $\mathcal{T}_2 = O_p(NTJ^{-2\kappa})$. The former holds by the Markov's inequality, since

$$\begin{aligned}
E[\mathcal{T}_1] &= E\left[\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\beta(z_{it})'\beta(z_{jt})\varepsilon_{it}\varepsilon_{jt}\right] \\
&= \sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}E[\beta(z_{it})'\beta(z_{jt})]E[\varepsilon_{it}\varepsilon_{jt}] \\
&\leq T\max_{i\leq N,t\leq T}E[\|\beta(z_{it})\|^2]\max_{t\leq T}\sum_{i=1}^{N}\sum_{j=1}^{N}|E[\varepsilon_{it}\varepsilon_{jt}]| = O(NT), \qquad \text{(H.7)}
\end{aligned}$$

where the second equality follows by the independence in Assumption 4.3(i), the inequality is due to the Cauchy-Schwartz inequality, and the last equality follows from $\max_{i\leq N,t\leq T}E[\|\beta(z_{it})\|^2] < \infty$ and Assumption 4.3(iii). The latter also holds by the Markov's inequality, since

$$\begin{aligned}
E[\mathcal{T}_2] &= E\left[\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\delta(z_{it})'\delta(z_{jt})\varepsilon_{it}\varepsilon_{jt}\right] \\
&= \sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}E[\delta(z_{it})'\delta(z_{jt})]E[\varepsilon_{it}\varepsilon_{jt}] \\
&\leq TKM^2\max_{k\leq K,m\leq M}\sup_z|\delta_{km,J}(z)|^2\sum_{i=1}^{N}\sum_{j=1}^{N}|E[\varepsilon_{it}\varepsilon_{jt}]| = O(NTJ^{-2\kappa}), \qquad \text{(H.8)}
\end{aligned}$$

where the second equality follows by the independence in Assumption 4.3(i), the inequality follows by the Cauchy-Schwartz inequality and the fact that $\max_{i\leq N,t\leq T}E[\|\delta(z_{it})\|^2] \leq KM^2\max_{k\leq K,m\leq M}\sup_z|\delta_{km,J}(z)|^2$, and the last equality follows from Assumptions 4.2(iv) and 4.3(iii). This completes the proof of (i).

(ii) Let $\Lambda(Z_t) \equiv (\beta(z_{1t}),\ldots,\beta(z_{Nt}))'$, then $\Phi(Z_t)B = \Lambda(Z_t) - \Delta(Z_t)$. By the fact that $\|C+D\|_F \leq \|C\|_F + \|D\|_F$,

$$\left\|\sum_{t=1}^{T}B'\Phi(Z_t)'\varepsilon_t f_t'\right\|_F^2 \leq 2\left\|\sum_{t=1}^{T}\Lambda(Z_t)'\varepsilon_t f_t'\right\|_F^2 + 2\left\|\sum_{t=1}^{T}\Delta(Z_t)'\varepsilon_t f_t'\right\|_F^2 \equiv 2\mathcal{T}_1 + 2\mathcal{T}_2. \qquad \text{(H.9)}$$

Therefore, it suffices to show that $\mathcal{T}_1 = O_p(NT)$ and $\mathcal{T}_2 = O_p(NTJ^{-2\kappa})$. Note that $\|A\|_F^2 = \mathrm{tr}(AA')$. The former holds by the Markov's inequality, since

$$E[\mathcal{T}_1] = E\left[\mathrm{tr}\left(\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\beta(z_{it})\varepsilon_{it}f_t'f_s\varepsilon_{js}\lambda(z_{js})'\right)\right]$$

$$
\begin{aligned}
&= \sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N} E[\beta(z_{it})'\beta(z_{js})] f_t' f_s E[\varepsilon_{it}\varepsilon_{js}] \\
&\leq \max_{t\leq T} \|f_t\|^2 \max_{i\leq N, t\leq T} E[\|\beta(z_{it})\|^2] \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} |E[\varepsilon_{it}\varepsilon_{js}]| = O(NT), \qquad \text{(H.10)}
\end{aligned}
$$

where the second equality follows by the independence in Assumption 4.3(i) and the fact that both expectation and trace operators are linear, the inequality is due to the Cauchy-Schwartz inequality, and the last equality follows from $\max_{i\leq N,t\leq T} E[\|\beta(z_{it})\|^2] < \infty$ and Assumptions 4.2(ii) and 4.3(iii). The latter also holds by the Markov's inequality, since

$$
\begin{aligned}
E[\mathcal{T}_2] &= E\left[\operatorname{tr}\left(\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N} \delta(z_{it})\varepsilon_{it} f_t' f_s \varepsilon_{js}\delta(z_{js})'\right)\right] \\
&= \sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N} E[\delta(z_{it})'\delta(z_{js})] f_t' f_s E[\varepsilon_{it}\varepsilon_{js}] \\
&\leq C_{NT} \max_{k\leq K, m\leq M} \sup_z |\delta_{km,J}(z)|^2 \sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N} |E[\varepsilon_{it}\varepsilon_{js}]| = O(NTJ^{-2\kappa}), \qquad \text{(H.11)}
\end{aligned}
$$

where $C_{NT} = KM^2 \max_{t\leq T} \|f_t\|^2$, the second equality follows by the independence in Assumption 4.3(i) and the fact that both expectation and trace operators are linear, the inequality follows by the Cauchy-Schwartz inequality and the fact that $\max_{i\leq N,t\leq T} E[\|\delta(z_{it})\|^2] \leq KM^2 \max_{k\leq K,m\leq M} \sup_z |\delta_{km,J}(z)|^2$, and the last equality follows from Assumptions 4.2(ii), (iv) and 4.3(iii). This completes the proof the first result of (ii), and the proofs of the other three are similar. ■

## Appendix I - Justification for Some Assumptions

We provide preliminary conditions for Assumptions 4.1(i) and 6.1(ii) in the following two propositions, justifying that the two assumptions are not restrictive.

**Proposition I.1** (Justification of Assumption 4.1(i))**.** *Suppose Assumptions 4.5(ii) and (iii) hold. Assume $J \geq 2$ and $\sqrt{T}\xi_J^2 \log J = o(N)$, where $\xi_J$ is given above Theorem 4.2. Then Assumption 4.1(i) holds.*

Proof: Let $Q_t \equiv E[\hat{Q}_t]$. By Lemma B.6 and the condition that $\sqrt{T}\xi_J^2 \log J = o(N)$,

$$
\max_{t\leq T} \|\hat{Q}_t - Q_t\|_2 \leq \left(\sum_{t=1}^{T} \|\hat{Q}_t - Q_t\|_2^4\right)^{1/4} = O_p\left(\frac{T^{1/4}\xi_J \log^{1/2} J}{\sqrt{N}}\right) = o_p(1). \qquad \text{(I.1)}
$$

By (I.1) and the Weyl's inequality,

$$\left|\min_{t\leq T}\lambda_{\min}(\hat{Q}_t)-\min_{t\leq T}\lambda_{\min}(Q_t)\right|\leq\max_{t\leq T}\|\hat{Q}_t-Q_t\|_2=o_p(1) \tag{I.2}$$

and

$$\left|\max_{t\leq T}\lambda_{\max}(\hat{Q}_t)-\max_{t\leq T}\lambda_{\max}(Q_t)\right|\leq\max_{t\leq T}\|\hat{Q}_t-Q_t\|_2=o_p(1). \tag{I.3}$$

The result of the lemma thus follows from (I.2) and (I.3) and Assumption 4.5(ii) by noting that $\min_{t\leq T}\lambda_{\min}(Q_t)\geq\min_{i\leq N,t\leq T}\lambda_{\min}(Q_{it})$ and $\max_{t\leq T}\lambda_{\max}(Q_t)\leq\max_{i\leq N,t\leq T}\lambda_{\max}(Q_{it})$. ■

**Proposition I.2** (Justification of Assumption 6.1(ii))**.** *Suppose Assumptions 4.3(ii) and 4.6(ii) hold. Assume* $\max_{i\leq N,t\leq T}E[\varepsilon_{it}^4]<\infty$ *and there is* $0<C_5<\infty$ *such that*

$$\max_{i\leq N}\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}|E[\varepsilon_{it}\varepsilon_{is}]|^2<C_5.$$

*Then Assumption 6.1(ii) holds.*

Proof: By the independence condition and Assumption 4.3(ii), $E[\varepsilon_{it}\varepsilon_{jt}]=0$ for $i\neq j$. Thus, we may have the following decomposition

$$\begin{aligned}
&\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{\ell=1}^{N}|cov(\varepsilon_{it}\varepsilon_{jt},\varepsilon_{ks}\varepsilon_{\ell s})|\\
&=\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{k=1}^{N}|cov(\varepsilon_{it}^2,\varepsilon_{ks}^2)|+\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{k=1}^{N}\sum_{\ell\neq k}^{N}|E[\varepsilon_{it}\varepsilon_{jt}\varepsilon_{ks}\varepsilon_{\ell s}]|\\
&\quad+2\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{k=1}^{N}\sum_{\ell\neq k}^{N}|E[(\varepsilon_{it}^2-E[\varepsilon_{it}^2])\varepsilon_{ks}\varepsilon_{\ell s}]|\\
&\equiv\mathcal{T}_1+\mathcal{T}_2+\mathcal{T}_3.
\end{aligned} \tag{I.4}$$

We next establish bound for $\mathcal{T}_1,\mathcal{T}_2$ and $\mathcal{T}_3$. By the independence condition,

$$\mathcal{T}_1=\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}var(\varepsilon_{it}^2)\leq\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}E[\varepsilon_{it}^4]\leq NT^2\max_{i\leq N,t\leq T}E[\varepsilon_{it}^4], \tag{I.5}$$

where the first inequality follows from $var(\varepsilon_{it}^2)\leq E[\varepsilon_{it}^4]$. By the independence condition and Assumption 4.3(ii), $E[\varepsilon_{it}\varepsilon_{jt}\varepsilon_{ks}\varepsilon_{\ell s}]=0$ unless $i=k$ and $j=\ell$ or $i=\ell$ and $j=k$ given $i\neq j$. It then follows that

$$\mathcal{T}_2=2\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{j\neq i}^{N}|E[\varepsilon_{it}\varepsilon_{is}\varepsilon_{jt}\varepsilon_{js}]|=2\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{j\neq 1}^{N}|E[\varepsilon_{it}\varepsilon_{is}]||E[\varepsilon_{jt}\varepsilon_{js}]|$$

$$\leq 2\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}|E[\varepsilon_{it}\varepsilon_{is}]||E[\varepsilon_{jt}\varepsilon_{js}]| = 2\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\sum_{i=1}^{N}|E[\varepsilon_{it}\varepsilon_{is}]|\right)^2$$

$$\leq 2N\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}|E[\varepsilon_{it}\varepsilon_{is}]|^2 \leq 2N^2\max_{i\leq N}\sum_{t=1}^{T}\sum_{s=1}^{T}|E[\varepsilon_{it}\varepsilon_{is}]|^2, \tag{I.6}$$

where the second equality follows by the independence condition, the first inequality follows since $|E[\varepsilon_{it}\varepsilon_{is}]|^2 \geq 0$, the second inequality is due to the Cauchy-Schwartz inequality. Again by the independence condition and Assumption 4.3(ii), $E[(\varepsilon_{it}^2 - E[\varepsilon_{it}^2])\varepsilon_{ks}\varepsilon_{\ell s}] = 0$ for $k \neq \ell$, so $\mathcal{T}_3 = 0$. This together with (I.4)-(I.6) and the assumptions thus concludes the result of the proposition. ■

## Appendix J - Additional Results for Section 8

# References


Ahn, S. and Horenstein, A. (2013). Eigenvalue ratio test for the number of factors. *Econometrica*, **81** 1203–1227.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, **71** 135–171.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, **70** 191–221.

Belloni, A., Chernozhukov, V., Chetverikov, D. and Kato, K. (2015). Some new asymptotic theory for least squares series: Pointwise and uniform results. *Journal of Econometrics*, **186** 345–366.

Bhatia, R. (1997). *Matrix Analysis.* Graduate Texts in Mathematics, Springer.

Chamberlain, G. and Rothschild, M. (1982). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica*, **51** 1281–1304.

Chen, L., Pelger, M. and Zhu, J. (2020). Deep learning in asset pricing. Tech. rep., Available at SSRN 3350138.

Chen, Q. and Fang, Z. (2019). Improved inference on the rank of a matrix. *Quantitative Economics*, **10** 1787–1824.

Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. *Handbook of Econometrics*, **6** 5549–5632.

Cochrane, J. H. (2011). Presidential address: Discount rates. *The Journal of finance*, **66** 1047–1108.

Table 12: Results under linear specifications of $\alpha(\cdot)$ and $\beta(\cdot)$ with 18 characteristics[†]

| Unrestricted ($\alpha(\cdot) \neq 0$) | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $K$ | $R_K^2$ | $R^2$ | $R_{T,N}^2$ | $R_{N,T}^2$ | $R_f^2$ | $R_{f,T,N}^2$ | $R_{f,N,T}^2$ | $R_O^2$ | $R_{T,N,O}^2$ | $R_{N,T,O}^2$ | Mean | Std | SR |
| 1 | 44.90 | 3.23 | 1.85 | 1.30 | 2.78 | 1.24 | 1.03 | 0.28 | 0.26 | 0.21 | 1.81 | 0.46 | 3.93 |
| 2* | 62.38 | 14.66 | 11.21 | 13.14 | 14.30 | 10.93 | 12.62 | 0.28 | 0.26 | 0.21 | 1.74 | 0.44 | 4.00 |
| 3 | 68.14 | 15.46 | 12.15 | 13.74 | 15.10 | 11.79 | 13.23 | 0.28 | 0.26 | 0.21 | 1.67 | 0.42 | 3.97 |
| 4 | 72.63 | 16.36 | 13.46 | 14.50 | 16.00 | 12.94 | 13.95 | 0.28 | 0.26 | 0.21 | 1.52 | 0.38 | 3.99 |
| 5 | 76.85 | 17.03 | 14.11 | 15.17 | 16.67 | 13.72 | 14.64 | 0.28 | 0.26 | 0.21 | 1.49 | 0.37 | 3.99 |
| 6 | 80.01 | 17.53 | 14.75 | 15.62 | 17.21 | 14.35 | 15.18 | 0.28 | 0.26 | 0.21 | 1.28 | 0.32 | 3.95 |
| 7 | 82.65 | 17.78 | 14.94 | 15.83 | 17.46 | 14.59 | 15.40 | 0.28 | 0.26 | 0.21 | 1.22 | 0.31 | 3.99 |
| 8 | 85.21 | 17.91 | 15.07 | 15.97 | 17.63 | 14.81 | 15.60 | 0.28 | 0.26 | 0.21 | 1.08 | 0.29 | 3.67 |
| 9 | 88.27 | 18.20 | 15.36 | 16.31 | 17.96 | 15.17 | 15.99 | 0.28 | 0.26 | 0.21 | 0.93 | 0.28 | 3.33 |
| 10 | 90.54 | 18.35 | 15.58 | 16.47 | 18.17 | 15.34 | 16.22 | 0.28 | 0.26 | 0.21 | 0.76 | 0.23 | 3.36 |
| Restricted ($\alpha(\cdot) = 0$) | | | | | | | | | | | | | |
| $K$ | $R_K^2$ | $R^2$ | $R_{T,N}^2$ | $R_{N,T}^2$ | $R_f^2$ | $R_{f,T,N}^2$ | $R_{f,N,T}^2$ | $R_O^2$ | $R_{T,N,O}^2$ | $R_{N,T,O}^2$ | Mean | Std | SR |
| 1 | 45.00 | NA | NA | NA | 2.88 | 1.21 | 1.04 | 0.19 | 0.21 | 0.02 | NA | NA | NA |
| 2* | 62.52 | NA | NA | NA | 14.27 | 10.84 | 12.58 | 0.34 | 0.49 | -0.21 | NA | NA | NA |
| 3 | 69.18 | NA | NA | NA | 14.82 | 11.44 | 13.25 | 0.50 | 0.70 | 0.09 | NA | NA | NA |
| 4 | 74.34 | NA | NA | NA | 15.75 | 12.46 | 13.95 | 0.49 | 0.73 | 0.15 | NA | NA | NA |
| 5 | 78.48 | NA | NA | NA | 16.72 | 13.96 | 14.90 | 0.50 | 0.64 | 0.17 | NA | NA | NA |
| 6 | 81.87 | NA | NA | NA | 17.29 | 14.48 | 15.42 | 0.51 | 0.61 | 0.17 | NA | NA | NA |
| 7 | 84.71 | NA | NA | NA | 17.78 | 15.09 | 15.90 | 0.50 | 0.60 | 0.18 | NA | NA | NA |
| 8 | 87.20 | NA | NA | NA | 17.99 | 15.20 | 16.08 | 0.50 | 0.63 | 0.22 | NA | NA | NA |
| 9 | 89.48 | NA | NA | NA | 18.10 | 15.31 | 16.19 | 0.50 | 0.61 | 0.22 | NA | NA | NA |
| 10 | 91.35 | NA | NA | NA | 18.30 | 15.51 | 16.40 | 0.51 | 0.62 | 0.22 | NA | NA | NA |

[†] $K$: the number of factor specified (* denotes the estimated one by our methods); Fama-MacBeth cross sectional regression $R^2$: $R^2_{\bar{Y}} = 19.68\%$; $R^2_K$ meassures the variations of managed portfolios captured by different numbers of factors from PCA; $R^2$, $R^2_{T,N}$, $R^2_{N,T}$: various in-sample $R^2$'s (%), see (25)-(27); $R^2_f$, $R^2_{f,T,N}$, $R^2_{f,N,T}$: various in-sample $R^2$'s (%) without $\alpha$, see (28)-(30); $R^2_O$, $R^2_{T,N,O}$, $R^2_{N,T,O}$: various out-sample predictive $R^2$'s (%), see (31)-(33); Mean: out-of-sample annualized means of the pure-alpha arbitrage strategy(%); Std: out-of-sample annualized standard deviations of the pure-alpha arbitrage strategy(%); SR: out-of-sample annualized Sharpe ratios of the pure-alpha arbitrage strategy.

Table 13: Results under linear specifications of $\alpha(\cdot)$ and $\beta(\cdot)$ with 12 characteristics†

| Unrestricted ($\alpha(\cdot) \neq 0$) | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $K$ | $R^2_K$ | $R^2$ | $R^2_{T,N}$ | $R^2_{N,T}$ | $R^2_f$ | $R^2_{f,T,N}$ | $R^2_{f,N,T}$ | $R^2_O$ | $R^2_{T,N,O}$ | $R^2_{N,T,O}$ | Mean | Std | SR |
| 1 | 47.80 | 4.31 | 2.52 | 2.32 | 3.86 | 2.01 | 1.98 | 0.26 | 0.13 | 0.20 | 1.63 | 0.43 | 3.76 |
| 2* | 67.27 | 14.28 | 10.66 | 12.80 | 13.95 | 10.37 | 12.31 | 0.26 | 0.13 | 0.20 | 1.54 | 0.41 | 3.78 |
| 3 | 74.25 | 14.94 | 11.30 | 13.39 | 14.61 | 10.91 | 12.90 | 0.26 | 0.13 | 0.20 | 1.49 | 0.40 | 3.75 |
| 4 | 79.11 | 16.25 | 13.41 | 14.48 | 15.92 | 12.88 | 13.98 | 0.26 | 0.13 | 0.20 | 1.34 | 0.35 | 3.83 |
| 5 | 83.51 | 16.74 | 13.98 | 15.04 | 16.41 | 13.61 | 14.56 | 0.26 | 0.13 | 0.20 | 1.32 | 0.34 | 3.89 |
| 6 | 86.46 | 17.46 | 14.83 | 15.68 | 17.13 | 14.47 | 15.22 | 0.26 | 0.13 | 0.20 | 0.99 | 0.31 | 3.17 |
| 7 | 89.09 | 17.93 | 15.08 | 16.13 | 17.61 | 14.96 | 15.68 | 0.26 | 0.13 | 0.20 | 0.54 | 0.21 | 2.59 |
| 8 | 93.73 | 18.20 | 15.44 | 16.44 | 18.12 | 15.39 | 16.32 | 0.26 | 0.13 | 0.20 | 0.46 | 0.17 | 2.75 |
| 9 | 95.53 | 18.48 | 15.85 | 16.78 | 18.40 | 15.77 | 16.67 | 0.26 | 0.13 | 0.20 | 0.30 | 0.11 | 2.68 |
| 10 | 96.84 | 18.80 | 16.54 | 17.21 | 18.72 | 16.45 | 17.10 | 0.26 | 0.13 | 0.20 | 0.15 | 0.07 | 2.02 |
| Restricted ($\alpha(\cdot) = 0$) | | | | | | | | | | | | | |
| $K$ | $R^2_K$ | $R^2$ | $R^2_{T,N}$ | $R^2_{N,T}$ | $R^2_f$ | $R^2_{f,T,N}$ | $R^2_{f,N,T}$ | $R^2_O$ | $R^2_{T,N,O}$ | $R^2_{N,T,O}$ | Mean | Std | SR |
| 1 | 47.84 | NA | NA | NA | 3.97 | 2.02 | 2.01 | 0.17 | 0.21 | 0.03 | NA | NA | NA |
| 2* | 67.39 | NA | NA | NA | 13.91 | 10.30 | 12.23 | 0.33 | 0.53 | -0.25 | NA | NA | NA |
| 3 | 74.74 | NA | NA | NA | 14.58 | 10.87 | 12.95 | 0.47 | 0.52 | 0.07 | NA | NA | NA |
| 4 | 80.44 | NA | NA | NA | 15.18 | 11.48 | 13.63 | 0.48 | 0.61 | 0.16 | NA | NA | NA |
| 5 | 85.23 | NA | NA | NA | 16.51 | 13.76 | 14.79 | 0.49 | 0.61 | 0.18 | NA | NA | NA |
| 6 | 88.69 | NA | NA | NA | 17.00 | 14.21 | 15.33 | 0.51 | 0.64 | 0.19 | NA | NA | NA |
| 7 | 91.62 | NA | NA | NA | 17.69 | 15.05 | 15.96 | 0.50 | 0.60 | 0.23 | NA | NA | NA |
| 8 | 94.23 | NA | NA | NA | 18.17 | 15.41 | 16.41 | 0.50 | 0.62 | 0.22 | NA | NA | NA |
| 9 | 96.04 | NA | NA | NA | 18.45 | 15.84 | 16.75 | 0.50 | 0.60 | 0.23 | NA | NA | NA |
| 10 | 97.35 | NA | NA | NA | 18.78 | 16.52 | 17.19 | 0.49 | 0.62 | 0.21 | NA | NA | NA |

† $K$: the number of factor specified (* denotes the estimated one by our methods); Fama-MacBeth cross sectional regression $R^2$: $R^2_{\bar{Y}} = 19.07\%$; $R^2_K$ meassures the variations of managed portfolios captured by different numbers of factors from PCA; $R^2$, $R^2_{T,N}$, $R^2_{N,T}$: various in-sample $R^2$'s (%), see (25)-(27); $R^2_f$, $R^2_{f,T,N}$, $R^2_{f,N,T}$: various in-sample $R^2$'s (%) without $\alpha$, see (28)-(30); $R^2_O$, $R^2_{T,N,O}$, $R^2_{N,T,O}$: various out-sample predictive $R^2$'s (%), see (31)-(33); Mean: out-of-sample annualized means of the pure-alpha arbitrage strategy(%); Std: out-of-sample annualized standard deviations of the pure-alpha arbitrage strategy(%); SR: out-of-sample annualized Sharpe ratios of the pure-alpha arbitrage strategy.

Table 14: Results under continuous piecewise linear specifications of $\alpha(\cdot)$ and $\beta(\cdot)$ with 12 characteristics and one internal knot†

| Unrestricted ($\alpha(\cdot) \neq 0$) | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $K$ | $R^2_K$ | $R^2$ | $R^2_{T,N}$ | $R^2_{N,T}$ | $R^2_f$ | $R^2_{f,T,N}$ | $R^2_{f,N,T}$ | $R^2_O$ | $R^2_{T,N,O}$ | $R^2_{N,T,O}$ | Mean | Std | SR |
| 1 | 47.93 | 5.94 | 3.28 | 3.68 | 5.57 | 2.90 | 3.22 | 0.55 | 0.59 | 0.27 | 2.23 | 0.66 | 3.37 |
| 2* | 66.23 | 9.79 | 6.18 | 7.18 | 9.44 | 5.72 | 6.62 | 0.55 | 0.59 | 0.27 | 2.05 | 0.52 | 3.95 |
| 3 | 72.25 | 10.66 | 6.75 | 8.05 | 10.30 | 6.33 | 7.49 | 0.55 | 0.59 | 0.27 | 2.01 | 0.52 | 3.86 |
| 4 | 77.10 | 13.94 | 10.32 | 11.81 | 13.55 | 9.91 | 11.25 | 0.55 | 0.59 | 0.27 | 1.99 | 0.51 | 3.90 |
| 5 | 80.71 | 14.35 | 10.53 | 12.32 | 13.98 | 10.22 | 11.82 | 0.55 | 0.59 | 0.27 | 2.01 | 0.49 | 4.11 |
| 6 | 83.62 | 14.97 | 11.50 | 12.83 | 14.53 | 10.87 | 12.25 | 0.55 | 0.59 | 0.27 | 1.40 | 0.44 | 3.20 |
| 7 | 87.36 | 15.28 | 11.77 | 13.13 | 14.90 | 11.04 | 12.68 | 0.55 | 0.59 | 0.27 | 0.71 | 0.24 | 3.00 |
| 8 | 89.85 | 15.65 | 12.15 | 13.46 | 15.47 | 11.78 | 13.26 | 0.55 | 0.59 | 0.27 | 0.49 | 0.15 | 3.18 |
| 9 | 91.44 | 16.49 | 12.83 | 14.10 | 16.24 | 12.35 | 13.83 | 0.55 | 0.59 | 0.27 | 0.47 | 0.15 | 3.18 |
| 10 | 92.98 | 16.80 | 13.16 | 14.38 | 16.63 | 12.84 | 14.18 | 0.55 | 0.59 | 0.27 | 0.44 | 0.12 | 3.68 |
| Restricted ($\alpha(\cdot) = 0$) | | | | | | | | | | | | | |
| $K$ | $R^2_K$ | $R^2$ | $R^2_{T,N}$ | $R^2_{N,T}$ | $R^2_f$ | $R^2_{f,T,N}$ | $R^2_{f,N,T}$ | $R^2_O$ | $R^2_{T,N,O}$ | $R^2_{N,T,O}$ | Mean | Std | SR |
| 1 | 48.03 | NA | NA | NA | 5.66 | 2.86 | 3.26 | 0.28 | 0.32 | -0.06 | NA | NA | NA |
| 2* | 66.39 | NA | NA | NA | 9.39 | 5.67 | 6.53 | 0.34 | 0.32 | -0.43 | NA | NA | NA |
| 3 | 72.44 | NA | NA | NA | 10.27 | 6.22 | 7.49 | 0.60 | 0.75 | 0.23 | NA | NA | NA |
| 4 | 77.79 | NA | NA | NA | 10.87 | 7.08 | 8.22 | 0.52 | 0.70 | 0.25 | NA | NA | NA |
| 5 | 82.59 | NA | NA | NA | 14.25 | 10.54 | 12.16 | 0.49 | 0.58 | 0.25 | NA | NA | NA |
| 6 | 85.70 | NA | NA | NA | 14.59 | 10.75 | 12.58 | 0.52 | 0.53 | 0.24 | NA | NA | NA |
| 7 | 88.34 | NA | NA | NA | 15.13 | 11.51 | 13.03 | 0.52 | 0.55 | 0.27 | NA | NA | NA |
| 8 | 90.31 | NA | NA | NA | 15.56 | 11.98 | 13.37 | 0.52 | 0.54 | 0.28 | NA | NA | NA |
| 9 | 91.88 | NA | NA | NA | 16.32 | 12.65 | 13.93 | 0.52 | 0.54 | 0.28 | NA | NA | NA |
| 10 | 93.34 | NA | NA | NA | 16.74 | 13.05 | 14.29 | 0.53 | 0.53 | 0.28 | NA | NA | NA |

† $K$: the number of factor specified (* denotes the estimated one by our methods); Fama-MacBeth cross sectional regression $R^2$: $R^2_{\tilde{Y}} = 20.24\%$; $R^2_K$ meassures the variations of managed portfolios captured by different numbers of factors from PCA; $R^2$, $R^2_{T,N}$, $R^2_{N,T}$: various in-sample $R^2$'s (%), see (25)-(27); $R^2_f$, $R^2_{f,T,N}$, $R^2_{f,N,T}$: various in-sample $R^2$'s (%) without $\alpha$, see (28)-(30); $R^2_O$, $R^2_{T,N,O}$, $R^2_{N,T,O}$: various out-sample predictive $R^2$'s (%), see (31)-(33); Mean: out-of-sample annualized means of the pure-alpha arbitrage strategy(%); Std: out-of-sample annualized standard deviations of the pure-alpha arbitrage strategy(%); SR: out-of-sample annualized Sharpe ratios of the pure-alpha arbitrage strategy.

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

Connor, G. and Korajczyk, R. A. (1986). Performance measurement with the arbitrage pricing theory: A new framework for analysis. *Journal of financial economics*, **15** 373–394.

Connor, G. and Korajczyk, R. A. (1988). Risk and return in an equilibrium APT: Application of a new test methodology. *Journal of financial economics*, **21** 255–289.

Connor, G. and Linton, O. (2007). Semiparametric estimation of a characteristic-based factor model of common stock returns. *Journal of Empirical Finance*, **14** 694–717.

Daniel, K. and Titman, S. (1997). Evidence on the characteristics of cross sectional variation in stock returns. *Journal of Finance*, **52** 1–33.

Fama, E. F. (1976). *Foundations of Finance*. Basic Books, New York.

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

Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model. *Journal of financial economics*, **116** 1–22.

Fama, E. F. and MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *The Journal of Political Economy*, **81** 607–636.

Fan, J., Liao, Y. and Wang, W. (2016a). Projected principal component analysis in factor models. *The Annals of Statistics*, **44** 219–254.

Fan, J., Liao, Y. and Wang, W. (2016b). Supplement to projected principal component analysisin factor models". *DOI:10.1214/15-AOS1364SUPP*.

Fengler, M. R., Härdle, W. K. and Mammen, E. (2007). A semiparametric factor model for implied volatility surface dynamics. *Journal of Financial Econometrics*, **5** 189–218.

Ferson, W. E. and Harvey, C. R. (1999). Conditioning variables and the cross section of stock returns. *The Journal of Finance*, **54** 1325–1360.

Freyberger, J., Neuhierl, A. and Weber, M. (2020). Dissecting characteristics nonparametrically. *The Review of Financial Studies*, **33** 2326–2377.

Gagliardini, P. and Ma, H. (2019). Extracting statistical factors when betas are time-varying. Tech. rep., Available at SSRN 3426534.

Gagliardini, P., Ossola, E. and Scaillet, O. (2016). Time-varying risk premium in large crosssectional equity data sets. *Econometrica*, **84** 985–1046.

Gagliardini, P., Ossola, E. and Scaillet, O. (2020). Estimation of large dimensional conditional factor models in finance. In *Handbook of Econometrics*, vol. 7A, chap. 3.

Giglio, S. and Xiu, D. (2019). Asset pricing with omitted factors. Tech. rep., Chicago Booth Research Paper.

Green, J., Hand, J. R. and Zhang, X. F. (2017). The characteristics that provide independent information about average us monthly stock returns. *The Review of Financial Studies*, **30** 4389–4436.

Gu, S., Kelly, B. and Xiu, D. (2021). Autoencoder asset pricing models. *Journal of Econometrics*, **222** 429–450.

Harvey, C., Liu, Y. and Zhu, H. (2016). and the cross-section of expected returns. *The Review of Financial Studies*, **29** 5–68.

Huang, J. (1998). Projection estimation in multiple regression with application to functional ANOVA models. *The Annals of Statistics*, **26** 242–272.

Jacobs, B. I. and Levy, K. N. (1988). Disentangling equity return regularities: New insights and investment opportunities. *Financial Analysts Journal*, **44** 18–43.

Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, **48** 65–91.

Kelly, B. T., Pruitt, S. and Su, Y. (2019). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics*, **134** 501–524.

Kelly, B. T., Pruitt, S. and Su, Y. (2020). Instrumented principal component analysis. Tech. rep., Available at SSRN 2983919.

Kim, S., Korajczyk, R. A. and Neuhierl, A. (2020). Arbitrage portfolios. *Review of Financial Studies* Forthcoming.

Kirby, C. (2020). Firm characteristics, cross-sectional regression estimates, and asset pricing tests. *The Review of Asset Pricing Studies*, **10** 290–334.

Kozak, S., Nagel, S. and Santosh, S. (2018). Interpreting factor models. *The Journal of Finance*, **73** 1183–1223.

Lewbel, A. (1991). The rank of demand systems: Theory and nonparametric estimation. *Econometrica*, **59** 711–730.

Lewellen, J. (2015). The cross section of expected stock returns. *Critical Finance Review*, **4** 1–44.

Li, J. and Liao, Z. (2019). Uniform nonparametric inference for time series. Tech. rep., Duke University.

Lorentz, G. G. (1986). *Approximation of Functions*. Chelsea Publishing.

Motta, G., Hafner, C. M. and Von Sachs, R. (2011). Locally stationary factor models: Identification and nonparametric estimation. *Econometric Theory*, **27** 1279–1319.

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

Park, B. U., Mammen, E., Härdle, W. and Borak, S. (2009). Time series modelling with semiparametric factor dynamics. *Journal of the American Statistical Association*, **104** 284–298.

Pelger, M. and Xiong, R. (2021). State-varying factor models of large dimensions. *Journal of Business Economics & Statistics* Accpted.

Pollard, D. (2002). *A User's Guide to Measure Theoretic Probability*, vol. 8 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press.

Reisman, H. (1992). Intertemporal arbitrage pricing theory. *The Review of Financial Studies*, **5** 105–122.

Rosenberg, B. and McKibben, W. (1973). The prediction of systematic and specific risk in common stocks. *Journal of Financial and Quantitative Analysis* 317–333.

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, **13** 341–360.

Roussanov, N. (2014). Composition of wealth, conditioning information, and the cross-section of stock returns. *Journal of Financial Economics,*, **111** 352–380.

Stock, J. H. and Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, **97** 1167–1179.

Su, L. and Wang, X. (2017). On time-varying factor models: Estimation and testing. *Journal of Econometrics*, **198** 84–101.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. 1st ed. Springer Series in Statistics, Springer, New York.

Yu, Y., Wang, T. and Samworth, R. J. (2014). A useful variant of the Davis-Kahan theorem for statisticians. *Biometrika*, **102** 315–323.