Semiparametric Conditional Factor Models in Asset Pricing*

Qihui Chen† Nikolai Roussanov‡ Xiaoliang Wang§
CUHK-Shenzhen The Wharton School HKUST

March 27, 2026

Abstract

We introduce a simple and tractable methodology for estimating semiparametric conditional latent factor models. Our approach disentangles the roles of characteristics in capturing factor betas of asset returns from "alpha." We construct factors by extracting principal components from Fama-MacBeth managed portfolios. Applying this methodology to the cross-section of U.S. individual stock returns, we find compelling evidence of substantial nonzero pricing errors, even though our factors demonstrate superior performance in standard asset pricing tests. Unexplained "arbitrage" portfolios earn high Sharpe ratios, which decline over time. Combining factors with these orthogonal portfolios produces out-of-sample Sharpe ratios exceeding 4.

KEYWORDS: Characteristics, Fama-MacBeth regression, factor models, managed portfolios, nonlinearity, PCA, sieve estimation

*We thank Torben Andersen, Bin Chen, Xiaohong Chen, Xu Cheng, Greg Duffee, Robert Korajczyk (discussant), Yuan Liao, Ye Luo, Hao Ma (discussant), Seth Pruitt (discussant), Andres Santos, Frank Schorfheide, Liangjun Su, Dacheng Xiu, Chu Zhang, and Linyan Zhu (discussant) as well as conference audiences and seminar participants at Nanyang Technological University, Peking University, Shanghai University of Finance and Economics, UC Riverside, University of Lausanne (Swiss Finance Institute), University of Pennsylvania, Xiamen University, 2021 YEAP, 2022 SFS Cavalcade, 2022 WFA, 2022 CICF, 2023 AFA, Spring 2023 Rochester Conference in Econometrics, 2024 HKUST IAS-SBM Financial Econometrics conference, the Inaugural Meeting of the Greater Bay Econometrics Study Group for helpful discussions and comments. Shengqi Yang provided excellent research assistance.

†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 Finance, HKUST Business School; xlwangfin@ust.hk

1 Introduction

A central question in empirical asset pricing is why different assets earn different average returns. While asset pricing theory attributes cross-sectional differences in returns to variations in risk exposures, considerable evidence suggests that mispricing—captured by the dependence of returns on asset characteristics—also plays a significant role, suggesting potential market inefficiencies. Much of the debate centers on multi-factor models that aim to link average returns to factor loadings, building on the influential framework of Fama and French (1993), who introduced a portfolio-sorting approach to constructing asset pricing factors. Following their seminal work, researchers have proposed hundreds of factors, leading to what Cochrane (2011) memorably termed the "*factor zoo*," a concept further explored by Harvey et al. (2016). While some factor models have an explicit justification based on economic theory, many implicitly rely on the idea that factors capture common variation in returns, thus appealing to arbitrage pricing theory and its extensions (Ross, 1976; Chamberlain and Rothschild, 1982; Connor and Korajczyk, 1986, 1988; Reisman, 1992). Since implementing the latter requires estimating the conditional covariance matrix of returns, which becomes impractical when the number of assets (N) exceeds the number of time periods (T), most studies rely on asset characteristics to proxy for (imperfectly measured) risk exposures, often employing the portfolio-sorting approach. However, this makes distinguishing between risk-based explanations and those rooted in mispricing virtually impossible, as exemplified by the "characteristics versus covariances" debate (Daniel and Titman, 1997).

In our analysis we consider a canonical conditional factor model:

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

Here, y_{it} is the excess return of asset i at time period t, z_{it} is an $M \times 1$ vector of pre-specified asset characteristics (which may include a constant term) that is observed at the beginning of time period t,[1] f_t is a $K \times 1$ vector of unobserved *latent* factors, $\beta(\cdot)$ is a $K \times 1$ vector of unknown factor loading functions, $\alpha(\cdot)$ is an unknown intercept function, ε_{it} is the idiosyncratic component that is orthogonal to the common factors f_t.[2] The model describes a *conditional* factor model, in the sense that it captures time-variation

[1] In asset pricing, $z_{i,t-1}$, the characteristics observed at time period $t-1$, is usually used in (1). For notational simplicity, here we use z_{it} rather than $z_{i,t-1}$.

[2] While our main focus is on cross-sectional asset pricing, the model has other potential applications, which include modelling the implied volatility of options (Park et al., 2009) and describing consumer demand system (Lewbel, 1991), among others.

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. This model is well suited for resolving the "characteristics versus covariances" debate, since it potentially allows for distinguishing between the risk and mispricing explanations of the role of characteristics in predicting asset returns.[3] Meanwhile, the model allows for pooling the information in a multitude of characteristics and summarizing the common variation using a small number of factors, thereby helping to "tame the *factor zoo*." The challenge of using the model 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 sample 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, alongside 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 estimation involves two steps: (i) regressing y_{it} on sieve functions of z_{it} for each t; and (ii) applying principal component analysis (PCA) to the estimated coefficients obtained in step (i). We refer to this approach as the *regressed-PCA*. The first step aligns with the cross-sectional regressions of Fama and MacBeth (1973), where the estimated coefficients at each point in time represent returns of "pure play" characteristic-managed portfolios. The second step is effectively a standard PCA on a relatively small set of characteristic-managed portfolios constructed via Fama-MacBeth regressions. Second, we develop a bootstrap inference framework to assess the significance of $\alpha(\cdot)$ as well as test the linearity of $\alpha(\cdot)$ and $\beta(\cdot)$. Third, we establish large-sample properties of the estimators under mild conditions, including consistency, rate of convergence, and asymptotic normality, as well as the validity of the proposed tests. Notably, the asymptotic results possess several advantages: (i) they do not require large T; (ii) they accommodate time-varying and potentially nonstationary z_{it}; (iii) they apply to unbalanced panels, which is particularly beneficial for securities with varying lifespans. Our Monte Carlo simulations demonstrate that the proposed estimators and tests exhibit satisfactory finite-sample performance and remain robust even when T is small, provided N is large. In addition to offering formal inference procedures and well-founded asymptotic properties,

[3] 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).

regressed-PCA presents several advantages over existing methods such as instrumented PCA (IPCA) (Kelly et al., 2019) and projected-PCA (Fan et al., 2016a; Kim et al., 2020). Specifically, regressed-PCA is computationally efficient and accommodates nonzero alphas, time-varying characteristics, unbalanced panels, and short samples, making it particularly well-suited for empirical asset pricing applications.

We apply our new methodology to analyzing the cross-section of individual stock returns. Our analysis uses the same dataset as Kelly et al. (2019), the study most closely aligned with ours in terms of empirical aims. However, our econometric approach and empirical findings differ significantly from theirs. First, unlike Kelly et al. (2019, 2020), our method does not aim to simultaneously maximize the "fit" of the factor model to individual asset returns in both the time-series and cross-section. Instead, we extract factors that capture the most time-series comovement within a set of portfolios, which, in turn, reflect the most cross-sectional variation in individual asset returns. Second, our approach allows the $\alpha(\cdot)$ and $\beta(\cdot)$ functions to be nonlinear. We test—and reject—the validity of linear specifications empirically, which reveals the strong evidence of nonlinearity in factor loadings and pricing errors. Third, our inference procedure also enables us to test the significance of pricing errors. Our empirical results reveal that the pricing errors associated with many characteristics are statistically significant, leading to the rejection of the risk-based model. Lastly, our methodology facilitates rolling sub-sample analyses to accommodate evolving factor dynamics, as it does not rely on large T. We find that both in-sample and out-of-sample goodness-of-fit measures for all factor models decline from 1970 until roughly 2000 but improve thereafter. This pattern aligns with the findings in Campbell et al. (2001) and Campbell et al. (2022) on the time-variation in the amount of idiosyncratic volatility in the U.S. stock market. We also document a significant decline in pricing errors in more recent years, particularly since 2000. This decline may reflect the growing prevalence of quantitative investing, which reduces mispricing by exploiting characteristic-related anomalies, as suggested by McLean and Pontiff (2016) and Green et al. (2017).

Based on these findings, we construct trading strategies. The pure-*alpha* portfolios constructed based on nonzero pricing errors are associated with annualized Sharpe ratios typically above 3 (as is common in the literature, we refer to these as "arbitrage" portfolios, even though their returns are far from riskless). Meanwhile, the mean-variance efficient (MVE) portfolios constructed from the corresponding factors deliver substantially lower Sharpe ratios. This is different from the case in IPCA, where the Sharpe ratios of pure-*alpha* and MVE factor portfolios are comparable. Moreover, we approximate the

stock market's MVE portfolio with the combined MVE portfolios of the pure-*alpha* portfolios and factors. Regressed-PCA consistently yields higher Sharpe ratios than IPCA, highlighting the advantages of our method. The higher Sharpe ratios in regressed-PCA primarily stem from the pure-*alpha* portfolios, in contrast to the case in IPCA, where both the pure-*alpha* portfolios and factors contribute comparably. Furthermore, we document that the nonlinear specifications consistently produce MVE factor and combined MVE portfolios with higher Sharpe ratios than the linear specification, underscoring the significance of incorporating nonlinearity. Our results indicate that low-dimensional factors are unlikely to span the conditional efficient frontier. At the same time they demonstrate that imposing a factor structure on the conditional covariance matrix of returns yields robust estimates of the stochastic discount factor, as evidenced by the high Sharpe ratios of the out-of-sample MVE portfolios that we obtain using our approach.

In order to further validate our factors, we evaluate their performance in standard asset pricing tests. Our empirical results demonstrate that the regressed-PCA factors consistently outperform IPCA's factors in pricing a large set of testing portfolios, as evidenced by smaller pricing errors, t-statistics, and GRS statistics. IPCA factors' inferior performance primarily arises from its much larger regression R^2's, which indicates that these factors capture more time-series variation in returns well but less cross-sectional variation. Moreover, our factors from the nonlinear specifications also outperform Fama and French (2015)'s factors, which justifies the advantages of regressed-PCA over the traditional portfolio-sorting approach.

Our paper contributes to several strands of the literature. A number of studies have estimated models similar to (1) under the assumption that z_{it} are time-invariant, at least within subsamples. These include Connor and Linton (2007), Connor et al. (2012), Fan et al. (2016a), Kim et al. (2020), Ge et al. (2022), and Fan et al. (2022). Gagliardini and Ma (2019) and Gu et al. (2021) explore conditional latent factor models that impose the absence of arbitrage, i.e., $\alpha(\cdot) = 0$. There are numerous studies of conditional models with observed factors; see Gagliardini et al. (2016) and Gagliardini et al. (2020) for a comprehensive review. Another strand of literature studies time-varying factor models in which factor loadings evolve smoothly as functions of t/T or aggregate variables;[4] see, for example, Motta et al. (2011), Su and Wang (2017), and Pelger and Xiong (2021).

[4]The broader literature on conditional models with observable factors has extensively explored time-varying factor loadings that depend on aggregate variables rather than firm-specific characteristics. For example, Ferson and Harvey (1999) use a linear specification, while Roussanov (2014) employs nonparametric kernel-based approaches. Building on our methodology, Chen (2022) extends the estimation of conditional latent factor models to include heterogeneous *alpha* and *beta* functions, accommodating aggregate variables within z_{it}.

The literature on the cross-section of asset returns is vast; here we focus on multi-factor models motivated by the arbitrage pricing theory. Empirical analysis that exploits the ability of characteristics to predict asset returns typically follows either the portfolio-sorting approach (Fama and French, 1993, 2015; Daniel and Titman, 1997) 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, 2021; Kozak and Nagel, 2023). The significance of nonlinear relationships in asset pricing has been underscored by several empirical studies (Connor et al., 2012; Kirby, 2020) and more recently explored through machine learning methods (Gu et al., 2021; Chen et al., 2022).

The remainder of the paper is organized as follows. Section 2 introduces the estimation method—regressed-PCA—along with its key properties and advantages. Section 3 interprets the method in the context of asset pricing. Section 4 establishes large-sample properties of the estimators and develops bootstrap inference procedures. Section 5 applies the new methodology to analyze the cross-section of individual stock returns in the U.S. market. Finally, Section 6 briefly concludes. The Online Appendix includes estimators for the number of factors, assumptions, proofs of theoretical results, additional discussions, simulation results, and additional empirical findings.

2 Estimation Method

In this section, we introduce a method for estimating the model in (1), which we term regressed principal component analysis or regressed-PCA, along with its key properties and advantages.

To illustrate the underlying idea of our regressed-PCA approach, we begin by assuming that $\alpha(\cdot)$ is null and $\beta(\cdot)$ is linear, i.e., $\alpha(\cdot) = 0$ and $\beta(z_{it}) = \Gamma' z_{it}$ for some $M \times K$ matrix Γ. Let $Y_t \equiv (y_{1t}, \ldots, y_{Nt})'$, $Z_t \equiv (z_{1t}, \ldots, z_{Nt})'$, and $\varepsilon_t \equiv (\varepsilon_{1t}, \ldots, \varepsilon_{Nt})'$. The model in (1) can then be written in matrix form as:

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

A key challenge in applying PCA to estimate Γ and f_t is the presence of Z_t in the first term on the right-hand side of (2). To address this, we first regress Y_t on Z_t, yielding:

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

Heuristically, variation in the common component $Z_t\Gamma f_t$ over t comes from two sources: Z_t and f_t, and regressing Y_t on Z_t disentangles these sources by isolating Z_t from the common component. Given the factor structure on the right-hand side of (3), we can apply PCA to the series $\{(Z_t'Z_t)^{-1}Z_t'Y_t\}_{t\leq T}$ to obtain estimators for Γ and f_t.

Alternatively, the model in (2) can be viewed as a panel data model with time-varying slope coefficients Γf_t, which exhibit a factor structure. Essentially, regressed-PCA first estimates the time-varying slope coefficients by period-by-period cross-sectional regressions and then exploits the underlying factor structure by using PCA.

2.1 Regressed-PCA

Now, we consider the general case where $\alpha(\cdot)$ is nonzero and show how to estimate $\alpha(\cdot)$ and $\beta(\cdot) = (\beta_1(\cdot), \ldots, \beta_K(\cdot))'$ nonparametrically. To avoid the curse of dimensionality when z_{it} is multivariate, we assume $\alpha(\cdot)$ and $\beta_k(\cdot)$ are separable. Specifically, we assume there exist functions $\{\alpha_m(\cdot)\}_{m\leq M}$ and $\{\beta_{km}(\cdot)\}_{m\leq M}$ 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 mth entry of z_{it}. We adopt the sieve method to estimate $\alpha_m(\cdot)$ and $\beta_{km}(\cdot)$. Let $\{\phi_j(\cdot)\}_{j\geq 1}$ be a set of basis functions (e.g., B-splines, Fourier series, polynomials) that span a dense linear space of the functional space for $\alpha_m(\cdot)$ and $\beta_{km}(\cdot)$. Then, we can express:

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

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

Here, $\{a_{m,j}\}_{j\leq J}$ and $\{b_{km,j}\}_{j\leq J}$ 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.[5] 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 $\bar{\phi}(z_{it,m}) \equiv (\phi_1(z_{it,m}), \ldots, \phi_J(z_{it,m}))'$, $\phi(z_{it}) \equiv (\bar{\phi}(z_{it,1})', \ldots, \bar{\phi}(z_{it,M})')'$, and define the vectors $a \equiv (a_{1,1}, \ldots, a_{1,J}, \ldots, a_{M,1}, \ldots, a_{M,J})'$, $b_k \equiv (b_{k1,1}, \ldots, b_{k1,J}, \ldots,$

[5] For notational simplicity, we use the same basis functions in (5) for different m's and the same sieve size. Our results remain valid if different basis functions and different sieve sizes are used for each m.

$b_{kM,1}, \ldots, b_{kM,J})'$, and the matrix $B \equiv (b_1, \cdots, b_K)$. Let $r(z_{it}) \equiv \sum_{m=1}^{M} r_{m,J}(z_{it,m})$ and $\delta(z_{it}) \equiv (\sum_{m=1}^{M} \delta_{1m,J}(z_{it,m}), \ldots, \sum_{m=1}^{M} \delta_{Km,J}(z_{it,m}))'$. Thus, we have:

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

This shows that $\alpha(z_{it})$ and $\beta(z_{it})$ can be approximated by $a'\phi(z_{it})$ and $B'\phi(z_{it})$, respectively, and estimating $\alpha(\cdot)$ and $\beta(\cdot)$ reduces to estimating a and B.

Next, we adapt the regressed-PCA to estimate a, B, and f_t using the sieve approximation in (7). 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, we write the model in (1) in matrix form as:

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

Under the basic sieve assumption, the term $R(Z_t) + \Delta(Z_t)f_t$ is negligible. The main challenge in applying PCA to estimate a, B, and f_t lies in the presence of $\Phi(Z_t)$ in the first two terms on the right-hand side of (8). To address this, we regress Y_t on $\Phi(Z_t)$, yielding:

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

where $\tilde{Y}_t = (\Phi(Z_t)'\Phi(Z_t))^{-1}\Phi(Z_t)'Y_t$. Thus, we estimate a, B, and f_t as follows. First, since $\tilde{Y}_t \approx a + Bf_t$, we remove a by subtracting $\bar{\tilde{Y}} = \sum_{t=1}^{T} \tilde{Y}_t/T$ from $\tilde{Y}_t$ and estimate B by applying PCA to the demeaned series $\{\tilde{Y}_t - \bar{\tilde{Y}}\}_{t \leq T}$. Second, for identifying a (and thus $\alpha(\cdot)$), we impose the condition $a'B = 0$. Since $\bar{\tilde{Y}} \approx a + B\bar{f}$ (with $\bar{f} = \sum_{t=1}^{T} f_t/T$), we estimate a as $a \approx [I_{JM} - B(B'B)^{-1}B]\bar{\tilde{Y}}$. Finally, f_t is estimated as $f_t \approx (B'B)^{-1}B'\tilde{Y}_t$.

The formal estimators for a, B, $\alpha(\cdot)$, $\beta(\cdot)$, and $F = (f_1, \ldots, f_T)'$ are defined as follows. Let $\hat{a}$, $\hat{B}$, $\hat{\alpha}(\cdot)$, $\hat{\beta}(\cdot)$, and $\hat{F}$ denote the respective estimators. Let $\tilde{Y} \equiv (\tilde{Y}_1, \ldots, \tilde{Y}_T)$ and $M_T \equiv I_T - 1_T 1_T'/T$, where 1_T is a $T \times 1$ vector of ones. Using the normalization $B'B = I_K$ and $F'M_TF/T$ being diagonal with descending diagonal entries, the columns of $\hat{B}$ are the eigenvectors corresponding to the largest K eigenvalues of $\tilde{Y}M_T\tilde{Y}'/T$. We then have: $\hat{a} = (I_{JM} - \hat{B}\hat{B}')\bar{\tilde{Y}}$,

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

We assume that the number of factors, K, is fixed and known. In Section 4, we establish asymptotic properties of the estimators and develop inference methods. In Appendix

A, we propose two consistent estimators for K, ensuring that our results extend to the case of an unknown K by using a conditioning argument.

2.2 Key Properties

Our regressed-PCA method has several appealing properties and is straightforward to implement. First, as discussed in Section 4.1, it accommodates time-varying z_{it} and does not require a large time dimension T. This flexibility allows us to examine the evolving relationship between risk and return using both full-sample and sub-sample analyses.

Second, the estimation procedure is well-suited for unbalanced panels, which is particularly relevant in cross-sectional asset pricing applications. The key step of regressed-PCA is to compute $\tilde{Y}_t$. Specifically, we can express $\tilde{Y}_t$ as:

$$\tilde{Y}_t = \left(\sum_{i=1}^{N} \phi(z_{it})\phi(z_{it})'\right)^{-1} \sum_{i=1}^{N} \phi(z_{it}) y_{it}. \tag{11}$$

For unbalanced panels, $\tilde{Y}_t$ can still be calculated by summing only over the i's for which both z_{it} and y_{it} are observed at time period t. This approach is equivalent to treating missing data as zeros, allowing us to proceed as if working with a balanced panel. The asymptotic results we derive remain valid as long as $\min_{t\leq T} N_t \to \infty$, where N_t is the sample size at time period t.

Third, our method continues to be effective even when the pricing errors and risk exposures are not fully explained by z_{it}. Let $e_{\alpha,it}$ and $e_{\beta,it}$ be the error terms in the pricing errors and the risk exposures, respectively, which are orthogonal to z_{it}. In this case, the model becomes:

$$y_{it} = [\alpha(z_{it}) + e_{\alpha,it}] + [\beta(z_{it}) + e_{\beta,it}]' f_t + \varepsilon_{it} = \alpha(z_{it}) + \beta(z_{it})' f_t + \varepsilon^*_{it}, \tag{12}$$

where $\varepsilon^*_{it} = \varepsilon_{it} + e_{\alpha,it} + e'_{\beta,it} f_t$. Since we are not interested in estimating $e_{\alpha,it}$ and $e_{\beta,it}$, our asymptotic results remain valid if we replace ε_{it} in the original model with ε^*_{it}.

Finally, efficiency of our estimation procedure could be improved by using generalized least squares in its first step. Our asymptotic results continue to hold if we replace $\Phi(Z_t)$ and ε_t with their transformed counterparts $V_t^{-1/2}\Phi(Z_t)$ and $V_t^{-1/2}\varepsilon_t$, where V_t is the conditional covariance matrix of Y_t at each time period t. While V_t is usually unknown, Hoberg and Welch (2009) suggest some practical guidance to account for the cross-correlation and heteroskedasticity of the idiosyncratic noise ε_{it}.

2.3 Comparing Methods

How does our regressed-PCA compare with existing methods that have been proposed in the literature? What are the advantages of our regressed-PCA?

First, the projected-PCA proposed by Fan et al. (2016a) applies PCA to the series $\{\Phi(Z_t)(\Phi(Z_t)'\Phi(Z_t))^{-1}\Phi(Z_t)'Y_t\}_{t\leq T}$. In contrast, our regressed-PCA applies PCA to the series $\{(\Phi(Z_t)'\Phi(Z_t))^{-1}\Phi(Z_t)'Y_t\}_{t\leq T}$. The two methods are fundamentally different: regressed-PCA applies PCA to estimated coefficients, while projected-PCA focuses on fitted values. The regression step in regressed-PCA is designed to extract Z_t from the common component for consistent estimation, whereas projected-PCA aims to remove noise in non-time-varying factor loadings to achieve efficiency in estimation. Consequently, projected-PCA may yield inconsistent estimates when Z_t is time-varying.[6] As noted by Fan et al. (2016b) and further investigated by Cheng et al. (2023), ensuring the consistency of projected-PCA may necessitate imposing smoothness conditions on how Z_t varies with t, which our regressed-PCA does not require. Additionally, projected-PCA often requires dropping certain observations to maintain a balanced panel, whereas regressed-PCA is applicable to unbalanced panels.

While Kim et al. (2020) extend projected-PCA to accommodate nonzero $\alpha(\cdot)$, they do not develop an inference procedure. Similarly, Fan et al. (2022) extend projected-PCA by employing deep neural networks and propose a local version to capture slowly changing alphas and betas. However, while neural networks can alleviate the curse of dimensionality for prediction tasks, they are not typically well-suited for inference—a key focus of our paper. Our regressed-PCA not only allows for nonzero $\alpha(\cdot)$ but also provides formal testing procedures, which are crucial for evaluating and comparing factor models. Moreover, by incorporating rapidly varying Z_t, our regressed-PCA is able to capture abrupt changes in both alphas and betas effectively.

Second, consider the least squares estimation approach introduced by Park et al. (2009), which is at the core of the IPCA of Kelly et al. (2019). The least squares method minimizes the following objective function:

$$\sum_{t=1}^{T}(Y_t - \Phi(Z_t)a - \Phi(Z_t)Bf_t)'(Y_t - \Phi(Z_t)a - \Phi(Z_t)Bf_t), \tag{13}$$

[6] For instance, as illustrated in (2), we have $Z_t(Z_t'Z_t)^{-1}Z_t'Y_t \approx Z_t\Gamma f_t$, which does not conform to a factor structure unless Z_t remains constant over t. Therefore, applying PCA to $\{Z_t(Z_t'Z_t)^{-1}Z_t'Y_t\}_{t=1}^{T}$ can result in failure to estimate f_t.

while regressed-PCA minimizes:[7]

$$\begin{aligned}&\sum_{t=1}^{T}(\tilde{Y}_t - a - Bf_t)'(\tilde{Y}_t - a - Bf_t)\\ =&\sum_{t=1}^{T}(Y_t - \Phi(Z_t)a - \Phi(Z_t)Bf_t)'S_t(Y_t - \Phi(Z_t)a - \Phi(Z_t)Bf_t),\end{aligned} \tag{14}$$

where $S_t = \Phi(Z_t)(\Phi(Z_t)'\Phi(Z_t))^{-1}(\Phi(Z_t)'\Phi(Z_t))^{-1}\Phi(Z_t)'$. The objective functions of these two approaches differ, except when $\Phi(Z_t)'\Phi(Z_t)/N = I_{JM}$.[8] Essentially, the least squares approach in (13) maximizes in-sample R^2, while regressed-PCA in (14) optimizes time-series comovement of $\tilde{Y}_t$, which captures the most cross-sectional variations of individual asset returns. A key challenge with the least squares approach is that its minimization problem is nonconvex and cannot be solved explicitly. While Park et al. (2009) develop a numerical algorithm to find the estimators, and Kelly et al. (2019) propose an alternating least squares procedure, both methods may require careful selection of initial values to ensure convergence to the correct solution. Furthermore, the asymptotic properties of these algorithms are not well understood. In addition to the asymptotic properties that we derive, our regressed-PCA provides estimators that can always be explicitly solved. Unlike IPCA, which relies on a long time series of returns, our regressed-PCA does not require a large time dimension T, enabling sub-sample analyses and capturing potential time variation in the coefficients (a and B).

Another advantage of our regressed-PCA approach lies in factor construction. As the number of factors, K, increases by one, regressed-PCA simply constructs an additional factor on top of existing ones through PCA. In contrast, IPCA requires reconstructing all factors by recomputing alternating least squares, making it sensitive to the number of factors. For example, the first factor under $K = 2$ may differ significantly from the first factor under $K = 1$, and the factor space under $K = 2$ does not necessarily nest the factor space under $K = 1$. Moreover, IPCA's factors can be correlated. In contrast, regressed-PCA produces factors that are uncorrelated by construction and remain stable, regardless of the number of factors chosen.

Overall, in addition to its formal inference procedures and well-established asymptotic properties, our regressed-PCA offers significant computational simplicity. It accommodates nonzero alphas, time-varying characteristics, unbalanced panels, and short samples, making it particularly suitable for empirical asset pricing. Moreover, it provides

[7] The equality in (14) follows because $\tilde{Y}_t = (\Phi(Z_t)'\Phi(Z_t))^{-1}\Phi(Z_t)'Y_t$.
[8] See Appendix D.1 for discussion.

stable and reliable factor construction.

3 Asset Pricing Interpretation

Regressed-PCA has deep roots in asset pricing. In a typical asset pricing application, y_{it} represents the realized returns on asset i at the end of time period t, while $z_{it,m}$ represents the m'th attribute or characteristic of asset i that is known at the *beginning* of time period t (or, alternatively, at the "end" of time period $t-1$). The regressed-PCA first estimates the time-varying slope coefficients by period-by-period cross-sectional regressions of returns on (functions of) characteristics, and then exploits the factor structure by using PCA. These period-by-period cross-sectional regressions are known as Fama-MacBeth regressions (Fama and MacBeth, 1973), which help transform a large unbalanced panel of noisy individual asset returns into a lower-dimensional balanced panel of portfolio returns that are largely free of idiosyncratic noise, $\tilde{Y}_t$. Furthermore, $\tilde{Y}_t$ can be interpreted as the time t realization of returns on a set of JM characteristic-managed portfolios, sometimes referred to as "optimized portfolios," "characteristic pure plays," or "cross-section factors" (e.g., as in Hoberg and Welch (2009), Back et al. (2015), and Fama and French (2020)). We also refer to them as Fama-MacBeth managed portfolios.

In particular, if the basis functions $\phi(z_{it})$ include a constant term (e.g., as the first element in $\phi(z_{it})$) and are standardized to have a zero mean in each cross-section, then the intercept in the Fama-MacBeth regressions (the first element in $\tilde{Y}_t$) represents a "level" return. This is essentially the equal-weighted average excess return across all individual assets, with weights summing up to unity or costing \$1, and it has no ex-ante loadings on any characteristics.[9] This is sometimes referred to as a "naively diversified" or $1/N$ portfolio. As shown by Fama (1976), the period-by-period slope coefficients corresponding to the time-varying basis functions (all elements of $\tilde{Y}_t$ starting from the second) are excess returns on zero-cost portfolios, as long as a constant term is included in the basis functions. These portfolios have weights on individual assets that set the weighted average value of the relevant basis function to one and those of all the remaining basis functions to zeros;[10] see Hoberg and Welch (2009) and Kirby (2020) for

[9] This is due to two properties. First, the intercept in $\tilde{Y}_t$ is equal to $\sum_{i=1}^{N} y_{it}/N$ when the nonconstant regressors have a zero mean. Second, the weights in $\tilde{Y}_t$, given by $W_t = (\Phi(Z_t)'\Phi(Z_t))^{-1}\Phi(Z_t)'$, satisfy the property that $W_t\Phi(Z_t) = I_{JM}$. The second property also clarifies the property of the weights for other portfolios in $\tilde{Y}_t$.

[10] See Footnote 9.

other attractive properties. In simpler terms, each portfolio has spread in only one basis function, or a pure play on a particular basis function. Moreover, adding one more basis function in the Fama-MacBeth regressions to a benchmark corresponds to introducing one additional zero-cost portfolio, which raises average exposure to that specific basis function by one unit while maintaining average exposures to other basis functions unchanged. Therefore, these portfolios can also be interpreted as "slope" returns with respect to the time-varying basis functions.

Moreover, Fama (1976) demonstrates that the portfolios in $\tilde{Y}_t$ exhibit minimum variance and often low correlations under OLS-like i.i.d. assumptions, rendering them maximally diversified.[11] Due to their low noise and correlations, several studies have explored the advantages of these portfolios in asset pricing tests. For instance, Hoberg and Welch (2009), Back et al. (2015), and Kirby (2020) highlight their utility as test assets (dependent variables), while Back et al. (2013) and Fama and French (2020) emphasize their effectiveness as pricing factors (independent variables), particularly when compared with the portfolio-sorting approach. However, these studies primarily focus on linear specifications (i.e., $\phi(z_{it}) = z_{it}$ which includes a constant term) with a small number of characteristics, and their findings may not generalize to more complex cases.

In the presence of a large number of characteristics or basis functions, the dimension of $\tilde{Y}_t$ can become substantial, and the portfolios in $\tilde{Y}_t$ may exhibit high correlations, even though each portfolio is maximally diversified. Understanding the factor structure of the portfolios in $\tilde{Y}_t$ as test assets is crucial to avoid spurious fit of misspecified asset pricing models that happen to be correlated with some of the latent factors, as emphasized by Lewellen et al. (2010). A straightforward solution is to apply PCA to extract uncorrelated principal components from $\tilde{Y}_t$, which is the essence of our approach.

3.1 Comparison with Portfolio Sorting

The extracted factors from regressed-PCA in (10) are linear combinations of characteristic-managed portfolios $\tilde{Y}_t$, and thus are themselves tradable portfolios. As a way of factor construction, regressed-PCA shares a strong connection with the portfolio-sorting approach. Sorting assets into portfolios can be equivalently framed as running cross-sectional regressions of returns on dummies that represent groups sorted by characteristics at each time period, whether the sorting is independent or dependent. Specifically, the return of each sorted portfolio (i.e., the average return of individual assets within a

[11] See Appendix D.2 for discussion.

group) is the coefficient of the corresponding group dummy in the regression. Using our notation, the returns of sorted portfolios can be represented by $\tilde{Y}_t$ when the basis functions $\phi(z_{it})$ are group dummies.[12] Thus, the key difference between the first steps of the regressed-PCA and portfolio-sorting approaches lies in their choice of basis functions. Sorting has been recognized as a nonparametric method for examining the relationship between average returns and characteristics, as highlighted by Fama and French (2008), Cochrane (2011), and Cattaneo et al. (2020).

However, regressed-PCA offers several advantages over sorting. First, sorting quickly encounters the curse of dimensionality and rarely handles more than four characteristics simultaneously. When multiple characteristics are present, double sorting is typically used for each pair of characteristics, making it difficult to infer which characteristics uniquely affect average returns. Regressed-PCA addresses this limitation by employing a separable additive specification (see (4)) and B-splines basis functions, which allow for a large number of characteristics and facilitate testing their significance. Second, sorting fails to fully exploit the variation in characteristics within each sorted group. In contrast, regressed-PCA takes advantage of the full variation in characteristics. Third, sorting struggles to effectively explore the nonlinear relationship between average returns and characteristics: sorting essentially uses step functions, which suffer from several well-known shortcomings, such as discontinuities at cutoffs, poor extrapolation, and unstable estimates that are highly sensitive to outlier assets (Hastie et al., 2011). These issues are mitigated by the use of B-splines, which our regressed-PCA incorporates.

Although long-short factors (such as high-minus-low and small-minus-big factors) are straightforward to interpret, regressed-PCA offers several advantages over them. First, the long-short approach focuses only on portfolios in extreme groups and ignores those in the middle. As a result, it may fail when factor loadings exhibit non-monotonicity with respect to characteristics, such as a "tent" shape. Regressed-PCA, on the other hand, uses all portfolios formed by Fama-MacBeth sieve regressions, making it more adaptable to capturing underlying nonlinearities. Second, long-short factors do not effectively distinguish between the risk and mispricing explanations of the role of characteristics in predicting asset returns, a distinction at the heart of the "characteristics versus covariances" debate. Regressed-PCA is based on a latent factor model for individual asset returns, which is ideally suited to resolve this debate. Third, when multiple characteristics are present, long-short methods can quickly lead to the "*factor zoo*" problem as distentangling the roles of correlated characteristics can be difficult in multi-way sorts

[12] See Appendix D.3 for discussion.

with potentially too few securities in each bin. Regressed-PCA mitigates this issue by utilizing Fama-Macbeth managed portfolios, which are maximally diversified "pure plays" on characteristics, together with PCA, which has also been proven effective in reducing the dimensionality of sorted portfolios (Kozak et al., 2018, 2020; Lettau and Pelger, 2020).

4 Econometric Analysis

In this section, we establish the asymptotic properties of our estimators, including consistency, the rate of convergence, and their asymptotic distribution. Additionally, we develop bootstrap inference procedures. We begin by defining some notation that will be used throughout the paper. For a symmetric matrix A, we denote its kth largest eigenvalue by $\lambda_k(A)$, and its smallest and largest eigenvalues by $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$, respectively. The operator norm of a matrix A is denoted by $\|A\|_2$, and its Frobenius norm by $\|A\|_F$. The vectorization of A is written as $\text{vec}(A)$. The Euclidian norm of a column vector x is denoted by $\|x\|$. Finally, for matrices A and B, we use $A \otimes B$ to denote their Kronecker product.

4.1 Asymptotic Properties

Before presenting formal theorems, we revisit (2) to briefly illustrate why a large T is not required and Z_t can be nonstationary over t. Consider the case when $T \geq K+1$ and $M \geq K$. Since the columns of $\hat{B}$ and Γ are the eigenvectors of $\tilde{Y} M_T \tilde{Y}'$ and $\Gamma F' M_T F \Gamma'$, respectively, corresponding to the first K largest eigenvalues, by the matrix perturbation theorem (see, for example, Yu et al. (2014)), the consistency of $\hat{B}$ to Γ (up to a rotational transformation) can be established if we can show that:

$$\|\tilde{Y} M_T - \Gamma F' M_T\|_F = o_p(1) \text{ as } N \to \infty. \tag{15}$$

Since $\tilde{Y} = \Gamma F' + ((Z_1'Z_1)^{-1} Z_1'\varepsilon_1, \ldots, (Z_T'Z_T)^{-1} Z_T'\varepsilon_T)$, (15) simplifies to:

$$\|((Z_1'Z_1)^{-1} Z_1'\varepsilon_1, \ldots, (Z_T'Z_T)^{-1} Z_T'\varepsilon_T) M_T\|_F = o_p(1) \text{ as } N \to \infty. \tag{16}$$

When T is fixed, (16) is equivalent to $(Z_t'Z_t)^{-1} Z_t'\varepsilon_t = o_p(1)$ for each t. Therefore, only regularity conditions on Z_t and ε_t for each t are needed to apply the law of large numbers. This result also implies that Z_t can vary over t in a nonstationary fashion.

Let $H \equiv (F'M_T\hat{F})(\hat{F}'M_T\hat{F})^{-1}$, which represents a rotational transformation matrix that governs the convergence limit of $\hat{B}$, $\hat{F}$, and $\hat{\beta}(\cdot)$. Define $\xi_J \equiv \sup_z \|\bar{\phi}(z)\|$, which scales as $O(\sqrt{J})$ for B-splines and Fourier series, and $O(J)$ for polynomials (see, for example, Belloni et al. (2015)).

Theorem 4.1. *Suppose Assumptions B.1-B.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) $T \geq K+1$ (T may stay fixed or grow simultaneously with N); (iii) $J \to \infty$ with $J^2\xi_J^2 \log J = o(N)$. Then*

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

where $\kappa > 1/2$ is a constant representing the smoothness of $\alpha(\cdot)$ and $\beta(\cdot)$.

All the assumptions are outlined in Appendix B. Theorem 4.1 establishes that a and $\alpha(\cdot)$ can be consistently estimated by $\hat{a}$ and $\hat{\alpha}(\cdot)$, while B, F, and $\beta(\cdot)$ can be consistently estimated by $\hat{B}$, $\hat{F}$, and $\hat{\beta}(\cdot)$, respectively, up to a rotational transformation. This consistency holds as long as $J \to \infty$ under both large N and either fixed or large T. Notably, the large J requirement differs from Fan et al. (2016a). This distinction highlights the importance of controlling sieve approximation errors in $\alpha(\cdot)$ and $\beta(\cdot)$ to ensure consistent estimation of F. Appendix D.4 illustrates how misspecifications in $\alpha(\cdot)$ and $\beta(\cdot)$ can result in inconsistent estimation of F, motivating the need for a specification test, which is addressed in Section 4.3.

Additionally, Theorem 4.1 shows that the estimators achieve fast convergence rates. In particular, $\hat{F}$ attains the optimal rate $1/N$ when $\alpha(\cdot)$ and $\beta(\cdot)$ are sufficiently smooth (i.e., κ is sufficiently large). This implies that the nonparametric modelling of $\alpha(\cdot)$ and $\beta(\cdot)$ do not degrade the convergence rate for estimating F, as long as the smoothness conditions are satisfied. This result is crucial for developing the specification test for $\alpha(\cdot)$ and $\beta(\cdot)$ in Section 4.3, and also important for utilizing $\hat{F}$ in subsequential asset pricing tests of Section 5.2. Notably, if the functional forms of $\alpha(\cdot)$ and $\beta(\cdot)$ are known and

correctly specified, sieve approximation errors can be avoided, and the asymptotic results hold for a fixed J. Theorem 4.1 allows for weak dependence of the errors $\{\varepsilon_{it}\}_{i\leq N,t\leq T}$ over both i and t, which is relevant in asset pricing.

Let $\Omega \equiv \sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} f_t^\dagger f_s^{\dagger\prime} Q_t^{-1} E[\phi(z_{it})\phi(z_{is})'] \times Q_s^{-1} E[\varepsilon_{it}\varepsilon_{is}]/NT$, where $f_t^\dagger = (1, (f_t - \bar{f})')'$ and $Q_t = \sum_{i=1}^{N} E[\phi(z_{it})\phi(z_{it})']/N$. This defines a variance-covariance matrix, which appears in the asymptotic distributions of $\hat{a}$ and $\hat{B}$.

Theorem 4.2. *Suppose Assumptions B.1-B.6 hold. Let $\hat{a}$ and $\hat{B}$ be given in (10). Assume (i) $N \to \infty$; (ii) $T \geq K+1$; (iii) $J \to \infty$ with $J^2\xi_J^2 \log J = o(N)$. Then there is 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 $\kappa > 1/2$ is a constant representing the smoothness of $\alpha(\cdot)$ and $\beta(\cdot)$, $\mathbb{G}_a = (I_{JM} - B\mathcal{H}\mathcal{H}'B')(\mathbb{N}_1 - \mathbb{G}_B\mathcal{H}^{-1}\bar{f}) - B\mathcal{H}\mathbb{G}_B'a$, and $\mathbb{G}_B = \mathbb{N}_2 B'B\mathcal{M}$. The matrices $\mathcal{H}$ and $\mathcal{M}$ are nonrandom, as given in Lemma C.15, while $\mathbb{N}_1$ and $\mathbb{N}_2$ are the first column and the last K columns of $\mathbb{N}$, respectively.

Theorem 4.2 establishes a strong approximation, demonstrating that $(\sqrt{NT}(\hat{a}-a), \sqrt{NT}(\hat{B}-BH))$ can be well approximated by a normal random matrix $(\mathbb{G}_a, \mathbb{G}_B)$. Specifically, the difference between them converges in probability to zero under the conditions $T = o(N)$, $NTJ^{-2\kappa} = o(1)$, and $J = o(\min\{N^{1/5}, N/T\})$. Since the dimensions of $\sqrt{NT}(\hat{a}-a)$ and $\sqrt{NT}(\hat{B}-BH)$ grow with N, rendering the classical central limit theorem inapplicable, we employ YurinskiiâĂŹs coupling to establish this strong approximation. This approach accommodates weak temporal dependence in the errors $\{\varepsilon_{it}\}_{i\leq N,t\leq T}$. Furthermore, the result can be readily extended to allow for cluster-type dependence across i in the errors $\{\varepsilon_{it}\}_{i\leq N,t\leq T}$; see the discussion following Assumption B.6. Notably, distributional results of this kind are not provided in Fan et al. (2016a).

4.2 Weighted Bootstrap

We develop a weighted bootstrap approach to estimating the distribution of $(\mathbb{G}_a, \mathbb{G}_B)$. Let $\{w_i\}_{i\leq N}$ be a sequence of i.i.d. positive random variables, with $E[w_i] = 1$ and

$var(w_i) = \omega_0 > 0$. For instance, the w_i's can be drawn from a standard exponential distribution, where $\omega_0 = 1$. To preserve the time dependence, we assign the same weight w_i to all observations over t. Define $\Phi(Z_t)^* \equiv (\phi(z_{1t})w_1, \ldots, \phi(z_{Nt})w_N)'$ and $\tilde{Y}_t^* \equiv (\Phi(Z_t)^{*\prime}\Phi(Z_t))^{-1}\Phi(Z_t)^{*\prime}Y_t$, which is the bootstrap version of $\tilde{Y}_t$. To define the bootstrap estimators of a and B, let $\tilde{Y}^* \equiv (\tilde{Y}_1^*, \ldots, \tilde{Y}_T^*)$ and $\bar{\tilde{Y}}^* \equiv \sum_{t=1}^T \tilde{Y}_t^*/T$. The bootstrap estimators are given by:

$$\hat{B}^* = \tilde{Y}^* M_T \hat{F}(\hat{F}' M_T \hat{F})^{-1} \text{ and } \hat{a}^* = (I_{JM} - \hat{B}^*(\hat{B}^{*\prime}\hat{B}^*)^{-1}\hat{B}^{*\prime})\bar{\tilde{Y}}^*, \tag{17}$$

which mimic the original estimators: $\hat{B} = \tilde{Y} M_T \hat{F}(\hat{F}' M_T \hat{F})^{-1}$ 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 estimating 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.[13]

The bootstrap procedure can be easily adapted for unbalanced panels. The key step is obtaining $\tilde{Y}_t^*$. For balanced panels, we write:

$$\tilde{Y}_t^* = \left(\sum_{i=1}^N \phi(z_{it})\phi(z_{it})' w_i\right)^{-1} \sum_{i=1}^N \phi(z_{it}) y_{it} w_i. \tag{18}$$

In unbalanced panels, we adjust by taking the sums over the i's for which both z_{it} and y_{it} are observed at time period t. This effectively replaces missing data with zeros, making the procedure identical to the balanced case. The asymptotic results established below remain valid as long as as $\min_{t \leq T} N_t \to \infty$, where N_t represents the sample size at time period t. Moreover, the bootstrap can easily accommodate cluster-type dependence across i by assigning the same weight within each cluster.

Theorem 4.3. *Suppose Assumptions B.1-B.7 hold. Let $\hat{a}$, $\hat{B}$, $\hat{a}^*$, and $\hat{B}^*$ be given in (10) and (17). Assume (i) $N \to \infty$; (ii) $T \geq K+1$; (iii) $J \to \infty$ with $J^2\xi_J^2 \log J = o(N)$. Then there is a $JM \times (K+1)$ random matrix $\mathbb{N}^*$ with $\text{vec}(\mathbb{N}^*) \sim N(0, \Omega)$ conditional on $\{Y_t, Z_t\}_{t \leq T}$ such that:*

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

[13] A more natural bootstrap estimator for B is given by the eigenvectors of $\tilde{Y}^* M_T \tilde{Y}^{*\prime}/T$ corresponding to its first K largest eigenvalues. However, the approach generally fails due to rational transformation matrices; see Appendix D.5 for discussion.

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 p^* is the probability measure with respect to $\{w_i\}_{i\leq N}$ conditional on $\{Y_t, Z_t\}_{t\leq T}$, $\kappa > 1/2$ is a constant representing the smoothness of $\alpha(\cdot)$ and $\beta(\cdot)$, $\mathbb{G}_a^* = (I_{JM} - B\mathcal{H}\mathcal{H}'B')(\mathbb{N}_1^* - \mathbb{G}_B^*\mathcal{H}^{-1}\bar{f}) - B\mathcal{H}\mathbb{G}_B^{*\prime}a$, and $\mathbb{G}_B^* = \mathbb{N}_2^* B'B\mathcal{M}$. The matrices $\mathcal{H}$ and $\mathcal{M}$ are nonrandom, as given in Lemma C.15, while $\mathbb{N}_1^*$ and $\mathbb{N}_2^*$ are the first column and the last K columns of $\mathbb{N}^*$, respectively.

Theorem 4.3 demonstrates that the distribution of $(\mathbb{G}_a, \mathbb{G}_B)$, which aligns with the distribution of $(\mathbb{G}_a^*, \mathbb{G}_B^*)$, can be 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, under the conditions $T = o(N)$, $NTJ^{-2\kappa} = o(1)$, and $J = o(\min\{N^{1/5}, N/T\})$. Theorems 4.2 and 4.3 can then be directly applied to conduct significance tests. To test whether $\alpha(\cdot) = 0$, we compare $NT\hat{a}'\hat{a}$ with the $1-\alpha$ quantile of $NT(\hat{a}^* - \hat{a})'(\hat{a}^* - \hat{a})/\omega_0$ conditional on the data for $0 < \alpha < 1$. Similarly, we can test whether each component of $\phi(z_{it})$ is significant in $\alpha(z_{it})$, which is equivalent to testing whether the corresponding element of a is zero. We can also test the joint significance of each component of $\phi(z_{it})$ in $\beta(z_{it})$, which corresponds to testing whether the associated row of BH is zero.

4.3 Specification Test

To test for linearity of $\alpha(\cdot)$ and $\beta(\cdot)$, we consider the following hypotheses:

$$\begin{aligned} &\mathrm{H}_0 : \alpha(z_{it}) = \gamma' z_{it} \text{ and } \beta(z_{it}) = \Gamma' z_{it} \text{ for some } \gamma \text{ and } \Gamma \text{ versus} \\ &\mathrm{H}_1 : \inf_{\pi} E[|\alpha(z_{it}) - \pi' z_{it}|^2] > 0 \text{ or } \inf_{\Pi} E[\|\beta(z_{it}) - \Pi' z_{it}\|^2] > 0. \end{aligned} \tag{19}$$

We develop a test by comparing the estimators under H_0 and H_1. The estimators of $\alpha(\cdot)$ and $\beta(\cdot)$ under H_1 are given by $\hat{\alpha}(\cdot)$ and $\hat{\beta}(\cdot)$, as defined 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 $\bar{\vec{Y}} \equiv \sum_{t=1}^T \vec{Y}_t/T$. The estimators of $\alpha(z_{it})$ and $\beta(z_{it})$ under 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 $\hat{\gamma} = \bar{\vec{Y}} - \hat{\Gamma}\sum_{t=1}^T \hat{f}_t/T$.[14]

[14] It is crucial to use the unrestricted estimator $\hat{F}$ in both $\hat{\Gamma}$ and $\hat{\gamma}$, rather than the restricted one under H_0. This ensures that $\hat{\Gamma}' z_{it}$ and $\hat{\beta}(z_{it})$ share a common rotational transformation matrix, justifying the validity of the test. It also avoids the full-rank requirement for Γ. Thanks to the optimal rate of $\hat{F}$ established in Theorem 4.1, this does not cause an issue.

Our test statistic is given by:

$$\mathcal{S} = \frac{1}{J}\sum_{i=1}^{N}\sum_{t=1}^{T}|\hat{\gamma}'z_{it} - \hat{\alpha}(z_{it})|^2 + \frac{1}{J}\sum_{i=1}^{N}\sum_{t=1}^{T}\|\hat{\Gamma}'z_{it} - \hat{\beta}(z_{it})\|^2. \tag{20}$$

To obtain critical values, we adopt a 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 the proof of Theorem 4.4 that under 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 estimate the null distribution of $\mathcal{S}$ by the distribution of

$$\begin{aligned}\mathcal{S}^* = &\frac{1}{J\omega_0}\sum_{i=1}^{N}\sum_{t=1}^{T}|(\hat{\gamma}^*-\hat{\gamma})'z_{it} - (\hat{a}^*-\hat{a})'\phi(z_{it})|^2 \\ &+ \frac{1}{J\omega_0}\sum_{i=1}^{N}\sum_{t=1}^{T}\|(\hat{\Gamma}^*-\hat{\Gamma})'z_{it} - (\hat{B}^*-\hat{B})'\phi(z_{it})\|^2\end{aligned} \tag{21}$$

conditional on the data. Here, $\hat{\Gamma}^* = \vec{Y}^*M_T\hat{F}(\hat{F}'M_T\hat{F})^{-1}$ 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 H_0 if $\mathcal{S} > c_{1-\alpha}$.

Theorem 4.4. *Suppose Assumptions B.1-B.8 hold. Let $\mathcal{S}$ be given in (20) and $c_{1-\alpha}$ be given after (21) for $0 < \alpha < 1$. Assume (i) $N \to \infty$; (ii) $T \geq K+1$; (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 $NTJ^{-2\kappa} = o(1)$, where $\kappa > 1/2$ is a constant representing the smoothness of $\alpha(\cdot)$ and $\beta(\cdot)$. Then*

$$P(\mathcal{S} > c_{1-\alpha}) \to \alpha \text{ under } \mathrm{H}_0 \text{ and } P(\mathcal{S} > c_{1-\alpha}) \to 1 \text{ under } \mathrm{H}_1.$$

5 Empirical Analysis

Our empirical analysis is based on the model specified in (1). Following standard practice in asset pricing, we depart from the notation used in previous sections and denote characteristics observed at time $t-1$ as $z_{i,t-1}$ instead of z_{it}. We begin by estimating the model and evaluating its overall performance, and then assess the performance of the extracted factors. The primary objective of our analysis is to investigate whether pricing errors are associated with characteristics and to evaluate the performance of our factors in asset pricing tests.

5.1 Data and Methodology

We use the same dataset as Kelly et al. (2019), which is originally from Freyberger et al. (2020). The dataset contains monthly returns of 12,813 individual stocks and 36 time-varying characteristics, covering the sample period from July 1962 to May 2014. The data is in the form of an unbalanced panel, for which our method is applicable. For the detailed descriptions, refer to these papers. To ensure comparability, we use the same 36 characteristics as those authors. Following the procedure in Kelly et al. (2019), we transform the values of the characteristics into relative rankings within the range $[-0.5, 0.5]$. This transformation standardizes the contributions of characteristics to pricing errors and risk exposures such that the estimation only depends on the rankings of characteristics and is robust to extreme values, sharing the similar logic with the sorting procedure as in Fama and French (1993, 2015). To meet the large N requirement, we select a sample period during which at least 1,000 individual stocks have observations for both returns and the 36 characteristics. This results in a sample spanning from September 1968 to May 2014. Based on this dataset, we construct the market factor and five long-short factors following Fama and French (2015). These factors exhibit close means and standard deviations and show high correlations with the corresponding factors from Kenneth R. FrenchâĂŹs website, as shown in Table F.I.

We implement regressed-PCA estimation by selecting the basis functions to be either linear (i.e., $\phi(z_{it}) = z_{it}$ including a constant term) or non-linear (via linear B-splines of z_{it}).[15] Using $\phi(z_{it}) = z_{it}$ leads to linear specifications of $\alpha(\cdot)$ and $\beta(\cdot)$, while setting $\phi(z_{it})$ as linear B-splines of z_{it} results in nonlinear specifications of $\alpha(\cdot)$ and $\beta(\cdot)$, where $\alpha(\cdot)$ and $\beta(\cdot)$ are continuous, piecewise linear functions.[16] For ease of comparison, we maintain the same parameter dimension across different specifications. Specifically, we consider 18 characteristics with one internal knot and 12 characteristics with two internal knots in the linear B-splines specifications. The most significant 18/12 characteristics are selected based on the linear specification, which are collected in Table F.II. To implement the weighted bootstrap, we let the bootstrap weights w_i's be i.i.d. random variables following the standard exponential distribution. For testing $\alpha(\cdot) = 0$ and linearity of $\alpha(\cdot)$ and $\beta(\cdot)$, we set the number of bootstrap draws to 499.

[15] Our econometric theory accommodates a variety of basis functions, such as Fourier series, polynomials, splines, and wavelets. Following Gu et al. (2020) and Freyberger et al. (2020), we employ splines due to their flexibility, which arises from increasing the number of knots. Unlike polynomials that require a higher degree for flexibility, splines generally produce more stable estimates (Hastie et al., 2011).

[16] The one dimensional linear B-splines $\{\psi_j(z)\}_{j=1}^{J}$ are defined over 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.

In order to evaluate the performance of the models estimated via regressed-PCA, we compute several measures of fit and prediction. First, we calculate Fama-MacBeth cross-sectional regression R^2, denoted as $R^2_{\tilde{Y}}$, which captures the variation in individual stock returns explained by the Fama-MacBeth managed portfolios $\tilde{Y}_t$ constructed from $\phi(z_{i,t-1})$. Next, we report the variation in these managed portfolios explained by the extracted factors $\hat{f}_t$, denoted as R^2_K. We then 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 measure (denoted as R^2) is total R^2 as used in Kelly et al. (2019). The second measure ($R^2_{T,N}$) calculates the cross-sectional average of time series R^2 across all stocks, which reflects the ability of the factors to capture common variation in stock returns. The third measure ($R^2_{N,T}$) computes the time-series average of cross-sectional goodness-of-fit measures, approximating the Fama-MacBeth cross-sectional regression R^2. This measure is particularly relevant for evaluating the model's capacity to explain the cross-section of average returns. The measures are defined as follows:[17]

$$R^2 = 1 - \frac{\sum_{i,t}[y_{it} - \hat{\alpha}(z_{i,t-1}) - \hat{\beta}(z_{i,t-1})'\hat{f}_t]^2}{\sum_{i,t} y_{it}^2}, \tag{22}$$

$$R^2_{T,N} = 1 - \frac{1}{N}\sum_i \frac{\sum_t[y_{it} - \hat{\alpha}(z_{i,t-1}) - \hat{\beta}(z_{i,t-1})'\hat{f}_t]^2}{\sum_t y_{it}^2}, \tag{23}$$

$$R^2_{N,T} = 1 - \frac{1}{T}\sum_t \frac{\sum_i[y_{it} - \hat{\alpha}(z_{i,t-1}) - \hat{\beta}(z_{i,t-1})'\hat{f}_t]^2}{\sum_i y_{it}^2}. \tag{24}$$

We also report 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 alphas $\hat{\alpha}(z_{i,t-1})$; see (F.1)-(F.3) for the corresponding formulas.

Finally, we assess the out-of-sample prediction and fit using expanding-window estimations. For $t \geq 120$, we use the data up to time period $t-1$ to implement the regressed-PCA and obtain estimates such as $\hat{a}_{t-1}$, $\hat{B}_{t-1}$, $\hat{\alpha}_{t-1}(\cdot)$, $\hat{\beta}_{t-1}(\cdot)$, and $\hat{F}_{t-1} \equiv (\hat{f}_1^{(t-1)}, \ldots, \hat{f}_{t-1}^{(t-1)})'$. Using these, we compute the out-of-sample prediction of y_{it} as $\hat{\alpha}_{t-1}(z_{i,t-1}) + \hat{\beta}_{t-1}(z_{i,t-1})'\hat{\lambda}_t$, where $\hat{\lambda}_t = \sum_{s\leq t-1} \hat{f}_s^{(t-1)}/(t-1)$, which is the average of factor estimates through time period $t-1$. The out-of-sample predictive R^2 is:

$$R^2_O = 1 - \frac{\sum_{i,t\geq 120}[y_{it} - \hat{\alpha}_{t-1}(z_{i,t-1}) - \hat{\beta}_{t-1}(z_{i,t-1})'\hat{\lambda}_t]^2}{\sum_{i,t\geq 120} y_{it}^2}; \tag{25}$$

[17] The differences among the three R^2's are provided in Appendix D.6.

see (F.4)-(F.6) for another two versions. We calculate the out-of-sample realized factor returns at time period t as: $\hat{f}_{t-1,t} = \hat{B}'_{t-1}\tilde{Y}_t = \hat{B}'_{t-1}(\Phi(Z_{t-1})'\Phi(Z_{t-1}))^{-1}\Phi'(Z_{t-1})Y_t$. Although the resulting factor returns are only known ex post at time period t, they represent returns on portfolios that are constructed ex ante, using weights based on estimates obtained at time period $t-1$. Using these, we can access how much of the cross-sectional variation of individual stock returns can be explained by the pre-estimated $\hat{\beta}_{t-1}(z_{i,t-1})$. We then define the out-of-sample fit R^2 as:

$$R^2_{f,O} = 1 - \frac{\sum_{i,t\geq 120}[y_{it} - \hat{\beta}_{t-1}(z_{i,t-1})'\hat{f}_{t-1,t}]^2}{\sum_{i,t\geq 120} y^2_{it}}; \tag{26}$$

see (F.7)-(F.9) for another two versions.

5.2 Empirical Results

5.2.1 Model Estimation

The main findings presented in Tables I-III can be summarized as follows. First, all of our measures of fit indicate that a low-dimensional factor model is unlikely to explain the time-series—or the cross-section—of individual stock returns (rather than the managed portfolios). In all specifications at least 5 or 6 factors are required for most of the in-sample R-squared to exceed 10%. That said, the total in-sample R^2's in our model are smaller than those of Kelly et al. (2019). This is not surprising, since the objective of their IPCA estimation is maximizing the total in-sample R^2, as discussed in (13). We extract factors that capture the most time-series comovement within a set of portfolios, which, in turn, reflect the most cross-sectional variation in individual asset returns. In contrast, our out-of-sample R^2_O's are 0.54% for the linear specification and 0.59% and 0.57% for the two nonlinear specifications,[18] which are comparable to the 0.60% in Kelly et al. (2019)'s linear specification with six factors. Similarly, the out-sample-sample fits are close to those of Kelly et al. (2019). With six factors, the out-sample-sample $R^2_{f,O}$'s are 15.38%, 16.20%, and 16.16% for the three specifications, which are comparable to the 17.80% in Kelly et al. (2019). Moreover, all in-sample and out-of-sample fits improve as the number of factors increases, as factor loadings soak up more of the variation in managed portfolio returns that is otherwise attributed to the alphas. In addition,

[18] Our out-of-sample predictive R^2's are invariant to the number of factors, because $\hat{\alpha}_{t-1}(z_{i,t-1}) + \hat{\beta}_{t-1}(z_{i,t-1})'\hat{\lambda}_t = \phi(z_{i,t-1})'[\hat{a}_{t-1} + \hat{B}_{t-1}\hat{F}'_{t-1}1_{t-1}/(t-1)] = \phi(z_{i,t-1})'\sum_{s=1}^{t-1}\tilde{Y}_t/t-1$, which does not depend on K.

compared to the linear specification, the nonlinear specifications based on linear B-splines significantly improve both in-sample and out-of-sample fits in most cases. They also slightly enhance the out-of-sample predictive R^2's. Both the improved fits and robustness highlight the advantage of the nonlinear specifications.

Second, the results of testing the linearity of $\alpha(\cdot)$ and $\beta(\cdot)$ explain why we observe lower R^2's in the linear specification. Linearity is strongly rejected at the 1% level in all factor models estimated with one to ten factors (Tables I-III report the p-values concisely to save space, all of which indicate rejection at the 1% level). Moreover, we also find robust evidence rejecting the null hypothesis $\alpha(\cdot) = 0$ across all cases. Additional empirical results are provided in Appendix F (see Tables F.III-F.V).

Before analyzing the contribution of each characteristic to pricing errors and risk exposures, we first determine the signs of the extracted factors. Under the normalization $B'B = I_K$ and $F'M_TF/T$ being diagonal with descending diagonal entries, the signs of the factors remain undetermined. To address this, we set the sample means of the factors to be positive, ensuring the unconditional risk premium on each factor is positive. To interpret the factors, we examine their correlations with the market factor and five long-short factors from Kenneth R. FrenchâĂŹs website as discussed in Table F.I, and conduct projection regressions, as detailed in Appendix F (see Tables F.VI-F.XII). We find the substantial correlations between these factors and our factors, with both sets explaining significant variations in each other.

Figures 1 and 2 illustrate the contribution of each characteristic to pricing errors and risk exposures under the linear specification. Figure 1 reveals that the 95% confidence intervals of characteristic coefficients in pricing errors remain relatively stable as the number of factors increases from one to six. Notably, 22 out of 36 characteristics remain significant at the 5% level for $K = 6$, in stark contrast to the results reported by Kelly et al. (2019). Figure 2 displays the characteristic coefficients in risk exposures for the first six factors. In contrast to Kelly et al. (2019)'s finding that 13 out 36 characteristics are significant in driving risk exposures, we find 24 significant characteristics for $K = 6$. Specifically, the coefficient of "market cap" in the first factor is negative and large in magnitude; the fourth and sixth factors exhibit substantial positive loadings on "market beta" (i.e., "beta"); the second and fifth factors display significant positive loadings on "book-to-market ratio" (i.e.,"bm"). These findings align with the traditional views of asset pricing anomalies as discussed in Fama and French (1993, 2015). More results for the two nonlinear specifications are provided in Appendix F (see Figures F.1-F.4).

Taking advantage of the fact that our regressed-PCA method does not require a large

time dimension (T), we perform subsample analyses using five-year intervals starting in January 1970. Figure 3 presents the key results. The left panel shows that average pricing errors (measured by $\|a\|^2$) under the nonlinear specifications are significantly smaller than those under the linear specification. For the linear case, average pricing errors are the highest during the initial subsample period (1970-1974), decline over time, rise again during the equity market "boom" of the 1990s, peak in the early 2000s, and then drop sharply. Under the nonlinear specifications, the patterns differ somewhat, with average pricing errors spiking around 1990-1994 and subsequently decreasing to levels comparable to those in the linear case by the end of the sample period. This decline may reflect the growing prevalence of quantitative investing, which reduces mispricing by exploiting characteristic-related anomalies, as suggested by McLean and Pontiff (2016) and Green et al. (2017).

The right panel of Figure 3 illustrates the proportion of time-series and cross-sectional variation in stock returns explained by common factors (measured by $R^2_{f,T,N}$ and $R^2_{f,N,T}$), which appears similar across model specifications. Notably, all the reported R^2 measures decline from 1970, reach a trough in the mid-1990s, and then steadily rise until the sample ends in 2014. This observation aligns with the empirical findings: Campbell et al. (2001) document a noticeable increase in firm-level volatility between 1962 and 1997; while extending this analysis to 2021, Campbell et al. (2022) find that idiosyncratic volatility declined after peaking in 1999-2000. Similar trends are evident in the out-of-sample fit measures shown in Figure F.5.

5.2.2 Trading Strategies

We construct trading strategies based on our estimated models. While constructing the MVE portfolio on individual stocks is usually infeasible due to the challenge in estimating a high-dimensional covariance matrix, the model in (1) enables us to devise trading strategies by leveraging the *alpha* (i.e., mispricing) and *beta* (i.e., risk) roles of characteristics. Specifically, the initial step of the regressed-PCA method (i.e., Fama-MacBeth regressions) reduces a large number of individual stock returns into a smaller set of characteristic-managed portfolios. These portfolios exhibit a classical factor structure as outlined in (9), facilitating the construction of a pure-*alpha* strategy and an MVE factor portfolio.

By (9) and Theorem 4.1, $\hat{a}'\tilde{Y}_t \xrightarrow{p} \|a\|^2$ for each t as $N \to \infty$, implying that $\hat{a}'\tilde{Y}_t$ represents a portfolio with positive returns (if $a \neq 0$) and no risk asymptotically. Using

the expanding-window procedure in Section 5.1, we construct a pure-*alpha* "arbitrage" portfolio for $t \geq 120$ as follows:

$$R_{\alpha,t} = \hat{a}_{t-1}' \tilde{Y}_t = \hat{a}_{t-1}' (\Phi(Z_{t-1})' \Phi(Z_{t-1}))^{-1} \Phi(Z_{t-1})' Y_t. \tag{27}$$

This portfolio can be equivalently constructed from individual stocks by using weights $\Phi(Z_{t-1})(\Phi(Z_{t-1})'\Phi(Z_{t-1}))^{-1}\hat{a}_{t-1}$. Since $R_{\alpha,t}$ relies on estimates obtained at time period $t-1$, it is tradable (ex ante), while $\hat{a}'\tilde{Y}_t$ is based on full-sample estimates. We consider $R_{\alpha,t}$, the out-of-sample version of $\hat{a}'\tilde{Y}_t$, and refer to their Sharpe ratios as the out-of-sample and in-sample Sharpe ratios of the pure-*alpha* portfolio, respectively.

Similarly, we construct the out-of-sample version of $\hat{f}_t = \hat{B}'\tilde{Y}_t$ as $\hat{f}_{t-1,t} = \hat{B}_{t-1}'\tilde{Y}_t$, which has been introduced in (26). This enables the construction of an MVE factor portfolio based on the same expanding-window procedure: for $t \geq 120$,

$$R_{\beta,t} = \hat{\mu}_{t-1}' \hat{\Sigma}_{t-1}^{-1} \hat{f}_{t-1,t} = \hat{\mu}_{t-1}' \hat{\Sigma}_{t-1} \hat{B}_{t-1}' (\Phi(Z_{t-1})' \Phi(Z_{t-1}))^{-1} \Phi'(Z_{t-1}) Y_t, \tag{28}$$

where $\hat{\mu}_{t-1}$ and $\hat{\Sigma}_{t-1}$ are estimates of the (conditional) mean and covariance matrix of f_t at time period $t-1$.[19] Weights for $R_{\beta,t}$ can also be derived from and applied on individual stocks, making it tradable. The in-sample counterpart of $R_{\beta,t}$ is $\hat{\mu}'\hat{\Sigma}^{-1}\hat{f}_t$, where $\hat{\mu}$ and $\hat{\Sigma}$ are the full-sample estimates of the (unconditional) mean and covariance matrix of f_t.[20] Their Sharpe ratios are referred to as the out-of-sample and in-sample Sharpe ratios of the MVE factor portfolio, respectively.

We further combine $R_{\alpha,t}$ and $\hat{f}_{t-1,t}$ to form a set of $K+1$ factor portfolios, constructing an MVE portfolio following the procedure outlined for $R_{\beta,t}$. The in-sample counterpart is derived from $\hat{a}'\tilde{Y}_t$ and $\hat{f}_t$. The resulting Sharpe ratios are referred to as the out-of-sample and in-sample Sharpe ratios of the combined MVE portfolio, respectively. By imposing a factor structure on the conditional covariance of individual asset returns as in (1), the combined MVE portfolio provides an approximation to the stock market's MVE portfolio. Tables IV and V present the annualized in-sample and out-of-sample Sharpe ratios for the pure-*alpha*, MVE factor, and combined MVE portfolios. In all subsequent tables, "Regressed-PCA" denotes the results under linear specifications of $\alpha(\cdot)$ and $\beta(\cdot)$ with 36 characteristics, while "Regressed-PCA S1" and "Regressed-PCA S2" correspond to nonlinear specifications with 18 and 12 characteristics, respectively. "IPCA" represents the results based on the linear specification using IPCA estimations

[19] Specifically, $\hat{\mu}_{t-1} = \sum_{s \leq t-1} \hat{f}_s^{(t-1)}/(t-1)$ and $\hat{\Sigma}_{t-1} = \sum_{s \leq t-1} (\hat{f}_s^{(t-1)} - \hat{\mu}_{t-1})(\hat{f}_s^{(t-1)} - \hat{\mu}_{t-1})'/(t-2)$.
[20] Specifically, $\hat{\mu} = \sum_{s \leq T} \hat{f}_s/T$ and $\hat{\Sigma} = \sum_{s \leq T} (\hat{f}_s - \hat{\mu})(\hat{f}_s - \hat{\mu})'/(T-1)$.

with the same 36 characteristics.

Two remarks are essential for understanding Tables IV and V. First, the in-sample Sharpe ratio of our MVE factor portfolio increases with K, whereas that of IPCA's MVE factor portfolio may not. This reflects the stability of our factor construction approach compared to IPCA, in particular the inherent orthogonality of regressed PCA factors (see Section 2.3). We report the out-of sample Sharpe ratio of each incremental regressed PCA factor in Table V along with the pure-*alpha* and MVE portfolios, but we do not report individual Sharpe ratios for IPCA factors as they are not incremental, since all of the factors are estimated jointly for each K. Second, for our regressed-PCA approach, the squared in-sample Sharpe ratio of the combined MVE portfolio equals the sum of those of the pure-*alpha* and MVE factor portfolios as the two are orthogonal by construction.[21] As a result, the in-sample Sharpe ratios of the pure-*alpha* and MVE factor portfolios are identical to their respective contributions to the combined MVE portfolio and are omitted in Table IV. Similarly, the out-of-sample Sharpe ratio of our MVE factor portfolio equals its contribution to the combined MVE portfolio, which is also omitted in Table V. The orthogonality property does not hold for IPCA, highlighting another advantage of regressed-PCA.

The main findings are summarized as follows. First, under the linear specification, we compare the Sharpe ratios of the pure-*alpha* and MVE factor portfolios constructed based on regressed-PCA and IPCA, separately. As K increases, the Sharpe ratio of our pure-*alpha* portfolio remains high (in-sample: 3.89 to 4.50; out-of-sample: 3.18 to 3.84), while that of the MVE factor portfolio is comparatively low (in-sample: 0.65 to 0.89; out-of-sample: 0.44 to 0.72). The observed increase in the Sharpe ratio of the pure-*alpha* portfolio as K grows suggests that factors play a crucial role in hedging common variation in stock returns, reducing the volatility of the pure-*alpha* portfolio at a rate exceeding the decline in alphas. The findings also align with the testing evidence of nonzero pricing errors in the linear specification found in Section 5.2.1, reinforcing our conclusions. While the Sharpe ratios of IPCA's pure-*alpha* and MVE factor portfolios are more comparable (in-sample: 1.61 to 3.13 vs. 1.07 to 2.85; out-of-sample: 1.31 to 2.84 vs. 0.92 to 1.72), the former is higher than the latter. Table V also shows that the higher Sharpe ratio of our pure-*alpha* portfolio compared with IPCA's is due to its low volatility, consistent with the idea of no risk asymptotically. The lower volatility of our pure-*alpha* portfolio underscores the superior hedging properties of our factors. Notably,

[21] This follows from $\sum_{t=1}^{T} \hat{a}'(\tilde{Y}_t - \bar{\tilde{Y}})\hat{f}_t/T = \hat{B}'\tilde{Y}M_T\tilde{Y}'\hat{a}/T = \hat{\Lambda}\hat{B}'\hat{a} = 0$, where $\hat{\Lambda}$ is a diagonal matrix with diagonal entries being the large K eigenvalues of $\tilde{Y}M_T\tilde{Y}'/T$.

regressed-PCA consistently yields higher in-sample and out-of-sample Sharpe ratios for the combined MVE portfolio than IPCA across all K. This indicates that the combined MVE portfolio constructed with regressed-PCA better approximates the stock market's MVE portfolio compared to IPCA. In particular, the high Sharpe ratios of the combined MVE portfolio primarily arise from the high Sharpe ratios of the pure-*alpha* portfolio. The better approximation is also attributed to the stability and orthogonality properties of our regressed-PCA as discussed above.

Second, the nonlinear specifications yield higher in-sample and out-of-sample Sharpe ratios for the MVE factor and combined MVE portfolios than the linear specification across all values of K (except $K = 1$ in Regressed-PCA S2 of Table IV). As K increases, the Sharpe ratio of the pure-*alpha* portfolio under the nonlinear specifications falls below that under the linear specification, indicating that nonlinear models yield smaller magnitudes of pricing errors. Meanwhile, the nonlinear specifications also yield better factors, reflected in higher Sharpe ratios of the MVE factor portfolio (in-sample: 0.61 to 3.96; out-of-sample: 0.51 to 3.33) than the linear specification (in-sample: 0.65 to 0.89; out-of-sample: 0.44 to 0.72). More importantly, the Sharpe ratios of the combined MVE portfolio from the nonlinear specifications are substantially higher than those from the linear specification, implying the potential nonlinearity of stochastic discount factor in the U.S. stock market. Similarly, the high Sharpe ratios of our combined MVE portfolio are primarily derived by the pure-*alpha* portfolio for $K \leq 5$, the Sharpe ratios of the MVE factor portfolio catch up and become comparable for larger K. Nevertheless, the Sharpe ratios of our pure-*alpha* portfolio remain consistently high (in-sample: 3.47 to 4.80; out-of-sample: 3.09 to 4.26) compared with those of the MVE factor portfolio (in-sample: 0.61 to 3.96; out-of-sample: 0.51 to 3.33), providing strong evidence of nonzero pricing errors. This finding corroborates the evidence of nonzero pricing errors presented in Section 5.2.1. All these findings align with the strong evidence of nonlinearity found in Section 5.2.1 and underscore the advantage of the flexible nonlinear specifications.

Lastly, we perform a subsample analysis of the pure-*alpha* strategy, using five-year intervals starting in January 1970. Within each subsample, we construct the pure-*alpha* "arbitrage" portfolio defined in (27), employing expanding window estimation from the second year onward. The key results are presented in Figure 4. Notably, the left panel of Figure 4 displays the decline in Sharpe ratios of the portfolio, and the right panel indicates that the decline is primarily driven by a reduction in the portfolioâĂŹs average returns, rather than an increase in its standard deviations. This is consistent with the findings in Figure 3, where we observe a significant decline in pricing errors since 2000.

In summary, we provide further evidence supporting the findings of nonlinearity and nonzero pricing errors, as well as their significant decline over time in Section 5.2.1. This section also reinforces the advantages of regressed-PCA. It is important to note that while nonzero pricing errors are observed, they do not necessarily imply market inefficiency unless the possibility of model misspecification is ruled out.

5.2.3 Asset Pricing Tests

We now evaluate the performance of our factors in asset pricing tests, considering a broad class of testing portfolios. Specifically, we examine three groups of Fama-MacBeth managed portfolios: Regressed-PCA, Regressed-PCA S1, and Regressed-PCA S2, as well as IPCA's managed portfolios. Additionally, we include two groups of single sorted portfolios based on 55 characteristics from Haddad et al. (2020) and our 36 characteristics, along with several groups of double sorted portfolios following Fama and French (2020). Table F.XIII reports the bilateral correlations and standard deviations of these portfolios. The Fama-MacBeth managed portfolios exhibit lower bilateral correlations than others, and smaller standard deviations compared to the sorted portfolios. This supports the optimality of the Fama-MacBeth managed portfolios, which are maximally diversified, as discussed in Section 3.

We compare our factors with several existing sets: IPCA's factors, the five factors from Fama and French (2015) (denoted as FF5), and the factors constructed following Kozak et al. (2018) (denoted as KNS). The comparison statistics from time series regressions are reported following the analysis in Fama and French (2020). The results for five groups of testing portfolios with $K = 5$ are presented in Tables VI and VII, with additional results provided in Appendix F (see Tables F.XIV-F.XXIII).

The main findings are summarized as follows. First, for the Fama-MacBeth managed portfolios in Group I, both our factors and IPCA's factors outperform FF5 and KNS's factors in terms of average absolute intercepts ($A|a|$). They also achieve larger average regression R^2's (AR^2), leading to smaller average standard errors ($As(a)$). The resulting average absolute t-statistics ($A|t(a)|$) and GRS statistics (GRS) are comparable across all factors. Notably, our factors under the nonlinear specifications do not improve performance, as the testing portfolios are derived from the linear specification.

Second, our factors consistently outperform IPCA's factors in pricing the sorted portfolios in Groups II, III, V, and VI, as evidenced by smaller average absolute intercepts, t-statistics, and GRS statistics. The higher average absolute t-statistics and GRS statis-

tics for IPCA's factors arise from their larger average regression R^2's (or smaller average residual standard deviations ($As(e)$) or standard errors). To investigate this, we project IPCA's factors onto our factors under the linear specification (without a constant term) and treat the resulting residuals as new factors (denoted as IPCA\Regressed-PCA). These residuals yield even larger average absolute intercepts and substantial average regression R^2's, indicating that IPCA's factors capture more time-series variation in returns but less cross-sectional variation, likely due to overfitting idiosyncratic noise rather than extracting true signals. This overfitting is much less pronounced in Group I, as the Fama-MacBeth managed portfolios are maximally diversified, unlike the sorted portfolios, as discussed in Section 3 and shown in Table F.XIII.

Third, our factors under the nonlinear specifications significantly reduce average absolute intercepts and t-statistics for sorted portfolios. This improvement stems from the nonparametric nature of sorting, as discussed in Section 3.1. However, due to the high correlations among sorted portfolios, as shown in Table F.XIII, no noticeable improvement in GRS statistics is observed. Moreover, factors under the nonlinear specifications with 12 characteristics outperform FF5, yielding smaller average absolute t-statistics and GRS statistics, with comparable average absolute intercepts. The higher t-statistics and GRS statistics for FF5 also arise from their larger regression R^2's, suggesting the presence of unpriced components similar to IPCA's factors (Daniel et al., 2020; Kozak and Nagel, 2023). Moreover, our factors also outperform KNS's factors, achieving smaller average absolute intercepts and t-statistics.

Lastly, for IPCA's managed portfolios in Group IV, IPCA's factors exhibit inferior performance compared to other factors, with larger average absolute intercepts, t-statistics, and GRS statistics. This finding is surprising given that IPCA factors are derived from its managed portfolios. As the testing portfolios are derived under the linear specification, our factors under the linear specification outperform those under the nonlinear specifications. The performance of our factors under the linear specification is comparable to FF5 and KNS's factors. In addition, the performance comparisons based on relative metrics (e.g., $Aa^2/V\overline{r}$ or $A\lambda^2/V\overline{r}$) align with those based on absolute metrics ($A|a|$), reinforcing the robustness of our results.

In summary, our factors demonstrate superior performance compared to IPCA's factors and long-short factors in asset pricing tests, with robustness across a wide range of testing portfolios.

6 Conclusion

In this paper, we considered semiparametric conditional latent factor models to address the "characteristics versus covariances" debate and the "*factor zoo*" problem in cross-sectional asset pricing. We proposed a simple and tractable sieve estimation approach combined with a weighted-bootstrap procedure for conducting inference on the *alpha* and *beta* functions. We established large-sample properties of the estimators and validity of the tests under large N, even when T is small. In addition to offering formal inference procedures and well-founded asymptotic properties, our approach presents several advantages over existing methods such as IPCA and projected-PCA. Specifically, it is computationally efficient and accommodates nonzero alphas, time-varying characteristics, unbalanced panels, and short samples, making it particularly suitable for empirical asset pricing applications. These results enable the estimation of conditional factor structures for a large set of individual assets by incorporating numerous characteristics, accounting for nonlinearity without requiring pre-specified factors. Moreover, our approach disentangles the role of risk from the purely predictive power of return characteristics that is unrelated to common risk exposures.

We applied this method to analyze the cross-sectional differences in individual stock returns in the U.S. market. The findings provide robust evidence of large nonzero pricing errors and nonlinearity in both *alpha* and *beta* functions, leading to the formation of "arbitrage" portfolios with exceptionally high Sharpe ratios (exceeding 3). Additionally, we documented a significant decline in pricing errors since 2000. Our method delivers stable and reliable factor construction without the risk of overfitting, yielding out-of-sample mean-variance efficient portfolios with Sharpe ratios in excess of 4. We also demonstrated that our factors outperform existing alternatives in explaining the cross-section of U.S. stock returns.

Table I. Results under linear specifications of $\alpha(\cdot)$ and $\beta(\cdot)$ with 36 characteristics[†]

Unrestricted ($\alpha(\cdot) \neq 0$)										
K	R^2_K	R^2	$R^2_{T,N}$	$R^2_{N,T}$	R^2_f	$R^2_{f,T,N}$	$R^2_{f,N,T}$	$R^2_{f,O}$	$R^2_{f,T,N,O}$	$R^2_{f,N,T,O}$
1	26.55	2.54	1.37	0.36	2.07	0.59	0.11	6.23	3.79	5.65
2	36.42	4.52	2.43	1.76	4.08	1.75	1.37	13.59	10.63	11.28
3	45.03	5.70	3.70	2.70	5.24	2.95	2.31	14.09	11.10	11.67
4	52.55	11.69	8.55	9.27	11.28	7.92	8.69	14.74	12.15	12.11
5	58.65	11.90	8.73	9.48	11.49	7.99	8.90	15.17	12.90	12.42
6	64.20	13.90	10.30	11.80	13.53	9.79	11.24	15.38	13.19	12.63
7	69.15	15.59	12.23	13.76	15.23	11.71	13.23	15.62	13.32	12.87
8	72.84	15.93	12.59	13.98	15.56	12.00	13.44	15.90	13.58	13.12
9	76.26	16.08	12.67	14.19	15.72	12.15	13.64	16.13	13.83	13.33
10	79.15	16.23	12.82	14.35	15.87	12.34	13.80	16.29	14.06	13.47
K	$R^2_{\tilde{Y}}$	R^2_O	$R^2_{T,N,O}$	$R^2_{N,T,O}$	p_α	p_{lin}				
1-10	20.89	0.54	0.64	0.21	$< 1\%$	$< 1\%$				

[†] K: the number of factors specified; $R^2_{\tilde{Y}}$: Fama-MacBeth cross-sectional regression R^2 (%); R^2_K: the variation of the Fama-MacBeth managed portfolios $\tilde{Y}_t$ captured by the extracted factors $\hat{f}_t$ (%); R^2, $R^2_{T,N}$, $R^2_{N,T}$: various in-sample R^2's (%), see (22)-(24); R^2_f, $R^2_{f,T,N}$, $R^2_{f,N,T}$: various in-sample R^2's without $\alpha(\cdot)$ (%), see (F.1)-(F.3); $R^2_{f,O}$, $R^2_{f,T,N,O}$, $R^2_{f,N,T,O}$: various out-of-sample fit R^2's (%), see (F.7)-(F.9); R^2_O, $R^2_{T,N,O}$, $R^2_{N,T,O}$: various out-of-sample predictive R^2's (%), see (F.4)-(F.6); p_α and p_{lin}: the p-values of *alpha* test ($\alpha(\cdot) = 0$) and model specification test (joint linearity of $\alpha(\cdot)$ and $\beta(\cdot)$), respectively.

Table II. Results under nonlinear specifications of $\alpha(\cdot)$ and $\beta(\cdot)$ with 18 characteristics[†]

Unrestricted ($\alpha(\cdot) \neq 0$)										
K	R^2_K	R^2	$R^2_{T,N}$	$R^2_{N,T}$	R^2_f	$R^2_{f,T,N}$	$R^2_{f,N,T}$	$R^2_{f,O}$	$R^2_{f,T,N,O}$	$R^2_{f,N,T,O}$
1	41.61	5.94	3.47	3.60	5.52	2.99	3.11	11.27	7.81	8.93
2	59.05	9.56	6.17	6.91	9.18	5.67	6.33	14.04	11.31	11.29
3	64.47	10.42	6.78	7.96	10.03	6.27	7.38	14.64	11.93	11.95
4	68.99	13.83	10.26	11.52	13.40	9.80	10.90	15.44	12.98	12.54
5	72.33	14.32	10.73	11.98	13.91	10.29	11.38	15.78	13.43	12.89
6	75.35	14.71	10.97	12.40	14.29	10.55	11.86	16.20	14.16	13.18
7	77.63	15.28	11.78	12.99	14.84	11.27	12.42	16.45	14.34	13.37
8	80.83	15.44	11.98	13.16	15.10	11.59	12.73	16.59	14.50	13.52
9	82.88	15.84	12.33	13.49	15.48	11.87	13.05	16.86	14.69	13.81
10	85.61	16.39	12.89	13.93	15.71	11.80	13.14	16.98	14.72	13.86
K	$R^2_{\tilde{Y}}$	R^2_O	$R^2_{T,N,O}$	$R^2_{N,T,O}$	p_α	p_{lin}				
1-10	21.11	0.59	0.64	0.28	$< 1\%$	$< 1\%$				

[†] K: the number of factors specified; $R^2_{\tilde{Y}}$: Fama-MacBeth cross-sectional regression R^2 (%); R^2_K: the variation of the Fama-MacBeth managed portfolios $\tilde{Y}_t$ captured by the extracted factors $\hat{f}_t$ (%); R^2, $R^2_{T,N}$, $R^2_{N,T}$: various in-sample R^2's (%), see (22)-(24); R^2_f, $R^2_{f,T,N}$, $R^2_{f,N,T}$: various in-sample R^2's without $\alpha(\cdot)$ (%), see (F.1)-(F.3); $R^2_{f,O}$, $R^2_{f,T,N,O}$, $R^2_{f,N,T,O}$: various out-of-sample fit R^2's (%), see (F.7)-(F.9); R^2_O, $R^2_{T,N,O}$, $R^2_{N,T,O}$: various out-of-sample predictive R^2's (%), see (F.4)-(F.6); p_α and p_{lin}: the p-values of *alpha* test ($\alpha(\cdot) = 0$) and model specification test (joint linearity of $\alpha(\cdot)$ and $\beta(\cdot)$), respectively.

Table III. Results under nonlinear 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	$R^2_{f,T,N}$	$R^2_{f,N,T}$	$R^2_{f,O}$	$R^2_{f,T,N,O}$	$R^2_{f,N,T,O}$
1	42.78	5.57	2.98	3.32	5.19	2.54	2.83	11.08	7.57	8.77
2	61.36	9.56	5.97	6.87	9.18	5.51	6.26	13.85	11.12	10.99
3	67.77	10.59	6.65	7.88	10.20	6.15	7.29	14.66	12.25	11.83
4	72.86	13.62	10.09	11.35	13.17	9.64	10.67	15.39	13.53	12.53
5	76.92	14.14	10.43	12.01	13.73	10.01	11.48	15.82	13.94	12.90
6	80.63	14.94	11.45	12.75	14.42	10.51	12.05	16.16	14.20	13.22
7	84.29	15.17	11.59	12.94	14.76	10.77	12.45	16.57	14.59	13.57
8	87.42	15.45	11.87	13.23	15.26	11.47	12.98	16.94	14.83	13.88
9	89.11	16.33	12.68	13.94	16.16	12.31	13.72	17.12	15.00	14.09
10	90.72	16.54	12.91	14.17	16.38	12.54	13.95	17.30	15.19	14.29
K	$R^2_{\tilde{Y}}$	R^2_O	$R^2_{T,N,O}$	$R^2_{N,T,O}$	p_α	p_{lin}				
1-10	20.72	0.57	0.57	0.27	$< 1\%$	$< 1\%$				

[†] K: the number of factors specified; $R^2_{\tilde{Y}}$: Fama-MacBeth cross-sectional regression R^2 (%); R^2_K: the variation of the Fama-MacBeth managed portfolios $\tilde{Y}_t$ captured by the extracted factors $\hat{f}_t$ (%); R^2, $R^2_{T,N}$, $R^2_{N,T}$: various in-sample R^2's (%), see (22)-(24); R^2_f, $R^2_{f,T,N}$, $R^2_{f,N,T}$: various in-sample R^2's without $\alpha(\cdot)$ (%), see (F.1)-(F.3); $R^2_{f,O}$, $R^2_{f,T,N,O}$, $R^2_{f,N,T,O}$: various out-of-sample fit R^2's (%), see (F.7)-(F.9); R^2_O, $R^2_{T,N,O}$, $R^2_{N,T,O}$: various out-of-sample predictive R^2's (%), see (F.4)-(F.6); p_α and p_{lin}: the p-values of *alpha* test ($\alpha(\cdot) = 0$) and model specification test (joint linearity of $\alpha(\cdot)$ and $\beta(\cdot)$), respectively.

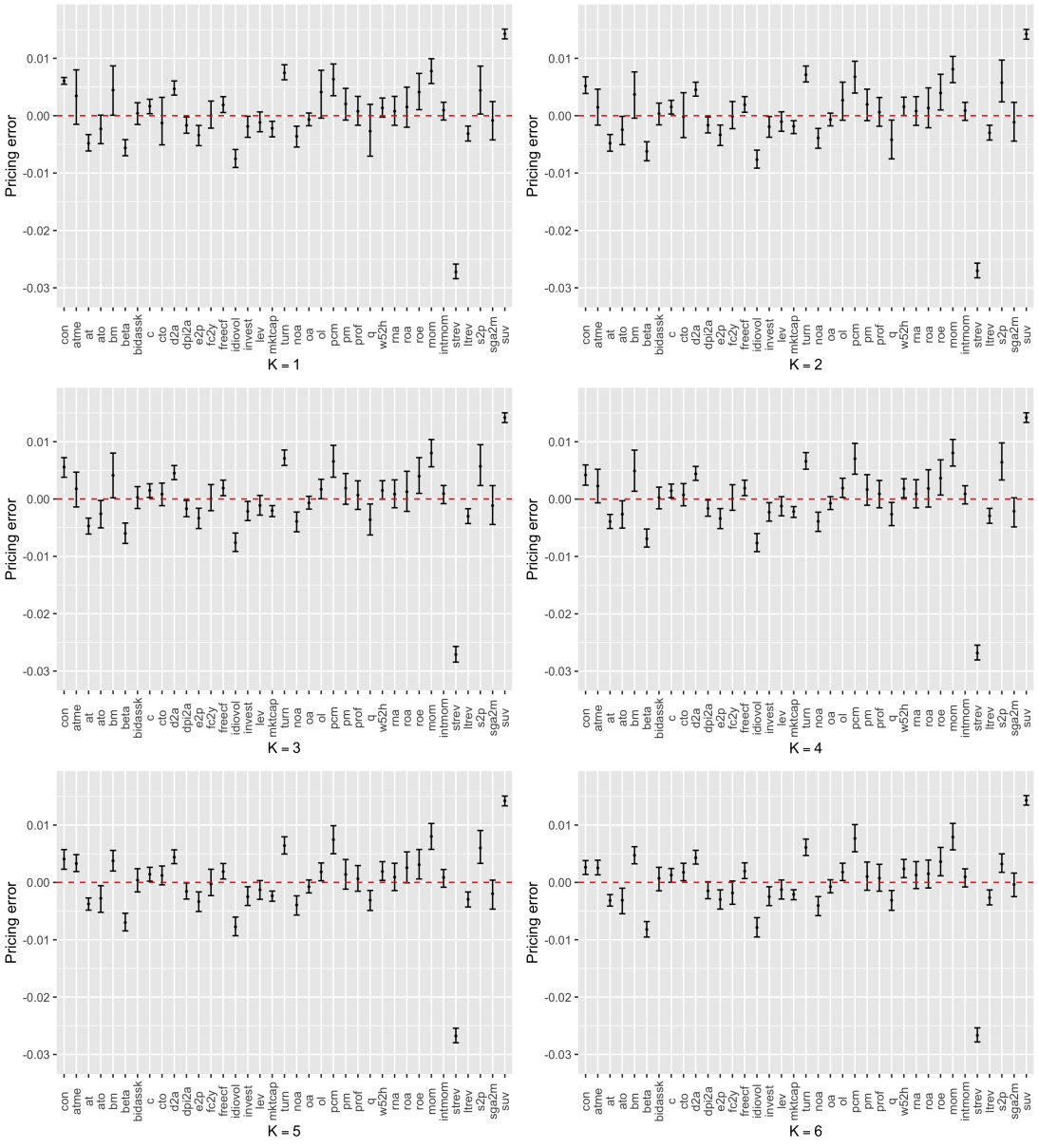


Figure 1. 95% confidence intervals for coefficients in $\alpha(\cdot)$ under linear specifications of $\alpha(\cdot)$ and $\beta(\cdot)$ with 36 characteristics

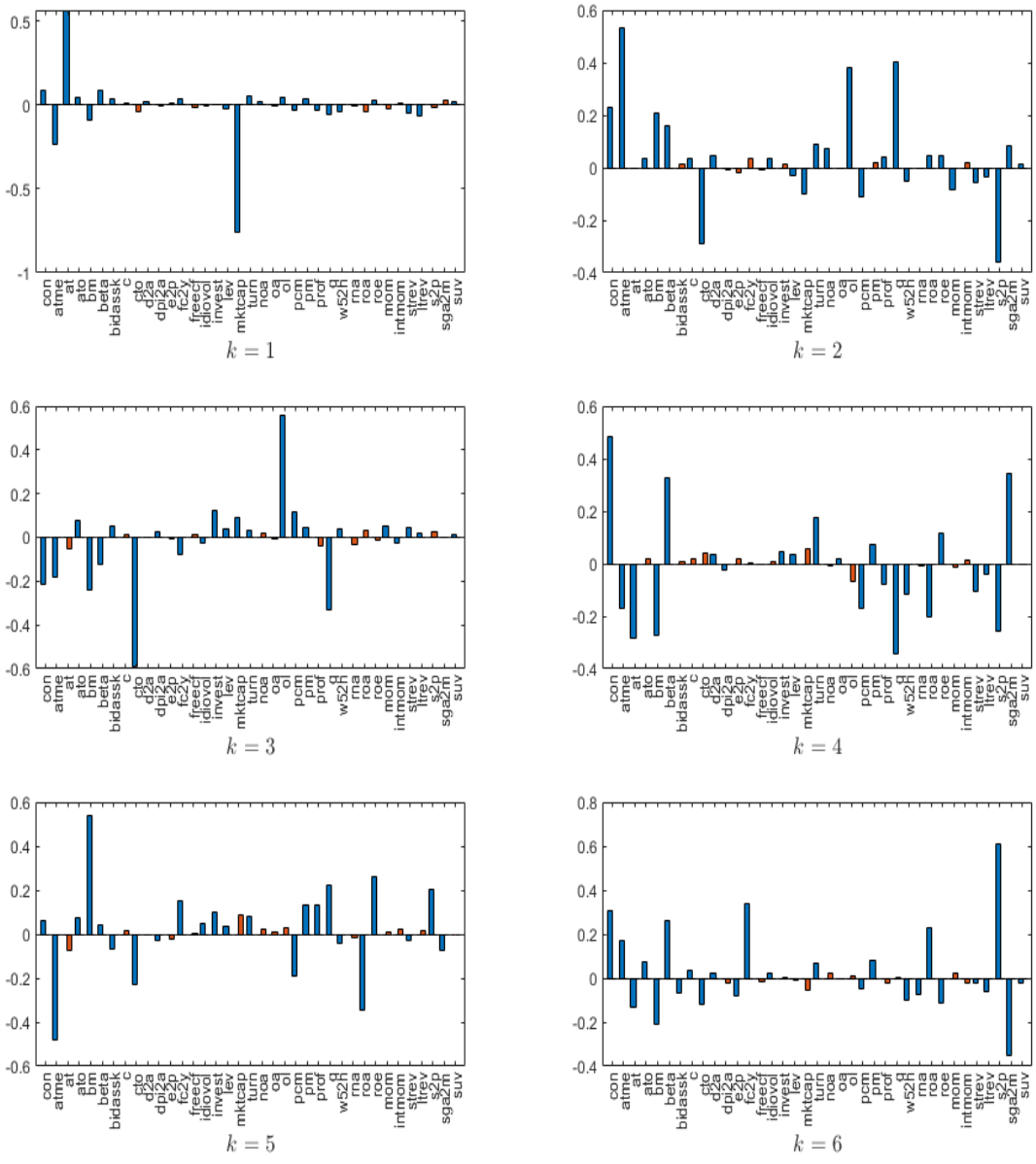


Figure 2. Estimates of coefficients in $\beta(\cdot)$ under linear specifications of $\alpha(\cdot)$ and $\beta(\cdot)$ with 36 characteristics (blue: significant at the 5% level; red: insignificant)

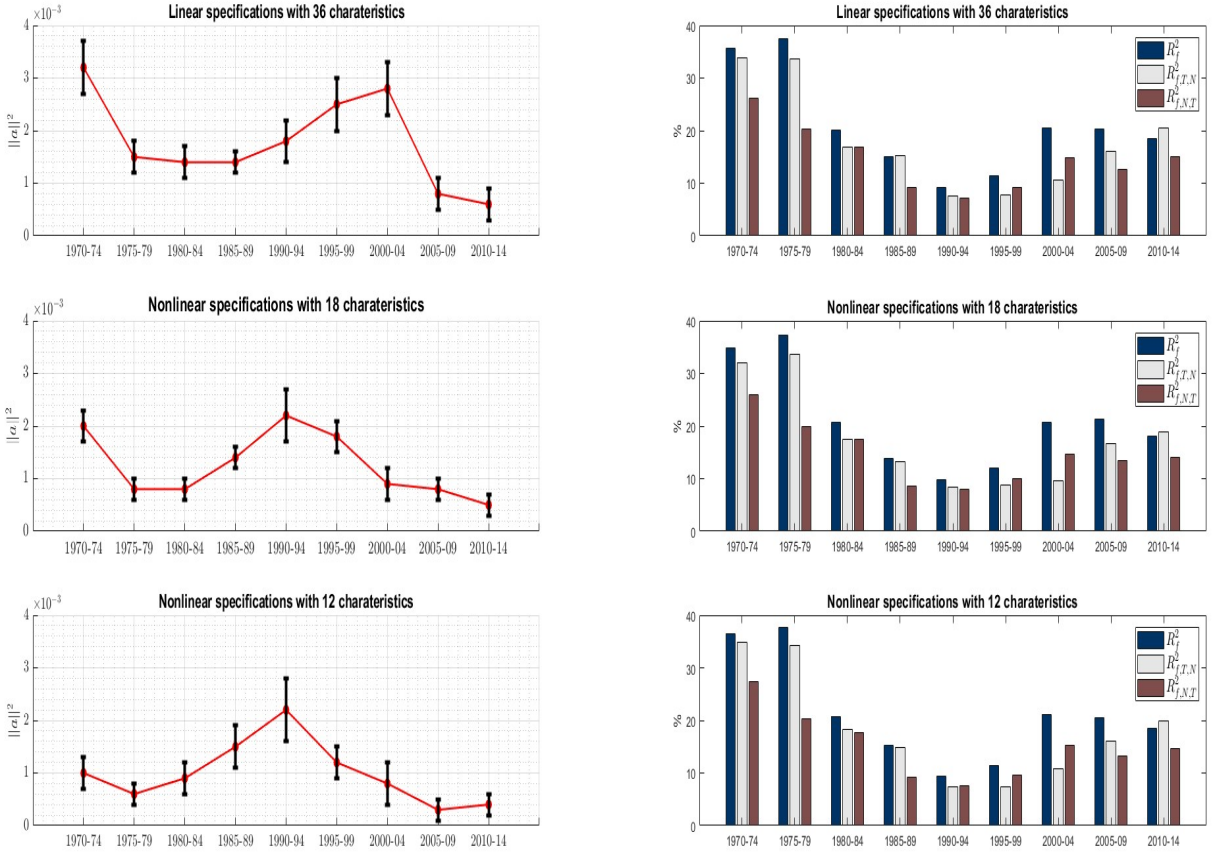


Figure 3. 95% confidence intervals for $\|a\|^2$, R^2_f, $R^2_{f,T,N}$, and $R^2_{f,N,T}$ ((F.1)-(F.3)) with $K = 10$: subsample analysis

Table IV. In-sample Sharpe ratios †

	Regressed-PCA			Regressed-PCA S1			Regressed-PCA S2			IPCA				
K	SR_α	SR_f	SR_M	SR_α	SR_f	SR_M	SR_α	SR_f	SR_M	SR_α	SR_f	SR_M	$SR_{M,\alpha}$	$SR_{M,f}$
1	3.89	0.65	3.94	4.31	0.61	4.35	3.67	0.66	3.73	1.61	1.07	2.10	1.61	1.07
2	4.00	0.67	4.06	4.76	0.68	4.81	4.31	0.79	4.39	1.98	1.36	2.22	1.98	1.36
3	4.02	0.67	4.08	4.78	0.73	4.84	4.32	0.80	4.39	2.64	1.07	2.85	2.64	1.05
4	4.07	0.68	4.13	4.80	0.87	4.88	4.32	0.89	4.41	3.13	1.07	3.32	3.13	1.03
5	4.10	0.69	4.15	4.75	1.38	4.95	4.28	1.24	4.45	3.00	1.11	3.21	3.00	1.07
6	4.19	0.72	4.25	4.71	1.63	4.98	3.90	2.25	4.51	2.57	1.99	3.20	2.57	1.97
7	4.37	0.80	4.44	4.69	1.69	4.99	3.47	3.07	4.63	2.74	2.09	3.41	2.74	2.08
8	4.48	0.88	4.56	4.16	2.95	5.10	3.72	3.74	5.28	2.50	2.85	3.69	2.50	2.84
9	4.49	0.89	4.58	4.07	3.11	5.12	3.74	3.75	5.29	2.41	2.77	3.56	2.41	2.76
10	4.50	0.89	4.58	3.78	3.96	5.48	3.78	3.76	5.34	2.40	2.85	3.62	2.40	2.84

† K: the number of factors specified; SR_α: annualized Sharpe ratios of $\hat{a}'\tilde{Y}_t$; SR_f: annualized Sharpe ratios of $\hat{\mu}'\hat{\Sigma}\hat{f}_t$; SR_M: annualized Sharpe ratios of the combined MVE portfolios on $\hat{a}'\tilde{Y}_t$ and $\hat{f}_t$; $SR_{M,\alpha}$: annualized Sharpe ratios of the component from $\hat{a}'\tilde{Y}_t$ in the combined MVE portfolios; $SR_{M,f}$: annualized Sharpe ratios of the component from $\hat{f}_t$ in the combined MVE portfolios.

Table V. Out-of-sample Sharpe ratios †

	Regressed-PCA							IPCA						
K	Mean	Std	SR_α	$SR_{f,K}$	SR_f	SR_M	$SR_{M,\alpha}$	Mean	Std	SR_α	SR_f	SR_M	$SR_{M,\alpha}$	$SR_{M,f}$
1	1.72	0.54	3.18	0.61	0.61	3.25	3.24	2.70	2.07	1.31	1.26	1.80	1.33	1.26
2	1.74	0.52	3.36	-0.12	0.55	3.39	3.38	2.38	1.41	1.69	1.44	2.20	1.87	1.54
3	1.77	0.50	3.56	-0.34	0.46	3.55	3.56	2.01	0.93	2.16	1.14	2.33	2.20	1.17
4	1.77	0.47	3.74	0.02	0.44	3.67	3.68	1.85	0.68	2.70	0.92	2.71	2.74	0.93
5	1.70	0.44	3.84	0.42	0.53	3.81	3.78	1.75	0.62	2.84	0.98	2.79	2.84	0.99
6	1.68	0.44	3.78	0.23	0.57	3.80	3.76	1.39	0.52	2.68	1.34	2.70	2.66	1.31
7	1.63	0.44	3.73	0.60	0.68	3.76	3.69	1.33	0.52	2.57	1.42	2.61	2.52	1.34
8	1.61	0.42	3.79	0.24	0.72	3.80	3.74	1.22	0.50	2.44	1.49	2.67	2.44	1.43
9	1.61	0.42	3.80	-0.06	0.69	3.83	3.76	1.23	0.49	2.51	1.53	2.73	2.52	1.47
10	1.60	0.42	3.82	0.11	0.67	3.87	3.79	1.19	0.48	2.49	1.72	2.74	2.45	1.66
	Regressed-PCA S1							Regressed-PCA S2						
K	Mean	Std	SR_α	$SR_{f,K}$	SR_f	SR_M	$SR_{M,\alpha}$	Mean	Std	SR_α	$SR_{f,K}$	SR_f	SR_M	$SR_{M,\alpha}$
1	2.46	0.69	3.54	0.51	0.51	3.57	3.52	3.29	0.99	3.33	0.54	0.54	3.27	3.24
2	2.39	0.57	4.22	0.18	0.53	4.16	4.13	3.01	0.80	3.78	0.47	0.70	3.77	3.72
3	2.36	0.57	4.17	0.45	0.64	4.13	4.08	2.94	0.80	3.69	0.51	0.78	3.75	3.64
4	2.19	0.53	4.12	0.85	1.04	4.21	4.08	2.97	0.78	3.81	-0.18	0.59	3.84	3.77
5	2.19	0.51	4.26	-0.03	0.93	4.29	4.21	2.98	0.76	3.91	-0.04	0.56	3.90	3.87
6	1.95	0.49	3.96	1.23	1.62	4.34	4.10	1.51	0.41	3.73	2.47	2.55	4.08	3.74
7	1.90	0.48	3.93	0.45	1.66	4.36	4.11	1.01	0.33	3.09	1.87	3.20	4.02	3.26
8	1.73	0.47	3.66	1.11	1.99	4.37	3.92	0.73	0.22	3.36	1.25	3.29	4.41	3.60
9	1.31	0.40	3.26	1.81	2.80	4.30	3.49	0.71	0.20	3.62	0.28	3.24	4.54	3.74
10	0.88	0.28	3.14	1.72	3.33	4.47	3.50	0.69	0.18	3.90	0.23	3.19	4.64	3.90

† K: the number of factors specified; Mean: annualized means of the pure-*alpha* portfolios $R_{\alpha,t}$ in (27) (%); Std: annualized standard deviations of $R_{\alpha,t}$ (%); SR_α: annualized Sharpe ratios of $R_{\alpha,t}$; $SR_{f,K}$: annualized Sharpe ratios of the Kth component in $\hat{f}_{t-1,t}$; SR_f: annualized Sharpe ratios of the MVE factor portfolios $R_{\beta,t}$ in (28); SR_M: annualized Sharpe ratios of the combined MVE portfolios on $R_{\alpha,t}$ and $\hat{f}_{t-1,t}$; $SR_{M,\alpha}$: annualized Sharpe ratios of the component from $R_{\alpha,t}$ in the combined MVE portfolios; $SR_{M,f}$: annualized Sharpe ratios of the component from $\hat{f}_{t-1,t}$ in the combined MVE portfolios.

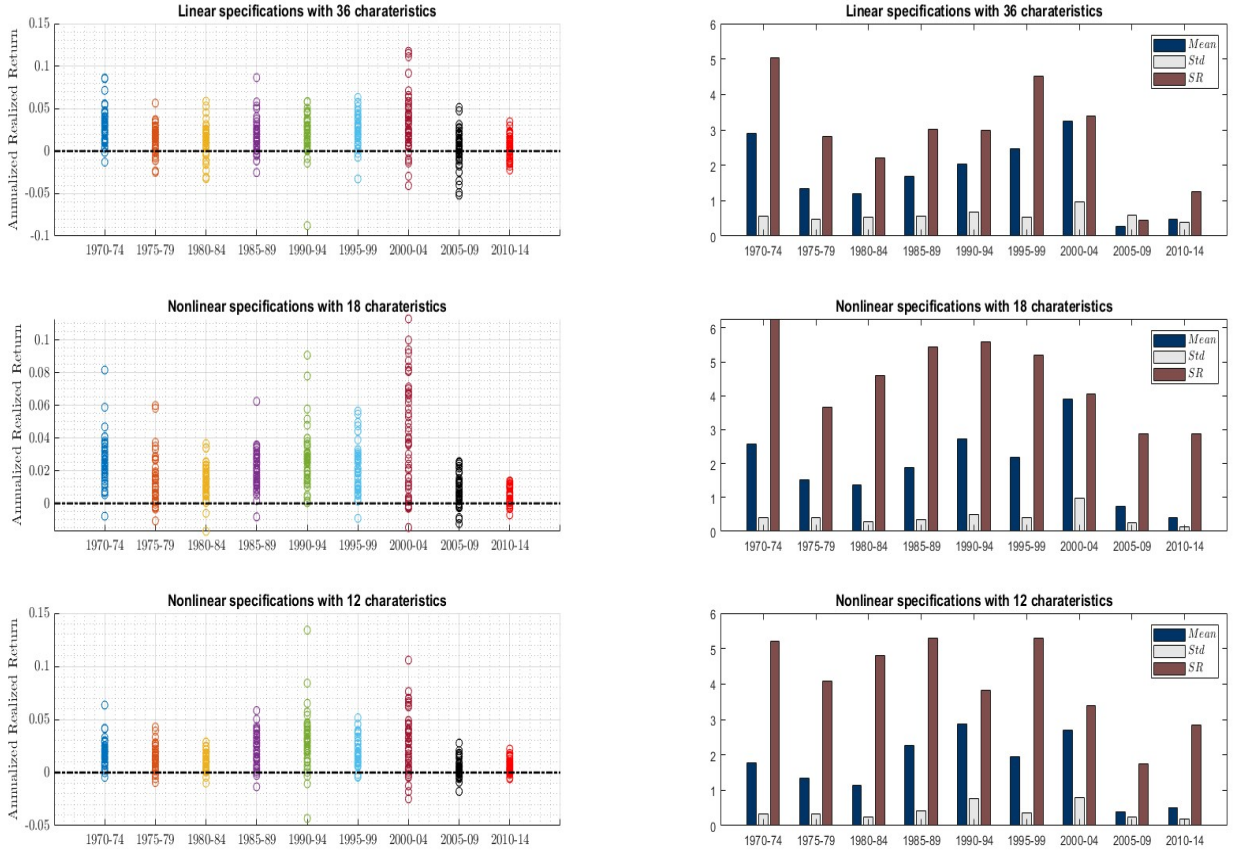


Figure 4. Annualized realized excess returns and Sharpe ratios of the pure-*alpha* portfolio with $K = 10$: subsample analysis

Table VI. Comparing asset pricing tests: $K = 5^{\dagger}$

Testing portfolios/Factors	$A\|a\|$	$A\|t(a)\|$	$Aa^2/V\overline{r}$	$A\lambda^2/V\overline{r}$	$As(a)$	$As(e)$	$Sh^2(a)$	$Sh^2(f)$	AR^2	GRS	$p(GRS)$
Group I: Regressed-PCA's 36 managed portfolios											
Regressed-PCA	0.40	3.40	0.64	0.61	0.13	2.93	2.57	0.04	28.93	34.89	0.00
Regressed-PCA S1	0.40	3.10	0.56	0.50	0.17	3.64	2.48	0.16	17.97	30.23	0.00
Regressed-PCA S2	0.43	3.28	0.59	0.53	0.17	3.73	2.71	0.13	16.08	33.86	0.00
IPCA	0.40	3.15	0.64	0.58	0.17	3.73	3.59	0.10	17.76	45.92	0.00
IPCA\Regressed-PCA	0.50	3.34	1.03	0.96	0.17	3.94	3.49	0.06	12.76	46.40	0.00
FF5	0.49	3.11	0.99	0.92	0.18	4.03	2.62	0.12	8.25	33.16	0.00
KNS	0.50	3.34	1.07	1.01	0.17	3.99	2.56	0.04	9.92	34.89	0.00
Group II: 100 sorted portfolios (double sort on Size and BM, OP, INV, and MOM)											
Regressed-PCA	0.85	5.04	14.14	13.60	0.17	3.97	1.00	0.04	50.81	4.26	0.00
Regressed-PCA S1	0.57	3.21	6.71	6.13	0.18	3.89	1.27	0.16	52.33	4.85	0.00
Regressed-PCA S2	0.44	2.42	4.18	3.59	0.18	4.00	1.35	0.13	49.67	5.32	0.00
IPCA	0.90	11.07	16.08	15.93	0.09	1.96	4.56	0.10	86.75	18.38	0.00
IPCA\Regressed-PCA	1.09	5.62	23.32	22.56	0.20	4.61	1.87	0.06	36.19	7.81	0.00
FF5	0.41	4.81	3.55	3.39	0.09	2.05	1.76	0.12	86.91	6.98	0.00
KNS	0.96	6.25	17.25	16.82	0.16	3.55	0.92	0.04	59.83	3.96	0.00
Group III: 110 sorted portfolios (double sort on Size and Beta, Accruals, NI, and Variance)											
Regressed-PCA	0.85	5.08	16.06	15.43	0.17	3.98	1.25	0.04	50.36	4.73	0.00
Regressed-PCA S1	0.55	3.09	7.42	6.74	0.18	3.94	1.55	0.16	50.83	5.29	0.00
Regressed-PCA S2	0.42	2.32	4.62	3.92	0.18	4.03	1.61	0.13	48.28	5.64	0.00
IPCA	0.88	10.63	17.68	17.50	0.09	2.00	4.56	0.10	85.72	16.31	0.00
IPCA\Regressed-PCA	1.09	5.70	26.13	25.21	0.21	4.66	2.25	0.06	35.61	8.36	0.00
FF5	0.43	4.88	4.22	4.03	0.09	2.08	2.09	0.12	86.06	7.37	0.00
KNS	0.95	6.15	19.41	18.90	0.16	3.59	1.11	0.04	58.32	4.22	0.00

† $A|a|$: average absolute intercept; $A|t(a)|$: average absolute t-statistic for the intercepts; $Aa^2/V\overline{r}$: average squared intercept over the cross-section variance of $\overline{r}$, average returns of the testing portfolios; $A\lambda^2/V\overline{r}$: average difference between each squared intercept and its squared standard error divided by the variance of $\overline{r}$; $As(a)$: average standard error of the intercepts; $As(e)$: average residual standard deviation; $Sh^2(a)$: maximized squared Sharpe ratio for the intercepts; $Sh^2(f)$: maximized squared Sharpe ratio for the factors; AR^2: average regression R^2 (%); GRS: GRS statistic of Gibbons et al. (1989); $p(GRS)$: p-value of GRS.

Table VII. Comparing asset pricing tests: $K = 5$ (continued)[†]

Testing portfolios/Factors	$A\|a\|$	$A\|t(a)\|$	$Aa^2/V\overline{r}$	$A\lambda^2/V\overline{r}$	$As(a)$	$As(e)$	$Sh^2(a)$	$Sh^2(f)$	AR^2	GRS	$p(GRS)$
Group IV: IPCA's 36 managed portfolios											
Regressed-PCA	0.05	3.44	0.87	0.82	0.01	0.33	1.50	0.04	25.22	20.30	0.00
Regressed-PCA S1	0.06	5.12	1.41	1.38	0.01	0.27	1.85	0.16	47.24	22.50	0.00
Regressed-PCA S2	0.06	5.30	1.32	1.29	0.01	0.25	1.83	0.13	51.44	22.90	0.00
IPCA	0.07	6.72	1.53	1.51	0.01	0.21	3.06	0.10	64.08	39.14	0.00
IPCA\Regressed-PCA	0.06	4.53	1.22	1.17	0.01	0.31	2.20	0.06	39.29	29.28	0.00
FF5	0.04	3.32	0.71	0.67	0.01	0.27	1.42	0.12	50.00	17.96	0.00
KNS	0.04	3.87	0.76	0.73	0.01	0.26	1.47	0.04	52.97	20.03	0.00
Group V: P1&10 of sorted portfolios (single sort on 55 characteristics in Kozak et al. (2018), 110 portfolios)											
Regressed-PCA	0.73	4.05	7.24	6.82	0.19	4.27	1.36	0.04	43.38	5.16	0.00
Regressed-PCA S1	0.47	2.61	3.78	3.37	0.19	4.01	1.42	0.16	49.01	4.82	0.00
Regressed-PCA S2	0.37	1.98	2.51	2.10	0.19	4.13	1.44	0.13	45.95	5.02	0.00
IPCA	0.73	6.43	7.71	7.53	0.12	2.66	3.15	0.10	77.14	11.26	0.00
IPCA\Regressed-PCA	0.90	4.68	10.96	10.45	0.21	4.67	1.85	0.06	34.10	6.87	0.00
FF5	0.40	4.50	2.40	2.25	0.11	2.34	2.54	0.12	82.37	8.97	0.00
KNS	0.92	5.77	10.54	10.24	0.16	3.66	1.31	0.04	56.83	5.01	0.00
Group VI: P1&10 of sorted portfolios (single sort on 36 characteristics, 72 portfolios)											
Regressed-PCA	0.69	3.55	5.72	5.28	0.21	4.71	0.81	0.04	41.77	5.13	0.00
Regressed-PCA S1	0.49	2.52	3.25	2.83	0.20	4.36	0.96	0.16	48.78	5.40	0.00
Regressed-PCA S2	0.39	1.89	2.35	1.91	0.21	4.50	1.02	0.13	45.53	5.94	0.00
IPCA	0.70	5.43	6.11	5.90	0.14	3.07	1.84	0.10	74.60	10.92	0.00
IPCA\Regressed-PCA	0.88	4.20	9.17	8.62	0.23	5.13	1.34	0.06	33.26	8.29	0.00
FF5	0.42	4.38	2.48	2.29	0.12	2.73	2.48	0.12	80.46	14.53	0.00
KNS	0.91	5.61	9.35	9.07	0.17	3.79	0.88	0.04	59.94	5.54	0.00

[†] $A|a|$: average absolute intercept; $A|t(a)|$: average absolute t-statistic for the intercepts; $Aa^2/V\overline{r}$: average squared intercept over the cross-section variance of $\overline{r}$, average returns of the testing portfolios; $A\lambda^2/V\overline{r}$: average difference between each squared intercept and its squared standard error divided by the variance of $\overline{r}$; $As(a)$: average standard error of the intercepts; $As(e)$: average residual standard deviation; $Sh^2(a)$: maximized squared Sharpe ratio for the intercepts; $Sh^2(f)$: maximized squared Sharpe ratio for the factors; AR^2: average regression R^2 (%); GRS: GRS statistic of Gibbons et al. (1989); $p(GRS)$: p-value of GRS.

References

Back, K., N. Kapadia, and B. Ostdiek (2013): "Slopes as factors: Characteristic pure plays," Tech. rep.

——— (2015): "Testing factor models on characteristic and covariance pure plays," Tech. rep.

Belloni, A., V. Chernozhukov, D. Chetverikov, and K. Kato (2015): "Some New Asymptotic Theory for Least Squares Series: Pointwise and Uniform Results," *Journal of Econometrics*, 186, 345–366.

Campbell, J. Y., M. Lettau, B. G. Malkiel, and Y. Xu (2001): "Have individual stocks become more volatile? An empirical exploration of idiosyncratic risk," *The Journal of Finance*, 56, 1–43.

——— (2022): "Idiosyncratic Equity Risk Two Decades Later," Tech. rep., NBER.

Cattaneo, M. D., R. K. Crump, M. H. Farrell, and E. Schaumburg (2020): "Characteristic-sorted portfolios: Estimation and inference," *Review of Economics and Statistics*, 102, 531–551.

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

Chen, L., M. Pelger, and J. Zhu (2022): "Deep learning in asset pricing," *Management Science*, forthcoming.

Chen, Q. (2022): "A Unified Framework for Estimation of High-dimensional Conditional Factor Models," Tech. rep., arXiv preprint arXiv:2209.00391.

Cheng, M., Y. Liao, and X. Yang (2023): "Uniform Predictive Inference for Factor Models with Instrumental and Idiosyncratic Betas," *Journal of Econometrics, Forthcoming*.

Cochrane, J. H. (2011): "Presidential address: Discount rates," *The Journal of finance*, 66, 1047–1108.

Connor, G., M. Hagmann, and O. Linton (2012): "Efficient semiparametric estimation of the FamaâĂŞFrench model and extensions," *Econometrica*, 80, 713–754.

Connor, G. and R. A. Korajczyk (1986): "Performance measurement with the arbitrage pricing theory: A new framework for analysis," *Journal of financial economics*, 15, 373–394.

——— (1988): "Risk and return in an equilibrium APT: Application of a new test methodology," *Journal of financial economics*, 21, 255–289.

Connor, G. and O. Linton (2007): "Semiparametric estimation of a characteristic-

based factor model of common stock returns," *Journal of Empirical Finance*, 14, 694–717.

Daniel, K., L. Mota, S. Rottke, and T. Santos (2020): "The cross-section of risk and returns," *The Review of Financial Studies*, 33, 1927–1979.

Daniel, K. and S. Titman (1997): "Evidence on the characteristics of cross sectional variation in stock returns," *Journal of Finance*, 52, 1–33.

Fama, E. and K. French (2008): "Dissecting anomalies," *The Journal of Finance*, 63, 1653–1678.

Fama, E. F. (1976): *Foundations of Finance*, New York: Basic Books.

Fama, E. F. and K. R. French (1993): "Common risk factors in the returns on stocks and bonds," *Journal of financial economics*, 33, 3–56.

——— (2015): "A five-factor asset pricing model," *Journal of financial economics*, 116, 1–22.

——— (2020): "Comparing cross-section and time-series factor models," *The Review of Financial Studies*, 33, 1891–1926.

Fama, E. F. and J. D. MacBeth (1973): "Risk, Return, and Equilibrium: Empirical Tests," *The Journal of Political Economy*, 81, 607–636.

Fan, J., Z. T. Ke, Y. Liao, and A. Neuhierl (2022): "Structural Deep Learning in Conditional Asset Pricing," Tech. rep., Available at SSRN 4117882.

Fan, J., Y. Liao, and W. Wang (2016a): "Projected principal component analysis in factor models," *The Annals of Statistics*, 44, 219–254.

——— (2016b): "Supplement to âĂIJProjected principal component analysisin factor models"," *DOI:10.1214/15-AOS1364SUPP*.

Ferson, W. E. and C. R. Harvey (1999): "Conditioning variables and the cross section of stock returns," *The Journal of Finance*, 54, 1325–1360.

Freyberger, J., A. Neuhierl, and M. Weber (2020): "Dissecting characteristics nonparametrically," *The Review of Financial Studies*, 33, 2326–2377.

Gagliardini, P. and H. Ma (2019): "Extracting statistical factors when betas are time-varying," Tech. rep., Available at SSRN 3426534.

Gagliardini, P., E. Ossola, and O. Scaillet (2016): "Time-varying risk premium in large crossâĂŘsectional equity data sets," *Econometrica*, 84, 985–1046.

——— (2020): "Estimation of large dimensional conditional factor models in finance," in *Handbook of Econometrics*, vol. 7A, chap. 3.

Ge, S., S. Li, and O. B. Linton (2022): "Dynamic Peer Groups of Arbitrage Characteristics," *Journal of Business & Economic Statistics*.

Gibbons, M. R., S. A. Ross, and J. Shanken (1989): "A test of the efficiency of a given portfolio," *Econometrica*, 57, 1121–1152.

Giglio, S. and D. Xiu (2021): "Asset pricing with omitted factors," *Journal of Political Economy*, 129, 1947–1990.

Green, J., J. R. Hand, and X. F. Zhang (2017): "The characteristics that provide independent information about average US monthly stock returns," *The Review of Financial Studies*, 30, 4389–4436.

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

——— (2021): "Autoencoder asset pricing models," *Journal of Econometrics*, 222, 429–450.

Haddad, V., S. Kozak, and S. Santosh (2020): "Factor timing," *The Review of Financial Studies*, 33, 1980–2018.

Harvey, C., Y. Liu, and H. Zhu (2016): "âĂę and the cross-section of expected returns," *The Review of Financial Studies*, 29, 5–68.

Hastie, T., R. Tibshirani, and J. Friedman (2011): *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer Series in Statistics, New York: Springer, 2nd edition ed.

Hoberg, G. and I. Welch (2009): "Optimized vs. sort-based portfolios," Tech. rep., AFA 2010 Atlanta Meetings Paper.

Jacobs, B. I. and K. N. Levy (1988): "Disentangling Equity Return Regularities: New Insights and Investment Opportunities," *Financial Analysts Journal*, 44, 18–43.

Kelly, B. T., S. Pruitt, and Y. Su (2019): "Characteristics are covariances: A unified model of risk and return," *Journal of Financial Economics*, 134, 501–524.

——— (2020): "Instrumented principal component analysis," Tech. rep., Available at SSRN 2983919.

Kim, S., R. A. Korajczyk, and A. Neuhierl (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. and S. Nagel (2023): "When do cross-sectional asset pricing factors span the stochastic discount factor?" Tech. rep., National Bureau of Economic Research.

Kozak, S., S. Nagel, and S. Santosh (2018): "Interpreting factor models," *The Journal of Finance*, 73, 1183–1223.

——— (2020): "Shrinking the cross-section," *Journal of Financial Economics,*, 135,

271–292.

Lettau, M. and M. Pelger (2020): "Factors that fit the time series and cross-section of stock returns," *The Review of Financial Studies*, 33, 2274–2325.

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.

Lewellen, J., S. Nagel, and J. Shanken (2010): "A skeptical appraisal of asset pricing tests," *Journal of Financial Economics*, 96, 175–194.

McLean, R. and J. Pontiff (2016): "Does academic research destroy stock return predictability?" *The Journal of Finance*, 71, 5–32.

Motta, G., C. M. Hafner, and R. Von Sachs (2011): "Locally stationary factor models: Identification and nonparametric estimation," *Econometric Theory*, 27, 1279–1319.

Park, B. U., E. Mammen, W. Härdle, and S. Borak (2009): "Time series modelling with semiparametric factor dynamics," *Journal of the American Statistical Association*, 104, 284–298.

Pelger, M. and R. Xiong (2021): "State-varying factor models of large dimensions," *Journal of Business Economics & Statistics*, Accpted.

Reisman, H. (1992): "Intertemporal Arbitrage Pricing Theory," *The Review of Financial Studies*, 5, 105–122.

Rosenberg, B. and W. McKibben (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.

Su, L. and X. Wang (2017): "On Time-Varying Factor Models: Estimation and Testing," *Journal of Econometrics*, 198, 84–101.

Yu, Y., T. Wang, and R. J. Samworth (2014): "A useful variant of the Davis-Kahan theorem for statisticians," *Biometrika*, 102, 315–323.

Online Appendix to "Semiparametric Conditional Factor Models in Asset Pricing"

Qihui Chen[†] CUHK-Shenzhen
Nikolai Roussanov[‡] The Wharton School
Xiaoliang Wang[§] HKUST

March 27, 2026

This online appendix is organized as follows. Appendix A introduces two estimators for the number of factors. Appendix B complies the assumptions. Appendix C provides proofs of the theoretical results. Appendix D presents additional discussions. Appendix E shows simulation results, and Appendix F collects additional empirical findings.

Appendix A - Determining the Number of Factors

In this appendix, we develop two estimators for the number of factors K: one based on maximizing the ratio of two adjacent eigenvalues (Ahn and Horenstein, 2013), and another by counting the number of "large" eigenvalues (Bai and Ng, 2002). To define the estimators, let $\lambda_k(\tilde{Y}M_T\tilde{Y}'/T)$ denote the kth largest eigenvalue of the $JM \times JM$ matrix $\tilde{Y}M_T\tilde{Y}'/T$. The first one is given by:

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

Here, $\hat{K}$ is constrained to between 1 and $JM/2$, which is not restrictive because we assume $K \ge 1$ is fixed and $J \to \infty$. The second one is defined as:

$$\tilde{K} = \#\{1 \le k \le JM : \lambda_k(\tilde{Y}M_T\tilde{Y}'/T) \ge \lambda_{NT}\}, \tag{A.2}$$

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

We differ from Ahn and Horenstein (2013) and Bai and Ng (2002) in two key aspects. First, in the presence of time-varying Z_t, methods based on the original data

[†]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 Finance, HKUST Business School; xlwangfin@ust.hk

$\{Y_t\}_{t\leq T}$ or the projected data $\{\Phi(Z_t)(\Phi(Z_t)'\Phi(Z_t))^{-1}\Phi(Z_t)'Y_t\}_{t\leq T}$ may fail to estimate K. Instead, we work on the regressed data $\{\tilde{Y}_t\}_{t\leq T}$, where the regressed data matrix $\tilde{Y}M_T$ is approximately equal to $BF'M_T$, whose rank is qual to the number of factors. Second, we allow $N/T \to \infty$.

Theorem A.1. *(A) Suppose Assumptions B.1-B.3, B.5(i), and B.9 hold. Let $\hat{K}$ be given in (A.1). 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 B.1-B.3 hold. Let $\tilde{K}$ be given in (A.2). Assume (i) $N \to \infty$; (ii) $T \geq K + 1$; (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

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

Theorem A.1 shows that both $\hat{K}$ and $\tilde{K}$ are consistent estimators of K. The consistency of $\hat{K}$ requires $T \to \infty$, while the consistency of $\tilde{K}$ does not. The latter relies on the choice of λ_{NT}. In practice, $\hat{K}$ is recommended when T is large, while $\tilde{K}$ is preferred when T is small.

Appendix B - Assumptions

Assumption B.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) $\max_{m\leq M, j\leq J, i\leq N, t\leq T} E[\phi_j^2(z_{it,m})] < \infty$.

Since $\hat{Q}_t = \sum_{i=1}^N \phi(z_{it})\phi(z_{it})'/N$ is a $JM \times JM$ matrix with JM much smaller than N, Assumption B.1(i) can follow from the law of large numbers for finite T and its uniform variant for $T \to \infty$; see Proposition D.1 for a set of sufficient conditions. The conditions can be easily verified for B-splines, Fourier series, and polynomials basis functions. In particular, we allow Z_t to be nonstationary over t. When Z_t is not changing over t, Assumption B.1 reduces to Assumptions 3.3 of Fan et al. (2016a).

Assumption B.2 (Factor 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)* $\lambda_{\min}(F'M_TF/T) > d_{\min}$*; (iv)* $\max_{k\leq K, m\leq M} \sup_z |\delta_{km,J}(z)| = O(J^{-\kappa})$ *and* $\max_{m\leq M} \sup_z |r_{m,J}(z)| = O(J^{-\kappa})$ *for some constant* $\kappa > 1/2$*.*

Assumption B.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 presentation, we assume f_t's are nonrandom fixed parameters by following Bai and Li (2012). Our analysis holds if they are random variables. In this case, we assume f_t's to be independent of all other variables, and all stochastic statements can then regarded as conditioning on f_t's realizations. Since the dimension of B is $JM \times K$, Assumption B.2(i) requires $JM \geq K$. Since the rank of M_T is $T-1$, Assumption B.2(iii) requires $T \geq K+1$, which implies $T \geq 2$. These two requirements are not restrictive, since we assume K is fixed. Assumption B.2(iv) is standard in the sieve literature. It can be easily satisfied by using B-splines or polynomials basis functions under certain smoothness of $\alpha(\cdot)$ and $\beta(\cdot)$; see, for example, Lorentz (1986) and Chen (2007).

Assumption B.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 B.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 B.3 reduces to Assumptions 3.4 (i) and (iii) of Fan et al. (2016a).

Assumption B.4 (Intercept function)**.** $a'B = 0$ *and* $\|a\| < C_0$ *for some* $0 < C_0 < \infty$*.*

Assumption B.4 is needed 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).

Assumption B.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

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

Assumptions B.1-B.4 allow us to establish a preliminary rate of the estimators in Theorem C.1. Assumption B.5 is an additional assumption that we need to establish a fast rate in Theorem 4.1. Assumption B.5(i) strengthens Assumption B.1(ii). Assumption B.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 widely used in the sieve literature; see, for example, Newey (1997) and Huang (1998). Assumption B.5(iii) is commonly imposed in the sieve literature, which is used to justify the asymptotic convergence of $\hat{Q}_t$. Assumption B.5(iv) allows for weak dependence of $\{\varepsilon_{it}\}_{i\leq N,t\leq T}$ over both i and t. The second condition is similar to the second condition in Assumption B.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.

Assumption B.6 (Asymptotic distribution)**.** *(i)* $(F'M_T F/T)B'B$ *has distinct eigenvalues; (ii)* $\{\varepsilon_{it}\}_{i\leq N,t\leq T}$ *are independent across* $i \leq N$*; (iii) there is* $0 < C_3 < \infty$ *such that*

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

Assumption B.6 is needed in Theorem 4.2. The distinct eigenvalue condition in Assumption B.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 B.6(ii) imposes independence of $\{\varepsilon_{it}\}_{i\leq N,t\leq T}$ across i for simplicity. Cross-sectional independence is also imposed in Fan et al. (2016a) for studying specification test (Theorem 5.1). It is straightforward to modify the proof of Theorem 4.2 to allow for clsuter-type dependence of $\{\varepsilon_{it}\}_{i\leq N,t\leq T}$ across i.[1] Assumption B.6(iii) allows for weak dependence of $\{\varepsilon_{it}\}_{i\leq N,t\leq T}$ over t.

[1] Assumption B.6(ii) permits the use of YurinskiiâĂŹs coupling. Alternatively, one could apply the coupling method from Li and Liao (2020), which accommodates mixingale-type dependence. However, it remains unclear which dependence structure is more suitable for asset pricing models. For this reason, we adhere to Assumption B.6(ii) in this context.

Assumption B.7 (Bootstrap). *(i)* $\{w_i\}_{i\leq N}$ *is a sequence of 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 $\hat{Q}_t^*=\Phi(Z_t)^{*\prime}\Phi(Z_t)/N$*; (iii)* $\lambda_{\min}(\Omega)>0$.

Assumption B.7 is needed in Theorem 4.3. Assumption B.7(i) defines the bootstrap weight w_i for each i. It is straightforward to extend the bootstrap to accommodate clsuter-type dependence of $\{\varepsilon_{it}\}_{i\leq N,t\leq T}$ across i by utilizing the same weight within each cluster. Since $\hat{Q}_t^*=\sum_{i=1}^N\phi(z_{it})\phi(z_{it})'w_i/N$ is a $JM\times JM$ matrix with JM much smaller than N, Assumption B.7(ii) can follow from the law of large numbers for finite T and its uniform variant for $T\to\infty$, similar to Assumption B.1(i). Assumption B.7(iii) requires nonsingularity of the variance-covariance matrix Ω.

Assumption B.8 (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* $\sup_z\|\beta(z)\|<\infty$.

Assumption B.8 is needed in Theorem 4.4. Assumptions B.8(i)-(iv) are analogous to Assumptions B.1(i), B.5(i), (ii), and B.7(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 former are implied by the latter ones. In this case, Assumptions B.8(i)-(iv) thus are redundant.

Assumption B.9 (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 B.9 is needed in Theorem A.1(i). Assumption B.9(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 B.9(ii) allows for weak dependence of $\{\varepsilon_{it}\}_{i\leq N,t\leq T}$ over both i and t; see Proposition D.2 for a set of sufficient conditions. Assumption B.9 is distinct from Assumption 6.1 in Fan et al. (2016a), which may not be easy to verify.

Appendix C - Proofs of Theoretical Results

C.1 A Preliminary Rate of Convergence

We first establish a preliminary convergence rate of $\hat{a}$, $\hat{B}$, $\hat{F}$, $\hat{\alpha}(\cdot)$, and $\hat{\beta}(\cdot)$, as an intermediate step toward proving Theorem 4.1.

Theorem C.1. *Suppose Assumptions B.1-B.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) $T\geq K+1$; (iii) $J\to\infty$ with $J=o(\sqrt{N})$. Then*

$$\begin{aligned}
\|\hat{a}-a\|^2 &= O_p\left(\frac{1}{J^{2\kappa}}+\frac{J^2}{N^2}+\frac{J}{NT}\right),\\
\|\hat{B}-BH\|_F^2 &= O_p\left(\frac{1}{J^{2\kappa}}+\frac{J^2}{N^2}+\frac{J}{NT}\right),\\
\frac{1}{T}\|\hat{F}-F(H')^{-1}\|_F^2 &= O_p\left(\frac{1}{J^{2\kappa}}+\frac{J}{N}\right),\\
\sup_z|\hat{\alpha}(z)-\alpha(z)|^2 &= O_p\left(\frac{1}{J^{2\kappa-1}}+\frac{J^3}{N^2}+\frac{J^2}{NT}\right)\max_{j\leq J}\sup_z|\phi_j(z)|^2,\\
\sup_z\|\hat{\beta}(z)-H'\beta(z)\|^2 &= O_p\left(\frac{1}{J^{2\kappa-1}}+\frac{J^3}{N^2}+\frac{J^2}{NT}\right)\max_{j\leq J}\sup_z|\phi_j(z)|^2.
\end{aligned}$$

Proof: Let us begin by defining some notation. For $A_t=\Delta_t\equiv R(Z_t)+\Delta(Z_t)f_t$ and ε_t, let $\tilde{A}_t\equiv(\Phi(Z_t)'\Phi(Z_t))^{-1}\Phi(Z_t)'A_t$. 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

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

where 1_T denote a $T\times 1$ vector of ones. Recall $M_T=I_T-1_T1_T'/T$. Post-multiplying

(C.1) by M_T to remove a, we thus obtain

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

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 (C.2) to $(\tilde{Y}M_T\tilde{Y}'/T)\hat{B} = \hat{B}V$ to obtain

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

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

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

where the equality follows by Lemmas C.1 and C.2(i), along with the fact that $\|D_3\|_F = \|D_6\|_F$. By the definition of $\hat{a}$,

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

where H^{-1} is well defined with probability approaching one by (C.4) and Lemma C.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 facts that $\|x + y\| \leq \|x\| + \|y\|$ and $\|Ax\| \leq \|A\|_2\|x\|$, (C.5) implies

$$\begin{aligned}\|\hat{a} - a\|^2 \leq 4\Big(&\|\hat{B} - BH\|_F^2\|a\|^2 + \|BH - \hat{B}\|_F^2\|H^{-1}\|_2^2\max_{t\leq T}\|f_t\|^2 \\ &+\frac{1}{T}\|\tilde{\Delta}\|_F^2 + \frac{1}{T^2}\|\tilde{E}1_T\|^2\Big) = O_p\left(\frac{1}{J^{2\kappa}} + \frac{J^2}{N^2} + \frac{J}{NT}\right),\end{aligned} \tag{C.6}$$

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

obtain

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

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

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

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

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

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

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

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

C.1.1 Technical Lemmas

Lemma C.1. *Let D_1, D_2, D_3, D_4, D_5 be given in the proof of Theorem C.1.*
(i) Under Assumptions B.1(i), B.2(i), (ii), and (iv), $\|D_1\|_F^2 = O_p(J^{-2\kappa})$.
(ii) Under Assumptions B.1(i), B.2(ii), and (iv), $\|D_2\|_F^2 = O_p(J^{-4\kappa})$.
(iii) Under Assumptions B.1, B.2(ii), (iv), and B.3, $\|D_3\|_F^2 = O_p(J^{-2\kappa}J/N)$.
(iv) Under Assumptions B.1, B.2(i), (ii), and B.3, $\|D_4\|_F^2 = O_p(J/NT)$.

(v) Under Assumptions B.1 and B.3, $\|D_5\|_F^2 = O_p(J^2/N^2)$.

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 B.2(i) and (ii) as well as Lemma C.3(i).

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

(iv) Since $\|D_4\|_F \leq \|B\|_2\|\tilde{E}M_TF\|_F/T$, the result then immediately follows from Assumption B.2(i) and Lemma C.3(iii).

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

Lemma C.2. *Suppose Assumptions B.1-B.3 hold. Let V be given in the proof of Theorem C.1. Assume (i) $N \to \infty$; (ii) $T \geq K+1$; (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 (C.2), $\tilde{Y}M_T\tilde{Y}'/T = BF'M_TFB'/T + \sum_{j=1}^8 D_j$, where $D_1, \ldots, D_6$ are given below (C.2). By the fact that $\|C+D\|_F \leq \|C\|_F + \|D\|_F$,

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

where the equality follows by Lemma C.1 and the facts 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 $(F'M_TF/T)B'B$, which are equal to the first K largest eigenvalues of $BF'M_TFB'/T$. By the Weyl's inequality and the fact that $\|A\|_2 \leq \|A\|_F$,

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

Thus, $\|V\|_2 = O_p(1)$ and $\|V^{-1}\|_2 = \lambda_{\min}^{-1}(V) = O_p(1)$ follow from (C.12) and Assumptions B.2(i)-(iii). Let $H^\diamond \equiv (F'M_TF/T)B'\hat{B}V^{-1}$. Recall that $H = (F'M_T\tilde{Y}'\hat{B}/T)V^{-1}$.

Then by the facts that $\|A\|_2 \leq \|A\|_F$ and $\|M_T\|_2 = 1$,

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

where the equality follows from the second result in (i), Assumption B.2(ii), and Lemmas C.3(i) and (iii). Since $\|H^\diamond\|_2 \leq \|F'M_TF/T\|_2\|B\|_2\|\hat{B}\|_2\|V^{-1}\|_2$, the third result in (i) follows from (C.13), the second result in (i), and Assumptions B.2(i) and (ii).

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

$$\|\hat{B}'\hat{B} - H'B'BH\|_F \leq \|\hat{B}\|_2\|\hat{B} - BH\|_F + \|\hat{B} - BH\|_F\|B\|_2\|H\|_2. \quad \text{(C.14)}$$

Thus, $I_K - H'B'BH = o_p(1)$ by Assumption B.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 B.2(i). ■

Lemma C.3. *Let $\tilde{\Delta}$ and $\tilde{E}$ be given in the proof of Theorem C.1.*
(i) Under Assumptions B.1(i), B.2(ii), and (iv), $\|\tilde{\Delta}\|_F^2/T = O_p(J^{-2\kappa})$.
(ii) Under Assumptions B.1 and B.3, $\|\tilde{E}\|_F^2/T = O_p(J/N)$.
(iii) Under Assumptions B.1, B.2 (ii), and B.3, $\|\tilde{E}M_TF\|_F^2/T = O_p(J/N)$.

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

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

where the last line follows from Assumptions B.1(i) and B.2(ii), as well as Lemma C.4(iii).

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

$$\frac{1}{T}\|\tilde{E}\|_F^2 = \frac{1}{T}\sum_{t=1}^T \|(\Phi(Z_t)'\Phi(Z_t))^{-1}\Phi(Z_t)'\varepsilon_t\|^2$$

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

where the last equality follows from Assumption B.1(i) and Lemma C.4(i).

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

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

where the equality follows from Assumption B.2(ii) and Lemma C.4(ii). ■

Lemma C.4. *(i) Under Assumptions B.1(ii) and B.3,*

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

(ii) Under Assumptions B.1, B.2(ii), and B.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 B.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

$$\begin{aligned}E\left[\sum_{t=1}^{T}\|\Phi(Z_t)'\varepsilon_t\|^2\right] &= E\left[\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\phi(z_{it})'\phi(z_{jt})\varepsilon_{it}\varepsilon_{jt}\right] \\ &= \sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}E[\phi(z_{it})'\phi(z_{jt})]E[\varepsilon_{it}\varepsilon_{jt}] \\ &\leq \max_{i\leq N,t\leq T}E[\|\phi(z_{it})\|^2]\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}|E[\varepsilon_{it}\varepsilon_{jt}]|\end{aligned}$$

$$\leq TJM \max_{m\leq M,j\leq J,i\leq N,t\leq T} E[\phi_j^2(z_{it,m})] \max_{t\leq T} \sum_{i=1}^{N}\sum_{j=1}^{N} |E[\varepsilon_{it}\varepsilon_{jt}]| = O(NTJ), \tag{C.18}$$

where the second equality follows by the independence in Assumption B.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 B.1(ii) and B.3(iii).

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

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

where the second equality follows from the independence in Assumption B.3(i) and the linearity of both expectation and trace operators, and the inequality follows by the fact that $\|Ax\| \leq \|A\|_2\|x\|$. Moreover,

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

where the first inequality is due to the Cauchy-Schwartz inequality, and the second one 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 (C.19) and (C.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 B.1, B.2(ii), and B.3(iii). Thus, the first result of the lemma follows by the Markov's inequality and Lemma C.5. The proof of the second result is similar.

(iii) The first result follows since

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

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 B.2(iv). The proof of the second result is similar. ■

Lemma C.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. ■

C.2 Proof of Theorem 4.1

Proof of Theorem 4.1: Theorem C.1 provides a preliminary rate of $\|\hat{a} - a\|^2$, $\|\hat{B} - BH\|_F^2$, and $\|\hat{F} - F(H')^{-1}\|_F^2$ by using rough bounds based on (C.3), (C.5), and (C.7). To improve the rate of $\|\hat{B} - BH\|_F^2$, we need to treat $D_5\hat{B}$ in as (C.3) a whole to establish its rate. By the Cauchy-Schwartz inequality and the facts that $\|C + D\|_F \leq \|C\|_F + \|D\|_F$ and $\|CD\|_F \leq \|C\|_2\|D\|_F$, (C.3) implies

$$\begin{aligned} \|\hat{B} - BH\|_F^2 &\leq 10\|\hat{B}\|_2^2\|V^{-1}\|_2^2 \left(\sum_{j\neq 5}^{6} \|D_j\|_F^2\right) + 2\|V^{-1}\|_2^2\|D_5\hat{B}\|_F^2, \\ &= O_p\left(\frac{1}{J^{2\kappa}} + \frac{J}{N^2} + \frac{J}{NT}\right), \end{aligned} \tag{C.22}$$

where the equality follows from $J = o(\sqrt{N})$, Lemmas C.1(i)-(iv), C.2(i), and C.6(ii), as well as the fact that $\|D_6\|_F = \|D_3\|_F$. Given the rate of $\|\hat{B} - BH\|_F^2$ in (C.22), the rate of $\|\hat{a} - a\|^2$ immediately follows from the same argument in (C.6). To improve the rate of $\|\hat{F} - F(H')^{-1}\|_F^2$, we need to plug in the expansion of $\hat{B} - BH$ to (C.5),

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 facts that $\|C+D\|_F \leq \|C\|_F + \|D\|_F$ and $\|CD\|_F \leq \|C\|_2\|D\|_F$, combining (C.3) and (C.7) implies

$$\begin{aligned}\|\hat{F} - F(H')^{-1}\|_F = &\left(\sum_{j\neq 4,5}^{6} \|D_j\|_F\|\hat{B}\|_2\|a\| + \|a'D_4\|\|\hat{B}\|_2 + \|a\|\|D_5\hat{B}\|_F\right)\times \\ &\|V^{-1}\|_2\|1_T\| + \left(\sum_{j\neq 4,5}^{6}\|D_j\|_F\|\hat{B}\|_2 + \|D_4'\hat{B}\|_F + \|D_5\hat{B}\|_F\right) \\ &\times \|F\|_2\|H^{-1}\|_2\|V^{-1}\|_2\|\hat{B}\|_2 + \|\tilde{\Delta}\|_F\|\hat{B}\|_2 + \|\tilde{E}'\hat{B}\|_F \\ = &\, O_p\left(\frac{\sqrt{T}}{J^\kappa} + \sqrt{\frac{T}{N}}\right), \end{aligned} \tag{C.23}$$

where the equality follows from $J = o(\sqrt{N})$, Assumptions B.2(ii) and B.4, Lemmas C.1(i)-(iii), C.2, C.3(i), C.6, and C.7(i), as well as the fact that $\|D_6\|_F = \|D_3\|_F$. Thus, the third result of the theorem follows from (C.23). The proofs of the last two results of the theorem are similar to the proofs of the last two results of Theorem C.1. ■

C.2.1 Technical Lemmas

Lemma C.6. *Let D_4 and D_5 be given in the proof of Theorem C.1. Assume (i) $N \to \infty$; (ii) $T \geq K+1$; (iii) $J \to \infty$ with $J^2\xi_J^2 \log J = o(N)$.*
(i) Under Assumptions B.1-B.5, $\|D_4'\hat{B}\|_F^2 = O_p(1/NT)$.
(ii) Under Assumptions B.1-B.5, $\|D_5\hat{B}\|_F^2 = O_p(J/N^2)$.
(iii) Under Assumptions B.1-B.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 B.2(i) and Lemma C.7(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 C.3(ii) and C.7(i).

(iii) Since $\|D_4'a\| \leq \|B\|_2\|a'\tilde{E}M_TF\|/T$, the result then immediately follows from Assumption B.2(i) and Lemma C.7(iii). ■

Lemma C.7. *Let $\tilde{E}$ be given in the proof of Theorem C.1. Assume (i) $N \to \infty$; (ii) $T \geq K+1$; (iii) $J \to \infty$ with $J^2\xi_J^2 \log J = o(N)$.*
(i) Under Assumptions B.1-B.5, $\|\hat{B}'\tilde{E}\|_F^2/T = O_p(1/N)$.

(ii) Under Assumptions B.1-B.5, $\|\hat{B}'\tilde{E}M_TF\|_F^2/T = O_p(1/N)$.
(iii) Under Assumptions B.1-B.5, $\|a'\tilde{E}M_TF\|^2/T = O_p(1/N)$.

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

$$\begin{aligned}\frac{1}{T}\|\hat{B}'\tilde{E}\|_F^2 &\leq \frac{2}{T}\|\tilde{E}\|_F^2\|\hat{B}-BH\|_F^2 + \frac{2}{T}\|H\|_2^2\|B'\tilde{E}\|_F^2 \\ &= \frac{2}{T}\|\tilde{E}\|_F^2\|\hat{B}-BH\|_F^2 + \frac{2}{N^2T}\|H\|_2^2\left(\sum_{t=1}^T\|B'\hat{Q}_t^{-1}\Phi(Z_t)'\varepsilon_t\|^2\right) \\ &= O_p\left(\frac{J}{N}\left(\frac{1}{J^{2\kappa}}+\frac{J^2}{N^2}+\frac{J}{NT}\right)+\frac{1}{N}\right) = O_p\left(\frac{1}{N}\right), \end{aligned} \tag{C.24}$$

where the second equality follows from $J^2\xi_J^2\log J = o(N)$, Lemmas C.2(i), C.3(ii), and C.8(i), as well as Theorem C.1, and the last line is due to $\kappa > 1/2$ and $J = o(\sqrt{N})$.

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

$$\begin{aligned}\frac{1}{T}\|\hat{B}'\tilde{E}M_TF\|_F^2 &\leq \frac{2}{T}\|\tilde{E}M_TF\|_F^2\|\hat{B}-BH\|_F^2 + \frac{2}{T}\|H\|_2^2\|B'\tilde{E}M_TF\|_F^2 \\ &\leq \frac{2}{T}\|\tilde{E}M_TF\|_F^2\|\hat{B}-BH\|_F^2 + \frac{4}{N^2T}\|H\|_2^2\left\|\sum_{t=1}^T B'\hat{Q}_t^{-1}\Phi(Z_t)'\varepsilon_t f_t'\right\|_F^2 \\ &\quad + \frac{4\|\bar{f}\|^2}{N^2T}\|H\|_2^2\left\|\sum_{t=1}^T B'\hat{Q}_t^{-1}\Phi(Z_t)'\varepsilon_t\right\|^2 \\ &= O_p\left(\frac{J}{N}\left(\frac{1}{J^{2\kappa}}+\frac{J^2}{N^2}+\frac{J}{NT}\right)+\frac{1}{N}\right) = O_p\left(\frac{1}{N}\right), \end{aligned} \tag{C.25}$$

where the first equality follows from $J^2\xi_J^2\log J = o(N)$, Assumption B.2(ii), Lemmas C.2(i), C.3(iii), and C.8(ii), as well as Theorem C.1, and the last equality is due to $\kappa > 1/2$ and $J = o(\sqrt{N})$.

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

$$\begin{aligned}\frac{1}{T}\|a'\tilde{E}M_TF\|^2 &\leq \frac{2}{N^2T}\left\|\sum_{t=1}^T a'\hat{Q}_t^{-1}\Phi(Z_t)'\varepsilon_t f_t'\right\|^2 + \frac{2\|\bar{f}\|^2}{N^2T}\left|\sum_{t=1}^T a'\hat{Q}_t^{-1}\Phi(Z_t)'\varepsilon_t\right|^2 \\ &= O_p\left(\frac{1}{N}\right), \end{aligned} \tag{C.26}$$

because $J^2\xi_J^2\log J = o(N)$, Assumption B.2(ii) and Lemma C.8(ii). ■

Lemma C.8. *Assume* $J \geq 2$ *and* $\xi_J^2\log J = o(N)$.

(i) Under Assumptions B.1(i), B.2(i), B.3, and B.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 B.1(i), B.2(i), (ii), B.3, B.4, and B.5,

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

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

$$\begin{aligned}
\sum_{t=1}^{T} \|B'\hat{Q}_t^{-1}\Phi(Z_t)'\varepsilon_t\|^2 &\le 2\sum_{t=1}^{T} \|B'Q_t^{-1}\Phi(Z_t)'\varepsilon_t\|^2\\
&\quad + 2\sum_{t=1}^{T} \|B'(\hat{Q}_t^{-1}-Q_t^{-1})\Phi(Z_t)'\varepsilon_t\|^2 \equiv 2\mathcal{T}_1 + 2\mathcal{T}_2. \qquad \text{(C.27)}
\end{aligned}$$

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

$$\begin{aligned}
E[\mathcal{T}_1] &= E\left[\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N} \phi(z_{it})'Q_t^{-1}BB'Q_t^{-1}\phi(z_{jt})\varepsilon_{it}\varepsilon_{jt}\right]\\
&= \sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N} E[\phi(z_{it})'Q_t^{-1}BB'Q_t^{-1}\phi(z_{jt})]E[\varepsilon_{it}\varepsilon_{jt}]\\
&\le T \max_{i\le N, t\le T} E[\|B'Q_t^{-1}\phi(z_{it})\|^2] \max_{t\le T}\sum_{i=1}^{N}\sum_{j=1}^{N} |E[\varepsilon_{it}\varepsilon_{jt}]| = O(NT), \qquad \text{(C.28)}
\end{aligned}$$

where the second equality follows by the independence in Assumption B.3(i), the inequality is due to the Cauchy-Schwartz inequality, and the last equality follows from

Assumption B.3(iii) and Lemma C.9. The latter also holds, since

$$\begin{aligned}\mathcal{T}_2 &\leq C_{NT}\sum_{t=1}^{T}\|\hat{Q}_t - Q_t\|_2^2\|\Phi(Z_t)'\varepsilon_t\|^2 \\ &\leq C_{NT}\left(\sum_{t=1}^{T}\|\hat{Q}_t - Q_t\|_2^4\right)^{1/2}\left(\sum_{t=1}^{T}\|\Phi(Z_t)'\varepsilon_t\|^4\right)^{1/2} = O_p(TJ\xi_J^2\log J), \qquad \text{(C.29)}\end{aligned}$$

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 B.1(i), B.2(i), and B.5(ii), as well as Lemmas C.10 and C.11.

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

$$\begin{aligned}\left\|\sum_{t=1}^{T}B'\hat{Q}_t^{-1}\Phi(Z_t)'\varepsilon_t f_t'\right\|_F^2 &\leq 2\left\|\sum_{t=1}^{T}B'Q_t^{-1}\Phi(Z_t)'\varepsilon_t f_t'\right\|_F^2 \\ &\quad + 2\left\|\sum_{t=1}^{T}B'(\hat{Q}_t^{-1}-Q_t^{-1})\Phi(Z_t)'\varepsilon_t f_t'\right\|_F^2 \equiv 2\mathcal{T}_1 + 2\mathcal{T}_2. \qquad \text{(C.30)}\end{aligned}$$

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 $\|A\|_F^2 = \mathrm{tr}(AA')$. The former holds by the Markov's inequality, since

$$\begin{aligned}E[\mathcal{T}_1] &= E\left[\mathrm{tr}\left(\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}B'Q_t^{-1}\phi(z_{it})\varepsilon_{it}f_t'f_s\varepsilon_{js}\phi(z_{js})'Q_s^{-1}B\right)\right] \\ &= \sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}E\left[\phi(z_{it})'Q_t^{-1}BB'Q_s^{-1}\phi(z_{js})\right]f_t'f_sE[\varepsilon_{it}\varepsilon_{js}] \\ &\leq C_{NT}\max_{i\leq N,t\leq T}E[\|B'Q_t^{-1}\phi(z_{it})\|^2]\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}|E[\varepsilon_{it}\varepsilon_{js}]| = O(NT), \qquad \text{(C.31)}\end{aligned}$$

where $C_{NT} = \max_{t\leq T}\|f_t\|^2$, the second equality follows from the independence in Assumption B.3(i) and the linearity of both expectation and trace operators, the inequality is due to the Cauchy-Schwartz inequality, and the last equality follows from Assumptions B.2(ii) and B.3(iii), as well as Lemma C.9. Let E_ε denote the expectation with

respect to $\{\varepsilon_t\}_{t\leq T}$. For the latter, we have

$$\begin{aligned}
E_\varepsilon[\mathcal{T}_2] &= E_\varepsilon\left[\mathrm{tr}\left(\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N} B'(\hat{Q}_t^{-1}-Q_t^{-1})\phi(z_{it})\varepsilon_{it}f_t'f_s\varepsilon_{js}\phi(z_{js})'(\hat{Q}_s^{-1}-Q_s^{-1})B\right)\right] \\
&= \sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N} \phi(z_{it})'(\hat{Q}_t^{-1}-Q_t^{-1})BB'(\hat{Q}_s^{-1}-Q_s^{-1})\phi(z_{js})f_t'f_sE[\varepsilon_{it}\varepsilon_{js}] \\
&\leq C_{NT}^{*}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N} \|\hat{Q}_t-Q_t\|_2\|\phi(z_{it})\|\|\phi(z_{js})\||E[\varepsilon_{it}\varepsilon_{js}]| \\
&\leq C_{NT}^{**}\left(\sum_{t=1}^{T}\left(\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\|\phi(z_{it})\|\|\phi(z_{js})\||E[\varepsilon_{it}\varepsilon_{js}]|\right)^2\right)^{1/2}, \qquad \text{(C.32)}
\end{aligned}$$

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 from the independence in Assumption B.3(i) and the linearity of both expectation and trace operators, 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

$$\begin{aligned}
&E\left[\sum_{t=1}^{T}\left(\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\|\phi(z_{it})\|\|\phi(z_{js})\||E[\varepsilon_{it}\varepsilon_{js}]|\right)^2\right] \\
&\leq \max_{i\leq N,t\leq T}E[\|\phi(z_{it})\|^4]\sum_{t=1}^{T}\left(\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}|E[\varepsilon_{it}\varepsilon_{js}]|\right)^2 \\
&\leq J^2M^2\max_{m\leq M,j\leq J,i\leq N,t\leq T}E[\phi_j^4(z_{it,m})]\sum_{t=1}^{T}\left(\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}|E[\varepsilon_{it}\varepsilon_{js}]|\right)^2, \qquad \text{(C.33)}
\end{aligned}$$

where the first inequality is due to the Cauchy-Schwartz inequality, the second one 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 B.1(i), B.2(i), (ii), and B.5(ii), as well as Lemma C.11, we have $C_{NT}^{**} = O_p(\sqrt{T}\xi_J\sqrt{\log J}/\sqrt{N})$. Combining this, (C.32) and (C.33) implies that $E_\varepsilon[\mathcal{T}_2] = O_p(\sqrt{N}TJ\xi_J\sqrt{\log J})$ by Assumptions B.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 C.5. This proves the first result, and the proofs of other results are similar. ■

Lemma C.9. *Suppose Assumptions B.2(i), B.4, and B.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 $\|x\|^2 = \text{tr}(xx')$,

$$\begin{aligned} E[\|B'Q_t^{-1}\phi(z_{it})\|^2] &= E[\text{tr}(B'Q_t^{-1}\phi(z_{it})\phi(z_{it})'Q_t^{-1}B)] = \text{tr}(B'Q_t^{-1}Q_{it}Q_t^{-1}B) \\ &\leq \max_{i\leq N,t\leq T} \lambda_{\max}(Q_{it}) \left(\min_{t\leq T} \lambda_{\min}(Q_t)\right)^{-1} K\|B\|_2^2 \\ &\leq \max_{i\leq N,t\leq T} \lambda_{\max}(Q_{it}) \left(\min_{i\leq N,t\leq T} \lambda_{\min}(Q_{it})\right)^{-1} K\|B\|_2^2, \end{aligned} \tag{C.34}$$

where the second equality follows from the linearity of both expectation and trace operators, the first inequality follows since $\text{tr}(B'B) = \|B\|_F^2 \leq K\|B\|_2^2$, 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 (C.34), along with Assumptions B.2(i) and B.5(ii). The proof of the second result is similar. ■

Lemma C.10. *Under Assumptions B.3(i), B.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

$$\begin{aligned} E\left[\sum_{t=1}^{T} \|\Phi(Z_t)'\varepsilon_t\|^4\right] &= E\left[\sum_{t=1}^{T}\left(\sum_{i=1}^{N}\sum_{j=1}^{N} \phi(z_{it})'\phi(z_{jt})\varepsilon_{it}\varepsilon_{jt}\right)^2\right] \\ = E\left[\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{\ell=1}^{N} \phi(z_{it})'\phi(z_{jt})\phi(z_{kt})'\phi(z_{\ell t})\varepsilon_{it}\varepsilon_{jt}\varepsilon_{kt}\varepsilon_{\ell t}\right] \\ = \sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{\ell=1}^{N} E[\phi(z_{it})'\phi(z_{jt})\phi(z_{kt})'\phi(z_{\ell t})]E[\varepsilon_{it}\varepsilon_{jt}\varepsilon_{kt}\varepsilon_{\ell t}] \\ \leq \max_{i\leq N,t\leq T} E[\|\phi(z_{it})\|^4] \sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{\ell=1}^{N} |E[\varepsilon_{it}\varepsilon_{jt}\varepsilon_{kt}\varepsilon_{\ell t}]| \\ \leq J^2M^2 \max_{m\leq M,j\leq J,i\leq N,t\leq T} E[\phi_j^4(z_{it,m})] \sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{\ell=1}^{N} |E[\varepsilon_{it}\varepsilon_{jt}\varepsilon_{kt}\varepsilon_{\ell t}]| \\ = O(N^2TJ^2), \end{aligned} \tag{C.35}$$

where the third equality follows by the independence in Assumption B.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 B.5(i) and (iv). ■

Lemma C.11. *Suppose Assumptions B.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

$$\begin{aligned}
\mathcal{D}_t &\equiv E[\|\hat{Q}_t - Q_t\|_2^4] \\
&\leq 16E\left[\left\|\frac{1}{N}\sum_{i=1}^N \eta_i\phi(z_{it})\phi(z_{it})'\right\|_2^4\right] \\
&\leq 16C\frac{\log^2 JM}{N^2}\sup_z \|\phi(z)\|^4 E\left[\left\|\frac{1}{N}\sum_{i=1}^N \phi(z_{it})\phi(z_{it})'\right\|_2^2\right] \\
&\leq 16M^2C\frac{\xi_J^4 \log^2 JM}{N^2} E[\|\hat{Q}_t\|_2^2], \qquad \text{(C.36)}
\end{aligned}$$

where the first inequality follows from the independence in Assumption B.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 C.12 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 (C.36) leads to the inequality: $\mathcal{D}_t \leq 2A(\sqrt{\mathcal{D}_t} + \|Q_t\|_2^2)$. Solving the inequality yields

$$E[\|\hat{Q}_t - Q_t\|_2^4] \leq \left(A + \sqrt{A^2 + 2A\|Q_t\|_2^2}\right)^2. \qquad \text{(C.37)}$$

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 (C.37) and Assumption B.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 C.12 (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 constant C, *where* E_η *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

$$\begin{aligned} & E_\eta \left[\left\| \frac{1}{N} \sum_{i=1}^{N} \eta_i S_i \right\|_2^4 \right] \leq E_\eta \left[\left\| \frac{1}{N} \sum_{i=1}^{N} \eta_i S_i \right\|_{S_{\log k}}^4 \right] \\ \leq & \left(E_\eta \left[\left\| \frac{1}{N} \sum_{i=1}^{N} \eta_i S_i \right\|_{S_{\log k}}^{\log k} \right] \right)^{4/\log k} \\ \leq & C_0^4 \frac{\log^2 k}{N^2} \left\| \left(\frac{1}{N} \sum_{i=1}^{N} S_i^2 \right)^{1/2} \right\|_{S_{\log k}}^4 \leq C_0^4 e^4 \frac{\log^2 k}{N^2} \left\| \frac{1}{N} \sum_{i=1}^{N} S_i^2 \right\|_2^2, \end{aligned} \tag{C.38}$$

where the first inequality follows by (6.44) in Belloni et al. (2015) and the fact that $\|\cdot\|_{S_{\log k}}$ 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 $C = C_0^4 e^4$. ■

C.3 Proof of Theorem 4.2

PROOF OF THEOREM 4.2: Let us first look at (C.22). 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 facts that $\|C + D\|_F \leq \|C\|_F + \|D\|_F$ and $\|CD\|_F \leq \|C\|_2 \|D\|_F$, (C.3) implies

$$\begin{aligned} \|\sqrt{NT}(\hat{B} - BH) - \sqrt{NT} D_4 \hat{B} V^{-1}\|_F \leq & \sqrt{NT} \|V^{-1}\|_2 \|D_5 \hat{B}\|_F \\ & + \sqrt{NT} \|\hat{B}\|_2 \|V^{-1}\|_2 \sum_{j \neq 4,5}^{6} \|D_j\|_F = O_p\left(\frac{\sqrt{NT}}{J^\kappa} + \frac{\sqrt{TJ}}{\sqrt{N}} \right), \end{aligned} \tag{C.39}$$

where the equality follows by $J = o(\sqrt{N})$, Lemmas C.1(i)-(iii), C.2(i), and C.6(ii), along with 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.39) and Lemma C.13 implies

$$\|\sqrt{NT}(\hat{B} - BH) - \mathcal{L}_{NT}B'B\mathcal{M}\|_F = O_p\left(\frac{\sqrt{NT}}{J^{\kappa}} + \frac{\sqrt{TJ}}{\sqrt{N}} + \frac{\sqrt{J\xi_J}\log^{1/4} J}{N^{1/4}}\right). \quad \text{(C.40)}$$

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.40) and Lemma C.14. We now look at (C.5). By the fact that $\|x + y\| \leq \|x\| + \|y\|$, it implies

$$\begin{aligned}\|\sqrt{NT}(\hat{a} - a) - (I_{JM} - \hat{B}\hat{B}')[\sqrt{N/T}\tilde{E}1_T - \sqrt{NT}(\hat{B} - BH)H^{-1}\bar{f}] \\ + \hat{B}\sqrt{NT}(\hat{B} - BH)'a\| \leq \|(I_{JM} - \hat{B}\hat{B}')\sqrt{N/T}\tilde{\Delta}1_T\| = O_p\left(\frac{\sqrt{NT}}{J^{\kappa}}\right), \quad \text{(C.41)}\end{aligned}$$

where the equality follows by Lemma C.3(i). Given the rate of $\|\hat{B} - BH\|_F$ in Theorem 4.1 and the rate of $\|N\tilde{E}1_T\|$ in Lemma C.4(ii), we may replace all $\hat{B}$ except those in $\hat{B} - BH$ with BH to obtain

$$\begin{aligned}\|\sqrt{NT}(\hat{a} - a) - (I_{JM} - BHH'B')[\sqrt{N/T}\tilde{E}1_T - \sqrt{NT}(\hat{B} - BH)H^{-1}\bar{f}] \\ + BH\sqrt{NT}(\hat{B} - BH)'a\| = O_p\left(\frac{\sqrt{NT}}{J^{\kappa}} + \frac{\sqrt{TJ}}{\sqrt{N}} + \frac{J}{\sqrt{NT}}\right) \quad \text{(C.42)}\end{aligned}$$

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.15, we may replace all H except those in $\hat{B} - BH$ with $\mathcal{H}$ to obtain

$$\begin{aligned}\|\sqrt{NT}(\hat{a} - a) - (I_{JM} - B'\mathcal{H}\mathcal{H}'B')[\sqrt{N/T}\tilde{E}1_T - \sqrt{NT}(\hat{B} - BH)\mathcal{H}^{-1}\bar{f}] \\ + B\mathcal{H}\sqrt{NT}(\hat{B} - BH)'a\| = O_p\left(\frac{\sqrt{NT}}{J^{\kappa}} + \frac{\sqrt{TJ}}{\sqrt{N}} + \frac{J}{\sqrt{NT}}\right) \quad \text{(C.43)}\end{aligned}$$

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.13, we may replace $\sqrt{N/T}\tilde{E}1_T$ with ℓ_{NT} to obtain

$$\begin{aligned}\|\sqrt{NT}(\hat{a} - a) - (I_{JM} - B'\mathcal{H}\mathcal{H}'B')[\ell_{NT} - \sqrt{NT}(\hat{B} - BH)\mathcal{H}^{-1}\bar{f}] \\ + B\mathcal{H}\sqrt{NT}(\hat{B} - BH)'a\| = O_p\left(\frac{\sqrt{NT}}{J^{\kappa}} + \frac{\sqrt{TJ}}{\sqrt{N}} + \frac{\sqrt{J\xi_J}\log^{1/4} J}{N^{1/4}}\right) \quad \text{(C.44)}\end{aligned}$$

by noting that $J/\sqrt{NT} = o(\sqrt{J\xi_J}\log^{1/4} J/N^{1/4})$. The arguments in (C.42)-(C.44) are similar to those for the first result in Lemma C.13. 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.44), Lemma C.14, and the second result of the theorem. ■

C.3.1 Technical Lemmas

Lemma C.13. *Suppose Assumptions B.1-B.5 and B.6(i) and (ii) hold. Let $\tilde{E}$, D_4, and V be given in the proof of Theorem C.1, and ℓ_{NT} and $\mathcal{L}_{NT}$ be given in the proof of Theorem 4.2. Assume (i) $N \to \infty$; (ii) $T \geq K+1$; (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.15.

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

$$\begin{aligned}\sqrt{NT}D_4\hat{B}V^{-1} &= \sqrt{N/T}\tilde{E}M_T FB'B\mathcal{M} + \sqrt{N/T}\tilde{E}M_T FB'(\hat{B}-BH)V^{-1} \\ &\quad + \sqrt{N/T}\tilde{E}M_T FB'B(HV^{-1}-\mathcal{M}) \equiv \mathcal{T}_1 + \mathcal{T}_2 + \mathcal{T}_3. \qquad \text{(C.45)}\end{aligned}$$

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

$$\begin{aligned}\|\mathcal{T}_1 - \mathcal{L}_{NT}B'B\mathcal{M}\|_F &\leq \|B\|_2^2\|\mathcal{M}\|_2 \left\|\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}(\hat{Q}_t^{-1} - Q_t^{-1})\Phi(Z_t)'\varepsilon_t f_t'\right\|_F \\ &\quad + \|B\|_2^2\|\mathcal{M}\|_2\|\bar{f}\| \left\|\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}(\hat{Q}_t^{-1} - Q_t^{-1})\Phi(Z_t)'\varepsilon_t\right\| \\ &= O_p\left(\frac{\sqrt{J\xi_J}\log^{1/4} J}{N^{1/4}}\right), \qquad \text{(C.46)}\end{aligned}$$

where the equality follows from Assumptions B.2(i) and (ii), as well as Lemma C.16. The latter two follow by a similar argument. The second result also follows by a similar

argument as in (C.46). This completes the proof of the lemma. ■

Lemma C.14. *Suppose Assumptions B.2(ii), B.3(i), (ii), B.5(i)-(iii), and B.6(ii) and (iii) hold. Let ℓ_{NT} and $\mathcal{L}_{NT}$ be given in the proof of Theorem 4.2. Then there exists a $JM \times (K+1)$ random matrix $\mathbb{N}$ with $\text{vec}(\mathbb{N}) \sim N(0, \Omega)$ 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 $\text{vec}((\ell_{NT}, \mathcal{L}_{NT})) = \sum_{i=1}^N \zeta_i$. Note that $E[\zeta_i] = 0$ by Assumptions B.3(i) and (ii), and $\zeta_1, \ldots, \zeta_N$ are independent by Assumptions B.3(i), B.5(iii), and B.6(ii). Moreover,

$$\sum_{i=1}^N E[\|\zeta_i\|^3] \leq \sum_{i=1}^N (E[\|\zeta_i\|^4])^{3/4} = O\left(\frac{J^{3/2}}{\sqrt{N}}\right), \tag{C.47}$$

where the inequality follows by the LiapounovâĂŹs inequality, and the equality follows from Assumptions B.2(ii), B.5(i), (ii), and B.6(iii) because

$$\begin{aligned}
E[\|\zeta_i\|^4] &= \frac{1}{N^2T^2} E\left[\left(\sum_{t=1}^T \sum_{s=1}^T \phi(z_{it})' Q_t^{-1} Q_s^{-1} \phi(z_{is}) f_t^{\dagger\prime} f_s^\dagger \varepsilon_{it}\varepsilon_{is}\right)^2\right] \\
&\leq C_{NT} \max_{i\leq N, t\leq T} E[\|\phi(z_{it})\|^4] \frac{1}{N^2T^2} \sum_{t=1}^T \sum_{s=1}^T \sum_{u=1}^T \sum_{v=1}^T |E[\varepsilon_{it}\varepsilon_{is}\varepsilon_{iu}\varepsilon_{iv}]| \\
&\leq C_{NT} \max_{m\leq M, j\leq J, i\leq N, t\leq T} E[\phi_j^4(z_{it,m})] \frac{J^2M^2}{N^2T^2} \sum_{t=1}^T \sum_{s=1}^T \sum_{u=1}^T \sum_{v=1}^T |E[\varepsilon_{it}\varepsilon_{is}\varepsilon_{iu}\varepsilon_{iv}]|,
\end{aligned} \tag{C.48}$$

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 B.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.17 implies that there is a $JM \times (K+1)$ random matrix $\mathbb{N}$ with $\text{vec}(\mathbb{N}) \sim N(0, \Omega)$ such that

$$\|(\mathcal{L}_{NT}, \ell_{NT}) - \mathbb{N}\|_F = \|\text{vec}((\mathcal{L}_{NT}, \ell_{NT})) - \text{vec}(\mathbb{N})\| = O_p\left(\frac{J^{5/6}}{N^{1/6}}\right). \tag{C.49}$$

This completes the proof of the Lemma. ■

Lemma C.15. *Suppose Assumptions B.1-B.4 and B.6(i) hold. Let V be given in the*

proof of Theorem C.1. Assume (i) $N \to \infty$; (ii) $T \geq K+1$; (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 Υ 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 it on both sides by $(F'M_TF/T)^{1/2}B'$ to obtain

$$(F'M_TF/T)^{1/2}B'(\tilde{Y}M_T\tilde{Y}'/T)\hat{B} = (F'M_TF/T)^{1/2}B'\hat{B}V. \tag{C.50}$$

To simplify 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.50) as

$$[(F'M_TF/T)^{1/2}B'B(F'M_TF/T)^{1/2} + \delta_{NT}R_{NT}^{-1}]R_{NT} = R_{NT}V. \tag{C.51}$$

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.51) as

$$[(F'M_TF/T)^{1/2}B'B(F'M_TF/T)^{1/2} + \delta_{NT}R_{NT}^{-1}]\Upsilon_{NT} = \Upsilon_{NT}V, \tag{C.52}$$

which implies that $(F'M_TF/T)^{1/2}B'B(F'M_TF/T)^{1/2} + \delta_{NT}R_{NT}^{-1}$ has eigenvector matrix Υ_{NT} and eigenvalue matrix V. Since $R_{NT} = (F'M_TF/T)^{1/2}B'BH + o_p(1)$ by simple algebra and Theorem C.1, $R_{NT}^{-1} = O_p(1)$ by Assumptions B.2(i)-(iii) and Lemma C.2. This, along with (C.11) and Assumptions B.2(i) and (ii), implies

$$\delta_{NT}R_{NT}^{-1} = O_p\left(\frac{1}{J^\kappa} + \frac{J}{N} + \frac{\sqrt{J}}{\sqrt{NT}}\right). \tag{C.53}$$

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 B.6(i). By the eigenvector perturbation theory, there exists a unique eigen-

vector matrix Υ of $(F'M_TF/T)^{1/2}B'B(F'M_TF/T)^{1/2}$ such that

$$\Upsilon_{NT} = \Upsilon + O_p\left(\frac{1}{J^\kappa} + \frac{J}{N} + \frac{\sqrt{J}}{\sqrt{NT}}\right). \tag{C.54}$$

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

$$D_{NT} = V + O_p\left(\frac{1}{J^\kappa} + \frac{J}{N} + \frac{\sqrt{J}}{\sqrt{NT}}\right). \tag{C.55}$$

Recall that $H^\diamond = (F'M_TF/T)B'\hat{B}V^{-1}$ as given in the proof of Lemma C.2(i). Thus, by (C.54) and (C.55), $H^\diamond = (F'M_TF/T)^{1/2}R_{NT}V^{-1} = (F'M_TF/T)^{1/2}\Upsilon_{NT}D_{NT}^{1/2}V^{-1} = \mathcal{H} + O_p(J^{-\kappa} + J/N + \sqrt{J/NT})$, which together with (C.12) and (C.13) leads to the first result of the lemma. The second result of the lemma follows from (C.12), the first result of the lemma, and Lemma C.2(i). ■

Lemma C.16. *Suppose Assumptions B.1(i), B.2(ii), B.3(i), (ii), B.5, and B.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_ε denote the expectation with respect to $\{\varepsilon_t\}_{t\leq T}$. Since $\|A\|_F^2 = \mathrm{tr}(AA')$,

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

$$\leq C_{NT}^{**}\frac{1}{NT}\left(\sum_{t=1}^{T}\left(\sum_{i=1}^{N}\sum_{s=1}^{T}\|\phi(z_{it})\|\|\phi(z_{is})\||E[\varepsilon_{it}\varepsilon_{is}]|\right)^2\right)^{1/2}, \tag{C.56}$$

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 from the independence in Assumption B.3(i) and the linearity of both expectation and trace operators, the third equality follows by Assumption B.3(ii) and the independence in Assumption B.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

$$\begin{aligned}
&E\left[\sum_{t=1}^{T}\left(\sum_{i=1}^{N}\sum_{s=1}^{T}\|\phi(z_{it})\|\|\phi(z_{is})\||E[\varepsilon_{it}\varepsilon_{is}]|\right)^2\right] \\
&\leq \max_{i\leq N,t\leq T}E[\|\phi(z_{it})\|^4]\sum_{t=1}^{T}\left(\sum_{i=1}^{N}\sum_{s=1}^{T}|E[\varepsilon_{it}\varepsilon_{is}]|\right)^2 \\
&\leq J^2M^2\max_{m\leq M,j\leq J,i\leq N,t\leq T}E[\phi_j^4(z_{it,m})]\sum_{t=1}^{T}\left(\sum_{i=1}^{N}\sum_{s=1}^{T}|E[\varepsilon_{it}\varepsilon_{is}]|\right)^2,
\end{aligned} \tag{C.57}$$

where the first inequality is due to the Cauchy-Schwartz inequality, the second one 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 B.1(i), B.2(ii), and B.5(ii), along with Lemma C.11, we obtain that $C_{NT}^{**} = O_p(\sqrt{T}\xi_J\sqrt{\log J}/\sqrt{N})$. Combining this, (C.56) and (C.57) implies that $E_\varepsilon[\|\mathcal{T}\|_F^2] = O_p(J\xi_J\sqrt{\log J}/\sqrt{N})$ by Assumptions B.5(i) and (iv). Thus, the first result of the lemma follows by the Markov's inequality and Lemma C.5. The proof of the second result is similar. ■

Lemma C.17 (YurinskiiâĂŹs 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*

$$P\{\|S-\mathbb{S}\| > 3\delta\} \leq C_0D_0\left(1 + \frac{|\log(1/D_0)|}{k}\right)$$

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

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

C.4 Proof of Theorem 4.3

PROOF OF THEOREM 4.3: Let us begin by defining some notation. For $A_t = \Delta_t \equiv R(Z_t) + \Delta(Z_t)f_t$ and ε_t, let $\tilde{A}_t^* \equiv (\Phi(Z_t)^{*\prime}\Phi(Z_t))^{-1}\Phi(Z_t)^{*\prime}A_t$. 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

$$\tilde{Y}^* = a1_T' + BF' + \tilde{\Delta}^* + \tilde{E}^*, \tag{C.58}$$

where 1_T denotes a $T \times 1$ vector of ones. Recall $M_T = I_T - 1_T 1_T'/T$. Post-multiplying (C.58) by M_T to remove a, we thus obtain

$$\tilde{Y}^* M_T = B(M_T F)' + \tilde{\Delta}^* M_T + \tilde{E}^* M_T. \tag{C.59}$$

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 definitions of $\hat{B}^*$, $\hat{B}^* = \tilde{Y}^*M_T\hat{F}(\hat{F}'M_T\hat{F})^{-1}$. We may substitute (C.59) to it to obtain

$$\hat{B}^* - BH = [(\tilde{\Delta}^* + \tilde{E}^*)M_T\tilde{Y}'/T]\hat{B}V^{-1} = \sum_{j=1}^{6} D_j^*\hat{B}V^{-1}, \tag{C.60}$$

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 (C.2) into the equation, and $D_1^* = \tilde{\Delta}^*M_TFB'/T$, $D_2^* = \tilde{\Delta}^*M_T\tilde{\Delta}'/T$, $D_3^* = \tilde{\Delta}^*M_T\tilde{E}'/T$, $D_4^* = \tilde{E}^*M_TFB'/T$, $D_5^* = \tilde{E}^*M_T\tilde{E}'/T$, and $D_6^* = \tilde{E}^*M_T\tilde{\Delta}'/T$. We can conduct the same exercise as in (C.39) to obtain

$$\begin{aligned}\|\sqrt{NT}(\hat{B}^* - BH) - \sqrt{NT}D_4^*\hat{B}V^{-1}\|_F &\leq \sqrt{NT}\|V^{-1}\|_2\|D_5^*\hat{B}\|_F \\ &+ \sqrt{NT}\|\hat{B}\|_2\|V^{-1}\|_2\sum_{j\neq 4,5}^{6}\|D_j^*\|_F = O_p\left(\frac{\sqrt{NT}}{J^\kappa} + \frac{\sqrt{TJ}}{\sqrt{N}}\right),\end{aligned} \tag{C.61}$$

where the equality follows by $J = o(\sqrt{N})$, Lemmas C.18 and C.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 (C.61) and Lemma C.19 implies

$$\|\sqrt{NT}(\hat{B}^* - BH) - \mathcal{L}_{NT}^{**}B'B\mathcal{M}\|_F = O_p\left(\frac{\sqrt{NT}}{J^\kappa} + \frac{\sqrt{TJ}}{\sqrt{N}} + \frac{\sqrt{J\xi_J}\log^{1/4}J}{N^{1/4}}\right). \tag{C.62}$$

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.40) and (C.62) to obtain

$$\|\sqrt{NT}(\hat{B}^* - \hat{B}) - \mathcal{L}^*_{NT}B'B\mathcal{M}\|_F = O_p\left(\frac{\sqrt{NT}}{J^\kappa} + \frac{\sqrt{TJ}}{\sqrt{N}} + \frac{\sqrt{J\xi_J}\log^{1/4} J}{N^{1/4}}\right). \quad \text{(C.63)}$$

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.63) and Lemmas C.5 and C.20. We now show the first result of the theorem. By the definition of $\hat{a}^*$,

$$\begin{aligned}\hat{a}^* - a = &-\hat{B}^*(\hat{B}^{*\prime}\hat{B}^*)^{-1}(\hat{B}^* - BH)'a + (I_{JM} - \hat{B}^*(\hat{B}^{*\prime}\hat{B}^*)^{-1}\hat{B}^{*\prime})(BH - \hat{B}^*)H^{-1}\bar{f} \\ &+ (I_{JM} - \hat{B}^*(\hat{B}^{*\prime}\hat{B}^*)^{-1}\hat{B}^{*\prime})\tilde{\Delta}^* 1_T/T + (I_{JM} - \hat{B}^*(\hat{B}^{*\prime}\hat{B}^*)^{-1}\hat{B}^{*\prime})\tilde{E}^* 1_T/T, \quad \text{(C.64)}\end{aligned}$$

where H^{-1} is well defined with probability approaching one by (C.4) and Lemma C.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.41)-(C.44), we have

$$\begin{aligned}&\|\sqrt{NT}(\hat{a}^* - a) - (I_{JM} - B\mathcal{H}\mathcal{H}'B')[\ell^{**}_{NT} - \sqrt{NT}(\hat{B}^* - BH)\mathcal{H}^{-1}\bar{f} \\ &\quad + B\mathcal{H}\sqrt{NT}(\hat{B}^* - BH)'a\| = O_p\left(\frac{\sqrt{NT}}{J^\kappa} + \frac{\sqrt{TJ}}{\sqrt{N}} + \frac{\sqrt{J\xi_J}\log^{1/4} J}{N^{1/4}}\right) \quad \text{(C.65)}\end{aligned}$$

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.44) and (C.65) to obtain

$$\begin{aligned}&\|\sqrt{NT}(\hat{a}^* - \hat{a}) - (I_{JM} - B\mathcal{H}\mathcal{H}'B')[\ell^*_{NT} - \sqrt{NT}(\hat{B}^* - \hat{B})\mathcal{H}^{-1}\bar{f} \\ &\quad + B\mathcal{H}\sqrt{NT}(\hat{B}^* - \hat{B})'a\| = O_p\left(\frac{\sqrt{NT}}{J^\kappa} + \frac{\sqrt{TJ}}{\sqrt{N}} + \frac{\sqrt{J\xi_J}\log^{1/4} J}{N^{1/4}}\right). \quad \text{(C.66)}\end{aligned}$$

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.66), the second result of the theorem, and Lemmas C.5 and C.20. This completes the proof of the theorem. ■

C.4.1 Technical Lemmas

Lemma C.18. *Let $D_1^*, D_2^*, D_3^*, D_5^*, D_6^*$ be given in the proof of Theorem 4.3.*
(i) Under Assumptions B.2(i), (ii), (iv), B.7(i), and (ii), $\|D_1^*\|_F^2 = O_p(J^{-2\kappa})$.
(ii) Under Assumptions B.1(i), B.2(ii), (iv), B.7(i), and (ii), $\|D_2^*\|_F^2 = O_p(J^{-4\kappa})$.
(iii) Under Assumptions B.1, B.2(ii), (iv), B.3, B.7(i), (ii), $\|D_3^*\|_F^2 = O_p(J^{-2\kappa}J/N)$.
(iv) Assume (i) $N \to \infty$; (ii) $T \geq K+1$; (iii) $J \to \infty$ with $J^2\xi_J^2 \log J = o(N)$. Under Assumptions B.1-B.5, B.7(i), and (ii), $\|D_5^*\hat{B}\|_F^2 = O_p(J/N^2)$.
(v) Under Assumptions B.1, B.2(ii), (iv), B.3, B.7(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 B.2(i) and (ii), as well as Lemma C.21(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 C.3(i) and C.21(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 C.3(ii) and C.21(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 C.7(i) and C.21(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 C.3(i) and C.21(ii). ■

Lemma C.19. *Suppose Assumptions B.1-B.5, B.6(i), (ii), B.7(i), and (ii) hold. Let V be given in the proof of Theorem C.1, and $\tilde{E}^*$, D_4^*, ℓ_{NT}^{**}, and $\mathcal{L}_{NT}^{**}$ be given in the proof of Theorem 4.3. Assume (i) $N \to \infty$; (ii) $T \geq K+1$; (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.15.

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

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

$$+\sqrt{N/T}\tilde{E}^*M_TFB'B(HV^{-1}-\mathcal{M})\equiv\mathcal{T}_1+\mathcal{T}_2+\mathcal{T}_3. \tag{C.67}$$

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

$$\begin{aligned}\|\mathcal{T}_1-\mathcal{L}^{**}_{NT}B'B\mathcal{M}\|_F &\leq \|B\|_2^2\|\mathcal{M}\|_2\left\|\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}(\hat{Q}_t^{*-1}-Q_t^{-1})\Phi(Z_t)^{*\prime}\varepsilon_t f_t'\right\|_F\\ &\quad+\|B\|_2^2\|\mathcal{M}\|_2\|\bar{f}\|\left\|\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}(\hat{Q}_t^{*-1}-Q_t^{-1})\Phi(Z_t)^{*\prime}\varepsilon_t\right\|\\ &=O_p\left(\frac{\sqrt{J\xi_J}\log^{1/4}J}{N^{1/4}}\right),\end{aligned} \tag{C.68}$$

where the equality follows from Assumptions B.2(i) and (ii), along with Lemma C.23. The latter two follow by a similar argument. The second result also follows by a similar argument as in (C.68). This completes the proof of the lemma. ■

Lemma C.20. *Suppose Assumptions B.2(ii), B.3(i), (ii), B.5(i)-(iii), B.6(ii), (iii), B.7(i), and (iii) hold. Let ℓ^*_{NT} and $\mathcal{L}^*_{NT}$ be given in the proof of Theorem 4.3. Assume $J=o(\sqrt{N})$. Then there exists a $JM\times(K+1)$ random matrix $\mathbb{N}^*$ with $\text{vec}(\mathbb{N}^*)\sim N(0,\Omega)$ conditional on $\{Y_t,Z_t\}_{t\leq T}$ 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 $\text{vec}((\ell^*_{NT},\mathcal{L}^*_{NT}))=\sum_{i=1}^{N}\zeta_i$. Let E_w denote the expectation with respect to $\{w_i\}_{i\leq N}$. Then conditional on $\{Y_t,Z_t\}_{t\leq T}$, $E_w[\zeta_i]=0$ and $\zeta_1,\ldots,\zeta_N$ are independent by Assumption B.7(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 $E_w[\text{vec}((\ell^*_{NT},\mathcal{L}^*_{NT}))\text{vec}((\ell^*_{NT},\mathcal{L}^*_{NT}))']=\omega_0\Omega_{NT}$. We now apply Lemma C.17 to the independent random vectors $\zeta_1,\ldots,\zeta_N$ conditional on $\{Y_t,Z_t\}_{t\leq T}$. There exists a $JM\times(K+1)$ random matrix $\mathbb{N}^{**}$ with $\text{vec}(\mathbb{N}^{**})\sim N(0,\Omega_{NT})$ conditional on $\{Y_t,Z_t\}_{t\leq T}$ such that the following holds:

$$\|\text{vec}((\ell^*_{NT},\mathcal{L}^*_{NT}))-\sqrt{\omega_0}\text{vec}(\mathbb{N}^{**})\|=O_{p^*}\left((J\beta)^{1/3}\right), \tag{C.69}$$

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

$$E[\|\zeta_i\|^4] = E[(w_1-1)^4]\frac{1}{N^2T^2}E\left[\left(\sum_{t=1}^T\sum_{s=1}^T \phi(z_{it})'Q_t^{-1}Q_s^{-1}\phi(z_{is})f_t^{\dagger\prime}f_s^\dagger\varepsilon_{it}\varepsilon_{is}\right)^2\right]$$
$$\leq C_{NT}\max_{i\leq N,t\leq T}E[\|\phi(z_{it})\|^4]\frac{1}{N^2T^2}\sum_{t=1}^T\sum_{s=1}^T\sum_{u=1}^T\sum_{v=1}^T|E[\varepsilon_{it}\varepsilon_{is}\varepsilon_{iu}\varepsilon_{iv}]|$$
$$\leq C_{NT}\max_{m\leq M,j\leq J,i\leq N,t\leq T}E[\phi_j^4(z_{it,m})]\frac{J^2M^2}{N^2T^2}\sum_{t=1}^T\sum_{s=1}^T\sum_{u=1}^T\sum_{v=1}^T|E[\varepsilon_{it}\varepsilon_{is}\varepsilon_{iu}\varepsilon_{iv}]|, \quad \text{(C.70)}$$

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 from the independence in Assumption B.3(i) and the Cauchy-Schwartz inequality, as well as 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 one follows by $\max_{i\leq N,t\leq T}E[\|\phi(z_{it})\|^4] \leq J^2M^2\max_{m\leq M,j\leq J,i\leq N,t\leq T}E[\phi_j^4(z_{it,m})]$. Thus,

$$\beta = \sum_{i=1}^N E[\|\zeta_i\|^3] \leq \sum_{i=1}^N (E[\|\zeta_i\|^4])^{3/4} = O\left(\frac{J^{3/2}}{\sqrt{N}}\right), \quad \text{(C.71)}$$

where the inequality follows by the LiapounovâĂŹs inequality, and the last equality follows from (C.70) and Assumptions B.2(ii), B.5(i), (ii), B.6(iii), and B.7(i). We now may combine (C.69), (C.71), and Lemma C.5 to obtain

$$\|\text{vec}((\ell_{NT}^*, \mathcal{L}_{NT}^*)) - \sqrt{\omega_0}\text{vec}(\mathbb{N}^{**})\| = O_p\left(\frac{J^{5/6}}{N^{1/6}}\right). \quad \text{(C.72)}$$

By Assumption B.7(iii) and Lemma C.25, $\Omega_{NT}^{-1/2}$ 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

$$\|\text{vec}((\ell_{NT}^*, \mathcal{L}_{NT}^*)) - \sqrt{\omega_0}\mathbb{N}^*\|_F \leq \|\text{vec}((\ell_{NT}^*, \mathcal{L}_{NT}^*)) - \sqrt{\omega_0}\text{vec}(\mathbb{N}^{**})\|$$
$$+ \sqrt{\omega_0}\|\text{vec}(\mathbb{N}^*) - \text{vec}(\mathbb{N}^{**})\| = O_p\left(\frac{J^{5/6}}{N^{1/6}} + \frac{J^{3/2}}{\sqrt{N}}\right) = O_p\left(\frac{J^{5/6}}{N^{1/6}}\right), \quad \text{(C.73)}$$

where the first equality follows by (C.72) 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 C.25. This completes the proof of the lemma. ■

Lemma C.21. *Let $\tilde{\Delta}^*$ and $\tilde{E}^*$ be given in the proof of Theorem 4.3.*
(i) Under Assumptions B.2(ii), (iv), B.7(i), and (ii), $\|\tilde{\Delta}^*\|_F^2/T = O_p(J^{-2\kappa})$.

(ii) Under Assumptions B.1(ii), B.3, B.7(i), and (ii), $\|\tilde{E}^*\|_F^2/T = O_p(J/N)$.

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

$$\begin{aligned}\frac{1}{T}\|\tilde{\Delta}^*\|_F^2 &= \frac{1}{T}\sum_{t=1}^{T}\|(\Phi(Z_t)^{*\prime}\Phi(Z_t))^{-1}\Phi(Z_t)^{*\prime}(R(Z_t)+\Delta(Z_t)f_t)\|^2 \\ &\leq 2\max_{t\leq T}\|f_t\|^2\left(\min_{t\leq T}\lambda_{\min}(\hat{Q}_t^*)\right)^{-1}\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}w_i\|\delta(z_{it})\|^2 \\ &\quad + 2\left(\min_{t\leq T}\lambda_{\min}(\hat{Q}_t^*)\right)^{-1}\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}w_i|r(z_{it})|^2 = O_p\left(\frac{1}{J^{2\kappa}}\right), \qquad \text{(C.74)}\end{aligned}$$

where the last equality follows from Assumptions B.2(ii) and B.7(ii), along with Lemma C.22(ii).

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

$$\begin{aligned}\frac{1}{T}\|\tilde{E}^*\|_F^2 &= \frac{1}{T}\sum_{t=1}^{T}\|(\Phi(Z_t)^{*\prime}\Phi(Z_t))^{-1}\Phi(Z_t)^{*\prime}\varepsilon_t\|^2 \\ &\leq \left(\min_{t\leq T}\lambda_{\min}(\hat{Q}_t^*)\right)^{-2}\frac{1}{N^2T}\sum_{t=1}^{T}\|\Phi(Z_t)^{*\prime}\varepsilon_t\|^2 = O_p\left(\frac{J}{N}\right), \qquad \text{(C.75)}\end{aligned}$$

where the last equality follows from Assumption B.7(ii) and Lemma C.22(i). ■

Lemma C.22. *(i) Under Assumptions B.1(ii), B.3, and B.7(i),*

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

(ii) Under Assumption B.2(iv) and B.7(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

$$\begin{aligned}E\left[\sum_{t=1}^{T}\|\Phi(Z_t)^{*\prime}\varepsilon_t\|^2\right] &= E\left[\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\phi(z_{it})'\phi(z_{jt})\varepsilon_{it}\varepsilon_{jt}w_iw_j\right] \\ = \sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}E[\phi(z_{it})'\phi(z_{jt})]E[\varepsilon_{it}\varepsilon_{jt}]E[w_iw_j]&\end{aligned}$$

$$\leq E[w_1^2] \max_{i\leq N,t\leq T} E[\|\phi(z_{it})\|^2] \sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N} |E[\varepsilon_{it}\varepsilon_{jt}]|$$

$$\leq TJME[w_1^2] \max_{m\leq M,j\leq J,i\leq N,t\leq T} E[\phi_j^2(z_{it,m})] \max_{t\leq T} \sum_{i=1}^{N}\sum_{j=1}^{N} |E[\varepsilon_{it}\varepsilon_{jt}]| = O(NTJ), \quad \text{(C.76)}$$

where the second equality follows by the independence in Assumptions B.3(i) and B.7(i), the first inequality is due to the Cauchy Schwartz inequality, the second one follows by $\max_{i\leq N,t\leq T} E[\|\phi(z_{it})\|^2] \leq JM \max_{m\leq M,j\leq J,i\leq N,t\leq T} E[\phi_j^2(z_{it,m})]$, and the last equality follows from Assumptions B.1(ii), B.3(iii), and B.7(i).

(iii) The first result follows since

$$\sum_{t=1}^{T}\sum_{i=1}^{N} w_i\|\delta(z_{it})\|^2 \leq TKM^2 \max_{k\leq K,m\leq M} \sup_z |\delta_{km,J}(z)|^2 \sum_{i=1}^{N} w_i = O_p(NTJ^{-2\kappa}), \quad \text{(C.77)}$$

where the inequality follows since w_i's are positive and $\max_{i\leq N,t\leq T} \|\delta(z_{it})\|^2 \leq M^2K \sup_{k\leq K,m\leq M} \sup_z |\delta_{km,J}(z)|^2$, and the equality follows by the law of large numbers and Assumptions B.2(iv) and B.7(i). The proof of the second result is similar. ■

Lemma C.23. *Suppose Assumptions B.2(ii), B.3(i), (ii), B.5, B.6(ii), B.7(i), and (ii) hold. Assume $J \geq 2$ and $\xi_J^2 \log J = o(N)$. Then*

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

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

$$E_\varepsilon[\|\mathcal{T}\|_F^2] = \frac{1}{NT} E_\varepsilon\left[\mathrm{tr}\left(\sum_{t=1}^{T}\sum_{s=1}^{T} (\hat{Q}_t^{*-1} - Q_t^{-1})\Phi(Z_t)^{*\prime}\varepsilon_t f_t' f_s \varepsilon_s' \Phi(Z_s)^* (\hat{Q}_s^{*-1} - Q_s^{-1}) \right) \right]$$

$$= \frac{1}{NT} \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} w_i \phi(z_{it})' (\hat{Q}_t^{*-1} - Q_t^{-1})(\hat{Q}_s^{*-1} - Q_s^{-1})\phi(z_{js}) w_j f_t' f_s E[\varepsilon_{it}\varepsilon_{js}]$$

$$= \frac{1}{NT} \sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} w_i^2 \phi(z_{it})' (\hat{Q}_t^{*-1} - Q_t^{-1})(\hat{Q}_s^{*-1} - Q_s^{-1})\phi(z_{is}) f_t' f_s E[\varepsilon_{it}\varepsilon_{is}]$$

$$\leq C^*_{NT}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\|\hat{Q}^*_t - Q_t\|_2 w_i^2\|\phi(z_{it})\|\|\phi(z_{is})\||E[\varepsilon_{it}\varepsilon_{is}]|$$

$$\leq C^{**}_{NT}\frac{1}{NT}\left(\sum_{t=1}^{T}\left(\sum_{i=1}^{N}\sum_{s=1}^{T} w_i^2\|\phi(z_{it})\|\|\phi(z_{is})\||E[\varepsilon_{it}\varepsilon_{is}]|\right)^2\right)^{1/2}, \tag{C.78}$$

where $C^*_{NT} = (\min_{t\leq T}\lambda_{\min}(\hat{Q}^*_t))^{-1}[(\min_{t\leq T}\lambda_{\min}(\hat{Q}^*_t))^{-1} + (\min_{i\leq N,t\leq T}\lambda_{\min}(Q_{it}))^{-1}]$ $\times(\min_{i\leq N,t\leq T}\lambda_{\min}(Q_{it}))^{-1}\max_{t\leq T}\|f_t\|^2$ and $C^{**}_{NT} = C^*_{NT}(\sum_{t=1}^{T}\|\hat{Q}^*_t - Q_t\|_2^2)^{1/2}$, the second equality follows from the independence in Assumptions B.3(i) and B.7(i) and the linearity of both expectation and trace operators, the third equality follows by Assumption B.3(ii) and the independence in Assumption B.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

$$E\left[\sum_{t=1}^{T}\left(\sum_{i=1}^{N}\sum_{s=1}^{T} w_i^2\|\phi(z_{it})\|\|\phi(z_{is})\||E[\varepsilon_{it}\varepsilon_{is}]|\right)^2\right]$$

$$\leq E[w_1^4]\max_{i\leq N,t\leq T}E[\|\phi(z_{it})\|^4]\sum_{t=1}^{T}\left(\sum_{i=1}^{N}\sum_{s=1}^{T}|E[\varepsilon_{it}\varepsilon_{is}]|\right)^2$$

$$\leq J^2M^2E[w_1^4]\max_{m\leq M,j\leq J,i\leq N,t\leq T}E[\phi_j^4(z_{it,m})]\sum_{t=1}^{T}\left(\sum_{i=1}^{N}\sum_{s=1}^{T}|E[\varepsilon_{it}\varepsilon_{is}]|\right)^2, \tag{C.79}$$

where the first inequality follows by the Cauchy-Schwartz inequality and the independence in Assumption B.7(i), the second one 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 B.2(ii), B.5(ii), B.7(ii), and Lemma C.24, $C^{**}_{NT} = O_p(\sqrt{T}\xi_J\sqrt{\log J}/\sqrt{N})$. Combining this, (C.78) and (C.79) implies that $E_\varepsilon[\|\mathcal{T}\|_F^2] = O_p(J\xi_J\sqrt{\log J}/\sqrt{N})$ by Assumptions B.5(i), (iv), and B.7(i). Thus, the result of the lemma follows by the Markov's inequality and Lemma C.5. The proof of the second result is similar. ■

Lemma C.24. *Suppose Assumptions B.5(ii), (iii), and B.7(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 C.11, thus omitted for brevity. ■

Lemma C.25. *Suppose Assumptions B.2(ii), B.3(i), (ii), B.5(i)-(iii), B.6(ii), (iii),*

and B.7(iii) hold. Let Ω_{NT} be given in the proof of Lemma C.20. 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 $\|A\|_F^2 = \text{tr}(AA')$,

$$\begin{aligned} E[\|\Omega_{NT} - \Omega\|_F^2] &= E\left[\text{tr}\left(\sum_{i=1}^N\sum_{j=1}^N(\zeta_i\zeta_i' - E[\zeta_i\zeta_i'])(\zeta_j\zeta_j' - E[\zeta_j\zeta_j'])'\right)\right] \\ &= \sum_{i=1}^N\left(E[(\zeta_i'\zeta_i)^2] - \|E[\zeta_i\zeta_i']\|_F^2\right) \leq N\max_{i\leq N}E[\|\zeta_i\|^4] = O\left(\frac{J^2}{N}\right), \end{aligned} \tag{C.80}$$

where the second equality follows because $\zeta_1, \ldots, \zeta_N$ are independent by Assumptions B.3(i), B.5(iii), and B.6(ii), and because 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.48) and Assumptions B.2(ii), B.5(i), (ii), and B.6(iii). Thus, $\|\Omega_{NT} - \Omega\|_F^2 = O_p(J^2/N)$ follows from (C.80) by the Markov's inequality. The result of the lemma follows from Assumption B.7(iii) and Lemma A.2 of Belloni et al. (2015). ■

C.5 Proof of Theorem 4.4

PROOF OF THEOREM 4.4: In order to show the first result, we assume that 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

$$\begin{aligned} \mathcal{S} &= \frac{1}{J}\sum_{i=1}^N\sum_{t=1}^T |(\hat{\gamma} - \gamma)'z_{it} - (\hat{a} - a)'\phi(z_{it}) + r(z_{it})|^2 \\ &\quad + \frac{1}{J}\sum_{i=1}^N\sum_{t=1}^T \|(\hat{\Gamma} - \Gamma H)'z_{it} - (\hat{B} - BH)'\phi(z_{it}) + H'\delta(z_{it})\|^2 \\ &= \frac{1}{J}\sum_{i=1}^N\sum_{t=1}^T |(\hat{\gamma} - \gamma)'z_{it} - (\hat{a} - a)'\phi(z_{it})|^2 + \mathcal{S}_1 + 2\mathcal{S}_2 + 2\mathcal{S}_3 \\ &\quad + \frac{1}{J}\sum_{i=1}^N\sum_{t=1}^T \|(\hat{\Gamma} - \Gamma H)'z_{it} - (\hat{B} - BH)'\phi(z_{it})\|^2 + \mathcal{S}_4 + 2\mathcal{S}_5 + 2\mathcal{S}_6, \end{aligned} \tag{C.81}$$

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 C.26, (C.81) implies

$$\begin{aligned} &\mathcal{S} - \frac{1}{J}\mathcal{W}_{NT,a}'\hat{\mathcal{Q}}\mathcal{W}_{NT,a} - \frac{1}{J}\mathrm{tr}(\mathcal{W}_{NT,B}'\hat{\mathcal{Q}}\mathcal{W}_{NT,B}) \\ &= \mathcal{S} - \frac{1}{J}\mathrm{tr}(\mathcal{W}_{NT}'\hat{\mathcal{Q}}\mathcal{W}_{NT}) = O_p\left(\frac{\sqrt{NT}}{J^{\kappa+1/2}}\right). \end{aligned} \tag{C.82}$$

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 C.27. By Lemmas C.27 and C.28 and Theorem 4.2, (C.82) implies

$$\begin{aligned} &\mathcal{S} - \frac{1}{J}\mathbb{W}_a'\mathcal{Q}\mathbb{W}_a - \frac{1}{J}\mathrm{tr}(\mathbb{W}_B'\mathcal{Q}\mathbb{W}_B) \\ &= \mathcal{S} - \frac{1}{J}\mathrm{tr}(\mathbb{W}'\mathcal{Q}\mathbb{W}) = O_p\left(\frac{\sqrt{NT}}{J^{\kappa+1/2}} + \frac{J^{1/3}}{N^{1/6}} + \frac{\sqrt{\xi_J}\log^{1/4} J}{N^{1/4}} + \sqrt{\frac{T}{N}}\right). \end{aligned} \tag{C.83}$$

Let $\mathcal{W}_{NT,a}^* \equiv (\sqrt{NT/\omega_0}(\hat{\gamma}^* - \hat{\gamma})', -\sqrt{NT/\omega_0}(\hat{a}^* - \hat{a})')'$, $\mathcal{W}_{NT,B}^* \equiv (\sqrt{NT/\omega_0}(\hat{\Gamma}^* - \hat{\Gamma})', -\sqrt{NT/\omega_0}(\hat{B}^* - \hat{B})')'$, $\mathcal{W}_{NT}^* \equiv (\mathcal{W}_{NT,a}^*, \mathcal{W}_{NT,B}^*)$, $\mathbb{W}_a^* \equiv (\mathbb{G}_\gamma^{*\prime}, -\mathbb{G}_a^{*\prime})'$, $\mathbb{W}_B^* \equiv (\mathbb{G}_\Gamma^{*\prime}, -\mathbb{G}_B^{*\prime})'$, and $\mathbb{W}^* \equiv (\mathbb{W}_a^*, \mathbb{W}_B^*)$, where $\mathbb{G}_\gamma^*$ and $\mathbb{G}_\Gamma^*$ are given in Lemma C.29. Then (21) can be written as $\mathcal{S}^* = \mathcal{W}_{NT,a}^{*\prime}\hat{\mathcal{Q}}\mathcal{W}_{NT,a}^*/J + \mathrm{tr}(\mathcal{W}_{NT,B}^{*\prime}\hat{\mathcal{Q}}\mathcal{W}_{NT,B}^*)/J = \mathrm{tr}(\mathcal{W}_{NT}^{*\prime}\hat{\mathcal{Q}}\mathcal{W}_{NT}^*)/J$. By Lemmas C.5, C.28, and C.29, along with Theorem 4.3,

$$\begin{aligned} &\mathcal{S}^* - \frac{1}{J}\mathbb{W}_a^{*\prime}\mathcal{Q}\mathbb{W}_a^* - \frac{1}{J}\mathrm{tr}(\mathbb{W}_B^{*\prime}\mathcal{Q}\mathbb{W}_B^*) \\ &= \mathcal{S}^* - \frac{1}{J}\mathrm{tr}(\mathbb{W}^{*\prime}\mathcal{Q}\mathbb{W}^*) = O_p\left(\frac{\sqrt{NT}}{J^{\kappa+1/2}} + \frac{J^{1/3}}{N^{1/6}} + \frac{\sqrt{\xi_J}\log^{1/4} J}{N^{1/4}} + \sqrt{\frac{T}{N}}\right). \end{aligned} \tag{C.84}$$

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 (C.84), Lemma A.1 of Belloni et al. (2015) implies that there exists a sequence $\{\nu_{NT}\}$ such that $\nu_{NT} = o(1)$ and

$$P(c_{1-\alpha} < c_{0,1-\alpha-\nu_{NT}} - \gamma_{NT}/\sqrt{J}) = o(1), \tag{C.85}$$

$$P(c_{1-\alpha} > c_{0,1-\alpha+\nu_{NT}} + \gamma_{NT}/\sqrt{J}) = o(1). \tag{C.86}$$

Note that $\text{tr}(\mathbb{W}'\mathcal{Q}\mathbb{W}) = \text{vec}(\mathbb{W})'(I_K \otimes \mathcal{Q})\text{vec}(\mathbb{W})$. Since $\mathcal{Q}$ has rank not smaller than $JM - M$ and the variance of $\text{vec}(\mathbb{G}_B)$ has full rank, $\text{tr}(\mathbb{W}'\mathcal{Q}\mathbb{W})$ 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

$$\begin{aligned} P(\mathcal{S} \leq c_{1-\alpha}) &\leq P(\text{tr}(\mathbb{W}'\mathcal{Q}\mathbb{W})/J \leq c_{1-\alpha} + \gamma_{NT}/\sqrt{J}) + o(1) \\ &\leq P(\text{tr}(\mathbb{W}'\mathcal{Q}\mathbb{W})/J \leq c_{0,1-\alpha+\nu_{NT}} + 2\gamma_{NT}/\sqrt{J}) + o(1) \\ &\leq P(\text{tr}(\mathbb{W}'\mathcal{Q}\mathbb{W})/\sqrt{J} \leq \sqrt{J}c_{0,1-\alpha+\nu_{NT}} + 2\gamma_{NT}) + o(1) \\ &\leq P(\text{tr}(\mathbb{W}'\mathcal{Q}\mathbb{W})/\sqrt{J} \leq \sqrt{J}c_{0,1-\alpha+\nu_{NT}}) + o(1) \\ &\leq 1 - \alpha + \nu_{NT} + o(1) = 1 - \alpha + o(1), \end{aligned} \tag{C.87}$$

where the first inequality follows since $P(|\mathcal{S} - \text{tr}(\mathbb{G}'Q\mathbb{G})/J| > \gamma_{NT}/\sqrt{J}) = o(1)$ due to (C.83), the second inequality follows from (C.86), and the fourth inequality follows since $\gamma_{NT} = o(1)$ and $\text{tr}(\mathbb{W}'\mathcal{Q}\mathbb{W})$ is bounded by chi-squared random variables. By a similar argument, $P(\mathcal{S} > c_{1-\alpha}) \leq 1-\alpha+o(1)$. Therefore, the first result of the theorem follows. To show the second result, we now assume that H_1 is true. Since $(x+y)^2 \geq x^2/2 - y^2$,

$$\begin{aligned} \frac{2J}{NT}\mathcal{S} \geq \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\|\hat{\Gamma}'z_{it} - H'\beta(z_{it})\|^2 - \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\|\hat{\beta}(z_{it}) - H'\beta(z_{it})\|^2 \\ + \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}|\hat{\gamma}'z_{it} - \hat{\alpha}(z_{it})|^2 \geq c_0 + o_p(1) \text{ for some } c_0 > 0, \end{aligned} \tag{C.88}$$

where the second inequality follows from Lemmas C.30 and C.31. We have

$$\begin{aligned} &\frac{2J}{NT}\mathcal{S}^* \leq \frac{4}{NT\omega_0}\sum_{i=1}^{N}\sum_{t=1}^{T}|(\hat{\gamma}^* - \hat{\gamma})'z_{it}|^2 + \frac{4}{NT\omega_0}\sum_{i=1}^{N}\sum_{t=1}^{T}|(\hat{a}^* - \hat{a})'\phi(z_{it})|^2 \\ &+ \frac{4}{NT\omega_0}\sum_{i=1}^{N}\sum_{t=1}^{T}\|(\hat{\Gamma}^* - \hat{\Gamma})'z_{it}\|^2 + \frac{4}{NT\omega_0}\sum_{i=1}^{N}\sum_{t=1}^{T}\|(\hat{B}^* - \hat{B})'\phi(z_{it})\|^2 = o_p(1), \end{aligned} \tag{C.89}$$

where the equality follows from Lemma C.32. In view of (C.89), Lemma A.1 of Belloni et al. (2015) implies that $2c_{1-\alpha}J/(NT) = o_p(1)$. This together with (C.88) thus concludes the second result of the theorem. ■

C.5.1 Technical Lemmas

Lemma C.26. *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 4.4. Assume (i) $N \to \infty$; (ii) $T \geq K+1$ and $T = o(N)$; (iii) $J \to \infty$ with $J^2\xi_J^2 \log J = o(N)$ and $NTJ^{-(2\kappa+1)} = o(1)$. Assume that H_0 is true.*
(i) Under Assumption B.2(iv), $\mathcal{S}_1 = O_p(NTJ^{-(2\kappa+1)})$.
(ii) Under Assumptions B.1-B.6, B.8(i)-(iii), $\mathcal{S}_2 = O_p(\sqrt{NT}J^{-(\kappa+1)})$.
(iii) Under Assumptions B.1-B.5, $\mathcal{S}_3 = O_p(\sqrt{NT}J^{-(\kappa+1/2)})$.
(iv) Under Assumptions B.1-B.3, $\mathcal{S}_4 = O_p(NTJ^{-(2\kappa+1)})$.
(v) Under Assumptions B.1-B.6, B.8(i)-(iii), $\mathcal{S}_5 = O_p(\sqrt{NT}J^{-(\kappa+1)})$.
(vi) Under Assumptions B.1-B.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

$$\mathcal{S}_4 \leq \|H\|_2^2 \sum_{i=1}^{T}\sum_{t=1}^{T} \|\delta(z_{it})\|^2/J \leq (NT/J)\|H\|_2^2 M^2 K \sup_{k\leq K, m\leq M} \sup_z |\delta_{km,J}(z)|^2, \tag{C.90}$$

where the second inequality follows by $\max_{i\leq N, t\leq T}\|\delta(z_{it})\|^2 \leq M^2K \sup_{k\leq K,m\leq M} \sup_z |\delta_{km,J}(z)|^2$. Thus, the result of the lemma follows from (C.90), Assumption B.2(iv), and Lemma C.2(i).

(v) By Assumption B.8(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

$$\begin{aligned}\frac{1}{J}\sum_{i=1}^{T}\sum_{t=1}^{T}\|(\hat{\Gamma}-\Gamma H)'z_{it}\|^2 &\leq \|\hat{\Gamma}-\Gamma H\|_F^2 \frac{1}{J}\sum_{i=1}^{N}\sum_{t=1}^{T}\|z_{it}\|^2 \\ &= O_p\left(\frac{1}{J} + \frac{T}{J^{2\kappa+1}} + \frac{T}{NJ}\right) = O_p\left(\frac{1}{J}\right),\end{aligned} \tag{C.91}$$

where the first equality follows from Lemma C.27, 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,

$$|\mathcal{S}_5| \leq \mathcal{S}_4^{1/2}\left(\frac{1}{J}\sum_{i=1}^{T}\sum_{t=1}^{T}\|(\hat{\Gamma}-\Gamma H)'z_{it}\|^2\right)^{1/2}. \tag{C.92}$$

Thus, the result of the lemma follows from (C.91), (C.92), and Lemma C.26(iv).

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

$$\frac{1}{J}\sum_{i=1}^{T}\sum_{t=1}^{T}\|(\hat{B}-BH)'\phi(z_{it})\|^2 = \frac{N}{J}\sum_{t=1}^{T}\mathrm{tr}\left((\hat{B}-BH)'\hat{Q}_t(\hat{B}-BH)\right)$$
$$\leq \frac{NT}{J}\max_{t\leq T}\lambda_{\max}(\hat{Q}_t)\|\hat{B}-BH\|_F^2 = O_p\left(\frac{NT}{J^{2\kappa+1}}+\frac{T}{N}+1\right) = O_p(1), \tag{C.93}$$

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

$$|\mathcal{S}_6| \leq \mathcal{S}_4^{1/2}\left(\frac{1}{J}\sum_{i=1}^{T}\sum_{t=1}^{T}\|(\hat{B}-BH)'\phi(z_{it})\|^2\right)^{1/2}. \tag{C.94}$$

Thus, the result follows from (C.93), (C.94), and Lemma C.26(iv). ■

Lemma C.27. *Suppose Assumptions B.1-B.6 and B.8(i)-(iii) hold. Let $\hat{\gamma}$ and $\hat{\Gamma}$ be given in Section 4.3. Assume (i) $N \to \infty$; (ii) $T \geq K+1$; (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 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 4.2, 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 $\vec{E} \equiv (\vec{\varepsilon}_1,\ldots,\vec{\varepsilon}_T)$. Then (9) under H_0 can be written as

$$\vec{Y} = \gamma 1_T' + \Gamma F' + \vec{E}, \tag{C.95}$$

where 1_T denotes a $T \times 1$ vector of ones. Recall $M_T = I_T - 1_T 1_T'/T$. Post-multiplying (C.95) by M_T to remove γ, we thus obtain

$$\vec{Y} M_T = \Gamma (M_T F)' + \vec{E} M_T. \tag{C.96}$$

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 C.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 C.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 (C.96) to $\hat{\Gamma} = \vec{Y} M_T \hat{F} (\hat{F}' M_T \hat{F})^{-1}$ to obtain

$$\hat{\Gamma} - \Gamma H = (\vec{E} M_T \tilde{Y}'/T) \hat{B} V^{-1} = \sum_{j=1}^{3} \mathcal{D}_j \hat{B} V^{-1}, \tag{C.97}$$

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 (C.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.39) to obtain

$$\begin{aligned} &\|\sqrt{NT}(\hat{\Gamma} - \Gamma H) - \sqrt{NT}\mathcal{D}_1 \hat{B} V^{-1}\|_F \\ &\leq \sqrt{NT} \|V^{-1}\|_2 (\|\mathcal{D}_2 \hat{B}\|_F + \|\mathcal{D}_3\|_F \|\hat{B}\|_2) = O_p\left(\frac{\sqrt{T}}{J^\kappa} + \frac{\sqrt{T}}{\sqrt{N}}\right), \end{aligned} \tag{C.98}$$

where the equality follows by Lemmas C.2(i) and C.33. Thus, the second result of the lemma follows from (C.98) and Lemma C.34. We now show the first result of the lemma. By the definition of $\hat{\gamma}$,

$$\begin{aligned} \hat{\gamma} - \gamma = \vec{E} 1_T/T + (\Gamma H - \hat{\Gamma}) H^{-1} \bar{f} - \hat{\Gamma}(\hat{B} - BH)' a \\ - \hat{\Gamma} \hat{B}' (BH - \hat{B}) H^{-1} \bar{f} - \hat{\Gamma} \hat{B}' \tilde{E} 1_T/T - \hat{\Gamma} \hat{B}' \tilde{\Delta} 1_T/T, \end{aligned} \tag{C.99}$$

where H^{-1} is well defined with probability approaching one by (C.4) and Lemma C.2(ii), and we have used $a' B = 0$ and $\hat{B}' \hat{B} = I_K$. By a similar argument as in (C.41)-(C.43),

$$\begin{aligned} &\|\sqrt{NT}(\hat{\gamma} - \gamma) - [\sqrt{N/T} \vec{E} 1_T - \sqrt{NT}(\hat{\Gamma} - \Gamma H) \mathcal{H}^{-1} \bar{f}] \\ &\quad + \Gamma \mathcal{H} \mathcal{H}' B' [\sqrt{N/T} \tilde{E} 1_T - \sqrt{NT}(\hat{B} - BH) \mathcal{H}^{-1} \bar{f}] \\ &\quad + \Gamma \mathcal{H} \sqrt{NT} (\hat{B} - BH)' a\| = O_p\left(\frac{\sqrt{NT}}{J^\kappa} + \frac{\sqrt{TJ}}{\sqrt{N}} + \frac{J}{\sqrt{NT}}\right). \end{aligned} \tag{C.100}$$

Thus, the first result of the lemma follows from (C.100), along with Lemmas C.34, C.13, C.14, and Theorem 4.2, as well as the second result of the lemma. ■

Lemma C.28. *Suppose Assumptions B.5(i), (iii), and B.8(ii) hold. Let $\hat{\mathcal{Q}}$ and $\mathcal{Q}$ be given in the proof of Theorem 4.4. 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 B.5(iii). By the Cauchy-Schwartz inequality,

$$E[\|\hat{\mathcal{Q}} - \mathcal{Q}\|_F^2] \leq \frac{1}{T}\sum_{t=1}^T E[\|\hat{\mathcal{Q}}_t - \mathcal{Q}_t\|_F^2] = O\left(\frac{J^2}{N}\right), \tag{C.101}$$

where the equality follows from Assumptions B.5(i) and B.8(ii). By the Markov's inequality, the result of the lemma thus follows from (C.101). ■

Lemma C.29. *Suppose Assumptions B.1-B.6, B.7, and B.8(ii)-(iv) hold. Let $\hat{\gamma}$, $\hat{\Gamma}$, $\hat{\gamma}^*$ and $\hat{\Gamma}^*$ be given in Section 4.3. Assume (i) $N \to \infty$; (ii) $T \geq K+1$; (iii) $J \to \infty$ with $J^2\xi_J^2 \log J = o(N)$. Assume that 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 Ω_z is given in Lemma C.27, $\mathbb{G}_\gamma^* = \mathbb{N}_{z,1}^* - \mathbb{G}_\Gamma^*\mathcal{H}^{-1}\bar{f} - \Gamma\mathcal{H}\mathcal{H}'B'(\mathbb{N}_1^* - \mathbb{G}_B^*\mathcal{H}^{-1}\bar{f}) - \Gamma\mathcal{H}\mathbb{G}_B^{*\prime}a$, $\mathbb{G}_\Gamma^* = \mathbb{N}_{z,2}^* B'B\mathcal{M}$, $\mathcal{H}$, $\mathcal{M}$, $\mathbb{N}_1^*$ and $\mathbb{G}_B^*$ are given in Theorem 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 $\vec{\varepsilon}_t^* \equiv (Z_t^{*\prime}Z_t)^{-1}Z_t^{*\prime}\varepsilon_t$ and $\vec{E}^* \equiv (\vec{\varepsilon}_1^*, \ldots, \vec{\varepsilon}_T^*)$. Then under H_0, we have

$$\vec{Y}^* = \gamma 1_T' + \Gamma F' + \vec{E}^*. \tag{C.102}$$

where 1_T denotes a $T \times 1$ vector of ones. Recall $M_T = I_T - 1_T 1_T'/T$. Post-multiplying (C.102) by M_T to remove γ, we thus obtain

$$\vec{Y}^* M_T = \Gamma (M_T F)' + \vec{E}^* M_T. \tag{C.103}$$

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 C.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 C.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 (C.103) to $\hat{\Gamma}^* = \vec{Y}^* M_T \hat{F} (\hat{F}' M_T \hat{F})^{-1}$ to obtain

$$\hat{\Gamma}^* - \Gamma H = (\vec{E}^* M_T \tilde{Y}'/T) \hat{B} V^{-1} = \sum_{j=1}^{3} \mathcal{D}_j^* \hat{B} V^{-1}, \tag{C.104}$$

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 (C.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.39) to obtain

$$\begin{aligned} &\|\sqrt{NT}(\hat{\Gamma}^* - \Gamma H) - \sqrt{NT}\mathcal{D}_1^* \hat{B} V^{-1}\|_F \\ &\leq \sqrt{NT}\|V^{-1}\|_2 (\|\mathcal{D}_2^* \hat{B}\|_F + \|\mathcal{D}_3^*\|_F \|\hat{B}\|_2) = O_p\left(\frac{\sqrt{T}}{J^\kappa} + \frac{\sqrt{T}}{\sqrt{N}}\right), \end{aligned} \tag{C.105}$$

where the equality follows by Lemmas C.2(i) and C.35. By the fact that $\|C + D\|_F \leq \|C\|_F + \|D\|_F$, we may combine (C.98) and (C.105) to obtain

$$\|\sqrt{NT}(\hat{\Gamma}^* - \hat{\Gamma}) - \sqrt{NT}(\mathcal{D}_1^* - \mathcal{D}_1)\hat{B} V^{-1}\|_F = O_p\left(\frac{\sqrt{T}}{J^\kappa} + \frac{\sqrt{T}}{\sqrt{N}}\right). \tag{C.106}$$

Thus, the second result of the lemma follows from (C.106) and Lemmas C.5 and C.36. We now show the first result of the lemma. By the definition of $\hat{\gamma}^*$,

$$\begin{aligned} \hat{\gamma}^* - \gamma = {} & \vec{E}^* 1_T/T + (\Gamma H - \hat{\Gamma}^*) H^{-1} \bar{f} - \hat{\Gamma}^* (\hat{B}^{*\prime} \hat{B}^*)^{-1} (\hat{B}^* - BH)' a \\ & - \hat{\Gamma}^* (\hat{B}^{*\prime} \hat{B}^*)^{-1} \hat{B}^{*\prime} (BH - \hat{B}^*) H^{-1} \bar{f} - \hat{\Gamma}^* (\hat{B}^{*\prime} \hat{B}^*)^{-1} \hat{B}^{*\prime} \tilde{E}^* 1_T/T \\ & - \hat{\Gamma}^* (\hat{B}^{*\prime} \hat{B}^*)^{-1} \hat{B}^{*\prime} \tilde{\Delta}^* 1_T/T, \end{aligned} \tag{C.107}$$

where H^{-1} is well defined with probability approaching one by (C.4) and Lemma C.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.41)-(C.43),

$$\begin{aligned}&\|\sqrt{NT}(\hat{\gamma}^* - \gamma) - [\sqrt{N/T}\vec{E}^* 1_T - \sqrt{NT}(\hat{\Gamma}^* - \Gamma H)\mathcal{H}^{-1}\bar{f}] \\ &\quad + \Gamma\mathcal{H}\mathcal{H}'B'[\sqrt{N/T}\tilde{E}^* 1_T - \sqrt{NT}(\hat{B}^* - BH)\mathcal{H}^{-1}\bar{f}] \\ &\quad + \Gamma\mathcal{H}\sqrt{NT}(\hat{B}^* - BH)'a\| = O_p\left(\frac{\sqrt{NT}}{J^\kappa} + \frac{\sqrt{TJ}}{\sqrt{N}} + \frac{J}{\sqrt{NT}}\right). \qquad \text{(C.108)}\end{aligned}$$

By the fact that $\|x+y\| \leq \|x\| + \|y\|$, we may combine (C.100) and (C.108) to obtain

$$\begin{aligned}&\|\sqrt{NT}(\hat{\gamma}^* - \hat{\gamma}) - [\sqrt{N/T}(\vec{E}^* 1_T - \vec{E} 1_T) - \sqrt{NT}(\hat{\Gamma}^* - \hat{\Gamma})\mathcal{H}^{-1}\bar{f}] \\ &\quad + \Gamma\mathcal{H}\mathcal{H}'B'[\sqrt{N/T}(\tilde{E}^* 1_T - \tilde{E} 1_T) - \sqrt{NT}(\hat{B}^* - \hat{B})\mathcal{H}^{-1}\bar{f}] \\ &\quad + \Gamma\mathcal{H}\sqrt{NT}(\hat{B}^* - \hat{B})'a\| = O_p\left(\frac{\sqrt{NT}}{J^\kappa} + \frac{\sqrt{TJ}}{\sqrt{N}} + \frac{J}{\sqrt{NT}}\right). \qquad \text{(C.109)}\end{aligned}$$

Thus, the first result of the lemma follows from (C.109), along with Lemma C.36, C.13, C.19, C.20, and Theorem 4.3, as well as the second result of the lemma. ■

Lemma C.30. *Suppose Assumptions B.1-B.4 hold. Assume (i) $N \to \infty$; (ii) $T \geq K+1$; (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 $\beta(z_{it}) = B'\phi(z_{it}) + \delta(z_{it})$,

$$\frac{1}{J}\sum_{i=1}^{N}\sum_{t=1}^{T}\|\hat{\beta}(z_{it}) - H'\beta(z_{it})\|^2 \leq \frac{2}{J}\sum_{i=1}^{T}\sum_{t=1}^{T}\|(\hat{B} - BH)'\phi(z_{it})\|^2 + 2\mathcal{S}_4, \qquad \text{(C.110)}$$

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 4.4. Note that (C.93) and Lemma C.26(iv) continue to hold under H_1. Thus, the result of the lemma follows from (C.93) and Lemma C.26(iv). ■

Lemma C.31. *Suppose Assumptions B.1-B.4, B.5(iii), B.8(i), (ii), and (v) hold. Assume (i) $N \to \infty$; (ii) $T \geq K+1$; (iii) $J \to \infty$ with $J = o(\sqrt{N})$. Assume that H_1 is true. Then there exists positive constant c_0 such that*

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

PROOF: Let us begin by defining some notation. Let $\vec{A}_t \equiv (Z_t'Z_t)^{-1}Z_t'A_t$ for $A_t = Y_t, \Psi_t$, ε_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 easy to show that $\hat{\Gamma} = (\vec{\Psi}M_TF/T)(F'M_TF/T)^{-1}H + o_p(1)$ by Theorem C.1, and $\|(\vec{\Psi}M_TF/T)(F'M_TF/T)^{-1}\|_F \leq C^*$ for some C^* with probability approaching one. This together with Lemma C.2(ii) implies that $P(\|\hat{\Gamma}H^{-1}\|_F > C) = o(1)$. Therefore, under H_1,

$$\begin{aligned}
\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\|\hat{\Gamma}'z_{it} - H'\beta(z_{it})\|^2 &\geq \lambda_{\min}(H'H)\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\|(\hat{\Gamma}H^{-1})'z_{it} - \beta(z_{it})\|^2 \\
&= \lambda_{\min}(H'H)\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}E\left[\|\beta(z_{it}) - (\hat{\Gamma}H^{-1})'z_{it}\|^2\right] + o_p(1) \\
&\geq \lambda_{\min}(H'H)\inf_{i\leq N, t\leq T}\inf_{\Pi}E[\|\beta(z_{it}) - \Pi'z_{it}\|^2] + o_p(1) \\
&\geq c_0 + o_p(1) \text{ for some } c_0 > 0, \qquad \text{(C.111)}
\end{aligned}$$

where the equality follows from Lemma C.37 since $P(\|\hat{\Gamma}H^{-1}\|_F > C) = o(1)$, and the last inequality follows by Lemma C.2(ii). ■

Lemma C.32. *Suppose Assumptions B.1-B.4, B.5(iii), B.8 hold. Assume (i) $N \to \infty$; (ii) $T \geq K + 1$; (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 (C.93) continue to hold under 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

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\|(\hat{\Gamma}^* - \hat{\Gamma})'z_{it}\|^2 \leq \|\hat{\Gamma}^* - \hat{\Gamma}\|_F^2\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\|z_{it}\|^2. \qquad \text{(C.112)}$$

Let us define some notation. 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 $\Psi_t = (\alpha(z_{1t}) + \beta(z_{1t})'f_t, \ldots, \alpha(z_{Nt}) + \beta(z_{Nt})'f_t)'$. 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 easy to show that

$\hat{\Gamma}^* = (\vec{\Psi}^* M_T F/T)(F' M_T F/T)^{-1} H + o_p(1)$ by Theorem C.1. From the proof of Lemma C.31, $\hat{\Gamma} = (\vec{\Psi} M_T F/T)(F' M_T F/T)^{-1} H + o_p(1)$. Moreover, it can be easily shown that $(\vec{\Psi}^* - \vec{\Psi}) M_T F/T = o_p(1)$. Thus,

$$\hat{\Gamma}^* - \hat{\Gamma} = (\vec{\Psi}^* - \vec{\Psi}) F/T (F'F/T)^{-1} = o_p(1). \tag{C.113}$$

By Assumption B.8(ii), $\sum_{i=1}^N \sum_{t=1}^T \|z_{it}\|^2/NT = O_p(1)$ by the Markov's inequality. This together with (C.112) and (C.113) implies that the first term is also $o_p(1)$. ■

Lemma C.33. *Let $\mathcal{D}_2$ and $\mathcal{D}_3$ be given in the proof of Lemma C.27.*
(i) Assume (i) $N \to \infty$; (ii) $T \geq K+1$; (iii) $J \to \infty$ with $J^2 \xi_J^2 \log J = o(N)$. Under Assumptions B.1-B.5, B.8(i), and (ii), $\|\mathcal{D}_2 \hat{B}\|_F^2 = O_p(1/N^2)$.
(ii) Under Assumptions B.1(i), B.2(ii), (iv), B.3, B.8(i), (ii), $\|\mathcal{D}_3\|_F^2 = O_p(J^{-2\kappa}/N)$.

Proof: (i) By Assumptions B.3, B.8(i), and (ii), we may follow a similar argument as in the proof of Lemma C.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 C.7(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 C.3(i). ■

Lemma C.34. *Suppose Assumptions B.1-B.3, B.5(iii), (iv), B.6, B.8(i)-(iii) hold. Let V be given in the proof of Theorem C.1, $\mathcal{D}_1$ and $\vec{E}$ be given in the proof of Lemma C.27. Assume (i) $N \to \infty$; (ii) $T \geq K+1$; (iii) $J \to \infty$ with $J = o(\sqrt{N})$. Then there exists an $M \times (K+1)$ random matrix $\mathbb{N}_z$ with $\text{vec}(\mathbb{N}_z) \sim N(0, \Omega_z)$ 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 Ω_z is given in Lemma C.27, $\mathbb{G}_\Gamma = \mathbb{N}_{z,2} B' B \mathcal{M}$, $\mathcal{M}$ is a nonrandom matrix in Lemma C.15, 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.13,

$$\|\sqrt{NT} \mathcal{D}_1 \hat{B} V^{-1} - \mathcal{L}_{NT,z} B' B \mathcal{M}\|_F = O_p\left(\frac{1}{J^\kappa} + \frac{1}{N^{1/4}}\right) \tag{C.114}$$

and

$$\|\sqrt{N/T}\vec{E}1_T - \ell_{NT,z}\| = O_p\left(\frac{1}{N^{1/4}}\right). \tag{C.115}$$

By a similar argument as in the proof of Lemma C.14, there exists an $M \times (K+1)$ random matrix $\mathbb{N}_z$ with $\text{vec}(\mathbb{N}_z) \sim N(0, \Omega_z)$ such that

$$\|(\ell_{NT,z}, \mathcal{L}_{NT,z}) - \mathbb{N}_z\|_F = O_p\left(\frac{1}{N^{1/6}}\right). \tag{C.116}$$

Thus the result of the lemma follows from (C.114)-(C.116). ■

Lemma C.35. *Let $\mathcal{D}_2^*$ and $\mathcal{D}_3^*$ be given in the proof of Lemma C.29.*
(i) Assume (i) $N \to \infty$; (ii) $T \geq K+1$; (iii) $J \to \infty$ with $J^2\xi_J^2 \log J = o(N)$. Under Assumptions B.1-B.5, B.7(i), B.8(ii), and (iv), $\|\mathcal{D}_2^*\hat{B}\|_F^2 = O_p(1/N^2)$.
(ii) Under Assumptions B.1(i), B.2(ii), (iv), B.3, B.7(i), B.8(ii), and (iv), $\|\mathcal{D}_3^*\|_F^2 = O_p(J^{-2\kappa}/N)$.

Proof: (i) By Assumptions B.3, B.7(i), B.8 (ii), and (iv), we may follow a similar argument as in the proof of Lemma C.21(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 C.7(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 C.3(i). ■

Lemma C.36. *Suppose Assumptions B.1-B.3, B.5(iii), (iv), B.6, B.7(i), and B.8(ii)-(iv) hold. Let V be given in the proof of Theorem C.1, $\mathcal{D}_1$ and $\vec{E}$ be given in the proof of Lemma C.27, and $\mathcal{D}_1^*$ and $\vec{E}^*$ be given in the proof of Lemma C.29. Assume (i) $N \to \infty$; (ii) $T \geq K+1$; (iii) $J \to \infty$ with $J = o(\sqrt{N})$. Then there exists an $M \times (K+1)$ random matrix $\mathbb{N}_z^*$ with $\text{vec}(\mathbb{N}_z^*) \sim N(0, \Omega_z)$ 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 Ω_z is given in Lemma C.27, $\mathbb{G}_\Gamma^* = \mathbb{N}_{z,2}^* B'B\mathcal{M}$, $\mathcal{M}$ is a nonrandom matrix in Lemma C.15, 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 C.19,

$$\|\sqrt{NT}\mathcal{D}_1^* \hat{B} V^{-1} - \mathcal{L}^{**}_{NT,z} B' B \mathcal{M}\|_F = O_p\left(\frac{1}{J^\kappa} + \frac{1}{N^{1/4}}\right) \tag{C.117}$$

and

$$\|\sqrt{N/T}\vec{E}^* 1_T - \ell^{**}_{NT,z}\| = O_p\left(\frac{1}{N^{1/4}}\right). \tag{C.118}$$

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 C.20, there exists an $M \times (K+1)$ random matrix $\mathbb{N}_z^*$ with $\text{vec}(\mathbb{N}_z^*) \sim N(0, \Omega_z)$ conditional on $\{Y_t, Z_t\}_{t \le T}$ such that

$$\|(\ell^*_{NT,z}, \mathcal{L}^*_{NT,z}) - \sqrt{\omega_0}\mathbb{N}_z^*\|_F = O_p\left(\frac{1}{N^{1/6}}\right). \tag{C.119}$$

Thus, the result of the lemma follows from (C.114),(C.115), and (C.117)-(C.119). ■

Lemma C.37. *Suppose Assumptions B.5(iii), B.8(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 $\Pi_1, \Pi_2 \in \mathcal{A}_C$,

$$\begin{aligned}
&|\zeta(z_1, \cdots, z_T, \Pi_1) - \zeta(z_1, \cdots, z_T, \Pi_2)| \\
&\le \|\Pi_1 - \Pi_2\|_F \frac{1}{T} \sum_{t=1}^{T} \|z_t\| (\|\beta(z_t) - \Pi_1' z_t\| + \|\beta(z_t) - \Pi_2' z_t\|) \\
&\le \|\Pi_1 - \Pi_2\|_F \frac{2}{T} \sum_{t=1}^{T} (\|z_t\| \|\beta(z_t)\| + C\|z_t\|^2) \equiv \|\Pi_1 - \Pi_2\|_F G(z_1, \cdots, z_T).
\end{aligned} \tag{C.120}$$

By Assumptions B.8(ii) and (v), $\max_{i \le N} E[G(z_{i1}, \cdots, z_{iT})] < \infty$. This together with (C.120) 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)). ■

C.6 Proof of Theorem A.1

Proof of Theorem A.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 $\hat{K} \neq K$, then there exists some $1 \leq k \leq K-1$ or $K+1 \leq k \leq JM/2$ such that $\theta_k \geq \theta_K$. Let $\underline{JM/2}$ be the integer part of $JM/2$. Since $\lambda_1(\tilde{Y}M_T\tilde{Y}'/T)/\lambda_K(\tilde{Y}M_T\tilde{Y}'/T) \geq \theta_k$ for all $1 \leq k \leq K-1$ and $\lambda_{K+1}(\tilde{Y}M_T\tilde{Y}'/T)/\lambda_{\underline{JM/2}}(\tilde{Y}M_T\tilde{Y}'/T) \geq \theta_k$ for all $K+1 \leq k \leq JM/2$, the event of $\hat{K} \neq K$ implies the event of $\lambda_1(\tilde{Y}M_T\tilde{Y}'/T)/\lambda_K(\tilde{Y}M_T\tilde{Y}'/T) \geq \theta_K$ 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,

$$P(\hat{K} \neq K) \leq P\left(\frac{\lambda_1(\tilde{Y}M_T\tilde{Y}'/T)}{\lambda_K(\tilde{Y}M_T\tilde{Y}'/T)} \geq \theta_K\right) + P\left(\frac{\lambda_{K+1}(\tilde{Y}M_T\tilde{Y}'/T)}{\lambda_{\underline{JM/2}}(\tilde{Y}M_T\tilde{Y}'/T)} \geq \theta_K\right). \quad \text{(C.121)}$$

By Lemmas C.38 and C.39, $\lambda_1(\tilde{Y}M_T\tilde{Y}'/T)/\lambda_K(\tilde{Y}M_T\tilde{Y}'/T) = O_p(1)$, $\theta_K/N = C + o_p(1)$ 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, $P(\hat{K} \neq K) \to 0$.

(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,

$$P(\tilde{K} \neq K) \leq P\left(\lambda_{K-1}(\tilde{Y}M_T\tilde{Y}'/T) < \lambda_{NT}\right) + P\left(\lambda_{K+1}(\tilde{Y}M_T\tilde{Y}'/T) \geq \lambda_{NT}\right). \quad \text{(C.122)}$$

By Lemma C.38 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 kth largest singular value of A. Since $\lambda_k(AA') = \sigma_k^2(A)$,

$$\begin{aligned}\lambda_{K+1}(\tilde{Y}M_T\tilde{Y}'/T) &= \sigma_{K+1}^2(\tilde{Y}M_T/\sqrt{T}) = |\sigma_{K+1}(\tilde{Y}M_T/\sqrt{T}) - \sigma_{K+1}(BF'M_T/\sqrt{T})|^2 \\ &\leq \frac{1}{T}\|\tilde{Y}M_T - B(M_TF)'\|_F^2 \leq \frac{2}{T}\|\tilde{\Delta}\|_F^2 + \frac{2}{T}\|\tilde{E}\|_F^2 = O_p\left(\frac{1}{J^{2\kappa}} + \frac{J}{N}\right), \quad \text{(C.123)}\end{aligned}$$

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 (C.2) and the Cauchy-Schwartz inequality, and the last equality follows from Lemmas C.3(i) and (ii). Since $\lambda_{NT}\min\{N/J, J^{2\kappa}\} \to \infty$, (C.123) implies $P(\lambda_{K+1}(\tilde{Y}M_T\tilde{Y}'/T) \geq \lambda_{NT}) \to 0$. This completes the proof of the theorem. ■

C.6.1 Technical Lemmas

Lemma C.38. *Suppose Assumptions B.1-B.3 hold. Assume (i) $N \to \infty$; (ii) $T \geq K+1$; (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 (C.12), $\lambda_k(\tilde{Y} M_T \tilde{Y}'/T) = \lambda_k((F' M_T F/T)B'B) + o_p(1)$ for $k = 1, \ldots, K$. Thus, the result immediately follows from Assumptions B.2(i)-(iii). ■

Lemma C.39. *Suppose Assumptions B.1(i), B.2(ii), (iv), B.3(i), B.5(i), and B.9 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 kth largest singular value of A. Noting that $\lambda_k(AA') = \sigma_k^2(A)$, it follows that for $k = 1, \ldots, JM - K$,

$$\begin{aligned}
&|\lambda_{K+k}(\tilde{Y} M_T \tilde{Y}') - \lambda_{K+k}((BF' + \tilde{E}) M_T (BF' + \tilde{E})')| \\
&\leq |\sigma_{K+k}(\tilde{Y} M_T) - \sigma_{K+k}((BF' + \tilde{E}) M_T)|^2 + 2|\sigma_{K+k}(\tilde{Y} M_T) \\
&\quad - \sigma_{K+k}((BF' + \tilde{E}) M_T)| \sigma_{K+k}((BF' + \tilde{E}) M_T) \\
&\leq \|\tilde{Y} M_T - (BF' + \tilde{E}) M_T\|_F^2 + 2\|\tilde{Y} M_T - (BF' + \tilde{E}) M_T\|_F \\
&\quad \times \lambda_{K+k}^{1/2}((BF' + \tilde{E}) M_T (BF' + \tilde{E})') \\
&\leq \|\tilde{\Delta}\|_F^2 + 2\|\tilde{\Delta}\|_F \lambda_{K+1}^{1/2}((BF' + \tilde{E}) M_T (BF' + \tilde{E})'), \qquad \text{(C.124)}
\end{aligned}$$

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 (C.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 (C.124) is asymptotically negligible and study the behavior of $\lambda_{K+k}((BF' + \tilde{E}) M_T (BF' + \tilde{E})')$. Let $\tilde{B} = B + \tilde{E} M_T F (F' M_T F)^{-1}$ and $M_F = I_T - M_T F (F' M_T F)^{-1} (M_T F)'$. We may decompose $(BF' + \tilde{E}) M_T (BF' + \tilde{E})'$ by

$$(BF' + \tilde{E}) M_T (BF' + \tilde{E})' = \tilde{B} F' M_T F \tilde{B}' + \tilde{E} M_T M_F M_T \tilde{E}'. \qquad \text{(C.125)}$$

Then, (C.125) implies that for $k = 1, \ldots, JM - K$,

$$\begin{aligned}\lambda_{K+k}((BF' + \tilde{E})M_T(BF' + \tilde{E})') &\leq \lambda_{K+1}(\tilde{B}F'M_TF\tilde{B}') \\ &+ \lambda_k(\tilde{E}M_TM_FM_T\tilde{E}') \leq \lambda_k(\tilde{E}M_T\tilde{E}') \leq \lambda_k(\tilde{E}\tilde{E}'),\end{aligned} \tag{C.126}$$

where the first inequality follows by Lemma C.40(i), the second inequality follows by Lemma C.40(ii) as well as 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, (C.125) also implies that for $k = 1, \ldots, JM - 2K - 1$,

$$\begin{aligned}&\lambda_{K+k}((BF' + \tilde{E})M_T(BF' + \tilde{E})') \geq \lambda_{K+k}(\tilde{E}M_TM_FM_T\tilde{E}') \\ &= \lambda_{K+k}(E\tilde{M}_TM_FM_T\tilde{E}') + \lambda_{K+1}(\tilde{E}M_T(I - M_F)M_T\tilde{E}') \geq \lambda_{2K+k}(\tilde{E}M_T\tilde{E}') \\ &= \lambda_{2K+k}(\tilde{E}M_T\tilde{E}') + \lambda_2(\tilde{E}(I_T - M_T)\tilde{E}') \geq \lambda_{2K+k+1}(\tilde{E}\tilde{E}'),\end{aligned} \tag{C.127}$$

where the first inequality follows by Lemma C.40(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 C.40(i), and the second equality and the third inequality follow similarly. Putting (C.126) and (C.127) 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 C.41. By the Weyl's inequality and Lemma C.41,

$$\sup_{k \leq JM} |\lambda_k(N\tilde{E}\tilde{E}'/T) - \lambda_k(\mathcal{A}_{NT})| \leq \|N\tilde{E}\tilde{E}'/T - \mathcal{A}_{NT}\|_F = o_p(1). \tag{C.128}$$

This implies that the eigenvalues of $N\tilde{E}\tilde{E}'/T$ and $\mathcal{A}_{NT}$ are asymptotically equivalent. Then, it follows from (C.126) and (C.128) that

$$\begin{aligned}&\lambda_{K+1}(N(BF' + \tilde{E})M_T(BF' + \tilde{E})'/T) \\ &\leq \lambda_1(N\tilde{E}\tilde{E}'/T) \leq \lambda_1\left(\mathcal{A}_{NT}\right) + o_p(1) = O_p(1),\end{aligned} \tag{C.129}$$

because $\lambda_1(\mathcal{A}_{NT}) \leq (\min_{t \leq T} \lambda_{\min}(\hat{Q}_t))^{-1} \max_{t \leq T} \lambda_{\max}(E[\varepsilon_t\varepsilon_t']) = O_p(1)$ by Assumptions B.1(i) and B.9 (i). Combining (C.124), (C.129), and Lemma C.3(i) yields

$$\sup_{k \leq JM-K} |N\lambda_{K+k}(\tilde{Y}M_T\tilde{Y}'/T) - N\lambda_{K+k}((BF' + \tilde{E})M_T(BF' + \tilde{E})'/T)| = o_p(1). \tag{C.130}$$

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 (C.126)-(C.128) and (C.130) that

$$\begin{aligned}\lambda_{JM}(\mathcal{A}_{NT}) + o_p(1) &\leq N\lambda_{JM-K-1}(\tilde{Y}M_T\tilde{Y}'/T) \\ &\leq N\lambda_{K+1}(\tilde{Y}M_T\tilde{Y}'/T) \leq \lambda_1(\mathcal{A}_{NT}) + o_p(1). \qquad \text{(C.131)}\end{aligned}$$

Because $\lambda_1(\mathcal{A}_{NT}) \leq (\min_{t\leq T}\lambda_{\min}(\hat{Q}_t))^{-1}\max_{t\leq T}\lambda_{\max}(E[\varepsilon_t\varepsilon_t'])$ 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 (C.131) along with Assumptions B.1(i) and B.9(i). ■

Lemma C.40 (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 C.41. *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 C.1. Under Assumptions B.1(i), B.3(i), B.5(i), and B.9(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_ε denote the expectation with respect to $\{\varepsilon_t\}_{t\leq T}$. 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 $\|A\|_F^2=\text{tr}(AA')$,

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

$$\times |cov(\varepsilon_{it}\varepsilon_{jt}, \varepsilon_{ks}\varepsilon_{\ell s})|, \tag{C.132}$$

where the second equality follows from the independence in Assumption B.3 (i) and the linearity of both expectation and trace operators, and the inequality follows since $\|\hat{\psi}_{it}\| \leq (\lambda_{\min}(\hat{Q}_t))^{-1}\|\phi(z_{it})\|$. Moreover,

$$\begin{aligned}
&E\left[\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{\ell=1}^{N}\|\phi(z_{it})\|\|\phi(z_{jt})\|\|\phi(z_{ks})\|\|\phi(z_{\ell s})\||cov(\varepsilon_{it}\varepsilon_{jt}, \varepsilon_{ks}\varepsilon_{\ell s})|\right] \\
&\leq \max_{i\leq N, t\leq T} E[\|\phi(z_{it})\|^4]\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{\ell=1}^{N}|cov(\varepsilon_{it}\varepsilon_{jt}, \varepsilon_{ks}\varepsilon_{\ell s})| \\
&\leq J^2M^2 \max_{\ell\leq JM, i\leq N, t\leq T} E[\phi^4(z_{it,m})]\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{\ell=1}^{N}|cov(\varepsilon_{it}\varepsilon_{jt}, \varepsilon_{ks}\varepsilon_{\ell s})|, \qquad \text{(C.133)}
\end{aligned}$$

where the first inequality is due to the Cauchy-Schwartz inequality, and the second one follows since $\max_{i\leq N,t\leq T} E[\|\phi(z_{it})\|^4] = J^2M^2 \max_{\ell\leq JM,i\leq N,t\leq T} E[\phi^4(z_{it,m})]$. Combining (C.132) and (C.133) implies that $E_\varepsilon[\|\bar{E}\bar{E}'/NT - \mathcal{A}_{NT}\|_F^2] = O_p(J^2/N + J^2/T)$ by Assumptions B.1(i), B.5(i), and B.9(ii). Thus, the result of the lemma follows by the Markov's inequality and Lemma C.5. ■

Appendix D - Additional Discussions

D.1 Regressed-PCA: A Special Case

We consider the regressed-PCA approach when $N > JM$ and$\Phi(Z_t)'\Phi(Z_t)/N = I_{JM}$. There exists an $N \times (N - JM)$ matrix Ψ_t such that $(\Phi(Z_t)/\sqrt{N}, \Psi_t)$ is orthonormal, i.e., $(\Phi(Z_t)/\sqrt{N}, \Psi_t)(\Phi(Z_t)/\sqrt{N}, \Psi_t)' = I_N$. It follows that

$$\begin{aligned}
&(Y_t - \Phi(Z_t)a - \Phi(Z_t)Bf_t)'(Y_t - \Phi(Z_t)a - \Phi(Z_t)Bf_t) \\
&= (Y_t - \Phi(Z_t)a - \Phi(Z_t)Bf_t)'(\Phi(Z_t)/\sqrt{N}, \Psi_t)(\Phi(Z_t)/\sqrt{N}, \Psi_t)'(Y_t - \Phi(Z_t)a - \Phi(Z_t)Bf_t) \\
&= N[(Y_t - \Phi(Z_t)a - \Phi(Z_t)Bf_t)'\Phi(Z_t)\Phi(Z_t)'(Y_t - \Phi(Z_t)a - \Phi(Z_t)Bf_t)/N^2] + Y_t'\Psi_t\Psi_t'Y_t \\
&= N[(Y_t - \Phi(Z_t)a - \Phi(Z_t)Bf_t)'S_t(Y_t - \Phi(Z_t)a - \Phi(Z_t)Bf_t)] + Y_t'\Psi_t\Psi_t'Y_t. \qquad \text{(D.1)}
\end{aligned}$$

Thus, the objective function in (13) is equivalent to the objective function in (14), scaled by a factor of N and plus a constant term that does not depend on a, B, or f_t. Therefore, when the condition $\Phi(Z_t)'\Phi(Z_t)/N = I_{JM}$ holds, the regressed-PCA method reduces to the least squares approach.

D.2 Optimality of Fama-MacBeth Managed Portfolios

Fama (1976) demonstrates that the jth column of $\Phi(Z_t)(\Phi(Z_t)'\Phi(Z_t))^{-1}$ solves the following dynamic programming problem:

$$\min_{\omega} \omega'\omega \text{ such that } \Phi(Z_t)'\omega = e_j, \tag{D.2}$$

where e_j is the jth column of I_{JM}. The proof is straightforward, relying on the use of a lagrangian function. The first order conditions are:

$$2\omega + \Phi(Z_t)\lambda = 0, \tag{D.3}$$

$$\Phi(Z_t)'\omega = e_j, \tag{D.4}$$

where λ is a lagrangian multiplier. Solving these yields $\omega = \Phi(Z_t)(\Phi(Z_t)'\Phi(Z_t))^{-1}e_j$, which corresponds to the jth column of $\Phi(Z_t)(\Phi(Z_t)'\Phi(Z_t))^{-1}$. If individual asset returns are i.i.d. (conditional on Z_t), the objective function presents the variance of the portfolio $Y_t'\omega$ multiplied by a constant. Hence, under the i.i.d. assumption, each portfolio in $\tilde{Y}_t = (\Phi(Z_t)'\Phi(Z_t))^{-1}\Phi(Z_t)'Y_t$ is minimum-variance portfolio. Moreover, the covariance matrix of $\tilde{Y}_t$ is proportional to $(\Phi(Z_t)'\Phi(Z_t))^{-1}$, suggesting low correlations among the portfolios in $\tilde{Y}_t$ when the columns of $\Phi(Z_t)$ have low correlations.

D.3 Sorting: Regression on Dummies

At each time t, sorting divides the space of z_{it} into q distinct, non-overlapping regions $R_{t,1}, R_{t,2}, \ldots, R_{q,t}$, which may vary over time. Let $D_{it,1}, D_{it,2}, \ldots, D_{it,q}$ represent group dummy variables for these regions, where $D_{it,j} = 1\{z_{it} \in R_{j,t}\}$. The equally weighted returns of the sorted portfolios are then the coefficient estimates from a cross-sectional regression of y_{it} on $D_{it,1}, D_{it,2}, \ldots, D_{it,q}$ without an intercept. These are calculated as $\tilde{Y}_t = (\sum_{i=1}^N \phi(z_{it})\phi(z_{it})')^{-1} \sum_{i=1}^N \phi(z_{it})y_{it}$, where $\phi(z_{it}) = (D_{it,1}, D_{it,2}, \ldots, D_{it,q})'$. Specially, the return of the portfolio corresponding to $R_{j,t}$ is equal to the coefficient on $D_{it,j}$. For value weighted returns, book-to-market ratios can be used as weights in the regression. In the example of constructing high-minus-low and small-minus-big factors, z_{it} consists of capitalization (c_{it}) and book-to-market ratio (bm_{it}) with $q = 6$:

$$D_{it,1} = 1\{c_{it} \leq Q_{0.5,t}(c_{it})\}1\{bm_{it} \leq Q_{0.3,t}(bm_{it})\},$$
$$D_{it,2} = 1\{c_{it} \leq Q_{0.5,t}(c_{it})\}1\{Q_{0.3,t}(bm_{it}) < bm_{it} \leq Q_{0.7,t}(bm_{it})\},$$

$$
\begin{aligned}
D_{it,3} &= 1\{c_{it} \leq Q_{0.5,t}(c_{it})\}1\{bm_{it} > Q_{0.7,t}(bm_{it})\} \\
D_{it,4} &= 1\{c_{it} > Q_{0.5,t}(c_{it})\}1\{bm_{it} \leq Q_{0.3,t}(bm_{it})\}, \\
D_{it,5} &= 1\{c_{it} > Q_{0.5,t}(c_{it})\}1\{Q_{0.3,t}(bm_{it}) < bm_{it} \leq Q_{0.7,t}(bm_{it})\}, \\
D_{it,6} &= 1\{c_{it} > Q_{0.5,t}(c_{it})\}1\{bm_{it} > Q_{0.7,t}(bm_{it})\},
\end{aligned} \tag{D.5}
$$

where $Q_{0.5,t}(c_{it})$ is the 50% quantile of $\{c_{1t}, c_{2t}, \ldots, c_{nt}\}$ at each time t.

D.4 Misspecifications of Alpha and Beta Functions

We use simple examples to illustrate how misspefications of $\alpha(\cdot)$ and $\beta(\cdot)$ may result in inconsistent estimation of F. Consider the following two models:

$$Y_t = W_t\Pi + Z_t\Gamma f_t + \varepsilon_t, \tag{D.6}$$

$$Y_t = (Z_t\Gamma + W_t\Pi)f_t + \varepsilon_t, \tag{D.7}$$

where Z_t and W_t are $N \times 1$ vectors, f_t is a scalar factor, and ε_t is independent of Z_t and W_t. 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. In this case, model (D.6) and model (D.7) can be rewritten as:

$$Y_t = Z_t\Gamma f_t^\star + \varepsilon_t^\star, \tag{D.8}$$

$$Y_t = Z_t\Gamma f_t^{\star\star} + \varepsilon_t^{\star\star}, \tag{D.9}$$

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 $\varepsilon_t^{\star\star} = v_t\Gamma f_t + \varepsilon_t$. If only Z_t is used for estimating model (D.6) (i.e., $\alpha(\cdot)$ is misspecified), then $\hat{F}$ consistently estimates $F^\star = (f_1^\star, \ldots, f_T^\star)'$ up to a scalar. Similarly, if only Z_t is used for estimating model (D.7) (i.e., $\beta(\cdot)$ is misspecified), then $\hat{F}$ consistently estimates $F^{\star\star} = (f_1^{\star\star}, \ldots, f_T^{\star\star})'$ 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 first case or remains constant over time in the second case.

D.5 On Bootstrap Failure

A more natural bootstrap estimator for B is given by $\hat{B}^{**}$, whose columns are the eigenvectors of $\tilde{Y}^* M_T \tilde{Y}^{*\prime}/T$ corresponding to its first K largest eigenvalues. We notice that the distribution of $\sqrt{NT/\omega_0}(\hat{B}^{**} - \hat{B})$ conditional on the data may fail to estimate the distribution of $\mathbb{G}_B$. The key part of the proof for Theorem 4.3 is to show that $\sqrt{NT}(\hat{B}^* - BH)$ and $\sqrt{NT}(\hat{B} - BH)$ share a similar asymptotic expansion. Specifically,

we show

$$\left\|\sqrt{NT}(\hat{B}-BH)-\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}Q_t^{-1}\Phi(Z_t)'\varepsilon_t(f_t-\bar{f})'B'B\mathcal{M}\right\|_F=O_p(\delta_{NT}) \tag{D.10}$$

and

$$\left\|\sqrt{NT}(\hat{B}^*-BH)-\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}Q_t^{-1}\Phi(Z_t)^{*\prime}\varepsilon_t(f_t-\bar{f})'B'B\mathcal{M}\right\|_F=O_p(\delta_{NT}), \tag{D.11}$$

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 also show

$$\left\|\sqrt{NT}(\hat{B}^{**}-BH^*)-\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}Q_t^{-1}\Phi(Z_t)^{*\prime}\varepsilon_t(f_t-\bar{f})'B'B\mathcal{M}\right\|_F=O_p(\delta_{NT}). \tag{D.12}$$

Thus, the distribution of $\sqrt{NT/\omega_0}(\hat{B}^{**}-\hat{B})$ 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 crucial to use $\hat{F}$ rather than $\hat{F}^*$ in (17) to ensure that $\hat{B}^*$ and $\hat{B}$ share a common rotational transformation matrix and are centered around the same quantity BH, rendering the validity of the bootstrap.

D.6 Three Versions of R^2

From a mathematical perspective, $R^2_{T,N}$ and $R^2_{N,T}$ can be expressed as weighted versions of R^2. However, the variation in these weights is not due to an unbalanced panel. Let $\hat{\varepsilon}_{it}\equiv y_{it}-\hat{\alpha}(z_{i,t-1})-\hat{\beta}(z_{i,t-1})'\hat{f}_t$. It follows that

$$R^2_{T,N}=1-\frac{1}{N}\sum_i\frac{\sum_t\hat{\varepsilon}_{it}^2}{\sum_t y_{it}^2}=1-\frac{\sum_i\frac{\sum_{i,t}y_{it}^2}{N\sum_t y_{it}^2}\sum_t\hat{\varepsilon}_{it}^2}{\sum_{i,t}y_{it}^2}=1-\frac{\sum_i\omega_i\sum_t\hat{\varepsilon}_{it}^2}{\sum_{i,t}y_{it}^2}, \tag{D.13}$$

where $w_i=\sum_{i,t}y_{it}^2/N\sum_t y_{it}^2$. The variation in ω_i arises from differences in the total time variation across individual stocks. The same reasoning applies to $R^2_{N,T}$.

D.7 Sufficient Conditions for Assumptions

We provide sufficient conditions for Assumptions B.1(i) and B.9(ii) in the following two propositions, justifying that the two assumptions are not restrictive.

Proposition D.1 (Assumption B.1(i))**.** *Suppose Assumptions B.5(ii) and (iii) hold. Assume $J \geq 2$ and $\sqrt{T}\xi_J^2 \log J = o(N)$, where ξ_J is given above Theorem 4.1. Then Assumption B.1(i) holds.*

PROOF: Let $Q_t \equiv E[\hat{Q}_t]$ Since $\sqrt{T}\xi_J^2 \log J = o(N)$, by Lemma C.11,

$$\max_{t\leq T} \|\hat{Q}_t - Q_t\|_2 \leq \left(\sum_{t=1}^{T} \|\hat{Q}_t - Q_t\|_2^4\right)^{1/4} = O_p\left(\frac{T^{1/4}\xi_J \log^{1/2} J}{\sqrt{N}}\right) = o_p(1). \tag{D.14}$$

By (D.14) and the Weyl's inequality,

$$\left|\min_{t\leq T} \lambda_{\min}(\hat{Q}_t) - \min_{t\leq T} \lambda_{\min}(Q_t)\right| \leq \max_{t\leq T} \|\hat{Q}_t - Q_t\|_2 = o_p(1) \tag{D.15}$$

and

$$\left|\max_{t\leq T} \lambda_{\max}(\hat{Q}_t) - \max_{t\leq T} \lambda_{\max}(Q_t)\right| \leq \max_{t\leq T} \|\hat{Q}_t - Q_t\|_2 = o_p(1). \tag{D.16}$$

The result of the lemma thus follows from (D.15) and (D.16) as well as Assumption B.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 D.2 (Assumption B.9(ii))**.** *Suppose Assumptions B.3(ii) and B.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 B.9(ii) holds.

PROOF: By the independence condition and Assumption B.3(ii), $E[\varepsilon_{it}\varepsilon_{jt}] = 0$ for $i \neq j$. Thus, we may have the following decomposition

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

$$= \sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{k=1}^{N} |cov(\varepsilon_{it}^2, \varepsilon_{ks}^2)| + \sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{k=1}^{N}\sum_{\ell\neq k}^{N} |E[\varepsilon_{it}\varepsilon_{jt}\varepsilon_{ks}\varepsilon_{\ell s}]|$$
$$+ 2\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{k=1}^{N}\sum_{\ell\neq k}^{N} |E[(\varepsilon_{it}^2 - E[\varepsilon_{it}^2])\varepsilon_{ks}\varepsilon_{\ell s}]|$$
$$\equiv \mathcal{T}_1 + \mathcal{T}_2 + \mathcal{T}_3. \tag{D.17}$$

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

$$\mathcal{T}_1 = \sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N} var(\varepsilon_{it}^2) \leq \sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N} E[\varepsilon_{it}^4] \leq NT^2 \max_{i\leq N, t\leq T} E[\varepsilon_{it}^4], \tag{D.18}$$

where the first inequality follows from $var(\varepsilon_{it}^2) \leq E[\varepsilon_{it}^4]$. By the independence condition and Assumption B.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

$$\mathcal{T}_2 = 2\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{j\neq i}^{N} |E[\varepsilon_{it}\varepsilon_{is}\varepsilon_{jt}\varepsilon_{js}]| = 2\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{j\neq 1}^{N} |E[\varepsilon_{it}\varepsilon_{is}]||E[\varepsilon_{jt}\varepsilon_{js}]|$$
$$\leq 2\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N} |E[\varepsilon_{it}\varepsilon_{is}]||E[\varepsilon_{jt}\varepsilon_{js}]| = 2\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\sum_{i=1}^{N} |E[\varepsilon_{it}\varepsilon_{is}]|\right)^2$$
$$\leq 2N\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} |E[\varepsilon_{it}\varepsilon_{is}]|^2 \leq 2N^2 \max_{i\leq N}\sum_{t=1}^{T}\sum_{s=1}^{T} |E[\varepsilon_{it}\varepsilon_{is}]|^2, \tag{D.19}$$

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 B.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 (D.17)-(D.19) and the assumptions thus concludes the result of the proposition. ■

Appendix E - Monte Carlo Simulations

In this appendix, we conduct small-scale Monte Carlo simulations to evaluate the finite sample performance of our estimators and tests.

We consider the following data-generating process, assuming

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

for $\theta \geq 0$ and $\delta \geq 0$, where $K = 2$ and $M = 3$. Note that $\alpha(\cdot) = 0$ when $\theta = \delta = 0$, and both $\alpha(z_{it})$ and $\beta(z_{it})$ become nonlinear functions of z_{it} when $\delta > 0$. We define:

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

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, σ_t's are i.i.d. $U(1, 2)$ over t, and $z_{i0,2}$'s are i.i.d. $N(0, 1)$. All components of z_{it} vary over t, but in different ways. We also define $f_t = 0.3f_{t-1} + \eta_t$, where η_t's are i.i.d. $N(0, I_K)$ and $f_0 \sim N(0, I_K/0.91)$. For $0 \leq \rho < 1$, we specify:

$$\varepsilon_t = \rho\varepsilon_{t-1} + e_t, \tag{E.3}$$

where e_t's are i.i.d. $N(0, I_N)$ and $\varepsilon_0 \sim N(0, I_N/(1-\rho^2))$. The parameter ρ measures the weak dependence of ε_{it} over t. Here, u_{it}'s, σ_t's, z_{i0}'s η_t's, f_0, e_t's, and ε_0 are mutually independent. We generate y_{it} based on the model (1).

To implement regressed-PCA, we select $\phi(z_{it}) = (z_{it,1}, z_{it,1}^2, z_{it,2}, z_{it,2}^2, z_{it,3}, z_{it,3}^2)'$, so $J = 2$, and the sieve approximation error is zero. We let $\lambda_{NT} = 1/\log(N)$ when implementing $\tilde{K}$ in (A.2). For the weighted bootstrap, we let w_i's be i.i.d. standard exponential random variables. We first analyse the performance of $\hat{a}$, $\hat{B}$, $\hat{F}$, $\hat{K}$ in (A.1), and $\tilde{K}$ under varying (N, T) values, with $\theta = 1$, $\delta = 0.5$, and $\rho = 0, 0.3, 0.7$. We report the mean square errors of $\hat{a}$, $\hat{B}$, and $\hat{F}$ in Table E.I and the correct rates of $\hat{K}$ and $\tilde{K}$ in Table E.II. Figures E.1 and E.2 present histograms of $\sqrt{NT}(\hat{a} - a)$ and $\sqrt{NT}(\hat{B} - BH)$ and their bootstrap estimates (i.e., $\sqrt{NT}(\hat{a}^* - \hat{a})$ and $\sqrt{NT}(\hat{B}^* - \hat{B})$) for $\rho = 0.3$, with similar results for $\rho = 0$ and 0.7 available upon request. Due to space limitations, we only display one entry of $\sqrt{NT}(\hat{a} - a)$ and $\sqrt{NT}(\hat{B} - BH)$, with similar results for other entries available on request.

Next, we assess the performance of tests for $\alpha(\cdot) = 0$ and the linearity of $\alpha(\cdot)$ and $\beta(\cdot)$. To test $\alpha(\cdot) = 0$, we fix $\delta = 0$. Hence, $\alpha(\cdot) = 0$ if and only if $\theta = 0$. We report the rejection rates for $\theta = 0, 0.01, 0.02, \ldots, 0.1$ under $\rho = 0.3$, with similar results for $\rho = 0$ and 0.7 available on request. For the linearity test, fixing $\theta = 1$, $\alpha(\cdot)$ 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 $\rho = 0.3$, with similar results for other values of ρ available on request. We set the number of simulation replications to 1,000 and the number of bootstrap draws to 499 for each replication.

The main findings are as follows. First, as shown in Table E.I, the mean square errors of $\hat{a}$, $\hat{B}$, and $\hat{F}$ decrease as N increases, even for $T = 10$, indicating consistency

of the estimators as $N \to \infty$, even for small T. Increasing T further reduces the mean square errors of $\hat{a}$ and $\hat{B}$, but does not affect that of $\hat{F}$. Both results hold regardless of ρ, confirming that the estimators remain valid under weak dependence of ε_{it}. These findings align with Theorem 4.1. Second, as shown in Table E.II, $\hat{K}$ and $\tilde{K}$ correctly estimate K in all cases, except for small N and T, consistent with Theorem A.1. Third, Figures E.1 and E.2 show that both $\sqrt{NT}(\hat{a} - a)$ and $\sqrt{NT}(\hat{B} - BH)$, as well as their bootstrap estimates, follow bell-shaped distributions, even for $T = 10$, suggesting asymptotic normality as per Theorems 4.2 and 4.3. The two distributions converge as N increases, though the approximation may be unsatisfactory for $N = 50$.

Finally, both tests perform well. Table E.III shows that the first test may slightly overreject $\alpha(\cdot) = 0$ (which holds when $\theta = 0$) when $N = 50$, but this corrects as N increases, even for $T = 10$. The test is consistent as $N \to \infty$ for small T, and increasing T can improve power, though it may slightly affect the size (e.g., for $\theta = 0$, the rejection rate increases as T grows from 10 to 100 for $N = 200$). This is in line with the $T = o(N)$ requirement in Theorem 4.4 or underlying in Theorem 4.3. The second test exhibits similar performance, as shown in Table E.IV, and the details are omitted for brevity. The findings of the second test are consistent with Theorem 4.4.

In conclusion, our estimators and bootstrap inference methods show strong performance for large N, even when T is small.

Table E.I. 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

† 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 ℓth 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 E.II. 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

Figure E.1. Histograms of the 2nd entry in $\sqrt{NT}(\hat{a}-a)$ (*blue*) and $\sqrt{NT}(\hat{a}^*-\hat{a})$ (*yellow*, based on the first simulation replication) when $\theta = 1$, $\delta = 0.5$, and $\rho = 0.3$

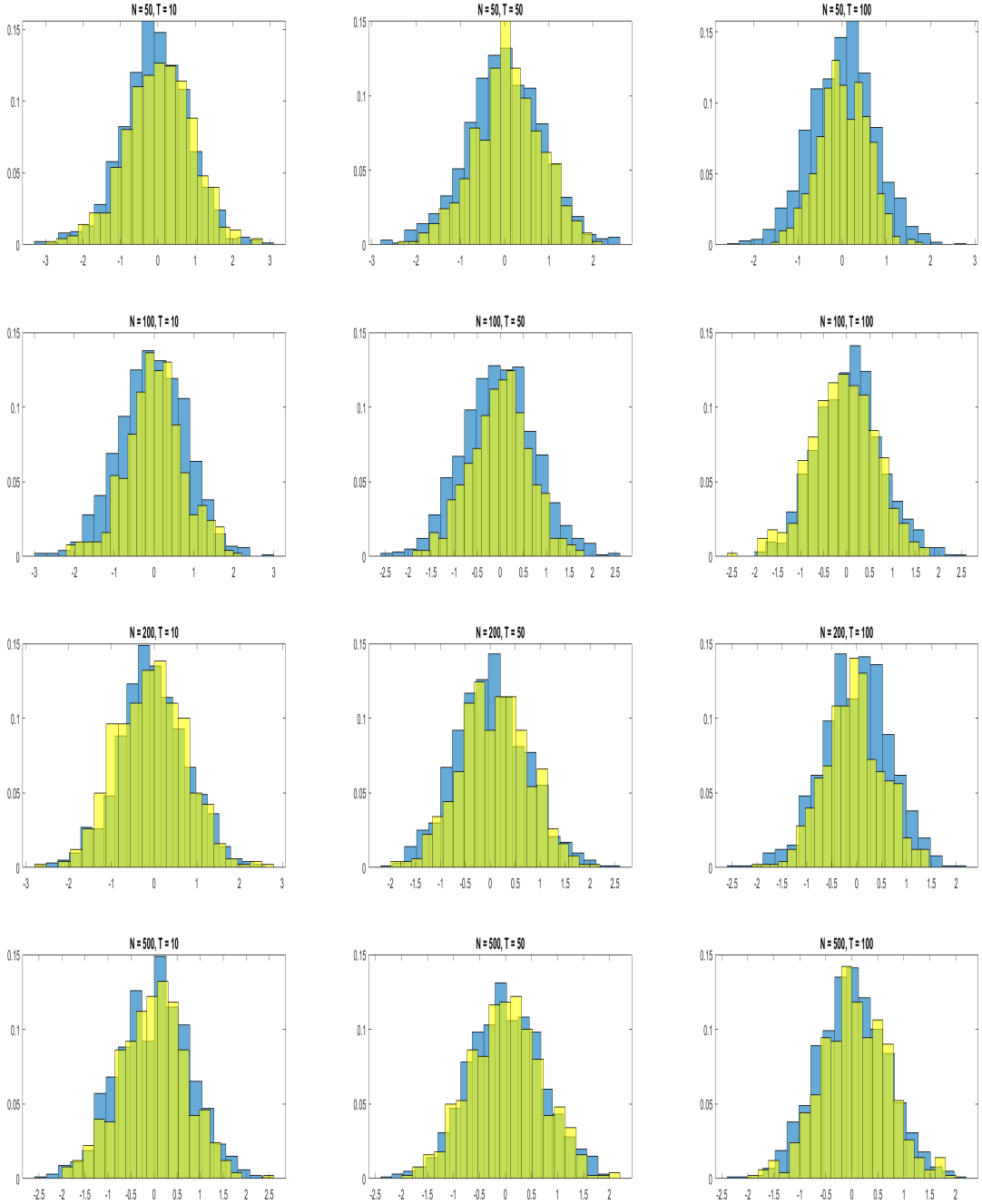


Figure E.2. Histograms of the $(1,2)$th entry in $\sqrt{NT}(\hat{B}-BH)$ (*blue*) and $\sqrt{NT}(\hat{B}^*-\hat{B})$ (*yellow*, based on the first simulation replication) when $\theta = 1$, $\delta = 0.5$, and $\rho = 0.3$

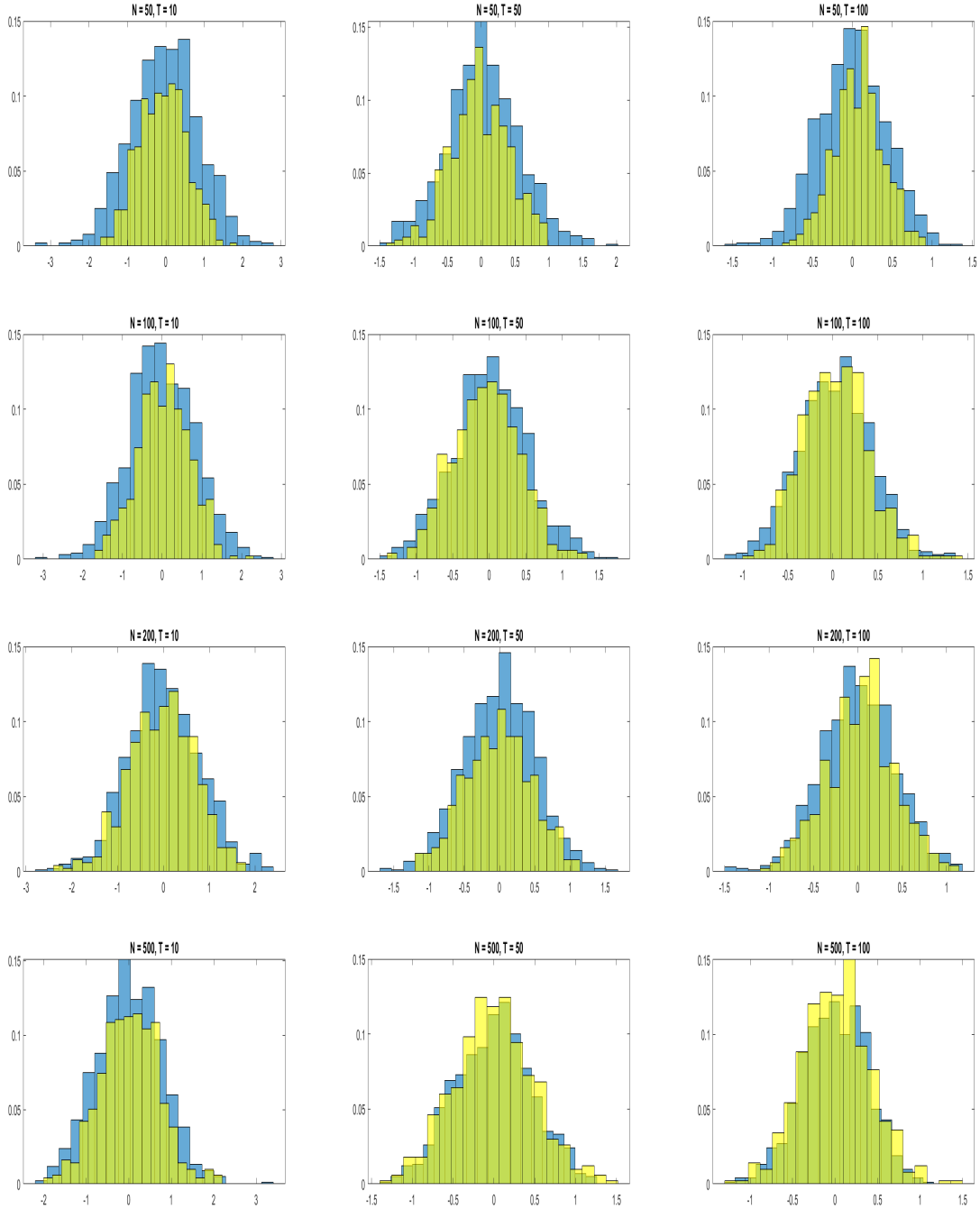


Table E.III. Rejection rates of testing $\alpha(\cdot) = 0$ when $\delta = 0$ and $\rho = 0.3^{\dagger}$

(N,T)	θ										
	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

† The significance level $\alpha = 5\%$.

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

(N,T)	δ										
	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

† The significance level $\alpha = 5\%$.

Appendix F - Additional Empirical Results

In this appendix, we provide additional results for Section 5. First, we provide formulas for various R^2's. The formulas for the in-sample fits by excluding $\hat{\alpha}(z_{i,t-1})$ are as follows:

$$R^2_f = 1 - \frac{\sum_{i,t}[y_{it} - \hat{\beta}(z_{i,t-1})'\hat{f}_t]^2}{\sum_{i,t} y^2_{it}}, \tag{F.1}$$

$$R^2_{f,T,N} = 1 - \frac{1}{N}\sum_i \frac{\sum_t[y_{it} - \hat{\beta}(z_{i,t-1})'\hat{f}_t]^2}{\sum_t y^2_{it}}, \tag{F.2}$$

$$R^2_{f,N,T} = 1 - \frac{1}{T}\sum_t \frac{\sum_i[y_{it} - \hat{\beta}(z_{i,t-1})'\hat{f}_t]^2}{\sum_i y^2_{it}}. \tag{F.3}$$

Three versions of out-of-sample predictive R^2's are as follows:

$$R^2_O = 1 - \frac{\sum_{i,t\geq 120}[y_{it} - \hat{\alpha}_{t-1}(z_{i,t-1}) - \hat{\beta}_{t-1}(z_{i,t-1})'\hat{\lambda}_t]^2}{\sum_{i,t\geq 120} y^2_{it}}, \tag{F.4}$$

$$R^2_{T,N,O} = 1 - \frac{1}{N}\sum_i \frac{\sum_{t\geq 120}[y_{it} - \hat{\alpha}_{t-1}(z_{i,t-1}) - \hat{\beta}_{t-1}(z_{i,t-1})'\hat{\lambda}_t]^2}{\sum_{t\geq 120} y^2_{it}}, \tag{F.5}$$

$$R^2_{N,T,O} = 1 - \frac{1}{T-120}\sum_{t\geq 120} \frac{\sum_i[y_{it} - \hat{\alpha}_{t-1}(z_{i,t-1}) - \hat{\beta}_{t-1}(z_{i,t-1})'\hat{\lambda}_t]^2}{\sum_i y^2_{it}}. \tag{F.6}$$

Three versions of out-of-sample fit R^2's are as follows:

$$R^2_{f,O} = 1 - \frac{\sum_{i,t\geq 120}[y_{it} - \hat{\beta}_{t-1}(z_{i,t-1})'\hat{f}_{t-1,t}]^2}{\sum_{i,t\geq 120} y^2_{it}}, \tag{F.7}$$

$$R^2_{f,T,N,O} = 1 - \frac{1}{N}\sum_i \frac{\sum_{t\geq 120}[y_{it} - \hat{\beta}_{t-1}(z_{i,t-1})'\hat{f}_{t-1,t}]^2}{\sum_{t\geq 120} y^2_{it}}, \tag{F.8}$$

$$R^2_{f,N,T,O} = 1 - \frac{1}{T-120}\sum_{t\geq 120} \frac{\sum_i[y_{it} - \hat{\beta}_{t-1}(z_{i,t-1})'\hat{f}_{t-1,t}]^2}{\sum_i y^2_{it}}. \tag{F.9}$$

F.1 Model Estimation

Table F.I show the first two moments of the long-short factors of Fama and French (2015) constructed based on the dataset used in this paper and the corresponding factors from Kenneth R. FrenchâĂŹs Website as well as their correlations.

Table F.II collects characteristics used in different specifications.

Tables F.III-F.V present the model estimation results under the restriction $\alpha(\cdot) = 0$. Similar to the unrestricted cases, the nonlinear specifications demonstrate both better in-sample and out-of-sample performance than the linear specification in most cases. The linearity of $\beta(\cdot)$ is also rejected at the 1% level in all cases. However, R^2, $R^2_{T,N}$, and $R^2_{N,T}$ are slightly smaller than those in the unrestricted cases. Unlike the unrestricted cases, increasing the number of factors may worsen the out-of-sample predictive R^2's. Specifically, $R^2_{N,T,O}$ can become negative for $K \geq 4$ in the linear specification. In the nonlinear specifications, R^2_O and $R^2_{O,T,N}$ exhibit a hump-shaped relationship with the number factors, peaking at three or four factors depending on the specification of $\beta(\cdot)$.

Table F.VI shows the correlations between our factors and the long-short factors of Fama and French (2015), and Tables F.VII-F.XII present the projection regression results. Our findings reveal substantial correlations between these factors and our factors, with both sets explaining significant variations in each other. However, the long-short factors fail to price most of our factors derived from our nonlinear specifications, while the pricing errors for most of the factors from our linear specification are not statistically significant. This suggests that the nonlinear specifications capture additional common variations in stock returns not accounted by the long-short factors. It is also observed that our factors are unable to fully explain the cross-sectional variations of some long-short factors. This observation arises because long-short factors are mixed with pricing errors, as they do not effectively distinguish between the risk and mispricing explanations of the role of characteristics in predicting stock returns, as discussed in Section 3.1.

Figures F.1-F.4 illustrate the contribution of each characteristic to pricing errors and risk exposures under the two nonlinear specifications. Almost all sieve coefficients are significant, which aligns with the strong evidence of nonlinearity found in Tables I-III.

Figure F.5 displays the patterns of $R^2_{f,O}$, $R^2_{f,T,N,O}$, and $R^2_{f,N,T,O}$ for ten factor models in subsample analysis. It shares the similar findings with Figure 3.

F.2 Asset Pricing Tests

The following characteristics are used in double sorted portfolios in Table VI and this section. Size: market capitalization; BM: book-to-market equity ratio; OP: operating profitability; INV: growth rate of assets; MOM: momentum; Beta: market beta; AC: accruals; NI: net stock issues; Var: variance of daily total returns.

Table F.XIII shows the bilateral correlations and standard deviations of the testing portfolios in our asset pricing tests.

Tables F.XIV and F.XV present the results for five groups of testing portfolios with $K = 6$ (corresponding to Tables VI and VII), where "MOM" represent the momentum factor Tables F.XVI-F.XXIII collect results for remaining testing portfolios, where Tables F.XVI-F.XIX correspond to $K = 5$ and Tables F.XX-F.XXIII correspond to $K = 6$. In explaining the Fama-MacBeth managed portfolios, our factors and IPCA's factors continue to outperform others for $K = 6$ in terms of average absolute intercepts. In explaining the sorted portfolios, our factors continue to outperform IPCA's factors for $K = 6$, as evidenced by smaller pricing errors, t-statistics, and GRS statistics. The finding of larger regression R^2's persists. In explaining IPCA's managed portfolios, the inferior performance of IPCA's factors is also observed for $K = 6$. The findings are robust for additional testing portfolios, as evidenced in Tables F.XVI-F.XXIII.

Table F.I. Summary statistics of factors[†]

Factors	MKT	SMB	HML	RMW	CMA	MOM
Based on the dataset in Kelly et al. (2019)						
Mean	0.50	0.16	0.63	0.27	0.45	0.85
Standard deviation	4.64	4.04	2.74	2.06	2.04	4.59
t	2.53	0.91	5.44	3.05	5.11	4.35
Kenneth R. French's Website						
Mean	0.48	0.18	0.38	0.30	0.37	0.67
Standard deviation	4.61	3.06	2.91	2.32	2.00	4.38
t	2.46	1.36	3.04	3.03	4.35	3.58
Correlation	0.99	0.84	0.70	0.24	0.85	0.95

[†] MKT: market excess return; SMB: small-minus-big factor; HML: high-minus-low factor; RMW: robust-minus-weak factor; CMA: conservative-minus-aggressive factor; MOM: momentum factor.

Table F.II. Characteristics used in different specifications[†]

	Linear specification	Nonlinear specification 1	Nonlinear specification 2
constant	✓	✓	✓
a2me	✓		
assets	✓	✓	✓
ato	✓	✓	
beta	✓	✓	✓
bidask	✓		
bm	✓	✓	✓
c	✓		
cto	✓		
d2a	✓	✓	
dpi2a	✓		
e2p	✓		
fc2y	✓		
freecf	✓		
idiovol	✓	✓	✓
intmom	✓		
invest	✓	✓	✓
lev	✓		
ltrevÂă	✓	✓	✓
mktcap	✓	✓	✓
mom	✓	✓	✓
noa	✓	✓	
oa	✓		
ol	✓		
pcm	✓	✓	
pm	✓		
prof	✓	✓	✓
q	✓		
rna	✓		
roa	✓		
roe	✓	✓	
s2p	✓		
sga2s	✓		
strevÂă	✓	✓	✓
suv	✓	✓	✓
turn	✓	✓	✓
w52h	✓	✓	

Table F.III. Results under linear specification of $\beta(\cdot)$ with 36 characteristics[†]

Restricted ($\alpha(\cdot) = 0$)										
K	R^2_K	R^2	$R^2_{T,N}$	$R^2_{N,T}$	$R^2_{f,O}$	$R^2_{f,T,N,O}$	$R^2_{f,N,T,O}$	R^2_O	$R^2_{T,N,O}$	$R^2_{N,T,O}$
1	26.62	2.14	0.58	0.06	6.79	4.10	5.98	0.20	0.09	0.07
2	36.48	4.18	1.72	1.37	13.66	10.55	11.33	0.28	0.34	0.02
3	45.10	5.32	2.98	2.30	14.20	11.17	11.77	0.26	0.31	0.01
4	52.62	11.45	8.03	8.86	14.74	12.16	12.16	0.31	0.39	-0.01
5	58.72	11.69	8.18	9.10	15.13	12.70	12.48	0.36	0.47	-0.04
6	64.28	13.85	10.06	11.58	15.32	12.96	12.69	0.38	0.47	-0.11
7	69.26	15.20	11.71	13.17	15.58	13.18	12.96	0.40	0.50	-0.13
8	72.98	15.53	11.99	13.44	15.90	13.46	13.24	0.41	0.53	-0.13
9	76.40	15.73	12.15	13.68	16.25	13.96	13.50	0.40	0.53	-0.08
10	79.29	15.90	12.37	13.85	16.42	14.21	13.70	0.41	0.51	-0.06
K	$R^2_{\tilde{Y}}$	R^2_f	$R^2_{f,T,N}$	$R^2_{f,N,T}$	p_α	p_{lin}				
1-10	20.89	R^2	$R^2_{T,N}$	$R^2_{N,T}$	NA	$< 1\%$				

[†] K: the number of factors specified; $R^2_{\tilde{Y}}$: Fama-MacBeth cross-sectional regression R^2 (%); R^2_K: the variation of the Fama-MacBeth managed portfolios $\tilde{Y}_t$ captured by the extracted factors $\hat{f}_t$ (%); R^2, $R^2_{T,N}$, $R^2_{N,T}$: various in-sample R^2's (%), see (22)-(24) with $\hat{\alpha}(\cdot) = 0$; R^2_f, $R^2_{f,T,N}$, $R^2_{f,N,T}$: various in-sample R^2's without $\alpha(\cdot)$ (%), see (F.1)-(F.3); $R^2_{f,O}$, $R^2_{f,T,N,O}$, $R^2_{f,N,T,O}$: various out-of-sample fit R^2's (%), see (F.7)-(F.9); R^2_O, $R^2_{T,N,O}$, $R^2_{N,T,O}$: various out-of-sample predictive R^2's (%), see (F.4)-(F.6) with $\hat{\alpha}_{t-1}(\cdot) = 0$; p_α and p_{lin}: the p-values of *alpha* test ($\alpha(\cdot) = 0$) and model specification test (linearity of $\beta(\cdot)$), respectively.

Table F.IV. Results under nonlinear specification of $\beta(\cdot)$ with 18 characteristics[†]

Restricted ($\alpha(\cdot)=0$)										
K	R^2_K	R^2	$R^2_{T,N}$	$R^2_{N,T}$	$R^2_{f,O}$	$R^2_{f,T,N,O}$	$R^2_{f,N,T,O}$	R^2_O	$R^2_{T,N,O}$	$R^2_{N,T,O}$
1	41.75	5.61	3.00	3.14	11.32	7.84	8.99	0.30	0.34	-0.12
2	59.20	9.14	5.56	6.26	14.01	11.40	11.34	0.34	0.30	-0.38
3	65.00	9.80	6.24	7.12	14.70	12.15	12.08	0.60	0.76	0.29
4	70.17	10.79	7.23	8.37	15.24	13.00	12.66	0.60	0.80	0.19
5	74.44	14.28	10.57	11.98	16.12	13.86	13.31	0.52	0.66	0.29
6	77.39	14.58	10.88	12.18	16.38	14.16	13.62	0.52	0.63	0.22
7	80.12	14.91	11.07	12.61	16.82	14.79	13.86	0.53	0.58	0.22
8	82.36	15.43	11.93	13.17	17.01	14.87	14.06	0.54	0.57	0.27
9	84.34	15.80	12.28	13.45	17.18	15.08	14.22	0.53	0.54	0.27
10	86.23	15.94	12.37	13.59	17.35	15.22	14.37	0.53	0.54	0.27
K	$R^2_{\tilde{Y}}$	R^2_f	$R^2_{f,T,N}$	$R^2_{f,N,T}$	p_α	p_{lin}				
1-10	21.11	R^2	$R^2_{T,N}$	$R^2_{N,T}$	NA	$< 1\%$				

[†] K: the number of factors specified; $R^2_{\tilde{Y}}$: Fama-MacBeth cross-sectional regression R^2 (%); R^2_K: the variation of the Fama-MacBeth managed portfolios $\tilde{Y}_t$ captured by the extracted factors $\hat{f}_t$ (%); R^2, $R^2_{T,N}$, $R^2_{N,T}$: various in-sample R^2's (%), see (22)-(24) with $\hat{\alpha}(\cdot)=0$; R^2_f, $R^2_{f,T,N}$, $R^2_{f,N,T}$: various in-sample R^2's without $\alpha(\cdot)$ (%), see (F.1)-(F.3); $R^2_{f,O}$, $R^2_{f,T,N,O}$, $R^2_{f,N,T,O}$: various out-of-sample fit R^2's (%), see (F.7)-(F.9); R^2_O, $R^2_{T,N,O}$, $R^2_{N,T,O}$: various out-of-sample predictive R^2's (%), see (F.4)-(F.6) with $\hat{\alpha}_{t-1}(\cdot)=0$; p_α and p_{lin}: the p-values of *alpha* test ($\alpha(\cdot)=0$) and model specification test (linearity of $\beta(\cdot)$), respectively.

Table F.V. Results under nonlinear specification of $\beta(\cdot)$ with 12 characteristics[†]

Restricted ($\alpha(\cdot)=0$)										
K	R^2_K	R^2	$R^2_{T,N}$	$R^2_{N,T}$	$R^2_{f,O}$	$R^2_{f,T,N,O}$	$R^2_{f,N,T,O}$	R^2_O	$R^2_{T,N,O}$	$R^2_{N,T,O}$
1	42.95	5.34	2.59	2.90	11.15	7.66	8.84	0.32	0.34	-0.10
2	61.58	9.12	5.45	6.15	13.79	11.11	11.04	0.33	0.21	-0.56
3	68.02	10.15	6.08	7.25	14.47	11.86	11.74	0.62	0.68	0.16
4	74.08	10.77	6.99	7.98	15.38	13.20	12.75	0.57	0.65	0.24
5	78.98	14.15	10.49	11.93	16.04	14.23	13.34	0.55	0.57	0.23
6	82.66	14.43	10.68	12.32	16.48	14.71	13.67	0.56	0.53	0.23
7	85.44	14.93	11.31	12.76	16.81	14.97	13.93	0.55	0.55	0.25
8	87.85	15.37	11.78	13.13	17.11	15.10	14.14	0.56	0.54	0.27
9	89.53	16.28	12.57	13.85	17.30	15.34	14.33	0.56	0.52	0.27
10	91.13	16.49	12.78	14.08	17.45	15.52	14.48	0.57	0.55	0.27

K	$R^2_{\tilde{Y}}$	R^2_f	$R^2_{f,T,N}$	$R^2_{f,N,T}$	p_α	p_{lin}
1-10	20.72	R^2	$R^2_{T,N}$	$R^2_{N,T}$	NA	$< 1\%$

[†] K: the number of factors specified; $R^2_{\tilde{Y}}$: Fama-MacBeth cross-sectional regression R^2 (%); R^2_K: the variation of the Fama-MacBeth managed portfolios $\tilde{Y}_t$ captured by the extracted factors $\hat{f}_t$ (%); R^2, $R^2_{T,N}$, $R^2_{N,T}$: various in-sample R^2's (%), see (22)-(24) with $\hat{\alpha}(\cdot)=0$; R^2_f, $R^2_{f,T,N}$, $R^2_{f,N,T}$: various in-sample R^2's without $\alpha(\cdot)$ (%), see (F.1)-(F.3); $R^2_{f,O}$, $R^2_{f,T,N,O}$, $R^2_{f,N,T,O}$: various out-of-sample fit R^2's (%), see (F.7)-(F.9); R^2_O, $R^2_{T,N,O}$, $R^2_{N,T,O}$: various out-of-sample predictive R^2's (%), see (F.4)-(F.6) with $\hat{\alpha}_{t-1}(\cdot)=0$; p_α and p_{lin}: the p-values of *alpha* test ($\alpha(\cdot)=0$) and model specification test (linearity of $\beta(\cdot)$), respectively.

Table F.VI. Factor correlations[†]

	MKT	SMB	HML	RMW	CMA	MOM
Linear specifications of $\alpha(\cdot)$ and $\beta(\cdot)$ with 36 characteristics						
Factor 1	0.02	0.11	0.10	-0.11	0.02	-0.36
Factor 2	0.26	0.26	0.05	-0.20	-0.02	-0.16
Factor 3	-0.26	-0.22	-0.05	0.11	0.07	0.18
Factor 4	0.58	0.46	-0.33	-0.30	-0.30	-0.06
Factor 5	0.10	0.05	-0.15	0.01	-0.16	0.05
Factor 6	0.32	0.24	-0.13	-0.20	-0.13	-0.04
Factor 7	0.30	0.19	0.02	-0.09	-0.10	-0.16
Factor 8	-0.04	-0.05	0.21	0.14	0.16	-0.41
Factor 9	0.02	0.03	-0.08	-0.06	-0.03	0.29
Factor 10	-0.03	0.01	0.13	0.09	0.08	0.01
Nonlinear specifications of $\alpha(\cdot)$ and $\beta(\cdot)$ with 18 characteristics						
Factor 1	0.24	0.37	0.07	-0.30	-0.02	-0.37
Factor 2	0.41	0.31	-0.37	-0.30	-0.29	-0.09
Factor 3	-0.22	-0.11	0.45	0.23	0.33	-0.50
Factor 4	0.48	0.21	-0.04	-0.15	-0.18	-0.34
Factor 5	-0.12	-0.01	-0.25	-0.06	-0.10	0.14
Factor 6	0.07	-0.21	-0.04	0.08	-0.08	0.09
Factor 7	-0.15	-0.22	-0.09	-0.06	-0.03	-0.12
Factor 8	-0.01	-0.08	0.14	0.09	0.05	0.05
Factor 9	-0.17	0.03	0.08	0.10	0.09	0.10
Factor 10	-0.21	-0.04	0.11	0.06	0.06	0.07
Nonlinear specifications of $\alpha(\cdot)$ and $\beta(\cdot)$ with 12 characteristics						
Factor 1	0.22	0.36	0.05	-0.32	-0.01	-0.32
Factor 2	0.39	0.31	-0.27	-0.26	-0.26	-0.34
Factor 3	-0.19	-0.19	0.49	0.25	0.30	-0.57
Factor 4	0.50	0.21	-0.12	-0.16	-0.24	-0.13
Factor 5	-0.04	-0.15	-0.11	-0.02	-0.07	0.12
Factor 6	-0.20	-0.20	-0.14	0.06	-0.07	0.08
Factor 7	-0.05	-0.04	0.16	0.12	0.04	0.00
Factor 8	-0.04	0.19	0.28	0.06	0.22	0.18
Factor 9	0.27	0.06	0.18	-0.14	0.16	0.11
Factor 10	0.17	-0.27	-0.22	0.21	-0.20	-0.06

[†] MKT: market excess return; SMB: small-minus-big factor; HML: high-minus-low factor; RMW: robust-minus-weak factor; CMA: conservative-minus-aggressive factor; MOM: momentum factor.

Table F.VII. Factor projections: linear specifications of $\alpha(\cdot)$ and $\beta(\cdot)$ with 36 characteristics[†]

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8	Factor 9	Factor 10
Constant	3.71***	0.14	0.39	-0.05	0.25	0.28	0.46	0.62***	0.04	-0.10
t	[5.83]	[0.36]	[1.03]	[-0.17]	[0.75]	[0.94]	[1.63]	[2.71]	[0.17]	[-0.42]
MKT	-0.20	0.43***	-0.42***	0.79***	0.09	0.40***	0.37***	0.02	0.04	0.02
t	[-1.31]	[4.61]	[-4.76]	[12.30]	[1.12]	[5.80]	[5.64]	[0.40]	[0.73]	[0.32]
SMB	0.42**	0.54***	-0.46***	0.80***	0.07	0.31***	0.27***	0.02	-0.00	0.09
t	[1.98]	[4.07]	[-3.68]	[8.73]	[0.64]	[3.11]	[2.87]	[0.24]	[-0.01]	[1.14]
HML	0.31	0.38**	-0.46**	-0.61***	-0.15	-0.13	0.37***	-0.04	0.03	0.25**
t	[1.02]	[2.03]	[-2.57]	[-4.69]	[-0.93]	[-0.89]	[2.72]	[-0.41]	[0.24]	[2.30]
RMW	-0.39	-0.33*	-0.02	-0.16	0.10	-0.24*	0.03	0.47***	-0.20*	0.19*
t	[-1.34]	[-1.83]	[-0.11]	[-1.31]	[0.68]	[-1.81]	[0.27]	[4.59]	[-1.88]	[1.83]
CMA	-0.17	0.01	0.28	0.26	-0.33	0.11	-0.31	0.57***	-0.11	-0.03
t	[-0.37]	[0.04]	[1.06]	[1.33]	[-1.41]	[0.51]	[-1.53]	[3.51]	[-0.62]	[-0.19]
MOM	-1.20***	-0.19**	0.22***	-0.02	0.08	0.02	-0.13**	-0.57***	0.38***	0.04
t	[-8.43]	[-2.14]	[2.64]	[-0.28]	[1.13]	[0.24]	[-1.98]	[-11.10]	[7.00]	[0.78]
R^2	14.82	13.77	12.47	46.37	3.23	13.87	12.74	23.66	9.23	2.56

[†] MKT: market excess return; SMB: small-minus-big factor; HML: high-minus-low factor; RMW: robust-minus-weak factor; CMA: conservative-minus-aggressive factor; MOM: momentum factor; ***: p-value $< 1\%$; **: p-value $< 5\%$; *: p-value $< 10\%$.

Table F.VIII. Factor projections: nonlinear specifications of $\alpha(\cdot)$ and $\beta(\cdot)$ with 18 characteristics[†]

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8	Factor 9	Factor 10
Constant	2.96***	1.03***	0.43**	0.61***	1.49***	0.95***	0.82***	2.25***	0.81***	2.26***
t	[5.21]	[2.79]	[2.42]	[3.31]	[8.57]	[5.51]	[5.53]	[14.77]	[5.70]	[16.25]
MKT	0.36***	0.55***	-0.14***	0.44***	-0.18***	0.10**	-0.15***	0.04	-0.10***	-0.15***
t	[2.68]	[6.36]	[-3.26]	[10.14]	[-4.38]	[2.54]	[-4.26]	[1.14]	[-3.02]	[-4.55]
SMB	1.33***	0.51***	0.03	0.15**	-0.02	-0.31***	-0.30***	-0.07	0.12***	0.02
t	[6.98]	[4.11]	[0.52]	[2.33]	[-0.34]	[-5.28]	[-5.98]	[-1.34]	[2.60]	[0.47]
HML	0.45*	-1.11***	0.35***	0.18**	-0.41***	0.08	-0.24***	0.25***	0.05	0.17**
t	[1.67]	[-6.33]	[4.19]	[2.02]	[-4.98]	[0.96]	[-3.38]	[3.45]	[0.73]	[2.58]
RMW	-0.95***	-0.51***	0.57***	-0.00	-0.16**	0.01	-0.26***	0.06	0.13**	-0.01
t	[-3.69]	[-3.08]	[7.04]	[-0.03]	[-2.01]	[0.17]	[-3.85]	[0.93]	[2.06]	[-0.12]
CMA	0.06	0.34	0.45***	-0.18	0.02	-0.19	0.04	-0.14	0.02	-0.21**
t	[0.15]	[1.30]	[3.53]	[-1.37]	[0.14]	[-1.58]	[0.37]	[-1.30]	[0.19]	[-2.13]
MOM	-1.03***	-0.20**	-0.63***	-0.28***	0.07*	0.09**	-0.14***	0.07**	0.06*	0.05
t	[-8.06]	[-2.44]	[-15.77]	[-6.75]	[1.68]	[2.35]	[-4.18]	[2.03]	[1.87]	[1.62]
R^2	29.29	30.86	48.20	31.51	11.52	7.45	13.83	3.81	5.15	5.93

[†] MKT: market excess return; SMB: small-minus-big factor; HML: high-minus-low factor; RMW: robust-minus-weak factor; CMA: conservative-minus-aggressive factor; MOM: momentum factor; ***: p-value $< 1\%$; **: p-value $< 5\%$; *: p-value $< 10\%$.

Table F.IX. Factor projections: nonlinear specifications of $\alpha(\cdot)$ and $\beta(\cdot)$ with 12 characteristics[†]

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8	Factor 9	Factor 10
Constant	3.57***	1.99***	0.27	0.47**	1.43***	2.67***	2.43***	1.93***	-0.16	0.36***
t	[5.47]	[4.83]	[1.37]	[2.08]	[6.36]	[14.11]	[12.86]	[12.03]	[-1.19]	[2.70]
MKT	0.36**	0.51***	-0.11**	0.56***	-0.05	-0.22***	-0.01	0.04	0.30***	0.16***
t	[2.38]	[5.23]	[-2.41]	[10.38]	[-1.03]	[-4.90]	[-0.16]	[1.17]	[9.24]	[5.17]
SMB	1.39***	0.68***	-0.17***	0.14*	-0.30***	-0.25***	0.02	0.33***	-0.05	-0.30***
t	[6.35]	[4.93]	[-2.62]	[1.89]	[-3.94]	[-3.88]	[0.30]	[6.22]	[-1.07]	[-6.72]
HML	0.39	-1.03***	0.64***	0.17	-0.12	-0.27***	0.34***	0.45***	0.27***	-0.28***
t	[1.25]	[-5.26]	[6.83]	[1.60]	[-1.10]	[-2.99]	[3.75]	[5.90]	[4.16]	[-4.46]
RMW	-1.28***	-0.32*	0.64***	-0.06	-0.21**	-0.08	0.19**	0.18**	-0.16***	0.28***
t	[-4.32]	[-1.72]	[7.16]	[-0.59]	[-2.02]	[-0.90]	[2.25]	[2.43]	[-2.60]	[4.66]
CMA	0.19	0.24	0.27*	-0.31*	-0.18	-0.11	-0.25*	0.03	0.24**	0.07
t	[0.40]	[0.80]	[1.91]	[-1.93]	[-1.09]	[-0.82]	[-1.86]	[0.25]	[2.43]	[0.71]
MOM	-1.01***	-0.86***	-0.83***	-0.06	0.12**	0.01	0.03	0.22***	0.16***	-0.07**
t	[-6.89]	[-9.29]	[-18.86]	[-1.17]	[2.36]	[0.28]	[0.79]	[6.01]	[5.34]	[-2.46]
R^2	26.23	33.78	57.02	26.17	5.62	10.69	4.49	19.40	21.20	20.97

[†] MKT: market excess return; SMB: small-minus-big factor; HML: high-minus-low factor; RMW: robust-minus-weak factor; CMA: conservative-minus-aggressive factor; MOM: momentum factor; ***: p-value $< 1\%$; **: p-value $< 5\%$; *: p-value $< 10\%$.

Table F.X. Factor projections: linear specifications of $\alpha(\cdot)$ and $\beta(\cdot)$ with 36 characteristics[†]

	MKT	SMB	HML	RMW	CMA	MOM
Constant	0.12	-0.05	0.32***	0.40***	0.39***	1.20***
t	[1.05]	[-0.46]	[2.83]	[4.45]	[4.84]	[8.46]
Factor 1	0.01	0.02***	0.02***	-0.02***	0.00	-0.11***
t	[0.82]	[3.30]	[2.77]	[-2.79]	[0.74]	[-11.65]
Factor 2	0.13***	0.08***	0.02	-0.05***	-0.00	-0.07***
t	[10.58]	[7.92]	[1.35]	[-5.21]	[-0.43]	[-5.06]
Factor 3	-0.14***	-0.08***	-0.02	0.03***	0.01	0.09***
t	[-10.75]	[-6.63]	[-1.30]	[2.77]	[1.63]	[5.72]
Factor 4	0.33***	0.17***	-0.12***	-0.08***	-0.07***	-0.03*
t	[23.52]	[13.95]	[-8.79]	[-7.83]	[-7.64]	[-1.79]
Factor 5	0.06***	0.02	-0.06***	0.00	-0.04***	0.03*
t	[4.11]	[1.34]	[-3.85]	[0.15]	[-3.99]	[1.66]
Factor 6	0.22***	0.11***	-0.05***	-0.07***	-0.04***	-0.02
t	[13.25]	[7.35]	[-3.43]	[-5.38]	[-3.23]	[-1.09]
Factor 7	0.21***	0.09***	0.01	-0.03**	-0.03**	-0.10***
t	[12.14]	[5.78]	[0.69]	[-2.40]	[-2.53]	[-4.98]
Factor 8	-0.03	-0.03	0.11***	0.06***	0.06***	-0.32***
t	[-1.63]	[-1.55]	[5.51]	[3.66]	[4.31]	[-13.25]
Factor 9	0.02	0.02	-0.04**	-0.03*	-0.01	0.23***
t	[1.00]	[0.83]	[-2.19]	[-1.66]	[-0.93]	[9.10]
Factor 10	-0.03	0.01	0.08***	0.04**	0.03**	0.01
t	[-1.31]	[0.29]	[3.45]	[2.20]	[2.03]	[0.45]
R^2	67.68	42.93	23.21	22.88	18.01	47.39

[†] MKT: market excess return; SMB: small-minus-big factor; HML: high-minus-low factor; RMW: robust-minus-weak factor; CMA: conservative-minus-aggressive factor; MOM: momentum factor; ***: p-value $< 1\%$; **: p-value $< 5\%$; *: p-value $< 10\%$.

Table F.XI. Factor projections: nonlinear specifications of $\alpha(\cdot)$ and $\beta(\cdot)$ with 18 characteristics[†]

	MKT	SMB	HML	RMW	CMA	MOM
Constant	1.04***	0.33**	0.06	0.19	0.32***	0.63***
t	[5.58]	[2.14]	[0.44]	[1.46]	[2.83]	[3.36]
Factor 1	0.07***	0.08***	0.01**	-0.05***	-0.00	-0.11***
t	[8.98]	[11.00]	[2.24]	[-8.34]	[-0.16]	[-12.95]
Factor 2	0.20***	0.10***	-0.11***	-0.07***	-0.06***	-0.04***
t	[15.60]	[9.20]	[-11.81]	[-8.19]	[-7.78]	[-3.27]
Factor 3	-0.19***	-0.06***	0.25***	0.10***	0.12***	-0.41***
t	[-8.27]	[-3.32]	[14.35]	[6.54]	[9.10]	[-17.69]
Factor 4	0.45***	0.13***	-0.02	-0.07***	-0.07***	-0.30***
t	[18.03]	[6.31]	[-0.90]	[-4.22]	[-4.90]	[-11.88]
Factor 5	-0.13***	-0.01	-0.18***	-0.03*	-0.06***	0.15***
t	[-4.44]	[-0.31]	[-8.09]	[-1.68]	[-3.12]	[5.03]
Factor 6	0.08**	-0.17***	-0.03	0.05**	-0.04**	0.10***
t	[2.47]	[-6.33]	[-1.18]	[2.33]	[-2.25]	[3.06]
Factor 7	-0.20***	-0.20***	-0.08***	-0.04*	-0.01	-0.15***
t	[-5.72]	[-6.62]	[-2.97]	[-1.81]	[-0.51]	[-4.12]
Factor 8	-0.01	-0.07**	0.12***	0.06**	0.03	0.07*
t	[-0.39]	[-2.45]	[4.35]	[2.39]	[1.44]	[1.83]
Factor 9	-0.24***	0.03	0.07**	0.07***	0.06**	0.14***
t	[-6.32]	[1.03]	[2.53]	[2.81]	[2.54]	[3.51]
Factor 10	-0.31***	-0.04	0.10***	0.05*	0.04*	0.10**
t	[-8.00]	[-1.24]	[3.49]	[1.82]	[1.74]	[2.51]
R^2	62.06	39.35	46.33	29.54	26.49	56.75

[†] MKT: market excess return; SMB: small-minus-big factor; HML: high-minus-low factor; RMW: robust-minus-weak factor; CMA: conservative-minus-aggressive factor; MOM: momentum factor; ***: p-value $< 1\%$; **: p-value $< 5\%$; *: p-value $< 10\%$.

Table F.XII. Factor projections: nonlinear specifications of $\alpha(\cdot)$ and $\beta(\cdot)$ with 12 characteristics[†]

	MKT	SMB	HML	RMW	CMA	MOM
Constant	0.45**	0.03	0.01	0.18	0.29***	0.44**
t	[2.53]	[0.20]	[0.11]	[1.50]	[2.83]	[2.58]
Factor 1	0.06***	0.07***	0.01*	-0.04***	0.00	-0.09***
t	[8.55]	[11.46]	[1.89]	[-9.17]	[0.01]	[-12.37]
Factor 2	0.16***	0.09***	-0.07***	-0.05***	-0.05***	-0.14***
t	[14.71]	[10.03]	[-9.30]	[-7.35]	[-7.39]	[-13.05]
Factor 3	-0.13***	-0.09***	0.22***	0.09***	0.10***	-0.38***
t	[-7.18]	[-6.19]	[16.92]	[7.23]	[8.85]	[-21.59]
Factor 4	0.40***	0.11***	-0.06***	-0.06***	-0.08***	-0.10***
t	[18.92]	[6.65]	[-3.94]	[-4.54]	[-6.67]	[-5.05]
Factor 5	-0.04	-0.09***	-0.07***	-0.01	-0.03**	0.11***
t	[-1.60]	[-4.85]	[-3.91]	[-0.34]	[-2.26]	[4.60]
Factor 6	-0.21***	-0.14***	-0.09***	0.03*	-0.03*	0.08***
t	[-7.53]	[-6.33]	[-4.86]	[1.78]	[-1.96]	[3.01]
Factor 7	-0.06**	-0.03	0.11***	0.07***	0.02	0.00
t	[-1.98]	[-1.35]	[5.39]	[3.80]	[1.36]	[0.17]
Factor 8	-0.05	0.15***	0.21***	0.03	0.11***	0.20***
t	[-1.45]	[6.11]	[9.56]	[1.61]	[6.05]	[6.75]
Factor 9	0.38***	0.05*	0.15***	-0.10***	0.09***	0.15***
t	[10.35]	[1.84]	[6.12]	[-4.24]	[4.32]	[4.22]
Factor 10	0.25***	-0.25***	-0.20***	0.15***	-0.13***	-0.07**
t	[6.57]	[-8.67]	[-7.56]	[6.02]	[-5.84]	[-2.03]
R^2	62.97	47.96	54.55	34.67	34.05	62.92

[†] MKT: market excess return; SMB: small-minus-big factor; HML: high-minus-low factor; RMW: robust-minus-weak factor; CMA: conservative-minus-aggressive factor; MOM: momentum factor; ***: p-value $< 1\%$; **: p-value $< 5\%$; *: p-value $< 10\%$.

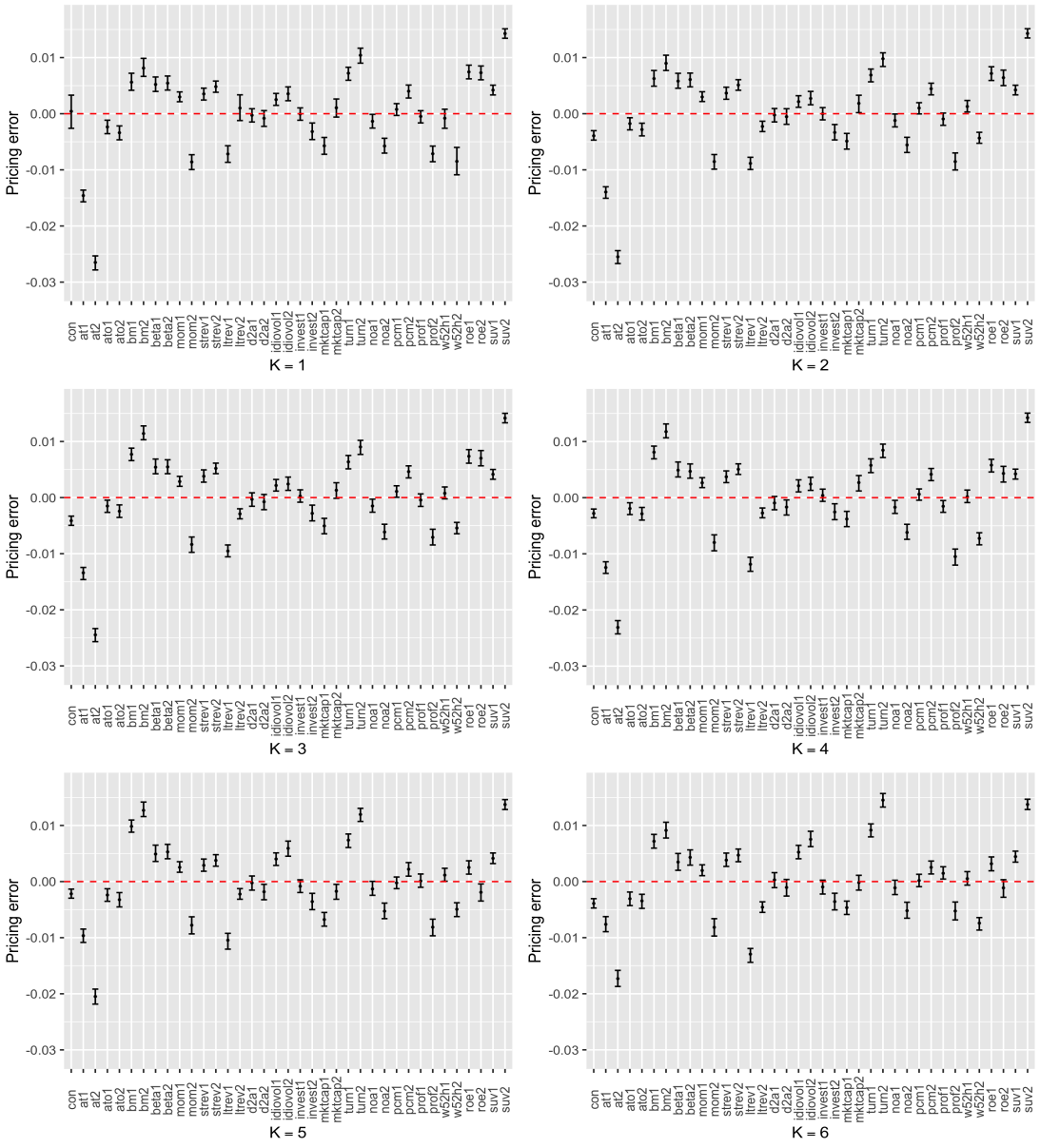


Figure F.1. 95% confidence intervals for coefficients in $\alpha(\cdot)$ under nonlinear specifications of $\alpha(\cdot)$ and $\beta(\cdot)$ with 18 characteristics

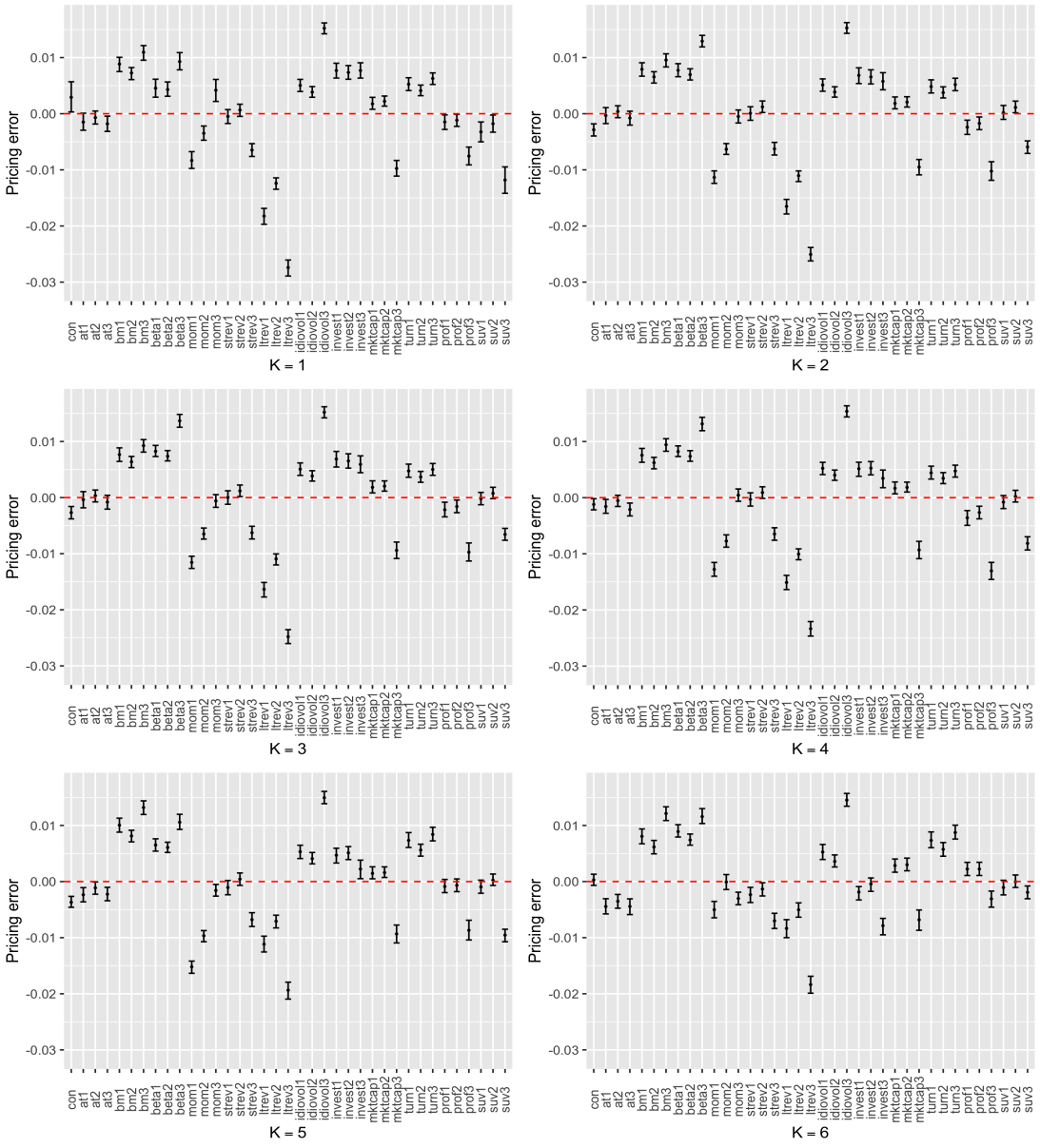


Figure F.2. 95% confidence intervals for coefficients in $\alpha(\cdot)$ under nonlinear specifications of $\alpha(\cdot)$ and $\beta(\cdot)$ with 12 characteristics

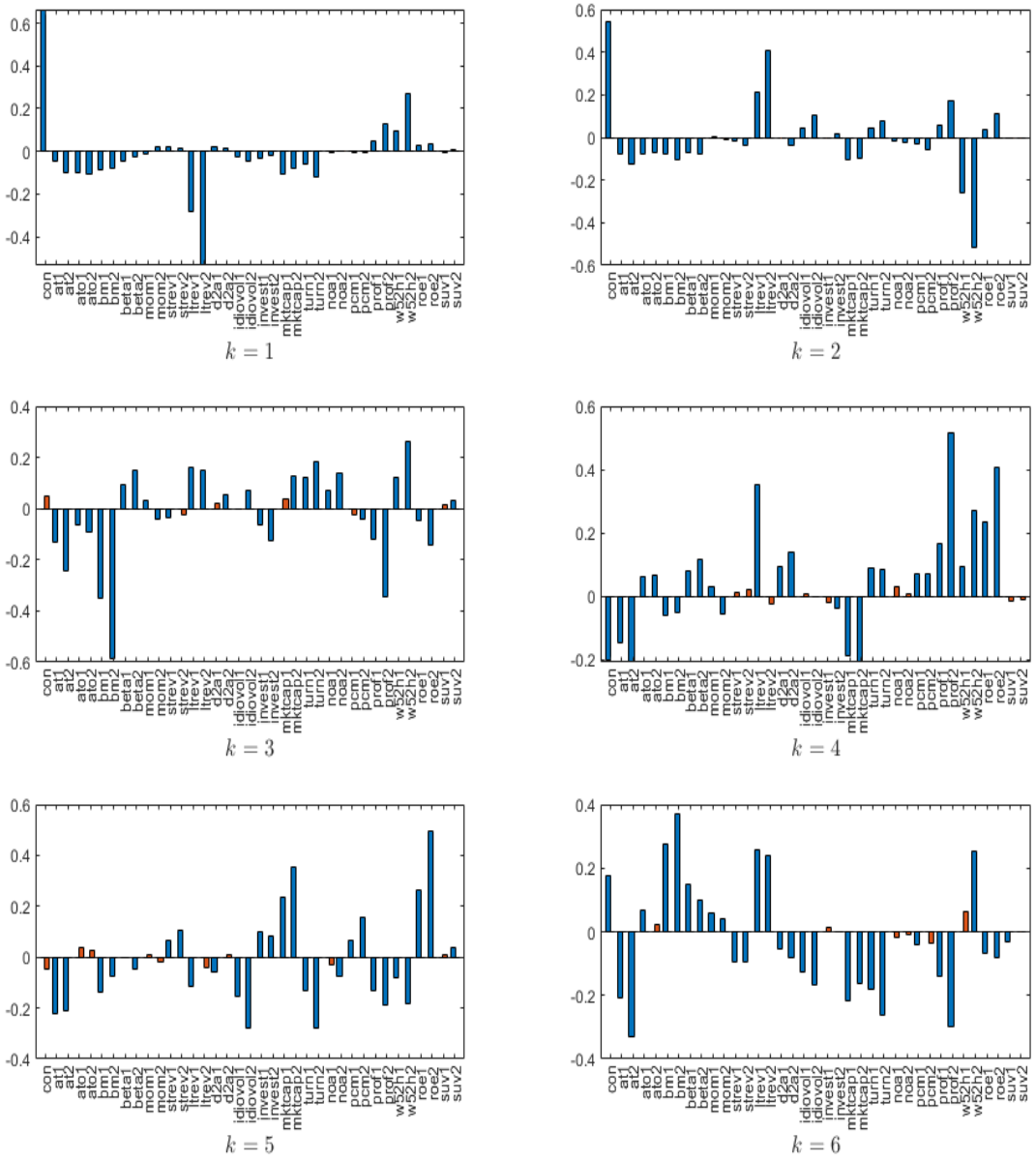


Figure F.3. Estimates of coefficients in $\beta(\cdot)$ under nonlinear specifications of $\alpha(\cdot)$ and $\beta(\cdot)$ with 18 characteristics (blue: significant at the 5% level; red: insignificant)

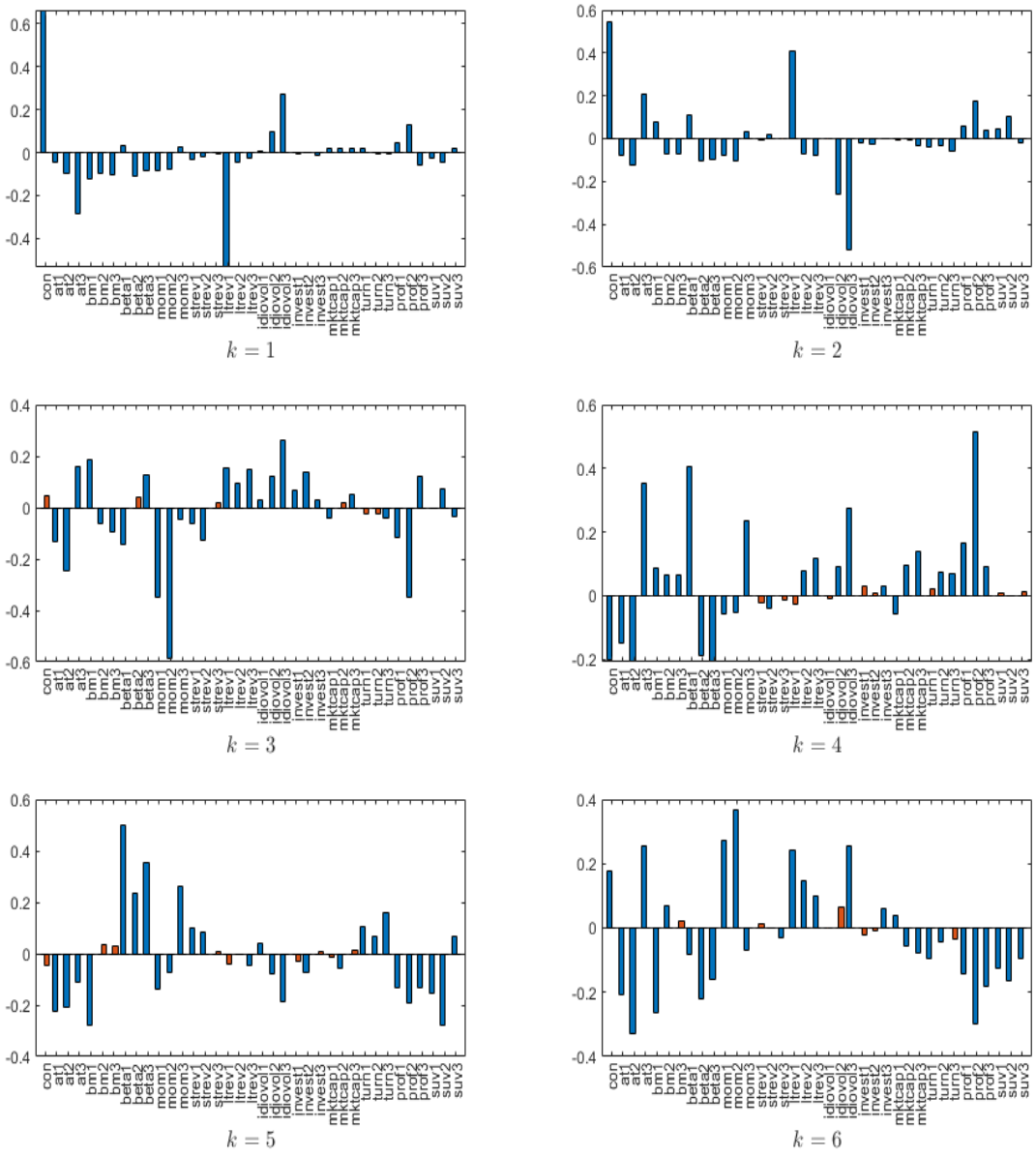


Figure F.4. Estimates of coefficients in $\beta(\cdot)$ under nonlinear specifications of $\alpha(\cdot)$ and $\beta(\cdot)$ with 12 characteristics (blue: significant at the 5% level; red: insignificant)

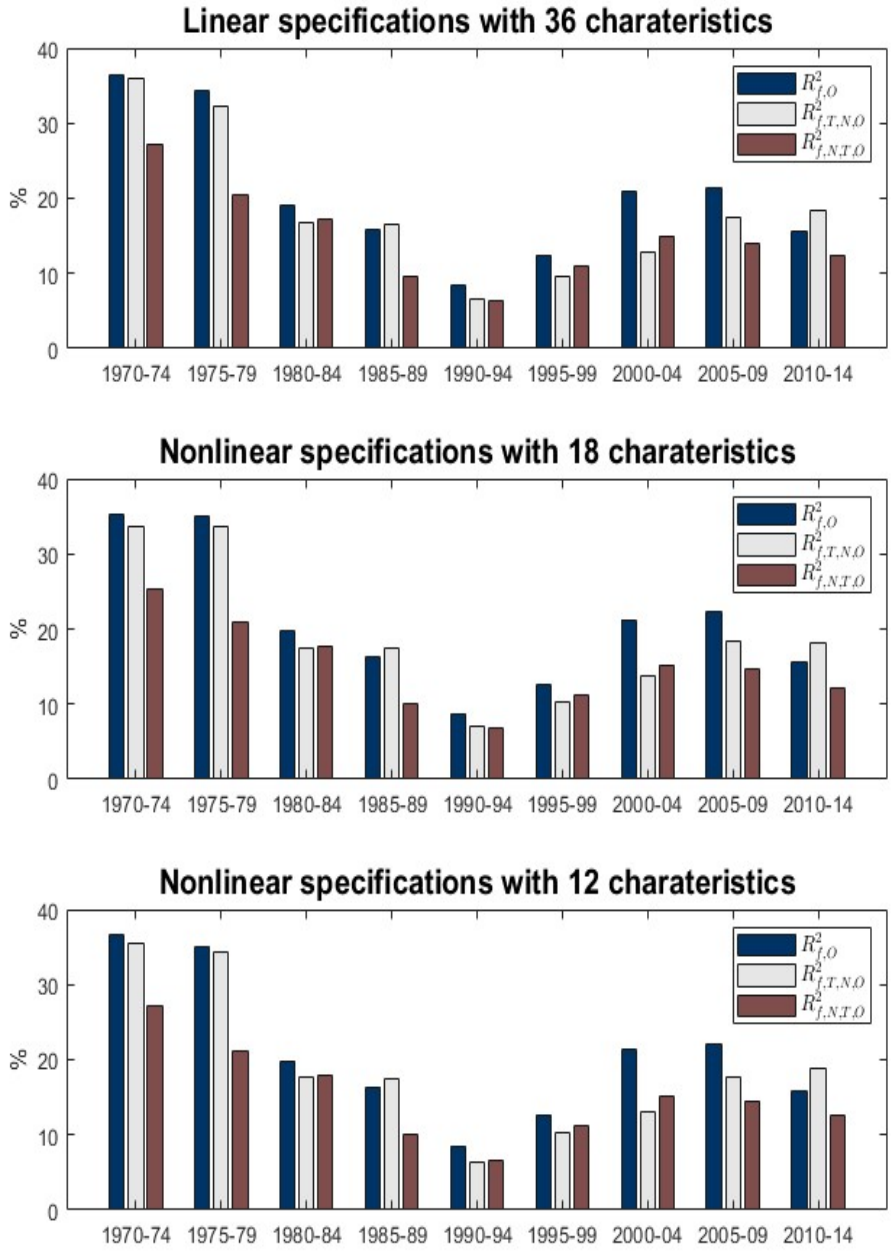


Figure F.5. $R^2_{f,O}$, $R^2_{f,T,N,O}$, and $R^2_{f,N,T,O}$ ((F.7)-(F.9)) with $K = 10$: subsample analysis

Table F.XIII. Bilateral correlations and standard deviations of testing portfolios[†]

	Correlation							Standard deviation		
Testing portfolios	Min	Median	Max	AMean	AMin	AMedian	AMax	Min	Median	Max
Regressed-PCA's 36 managed portfolios	-0.87	-0.01	0.63	0.09	0.00	0.06	0.87	0.01	0.04	0.12
100 double sorted portfolios on Size and BM, OP, INV, and MOM	0.53	0.85	0.99	0.84	0.53	0.85	0.99	0.04	0.06	0.08
110 double sorted portfolios on Size and Beta, AC, NI, and Var	0.42	0.85	0.98	0.83	0.42	0.85	0.98	0.04	0.06	0.09
IPCA's 36 managed portfolios	-0.97	-0.00	0.98	0.36	0.00	0.33	0.98	0.00	0.00	0.01
110 single sorted portfolios on 55 characteristics (P1&10)	0.34	0.80	1.00	0.79	0.34	0.80	1.00	0.04	0.06	0.09
72 single sorted portfolios on 36 characteristics (P1&10)	0.39	0.75	1.00	0.74	0.39	0.75	1.00	0.04	0.06	0.10
25 double sorted portfolios on Size and BM	0.53	0.84	0.97	0.82	0.53	0.84	0.97	0.04	0.06	0.08
25 double sorted portfolios on Size and OP	0.64	0.86	0.97	0.86	0.64	0.86	0.97	0.04	0.06	0.07
25 double sorted portfolios on Size and INV	0.63	0.87	0.97	0.85	0.63	0.87	0.97	0.04	0.06	0.07
25 double sorted portfolios on Size and MOM	0.54	0.82	0.97	0.81	0.54	0.82	0.97	0.04	0.06	0.08
25 double sorted portfolios on Size and Beta	0.65	0.89	0.97	0.86	0.65	0.89	0.97	0.04	0.06	0.07
25 double sorted portfolios on Size and AC	0.44	0.83	0.97	0.81	0.44	0.83	0.97	0.04	0.06	0.08
35 double sorted portfolios on Size and NI	0.47	0.84	0.97	0.83	0.47	0.84	0.97	0.04	0.05	0.08
25 double sorted portfolios on Size and Var	0.46	0.84	0.97	0.83	0.46	0.84	0.97	0.04	0.06	0.09
Regressed-PCA S1's 36 managed portfolios	-0.77	0.00	0.83	0.12	0.00	0.08	0.83	0.01	0.02	0.09
Regressed-PCA S2's 36 managed portfolios	-0.74	0.01	0.96	0.16	0.00	0.10	0.96	0.01	0.03	0.09

[†] Min: minimum value; Median: median value; Max: maximal value; AMean: average absolute value; AMin: minimum absolute value; AMedian: median absolute value; AMax: maximal absolute value.

Table F.XIV. Comparing asset pricing tests: $K = 6^{\dagger}$

Testing portfolios/Factors	$A\|a\|$	$A\|t(a)\|$	$Aa^2/V\overline{r}$	$A\lambda^2/V\overline{r}$	$As(a)$	$As(e)$	$Sh^2(a)$	$Sh^2(f)$	AR^2	GRS	$p(GRS)$
Group I: Regressed-PCA's 36 managed portfolios											
Regressed-PCA	0.39	3.51	0.64	0.61	0.12	2.75	1.97	0.04	33.27	26.64	0.00
Regressed-PCA S1	0.41	3.08	0.53	0.47	0.17	3.60	2.47	0.22	19.43	28.49	0.00
Regressed-PCA S2	0.40	2.76	0.50	0.43	0.19	3.68	2.48	0.42	17.68	24.52	0.00
IPCA	0.43	3.07	0.48	0.41	0.18	3.67	3.13	0.33	20.17	33.15	0.00
IPCA\Regressed-PCA	0.53	3.22	1.02	0.94	0.19	3.92	3.14	0.28	14.33	34.41	0.00
FF5+MOM	0.52	3.18	1.28	1.21	0.18	3.95	2.59	0.17	10.83	31.28	0.00
KNS	0.50	3.35	1.12	1.05	0.17	3.95	2.55	0.05	11.25	34.33	0.00
Group II: 100 double sorted portfolios on Size and BM, OP, INV, and MOM											
Regressed-PCA	0.74	4.95	11.01	10.58	0.15	3.52	0.99	0.04	60.32	4.22	0.00
Regressed-PCA S1	0.55	3.01	6.27	5.66	0.18	3.88	1.22	0.22	52.70	4.43	0.00
Regressed-PCA S2	0.99	5.19	19.77	19.09	0.19	3.78	1.37	0.42	54.71	4.28	0.00
IPCA	0.86	9.66	14.83	14.65	0.10	1.96	3.92	0.33	86.71	13.05	0.00
IPCA\Regressed-PCA	1.29	5.54	30.77	29.71	0.24	4.93	1.57	0.28	27.02	5.43	0.00
FF5+MOM	0.38	4.53	2.75	2.61	0.09	1.90	1.71	0.17	88.67	6.48	0.00
KNS	0.87	5.81	13.95	13.54	0.15	3.42	0.94	0.05	62.92	3.96	0.00
Group III: 110 double sorted portfolios on Size and Beta, AC, NI, and Var											
Regressed-PCA	0.75	4.96	12.51	12.02	0.15	3.51	1.25	0.04	59.88	4.70	0.00
Regressed-PCA S1	0.53	2.91	6.75	6.04	0.19	3.92	1.52	0.22	51.27	4.90	0.00
Regressed-PCA S2	0.98	5.10	21.83	21.03	0.20	3.82	1.66	0.42	53.13	4.59	0.00
IPCA	0.88	9.52	16.84	16.62	0.10	2.00	3.94	0.33	85.66	11.65	0.00
IPCA\Regressed-PCA	1.30	5.59	35.65	34.38	0.24	4.99	1.83	0.28	26.38	5.60	0.00
FF5+MOM	0.39	4.38	3.42	3.23	0.10	2.05	2.08	0.17	86.41	7.00	0.00
KNS	0.86	5.72	15.90	15.42	0.15	3.47	1.10	0.05	61.06	4.14	0.00

† $A|a|$: average absolute intercept; $A|t(a)|$: average absolute t-statistic for the intercepts; $Aa^2/V\overline{r}$: average squared intercept over the cross-section variance of $\overline{r}$, average returns of the testing portfolios; $A\lambda^2/V\overline{r}$: average difference between each squared intercept and its squared standard error divided by the variance of $\overline{r}$; $As(a)$: average standard error of the intercepts; $As(e)$: average residual standard deviation; $Sh^2(a)$: maximized squared Sharpe ratio for the intercepts; $Sh^2(f)$: maximized squared Sharpe ratio for the factors; AR^2: average regression R^2 (%); GRS: GRS statistic of Gibbons et al. (1989); $p(GRS)$: p-value of GRS.

Table F.XV. Comparing asset pricing tests: $K = 6$ (continued)[†]

Testing portfolios/Factors	$A\|a\|$	$A\|t(a)\|$	$Aa^2/V\overline{r}$	$A\lambda^2/V\overline{r}$	$As(a)$	$As(e)$	$Sh^2(a)$	$Sh^2(f)$	AR^2	GRS	$p(GRS)$
Group IV: IPCA's 36 managed portfolios											
Regressed-PCA	0.05	3.60	0.91	0.86	0.01	0.32	1.55	0.04	28.06	20.98	0.00
Regressed-PCA S1	0.06	5.36	1.47	1.44	0.01	0.25	1.86	0.22	51.86	21.46	0.00
Regressed-PCA S2	0.07	4.92	1.45	1.41	0.01	0.25	1.74	0.42	52.31	17.24	0.00
IPCA	0.06	5.77	1.24	1.22	0.01	0.21	2.29	0.33	66.51	24.27	0.00
IPCA\Regressed-PCA	0.06	5.06	1.38	1.33	0.01	0.31	1.92	0.28	40.25	21.06	0.00
FF5+MOM	0.04	3.39	0.79	0.76	0.01	0.25	1.39	0.17	54.97	16.76	0.00
KNS	0.04	3.83	0.76	0.73	0.01	0.25	1.46	0.05	55.27	19.69	0.00
Group V: 110 single sorted portfolios on 55 characteristics (P1&10)											
Regressed-PCA	0.63	3.81	5.61	5.27	0.17	3.87	1.36	0.04	52.76	5.12	0.00
Regressed-PCA S1	0.42	2.27	3.16	2.73	0.19	3.99	1.36	0.22	49.50	4.37	0.00
Regressed-PCA S2	0.73	3.67	8.01	7.51	0.20	4.00	1.44	0.42	49.28	3.99	0.00
IPCA	0.73	5.87	7.55	7.34	0.13	2.65	2.83	0.33	77.17	8.37	0.00
IPCA\Regressed-PCA	1.10	4.78	15.34	14.63	0.24	4.99	1.62	0.28	25.12	4.96	0.00
FF5+MOM	0.39	4.36	2.26	2.12	0.10	2.24	2.54	0.17	83.76	8.58	0.00
KNS	0.82	5.34	8.60	8.32	0.15	3.52	1.30	0.05	59.97	4.90	0.00
Group VI: 72 single sorted portfolios on 36 characteristics (P1&10)											
Regressed-PCA	0.59	3.32	4.33	3.96	0.19	4.28	0.81	0.04	51.08	5.07	0.00
Regressed-PCA S1	0.43	2.17	2.73	2.29	0.21	4.33	1.08	0.22	49.42	5.81	0.00
Regressed-PCA S2	0.71	3.32	6.75	6.24	0.22	4.34	0.89	0.42	49.14	4.12	0.00
IPCA	0.67	4.77	5.76	5.51	0.15	3.06	1.74	0.33	74.73	8.57	0.00
IPCA\Regressed-PCA	1.08	4.26	13.28	12.51	0.27	5.46	1.27	0.28	24.47	6.49	0.00
FF5+MOM	0.41	4.27	2.44	2.26	0.12	2.65	2.48	0.17	81.54	13.91	0.00
KNS	0.82	5.24	7.69	7.43	0.16	3.63	0.73	0.05	63.15	4.57	0.00

[†] $A|a|$: average absolute intercept; $A|t(a)|$: average absolute t-statistic for the intercepts; $Aa^2/V\overline{r}$: average squared intercept over the cross-section variance of $\overline{r}$, average returns of the testing portfolios; $A\lambda^2/V\overline{r}$: average difference between each squared intercept and its squared standard error divided by the variance of $\overline{r}$; $As(a)$: average standard error of the intercepts; $As(e)$: average residual standard deviation; $Sh^2(a)$: maximized squared Sharpe ratio for the intercepts; $Sh^2(f)$: maximized squared Sharpe ratio for the factors; AR^2: average regression R^2 (%); GRS: GRS statistic of Gibbons et al. (1989); $p(GRS)$: p-value of GRS.

Table F.XVI. Additional asset pricing tests: $K = 5^{\dagger}$

Testing portfolios/Factors	$A\|a\|$	$A\|t(a)\|$	$Aa^2/V\overline{r}$	$A\lambda^2/V\overline{r}$	$As(a)$	$As(e)$	$Sh^2(a)$	$Sh^2(f)$	AR^2	GRS	$p(GRS)$
Group I: 25 double sorted portfolios on Size and BM											
Regressed-PCA	0.86	5.03	17.75	17.07	0.17	3.99	0.32	0.04	49.64	6.42	0.00
Regressed-PCA S1	0.58	3.23	8.72	7.97	0.18	3.94	0.45	0.16	51.07	8.05	0.00
Regressed-PCA S2	0.44	2.40	5.20	4.44	0.18	4.05	0.44	0.13	47.97	8.14	0.00
IPCA	0.90	10.41	19.45	19.25	0.09	2.04	1.83	0.10	85.43	34.55	0.00
IPCA\Regressed-PCA	1.12	5.76	29.70	28.74	0.20	4.61	0.67	0.06	36.10	13.16	0.00
FF5	0.39	4.57	3.67	3.47	0.09	2.04	0.69	0.12	86.80	12.88	0.00
KNS	0.97	6.19	22.03	21.48	0.16	3.61	0.30	0.04	58.51	5.99	0.00
Group II: 25 double sorted portfolios on Size and OP											
Regressed-PCA	0.84	5.04	27.72	26.64	0.17	3.89	0.22	0.04	51.55	4.32	0.00
Regressed-PCA S1	0.50	2.86	10.83	9.63	0.18	3.87	0.28	0.16	52.03	4.93	0.00
Regressed-PCA S2	0.37	2.08	5.86	4.63	0.18	3.98	0.28	0.13	49.34	5.07	0.00
IPCA	0.86	11.19	30.45	30.16	0.09	1.90	2.23	0.10	87.16	42.02	0.00
IPCA\Regressed-PCA	1.09	5.59	46.29	44.74	0.20	4.55	0.54	0.06	35.88	10.64	0.00
FF5	0.40	4.96	6.47	6.17	0.09	1.93	0.73	0.12	88.20	13.65	0.00
KNS	0.94	6.16	33.83	32.94	0.15	3.53	0.22	0.04	59.64	4.41	0.00
Group III: 25 double sorted portfolios on Size and INV											
Regressed-PCA	0.88	5.35	21.28	20.54	0.17	3.83	0.36	0.04	51.90	7.24	0.00
Regressed-PCA S1	0.56	3.24	9.25	8.44	0.18	3.82	0.51	0.16	52.11	9.06	0.00
Regressed-PCA S2	0.42	2.42	5.47	4.63	0.18	3.92	0.49	0.13	49.49	9.09	0.00
IPCA	0.90	11.22	23.08	22.89	0.08	1.88	2.10	0.10	87.18	39.53	0.00
IPCA\Regressed-PCA	1.13	5.89	34.84	33.76	0.20	4.53	0.83	0.06	35.53	16.28	0.00
FF5	0.43	5.24	5.05	4.86	0.08	1.84	0.75	0.12	89.09	13.92	0.00
KNS	0.97	6.40	25.75	25.14	0.15	3.50	0.37	0.04	59.02	7.44	0.00

† $A|a|$: average absolute intercept; $A|t(a)|$: average absolute t-statistic for the intercepts; $Aa^2/V\overline{r}$: average squared intercept over the cross-section variance of $\overline{r}$, average returns of the testing portfolios; $A\lambda^2/V\overline{r}$: average difference between each squared intercept and its squared standard error divided by the variance of $\overline{r}$; $As(a)$: average standard error of the intercepts; $As(e)$: average residual standard deviation; $Sh^2(a)$: maximized squared Sharpe ratio for the intercepts; $Sh^2(f)$: maximized squared Sharpe ratio for the factors; AR^2: average regression R^2 (%); GRS: GRS statistic of Gibbons et al. (1989); $p(GRS)$: p-value of GRS.

Table F.XVII. Additional asset pricing tests: $K = 5$ (continued)[†]

Testing portfolios/Factors	$A\|a\|$	$A\|t(a)\|$	$Aa^2/V\overline{r}$	$A\lambda^2/V\overline{r}$	$As(a)$	$As(e)$	$Sh^2(a)$	$Sh^2(f)$	AR^2	GRS	$p(GRS)$
Group IV: 25 double sorted portfolios on Size and MOM											
Regressed-PCA	0.82	4.73	7.27	6.97	0.18	4.16	0.41	0.04	50.16	8.13	0.00
Regressed-PCA S1	0.64	3.53	4.17	3.87	0.18	3.94	0.58	0.16	54.13	10.40	0.00
Regressed-PCA S2	0.52	2.78	3.01	2.71	0.18	4.04	0.67	0.13	51.89	12.34	0.00
IPCA	0.94	11.45	9.20	9.12	0.09	2.03	2.29	0.10	87.24	43.22	0.00
IPCA\Regressed-PCA	1.01	5.24	11.77	11.36	0.21	4.75	0.81	0.06	37.26	15.76	0.00
FF5	0.44	4.47	2.36	2.25	0.11	2.40	0.47	0.12	83.53	8.68	0.00
KNS	0.97	6.27	8.79	8.57	0.16	3.57	0.26	0.04	62.15	5.15	0.00
Group V: 25 double sorted portfolios on Size and Beta											
Regressed-PCA	0.87	5.31	41.95	40.35	0.17	3.92	0.23	0.04	49.54	4.51	0.00
Regressed-PCA S1	0.53	3.12	18.66	16.91	0.18	3.91	0.28	0.16	48.81	4.97	0.00
Regressed-PCA S2	0.40	2.31	10.28	8.48	0.18	4.01	0.27	0.13	46.17	5.04	0.00
IPCA	0.90	11.17	47.69	47.25	0.09	1.97	1.92	0.10	85.50	36.20	0.00
IPCA\Regressed-PCA	1.13	6.12	70.09	67.78	0.20	4.55	0.63	0.06	36.26	12.40	0.00
FF5	0.41	4.70	9.73	9.24	0.09	2.07	0.37	0.12	85.48	6.94	0.00
KNS	0.96	6.27	49.66	48.36	0.16	3.56	0.19	0.04	56.91	3.72	0.00
Group VI: 25 double sorted portfolios on Size and AC											
Regressed-PCA	0.83	4.90	36.79	35.26	0.17	3.92	0.30	0.04	53.86	5.93	0.00
Regressed-PCA S1	0.42	2.40	10.30	8.62	0.18	3.89	0.28	0.16	54.78	5.00	0.00
Regressed-PCA S2	0.28	1.61	4.99	3.30	0.18	3.96	0.28	0.13	52.94	5.14	0.00
IPCA	0.81	9.39	34.66	34.25	0.09	1.95	0.96	0.10	87.20	18.04	0.00
IPCA\Regressed-PCA	1.05	5.04	57.71	55.31	0.21	4.80	0.35	0.06	33.49	6.78	0.00
FF5	0.46	5.09	11.74	11.29	0.09	2.04	0.83	0.12	87.79	15.34	0.00
KNS	0.94	6.10	46.85	45.60	0.16	3.56	0.31	0.04	61.58	6.30	0.00

[†] $A|a|$: average absolute intercept; $A|t(a)|$: average absolute t-statistic for the intercepts; $Aa^2/V\overline{r}$: average squared intercept over the cross-section variance of $\overline{r}$, average returns of the testing portfolios; $A\lambda^2/V\overline{r}$: average difference between each squared intercept and its squared standard error divided by the variance of $\overline{r}$; $As(a)$: average standard error of the intercepts; $As(e)$: average residual standard deviation; $Sh^2(a)$: maximized squared Sharpe ratio for the intercepts; $Sh^2(f)$: maximized squared Sharpe ratio for the factors; AR^2: average regression R^2 (%); GRS: GRS statistic of Gibbons et al. (1989); $p(GRS)$: p-value of GRS.

Table F.XVIII. Additional asset pricing tests: $K = 5$ (continued)[†]

Testing portfolios/Factors	$A\|a\|$	$A\|t(a)\|$	$Aa^2/V\overline{r}$	$A\lambda^2/V\overline{r}$	$As(a)$	$As(e)$	$Sh^2(a)$	$Sh^2(f)$	AR^2	GRS	$p(GRS)$
Group VII: 35 double sorted portfolios on Size and NI											
Regressed-PCA	0.84	4.85	16.08	15.43	0.18	4.02	0.40	0.04	49.84	5.63	0.00
Regressed-PCA S1	0.54	2.94	7.14	6.41	0.18	4.00	0.56	0.16	50.20	7.07	0.00
Regressed-PCA S2	0.41	2.18	4.11	3.37	0.19	4.10	0.53	0.13	47.68	6.84	0.00
IPCA	0.87	9.73	17.60	17.39	0.10	2.15	2.69	0.10	83.75	35.50	0.00
IPCA\Regressed-PCA	1.09	5.43	26.60	25.67	0.21	4.71	0.95	0.06	34.16	13.02	0.00
FF5	0.42	4.44	4.06	3.84	0.10	2.21	0.83	0.12	84.30	10.85	0.00
KNS	0.94	5.86	19.28	18.73	0.16	3.68	0.39	0.04	57.44	5.45	0.00
Group VIII: 25 double sorted portfolios on Size and Var											
Regressed-PCA	0.89	5.34	8.00	7.71	0.18	4.04	0.45	0.04	48.41	8.98	0.00
Regressed-PCA S1	0.70	3.99	5.19	4.88	0.18	3.92	0.70	0.16	49.80	12.47	0.00
Regressed-PCA S2	0.58	3.24	3.90	3.59	0.18	4.04	0.72	0.13	46.55	13.29	0.00
IPCA	0.95	12.60	9.63	9.57	0.08	1.86	2.21	0.10	87.24	41.58	0.00
IPCA\Regressed-PCA	1.11	6.32	12.81	12.40	0.20	4.57	1.15	0.06	39.10	22.45	0.00
FF5	0.45	5.46	2.07	1.99	0.09	1.96	0.71	0.12	87.38	13.12	0.00
KNS	0.99	6.48	9.51	9.30	0.15	3.53	0.43	0.04	57.72	8.57	0.00

[†] $A|a|$: average absolute intercept; $A|t(a)|$: average absolute t-statistic for the intercepts; $Aa^2/V\overline{r}$: average squared intercept over the cross-section variance of $\overline{r}$, average returns of the testing portfolios; $A\lambda^2/V\overline{r}$: average difference between each squared intercept and its squared standard error divided by the variance of $\overline{r}$; $As(a)$: average standard error of the intercepts; $As(e)$: average residual standard deviation; $Sh^2(a)$: maximized squared Sharpe ratio for the intercepts; $Sh^2(f)$: maximized squared Sharpe ratio for the factors; AR^2: average regression R^2 (%); GRS: GRS statistic of Gibbons et al. (1989); $p(GRS)$: p-value of GRS.

Table F.XIX. Additional asset pricing tests: $K = 5$ (continued)[†]

Testing portfolios/Factors	$A\|a\|$	$A\|t(a)\|$	$Aa^2/V\overline{r}$	$A\lambda^2/V\overline{r}$	$As(a)$	$As(e)$	$Sh^2(a)$	$Sh^2(f)$	AR^2	GRS	$p(GRS)$
Group IX: Regressed-PCA S1's 36 managed portfolios											
Regressed-PCA	0.54	5.20	0.84	0.82	0.11	2.44	3.17	0.04	18.11	43.05	0.00
Regressed-PCA S1	0.52	5.92	0.73	0.72	0.09	1.88	3.13	0.16	38.14	38.17	0.00
Regressed-PCA S2	0.53	5.80	0.71	0.69	0.09	2.07	3.18	0.13	33.15	39.73	0.00
IPCA	0.53	5.39	0.80	0.78	0.11	2.33	4.75	0.10	27.85	60.84	0.00
IPCA\Regressed-PCA	0.59	5.38	0.99	0.96	0.12	2.61	3.80	0.06	16.89	50.50	0.00
FF5	0.56	4.93	0.99	0.96	0.12	2.59	3.18	0.12	16.02	40.13	0.00
KNS	0.57	5.37	1.00	0.98	0.11	2.56	3.25	0.04	16.28	44.31	0.00
Group X: Regressed-PCA S2's 36 managed portfolios											
Regressed-PCA	0.63	5.42	0.83	0.81	0.12	2.82	2.88	0.04	20.45	39.12	0.00
Regressed-PCA S1	0.60	5.84	0.72	0.71	0.10	2.21	2.89	0.16	42.04	35.15	0.00
Regressed-PCA S2	0.62	7.27	0.73	0.72	0.08	1.86	1.46	0.13	50.54	18.20	0.00
IPCA	0.64	5.88	0.82	0.80	0.12	2.56	4.59	0.10	32.44	58.83	0.00
IPCA\Regressed-PCA	0.69	5.53	1.00	0.97	0.13	2.96	3.68	0.06	17.67	48.85	0.00
FF5	0.62	4.76	0.96	0.94	0.13	2.90	2.88	0.12	21.41	36.42	0.00
KNS	0.65	5.53	1.00	0.98	0.13	2.86	2.97	0.04	20.33	40.52	0.00

[†] $A|a|$: average absolute intercept; $A|t(a)|$: average absolute t-statistic for the intercepts; $Aa^2/V\overline{r}$: average squared intercept over the cross-section variance of $\overline{r}$, average returns of the testing portfolios; $A\lambda^2/V\overline{r}$: average difference between each squared intercept and its squared standard error divided by the variance of $\overline{r}$; $As(a)$: average standard error of the intercepts; $As(e)$: average residual standard deviation; $Sh^2(a)$: maximized squared Sharpe ratio for the intercepts; $Sh^2(f)$: maximized squared Sharpe ratio for the factors; AR^2: average regression R^2 (%); GRS: GRS statistic of Gibbons et al. (1989); $p(GRS)$: p-value of GRS.

Table F.XX. Additional asset pricing tests: $K = 6^{\dagger}$

Testing portfolios/Factors	$A\|a\|$	$A\|t(a)\|$	$Aa^2/V\overline{r}$	$A\lambda^2/V\overline{r}$	$As(a)$	$As(e)$	$Sh^2(a)$	$Sh^2(f)$	AR^2	GRS	$p(GRS)$
Group I: 25 double sorted portfolios on Size and BM											
Regressed-PCA	0.76	4.93	13.84	13.28	0.16	3.56	0.32	0.04	58.73	6.37	0.00
Regressed-PCA S1	0.57	3.10	8.41	7.63	0.19	3.92	0.46	0.22	51.46	7.87	0.00
Regressed-PCA S2	1.04	5.33	27.04	26.18	0.20	3.83	0.52	0.42	53.23	7.64	0.00
IPCA	0.89	9.35	18.91	18.66	0.10	2.04	1.83	0.33	85.34	28.54	0.00
IPCA\Regressed-PCA	1.31	5.64	39.81	38.48	0.24	4.92	0.66	0.28	27.28	10.60	0.00
FF5+MOM	0.36	4.22	3.06	2.86	0.09	2.00	0.65	0.17	87.25	11.61	0.00
KNS	0.87	5.75	17.68	17.16	0.15	3.47	0.30	0.05	61.83	5.92	0.00
Group II: 25 double sorted portfolios on Size and OP											
Regressed-PCA	0.72	4.88	21.02	20.17	0.15	3.43	0.21	0.04	61.43	4.21	0.00
Regressed-PCA S1	0.47	2.61	10.10	8.84	0.18	3.86	0.24	0.22	52.34	4.10	0.00
Regressed-PCA S2	0.98	5.12	38.54	37.15	0.19	3.76	0.39	0.42	54.34	5.69	0.00
IPCA	0.85	9.90	28.88	28.52	0.09	1.91	1.87	0.33	87.08	29.12	0.00
IPCA\Regressed-PCA	1.27	5.49	61.94	59.78	0.24	4.88	0.42	0.28	26.23	6.83	0.00
FF5+MOM	0.35	4.38	5.13	4.82	0.09	1.90	0.72	0.17	88.59	12.71	0.00
KNS	0.84	5.69	27.20	26.37	0.15	3.39	0.21	0.05	62.68	4.17	0.00
Group III: 25 double sorted portfolios on Size and INV											
Regressed-PCA	0.76	5.24	16.41	15.84	0.15	3.36	0.36	0.04	61.87	7.15	0.00
Regressed-PCA S1	0.55	3.11	9.11	8.26	0.18	3.80	0.49	0.22	52.50	8.33	0.00
Regressed-PCA S2	1.03	5.50	29.94	29.00	0.19	3.70	0.61	0.42	54.61	8.92	0.00
IPCA	0.91	10.23	23.31	23.08	0.09	1.88	1.85	0.33	87.11	28.89	0.00
IPCA\Regressed-PCA	1.34	5.82	48.05	46.54	0.24	4.86	0.73	0.28	25.88	11.76	0.00
FF5+MOM	0.37	4.66	3.83	3.64	0.08	1.80	0.72	0.17	89.56	12.86	0.00
KNS	0.87	5.95	20.66	20.10	0.15	3.35	0.39	0.05	62.54	7.66	0.00

† $A|a|$: average absolute intercept; $A|t(a)|$: average absolute t-statistic for the intercepts; $Aa^2/V\overline{r}$: average squared intercept over the cross-section variance of $\overline{r}$, average returns of the testing portfolios; $A\lambda^2/V\overline{r}$: average difference between each squared intercept and its squared standard error divided by the variance of $\overline{r}$; $As(a)$: average standard error of the intercepts; $As(e)$: average residual standard deviation; $Sh^2(a)$: maximized squared Sharpe ratio for the intercepts; $Sh^2(f)$: maximized squared Sharpe ratio for the factors; AR^2: average regression R^2 (%); GRS: GRS statistic of Gibbons et al. (1989); $p(GRS)$: p-value of GRS.

Table F.XXI. Additional asset pricing tests: $K = 6$ (continued)[†]

Testing portfolios/Factors	$A\|a\|$	$A\|t(a)\|$	$Aa^2/V\overline{r}$	$A\lambda^2/V\overline{r}$	$As(a)$	$As(e)$	$Sh^2(a)$	$Sh^2(f)$	AR^2	GRS	$p(GRS)$
Group IV: 25 double sorted portfolios on Size and MOM											
Regressed-PCA	0.73	4.72	5.84	5.59	0.16	3.73	0.40	0.04	59.23	8.00	0.00
Regressed-PCA S1	0.60	3.23	3.64	3.33	0.19	3.93	0.52	0.22	54.48	8.87	0.00
Regressed-PCA S2	0.91	4.80	9.25	8.90	0.20	3.82	0.62	0.42	56.68	9.03	0.00
IPCA	0.80	9.14	7.22	7.12	0.10	2.02	1.82	0.33	87.32	28.34	0.00
IPCA\Regressed-PCA	1.22	5.21	14.36	13.79	0.25	5.07	0.50	0.28	28.70	8.04	0.00
FF5+MOM	0.42	4.85	1.74	1.67	0.09	1.92	0.42	0.17	89.28	7.42	0.00
KNS	0.88	5.84	7.24	7.03	0.15	3.45	0.25	0.05	64.65	4.90	0.00
Group V: 25 double sorted portfolios on Size and Beta											
Regressed-PCA	0.75	5.15	32.54	31.29	0.15	3.46	0.22	0.04	58.93	4.41	0.00
Regressed-PCA S1	0.51	2.91	16.81	14.97	0.18	3.89	0.23	0.22	49.29	3.94	0.00
Regressed-PCA S2	1.02	5.47	59.59	57.55	0.19	3.79	0.37	0.42	51.55	5.36	0.00
IPCA	0.89	9.83	43.36	42.82	0.10	1.97	1.60	0.33	85.49	24.98	0.00
IPCA\Regressed-PCA	1.31	5.79	92.63	89.41	0.24	4.88	0.41	0.28	27.17	6.69	0.00
FF5+MOM	0.36	4.10	7.36	6.87	0.09	2.03	0.34	0.17	85.92	6.03	0.00
KNS	0.86	5.81	39.84	38.62	0.15	3.42	0.17	0.05	60.06	3.46	0.00
Group VI: 25 double sorted portfolios on Size and AC											
Regressed-PCA	0.71	4.77	26.92	25.77	0.15	3.40	0.29	0.04	64.24	5.80	0.00
Regressed-PCA S1	0.43	2.40	10.71	8.96	0.18	3.87	0.25	0.22	55.22	4.18	0.00
Regressed-PCA S2	0.90	4.69	43.07	41.14	0.19	3.77	0.36	0.42	56.97	5.18	0.00
IPCA	0.88	9.35	41.19	40.69	0.10	1.95	1.05	0.33	87.13	16.36	0.00
IPCA\Regressed-PCA	1.33	5.38	93.52	90.18	0.25	5.15	0.34	0.28	23.54	5.48	0.00
FF5+MOM	0.40	4.43	8.91	8.45	0.09	2.00	0.84	0.17	88.20	14.93	0.00
KNS	0.86	5.67	39.04	37.84	0.15	3.47	0.30	0.05	63.58	6.00	0.00

[†] $A|a|$: average absolute intercept; $A|t(a)|$: average absolute t-statistic for the intercepts; $Aa^2/V\overline{r}$: average squared intercept over the cross-section variance of $\overline{r}$, average returns of the testing portfolios; $A\lambda^2/V\overline{r}$: average difference between each squared intercept and its squared standard error divided by the variance of $\overline{r}$; $As(a)$: average standard error of the intercepts; $As(e)$: average residual standard deviation; $Sh^2(a)$: maximized squared Sharpe ratio for the intercepts; $Sh^2(f)$: maximized squared Sharpe ratio for the factors; AR^2: average regression R^2 (%); GRS: GRS statistic of Gibbons et al. (1989); $p(GRS)$: p-value of GRS.

Table F.XXII. Additional asset pricing tests: $K = 6$ (continued)[†]

Testing portfolios/Factors	$A\|a\|$	$A\|t(a)\|$	$Aa^2/V\overline{r}$	$A\lambda^2/V\overline{r}$	$As(a)$	$As(e)$	$Sh^2(a)$	$Sh^2(f)$	AR^2	GRS	$p(GRS)$
Group VII: 35 double sorted portfolios on Size and NI											
Regressed-PCA	0.73	4.73	12.35	11.84	0.16	3.57	0.40	0.04	59.32	5.56	0.00
Regressed-PCA S1	0.53	2.80	6.71	5.96	0.19	3.99	0.53	0.22	50.60	6.36	0.00
Regressed-PCA S2	0.97	4.92	22.54	21.70	0.20	3.89	0.69	0.42	52.43	7.01	0.00
IPCA	0.87	8.83	17.37	17.11	0.11	2.15	2.35	0.33	83.69	25.70	0.00
IPCA\Regressed-PCA	1.30	5.42	36.67	35.37	0.24	5.03	0.76	0.28	25.00	8.65	0.00
FF5+MOM	0.37	3.90	3.13	2.90	0.10	2.17	0.80	0.17	84.75	9.96	0.00
KNS	0.84	5.43	15.68	15.16	0.16	3.55	0.38	0.05	60.17	5.33	0.00
Group VIII: 25 double sorted portfolios on Size and Var											
Regressed-PCA	0.80	5.31	6.68	6.45	0.16	3.60	0.45	0.04	57.27	8.99	0.00
Regressed-PCA S1	0.64	3.59	4.38	4.06	0.19	3.90	0.64	0.22	50.26	10.91	0.00
Regressed-PCA S2	1.01	5.41	11.24	10.89	0.20	3.82	0.73	0.42	51.84	10.58	0.00
IPCA	0.87	10.34	7.72	7.64	0.09	1.87	1.74	0.33	87.10	27.04	0.00
IPCA\Regressed-PCA	1.26	5.84	15.16	14.59	0.24	4.90	0.79	0.28	30.37	12.67	0.00
FF5+MOM	0.44	5.28	1.97	1.89	0.09	1.96	0.68	0.17	87.45	12.00	0.00
KNS	0.90	6.07	7.90	7.69	0.15	3.41	0.42	0.05	60.78	8.31	0.00

[†] $A|a|$: average absolute intercept; $A|t(a)|$: average absolute t-statistic for the intercepts; $Aa^2/V\overline{r}$: average squared intercept over the cross-section variance of $\overline{r}$, average returns of the testing portfolios; $A\lambda^2/V\overline{r}$: average difference between each squared intercept and its squared standard error divided by the variance of $\overline{r}$; $As(a)$: average standard error of the intercepts; $As(e)$: average residual standard deviation; $Sh^2(a)$: maximized squared Sharpe ratio for the intercepts; $Sh^2(f)$: maximized squared Sharpe ratio for the factors; AR^2: average regression R^2 (%); GRS: GRS statistic of Gibbons et al. (1989); $p(GRS)$: p-value of GRS.

Table F.XXIII. Additional asset pricing tests: $K = 6$ (continued)[†]

Testing portfolios/Factors	$A\|a\|$	$A\|t(a)\|$	$Aa^2/V\overline{r}$	$A\lambda^2/V\overline{r}$	$As(a)$	$As(e)$	$Sh^2(a)$	$Sh^2(f)$	AR^2	GRS	$p(GRS)$
Group IX: Regressed-PCA S1's 36 managed portfolios											
Regressed-PCA	0.54	5.27	0.84	0.82	0.11	2.41	3.19	0.04	19.82	43.08	0.00
Regressed-PCA S1	0.51	6.27	0.69	0.68	0.08	1.77	5.39	0.22	41.99	62.20	0.00
Regressed-PCA S2	0.47	4.85	0.56	0.54	0.10	2.00	2.89	0.42	35.74	28.58	0.00
IPCA	0.50	4.86	0.60	0.57	0.11	2.26	3.85	0.33	31.41	40.80	0.00
IPCA\Regressed-PCA	0.57	5.05	0.77	0.74	0.12	2.56	3.50	0.28	19.43	38.43	0.00
FF5+MOM	0.56	4.92	1.09	1.07	0.12	2.51	3.13	0.17	19.79	37.82	0.00
KNS	0.57	5.36	1.03	1.00	0.11	2.53	3.25	0.05	18.03	43.79	0.00
Group X: Regressed-PCA S2's 36 managed portfolios											
Regressed-PCA	0.63	5.46	0.83	0.81	0.12	2.77	2.88	0.04	22.56	38.82	0.00
Regressed-PCA S1	0.62	6.29	0.72	0.70	0.10	2.09	2.85	0.22	46.56	32.87	0.00
Regressed-PCA S2	0.54	6.03	0.55	0.54	0.09	1.72	3.40	0.42	54.10	33.71	0.00
IPCA	0.64	5.61	0.69	0.67	0.12	2.43	3.63	0.33	37.52	38.48	0.00
IPCA\Regressed-PCA	0.67	5.35	0.79	0.76	0.14	2.89	3.26	0.28	20.86	35.72	0.00
FF5+MOM	0.63	4.97	1.07	1.05	0.13	2.78	2.85	0.17	25.91	34.35	0.00
KNS	0.65	5.51	1.02	1.00	0.12	2.84	2.96	0.05	21.88	39.89	0.00

[†] $A|a|$: average absolute intercept; $A|t(a)|$: average absolute t-statistic for the intercepts; $Aa^2/V\overline{r}$: average squared intercept over the cross-section variance of $\overline{r}$, average returns of the testing portfolios; $A\lambda^2/V\overline{r}$: average difference between each squared intercept and its squared standard error divided by the variance of $\overline{r}$; $As(a)$: average standard error of the intercepts; $As(e)$: average residual standard deviation; $Sh^2(a)$: maximized squared Sharpe ratio for the intercepts; $Sh^2(f)$: maximized squared Sharpe ratio for the factors; AR^2: average regression R^2 (%); GRS: GRS statistic of Gibbons et al. (1989); $p(GRS)$: p-value of GRS.

References

Ahn, S. and A. Horenstein (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 K. Li (2012): "Statistical analysis of factor models of high dimension," *Annals of Statistics*, 40, 436–465.

Bai, J. and S. Ng (2002): "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70, 191–221.

Belloni, A., V. Chernozhukov, D. Chetverikov, and K. Kato (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.

Chen, Q. and Z. Fang (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.

Connor, G., M. Hagmann, and O. Linton (2012): "Efficient semiparametric estimation of the FamaâĂŞFrench model and extensions," *Econometrica*, 80, 713–754.

Fama, E. F. (1976): *Foundations of Finance*, New York: Basic Books.

Fama, E. F. and K. R. French (2015): "A five-factor asset pricing model," *Journal of financial economics*, 116, 1–22.

Fan, J., Y. Liao, and W. Wang (2016a): "Projected principal component analysis in factor models," *The Annals of Statistics*, 44, 219–254.

——— (2016b): "Supplement to âĂIJProjected principal component analysisin factor models"," *DOI:10.1214/15-AOS1364SUPP*.

Gibbons, M. R., S. A. Ross, and J. Shanken (1989): "A test of the efficiency of a given portfolio," *Econometrica*, 57, 1121–1152.

Huang, J. (1998): "Projection Estimation in Multiple Regression with Application to Functional ANOVA Models," *The Annals of Statistics*, 26, 242–272.

Kelly, B. T., S. Pruitt, and Y. Su (2019): "Characteristics are covariances: A unified model of risk and return," *Journal of Financial Economics*, 134, 501–524.

Kim, S., R. A. Korajczyk, and A. Neuhierl (2020): "Arbitrage portfolios," *Review of Financial Studies*, Forthcoming.

Li, J. and Z. Liao (2020): "Uniform Nonparametric Inference for Time Series," *Journal of Econometrics*, 219, 38–51.

Lorentz, G. G. (1986): *Approximation of Functions*, Chelsea Publishing.

Newey, W. (1997): "Convergence Rates and Asymptotic Normality for Series Estimators," *Journal of Econometrics*, 79, 147–168.

Pollard, D. (2002): *A User's Guide to Measure Theoretic Probability*, vol. 8 of *Cambridge Series in Statistical and Probabilistic Mathematics*, Cambridge University Press.

Stock, J. H. and M. W. Watson (2002): "Forecasting using principal components from a large number of predictors," *Journal of the American Statistical Association*, 97, 1167–1179.

van der Vaart, A. W. and J. A. Wellner (1996): *Weak Convergence and Empirical Processes: With Applications to Statistics*, Springer Series in Statistics, New York: Springer, 1st edition ed.