Arbitrage Portfolios in Large Panels*

Soohun Kim† Robert A. Korajczyk‡ Andreas Neuhierl§

July 31, 2018

Abstract

We propose new methodology to estimate arbitrage portfolios by utilizing information contained in firm characteristics for both abnormal returns and factor loadings. The methodology gives maximal weight to risk-based interpretations of characteristic predictive power before any attribution to abnormal returns. We apply the methodology in simulated factor economies and on a large panel of U.S. stock returns from 1965-2014. The methodology works well in simulation and in out-of-sample portfolios of U.S. stocks. Empirically, we find the arbitrage portfolio has significant (statistically and economically) alphas relative to several popular asset pricing models and annualized Sharpe ratios ranging from 0.67 to 1.16. Data mining driven alphas imply that performance of the strategy should decline after the discovery of pricing anomalies. However, we find that the abnormal returns on the arbitrage portfolio do not decrease significantly over time.

*PRELIMINARY AND INCOMPLETE. We gratefully acknowledge Unigestion Alternative Risk Premia Research Academy for providing financial support.

†Scheller College of Business, Georgia Institute of Technology, Atlanta, GA 30308. E-mail: soohun.kim@scheller.gatech.edu.

‡Kellogg School of Management, Northwestern University, Evanston, IL 60208. E-mail: r-korajczyk@kellogg.northwestern.edu.

§University of Notre Dame, Notre Dame, IN 46556. E-Mail: aneuhier@nd.edu.

1 Introduction

Over the past 50 years, the literature in empirical asset pricing has produced a large number of variables that seem to explain cross-sectional variation in expected returns. While this quest has possibly led to a number of spurious explanatory variables, the literature aims to answer a critical underlying economic question, namely what are determinants of expected stock returns? Standard economic models prescribe that larger expected returns go hand in hand with higher exposure to the underlying sources of risk. Testing this proposition empirically is a delicate task for a number of reasons. We do not observe the underlying sources of risk directly and risk exposures have to be estimated from noisy data. Observable firm characteristics play a prominent role in explaining cross-sectional return differences, however their interpretation is not undisputed. Characteristics with explanatory power for returns can be interpreted as a proxy for factor loadings, possible mispricing, or the result of data mining.

Researchers often take a standard factor model, such as the Fama and French (1993) as given and then construct zero-investment portfolios which have estimated factor loadings equal to zero but non-zero values of their characteristics to show that the characteristics explain returns beyond that captured by factor loadings (e.g., Daniel and Titman (1997)). A canonical example is that firms with similar loadings on the size factor are often very different in terms of their actual market capitalization and expected returns, which raises question about the interpretation of size as a risk factor. One shortcoming of such an approach is that is not a very precise test of the "betas vs. characteristics" question since it assumes a particular factor model directly at the outset. Analyses like these are both tests of a particular factor model (plus ancillary assumptions) jointly with tests of the "betas vs. characteristics" question.

We propose a new methodology which allows us to jointly estimate factors, factor loadings, and the possible mispricing of assets. The proposed methodology allows us to estimate factor realizations using all the information contained in firm characteristics and is, therefore, giving risk-based models "a fair chance" to successfully explain the cross-section of expected returns, in the spirit of Ferson and Harvey (1997). As a result, we interpret any return predictability beyond that embedded in risk premia, as either "true alpha" or data snooping.[1]

[1]Kozak et al. (2018) argue that the distinction between risk premia and abnormal returns is not totally clear because abnormal returns correlated with risk exposures are the only ones that would

Our approach allows us to make a number of contributions to empirical asset pricing. First, we provide useful guidance in portfolio construction for practitioners who want to eliminate exposure to the common risks and focus on exploiting mispricing, or alphas, of traded securities. Second, we address, in a unified manner, the question of "betas vs. characteristics" in a statistical factor pricing model (a long standing issue since Fama and French (1993) and Daniel and Titman (1997)). The typical use of unconditional, rolling estimates of betas may inflate estimated alpha if the predictive characteristics are correlated with estimation errors in betas. This can happen if these firm-specific characteristics help predict assets' true betas over and above the information included in the unconditional estimates. For example, if a firm's underlying assets have constant betas, leverage in a firm's capital structure implies equity betas are time varying and that time-series changes in equity betas will be related to changes in the firm's leverage. Since changes in firm size, book-to-market equity ratio, and the firm's past price movements are correlated with leverage changes, commonly used characteristics (such as size, market-to-book equity, and momentum) might help us predict conditional betas, over and above the predictive power of unconditional betas. Our approach incorporates the cross-sectional predictive power of asset characteristics for factors betas, as in Ferson and Harvey (1997), Connor and Linton (2007), and Connor et al. (2012) for prespecified factor models and Fan et al. (2016) and Kelly et al. (2018) for statistical factor models.

1.1 Related Literature

The early literature on risk-based determinants of cross-sectional expected returns is closely linked to the Capital Asset Pricing model (CAPM) of Treynor (1962, 1999), Sharpe (1964), Lintner (1965), and Mossin (1966), the Intertemporal CAPM (ICAPM) of Merton (1973), and the Arbitrage Pricing Theory (APT) of Ross (1976). One strand focuses on models in which the risk factors are known ex ante, at least in theory, as in the CAPM and ICAPM. A second strand, as in the APT, focuses on latent factors, estimated through statistical models.

There is a large and growing literature that relates observable firm characteristics to expected returns, over and above those implied by extant asset pricing models. Classical

survive arbitrage activities by investors.

early contributions to this literature were made by Banz (1981) (size effect), Stattman (1980) and Rosenberg et al. (1985) (the book-to-market ratio), and Fama and French (1992) which provides an early synthesis of findings of multiple characteristics. The explanatory power of firm characteristics has led to alternative specifications of asset pricing models (e.g., Fama and French (1993, 1996)) and further testing of characteristics' ability to explain the cross-section of returns beyond that implied by the expanded set of asset pricing models. Daniel and Titman (1997) propose using double sorts on factor betas and characteristics to address the debate about whether factor loadings or characteristics have greater explanatory power for expected returns. The recent meta study by Harvey et al. (2016) provides an extensive overview of many of the variables that the literature has produced and also raises important statistical concerns related to multiple hypothesis testing.

A large portion of the earlier empirical literature works at the portfolio level, i.e. rather than using individual assets to test models, researchers group assets into portfolio and conduct tests on these portfolios. Gagliardini et al. (2016) stress that the "pre-grouping" possibly masks considerable important variation and develop a new methodology to test asset pricing models on individual assets. Kim and Skoulakis (2018a,b) argue in a similar fashion and propose various asset pricing tests using a large cross-sectional individual stock data over a short time horizon. In particular, Kim and Skoulakis (2018b) estimate the rewards of firm characteristics after controlling for the risk of a given asset pricing model. While their interest is in the evaluation of a specific asset pricing model, we provide a methodology to form arbitrage portfolios in a general, latent factor structure of returns.

Fan et al. (2016) make a methodological contribution by bridging the gap between purely statistical factor models and characteristic based models. We use their contribution as the basis for our analysis and extend the method to explicitly estimate and test for possible characteristic-related mispricing. Kelly et al. (2017, 2018) develop and apply similar methodology, instrumented principal component analysis (IPCA). Our work is particularly closely related to Kelly et al. (2018) who also investigate the question of whether characteristics contain information on risk loadings, mispricing, or both. They conclude that firm-level characteristics ability to predict the cross-section of returns is due to their ability to predict the cross-section of risk loadings rather than mispricing. We use the characteristics at the beginning of each sub-interval of time to estimate the cross-sectional relation between beta, alphas and characteristics. We allow the cross-sectional relation to vary across sub-intervals and we use updated characteristics to form our arbitrage portfolio. Kelly et al. (2018) allow the characteristics to change month-by-month but hold the cross-sectional relation constant. Thus the dynamics in our procedure are coming from changes in the cross-sectional relation between alphas, betas, and characteristics, with changes in characteristics across sub-intervals of time. The dynamics in Kelly et al. (2018) come from the time series of characteristics, holding the cross-sectional relation constant. Our procedure will tend to perform better in situations where characteristics are relatively stable (e.g., market capitalization and book-to-market equity) but whose relation to risk and alpha changes over time, as one might expect if anomalies are arbitraged away after discovery. Their procedure will tend to perform better in situations where characteristics have important short-term dynamics (e.g., short-term reversal and the January seasonal) but whose relation to risk and alpha are stable over time. When we apply their procedure to form out-of-sample arbitrage portfolios in simulated economies, we find the abnormal returns on the arbitrage portfolio to be noisier than those from our procedure.

The rest of the paper is organized as follows. In Section 2, we describe our large cross-sectional economy and propose an estimator of an arbitrage portfolio. In Section 3, we simulate an economy in which asset risks match those in the U.S. equity markets and examine the performance of our estimator of arbitrage portfolio. The estimator performs well with empirically relevant sample sizes. In Section 4, we apply our methodology to a large cross section of individual stocks in U.S. equity market and provide evidence that our arbitrage portfolio indeed generates strong profitability after controlling for commonly used risk factors. We also test for time trends in the abnormal returns on the arbitrage portfolio. One would expect that data mining would lead to returns that dissipate over time. While we find a slight negative time trend, it is not economically significant.

2 The Model

We assume that there exists a large number of securities indexed by $i = 1, \cdots, N$, and the return generating processes for those individual securities are is stable for relatively short time periods (e.g., dozens of months) $t = 1, \cdots, T$. We allow the return generating process to change across time periods. The return generating process of each individual security follows a K-factor model in which the factors are unobservable, latent factors. In particular, the excess return of i-th asset at time t is generated by a factor model,

$$R_{i,t} = \alpha_i + \boldsymbol{\beta}_i' \mathbf{f}_t + e_{i,t}, \quad i = 1, \cdots, N \text{ and } t = 1, \cdots, T, \tag{2.1}$$

where $\boldsymbol{\beta}_i = [\beta_{i,1} \ \cdots \ \beta_{i,K}]'$ is the $(K \times 1)$ factor loadings of the i-th asset, $\mathbf{f}_t$ is the $(K \times 1)$ systematic factor realization in period t, and $e_{i,t}$ is the zero-mean idiosyncratic residual return of asset i at time t. Since our objective is to extract possible mispricing from a large cross section of assets and construct an arbitrage portfolio, we explicitly add a mispricing term, α_i, to the return generating process (2.1). Throughout, we use $\mathbf{0}_m$, $\mathbf{1}_m$ and $\mathbf{0}_{m\times l}$ denote the $(m \times 1)$ vectors of zeros and ones and the $(m \times l)$ matrix of zeros, respectively. The return generating process of (2.1) is expressed compactly in matrices:

$$\mathbf{R} = \boldsymbol{\alpha} \mathbf{1}_T' + \mathbf{B}\mathbf{F}' + \mathbf{E}, \tag{2.2}$$

where the (i, t) element of the $(N \times T)$ matrix of $\mathbf{R}$ is $R_{i,t}$, respectively, $\boldsymbol{\alpha}$ is the $(N \times 1)$ vector of $[\alpha_1 \cdots \alpha_N]'$, the i-th row of the $(N \times K)$ matrix of $\mathbf{B}$ is $\boldsymbol{\beta}_i'$, the t-th row of the $(T \times K)$ matrix of $\mathbf{F}$ is $\mathbf{f}_t'$=$[f_{1,t} \ \cdots \ f_{K,t}]$, and the (i, t) element of the $(N \times T)$ matrix of $\mathbf{E}$ is $e_{i,t}$.

Our estimator is an extension of the Projected Principal Components Analysis (PPCA) approach of Fan et al. (2016). While they allow the factor loading matrix, $\mathbf{B}$, to be a non-parametric function of firm characteristics, we allow both the mispricing, $\boldsymbol{\alpha}$, and the systematic risk, $\mathbf{B}$, to be functions of asset-specific characteristics. Let $\mathbf{x}_i = [x_{i,1} \ \cdots \ x_{i,L}]'$ be the $(L \times 1)$ vector of the characteristics associated with stock i. Define the $(N \times L)$ matrix of $\mathbf{X}$, the i-th row of which is $\mathbf{x}_i'$. We assume the following structure for $\boldsymbol{\alpha}$ and $\mathbf{B}$:

$$\boldsymbol{\alpha} = \mathbf{G}_\alpha(\mathbf{X}) + \Gamma_\alpha$$
$$\mathbf{B} = \mathbf{G}_\beta(\mathbf{X}) + \Gamma_\beta,$$

where $\mathbf{G}_\alpha(\mathbf{X}) : \ \mathbb{R}^{N\times L} \to \mathbb{R}^N$, $\mathbf{G}_\beta(\mathbf{X}) : \ \mathbb{R}^{N\times L} \to \mathbb{R}^{N\times K}$, and the $(N \times 1)$ vector, Γ_α, and the $(N \times K)$ matrix, Γ_β, are cross-sectionally orthogonal to the characteristic space of $\mathbf{X}$. We call $\mathbf{G}_\alpha(\mathbf{X})$ the "mispricing function" and $\mathbf{G}_\beta(\mathbf{X})$ the "factor loading function." There are a number of ways in which one could incorporate non-linearity into the mispricing and factor loading functions. One could treat $\mathbf{X}^*$ as a relatively small set of characteristics and allow $\mathbf{G}_\alpha(\mathbf{X}^*)$ and $\mathbf{G}_\beta(\mathbf{X}^*)$ to be explicit non-linear functions of $\mathbf{X}^*$. However, for the ease of exposition, we chose $\mathbf{X}$ to be a large set of characteristics, possibly containing suitable polynomials of the underlying characteristics, $\mathbf{X}^*$. Hence, we treat $\mathbf{G}_\alpha(\mathbf{X})$ and $\mathbf{G}_\beta(\mathbf{X})$ as linear functions of a large set of characteristics $\mathbf{X}$. We then rewrite the return generating process (2.2) as follows:

$$\mathbf{R} = \left(\mathbf{G}_\alpha(\mathbf{X}) + \Gamma_\alpha\right)\mathbf{1}_T' + \left(\mathbf{G}_\beta(\mathbf{X}) + \Gamma_\beta\right)\mathbf{F}' + \mathbf{E}. \tag{2.3}$$

Note that the Arbitrage Pricing Theory (APT, Ross (1976)) implies that $\frac{1}{N}\boldsymbol{\alpha}'\boldsymbol{\alpha} \to 0$. Hence, in an economy governed by the APT, it follows that $\frac{\mathbf{G}_\alpha(\mathbf{X})'\mathbf{G}_\alpha(\mathbf{X})}{N} \to 0$ because $0 \le \frac{\mathbf{G}_\alpha(\mathbf{X})'\mathbf{G}_\alpha(\mathbf{X})}{N} \le \frac{1}{N}\boldsymbol{\alpha}'\boldsymbol{\alpha}$. Allowing for significant mispricing of assets implies the cross-sectional average of the squared mispricing function $\mathbf{G}_\alpha(\mathbf{X})$ may be non-zero:

Assumption 1. *As $N \to \infty$,*

$$\frac{\mathbf{G}_\alpha(\mathbf{X})'\mathbf{G}_\alpha(\mathbf{X})}{N} \to \delta \ge 0.$$

The above assumption specifies that the characteristics in $\mathbf{X}$ may contain information about asset mispricing, $\boldsymbol{\alpha}$. It is beyond the scope of this paper to examine the underlying cause of a non-trivial relation between the characteristics, $\mathbf{X}$, and $\boldsymbol{\alpha}$. The main objective of this paper is providing a method to detect the relation between $\mathbf{X}$ and $\boldsymbol{\alpha}$ while also allowing the characteristics to predict differences in systematic risk across assets. Using the relation between $\mathbf{X}$ and both $\boldsymbol{\alpha}$ and $\mathbf{B}$ possibly allows us to form portfolios that yield abnormal returns (if $\delta>0$) while hedging out the risk associated with the firm characteristics.

The following are standard regularity conditions on the characteristics and residual returns.

Assumption 2. *As $N \to \infty$, it holds that*
(i) $\frac{\mathbf{R}'\mathbf{R}}{N} \xrightarrow{p} \mathbf{V}_R$ *and* $\frac{\mathbf{X}'\mathbf{X}}{N} \to \mathbf{V}_X$*, where* $\mathbf{V}_R$ *and* $\mathbf{V}_X$ *are positive definite matrices,*
(ii) $\frac{\mathbf{G}_\beta(\mathbf{X})'\Gamma_\alpha}{N} \xrightarrow{p} \mathbf{0}_K$, $\frac{\mathbf{G}_\beta(\mathbf{X})'\Gamma_\beta}{N} \xrightarrow{p} \mathbf{0}_{K\times K}$, $\frac{\mathbf{X}'\Gamma_\alpha}{N} \xrightarrow{p} \mathbf{0}_L$, $\frac{\mathbf{X}'\Gamma_\beta}{N} \xrightarrow{p} \mathbf{0}_{L\times K}$, $\frac{\mathbf{G}_\beta(\mathbf{X})'\mathbf{E}}{N} \xrightarrow{p} \mathbf{0}_{K\times T}$ *and* $\frac{\mathbf{X}'\mathbf{E}}{N} \xrightarrow{p} \mathbf{0}_{L\times T}$.

Condition (i) simply states that the cross-section of returns and characteristics are not redundant but well-spread over individual stocks. Condition (ii) imposes the various cross-sectional orthogonality conditions between the mispricing function, mispricing function residuals, factor loading function, factor loading function residuals, and residual returns.

Lastly, we assume mild restrictions to separately identify $\mathbf{G}_\alpha(\mathbf{X})$ and $\mathbf{G}_\beta(\mathbf{X})$. To ease notation, we wefine the $(T \times T)$ matrix $\mathbf{J}_T = \mathbf{I}_T - \frac{1}{T}\mathbf{1}_T\mathbf{1}_T'$ which corresponds to time-series demeaning.

Assumption 3. *As $N \to \infty$, we assume*
(i) $\frac{\mathbf{G}_\beta(\mathbf{X})'\mathbf{G}_\alpha(\mathbf{X})}{N} \to \mathbf{0}_K$,
(ii) $\frac{\mathbf{G}_\beta(\mathbf{X})'\mathbf{G}_\beta(\mathbf{X})}{N} = \mathbf{I}_K$ *and*
(iii) $\mathbf{F}\mathbf{J}_T\mathbf{F}'$ *is a full rank $(K \times K)$ diagonal matrix with distinct diagonal elements.*

Condition (i) restricts the mispricing function of $\mathbf{G}_\alpha(\mathbf{X})$ to be cross-sectionally orthogonal to the factor loading function of $\mathbf{G}_\beta(\mathbf{X})$. This assumption is without loss of generality. If there is any correlation between $\mathbf{G}_\alpha(\mathbf{X})$ and $\mathbf{G}_\beta(\mathbf{X})$, the correlated component can be assigned to the risk-based component reflected in $\mathbf{G}_\beta(\mathbf{X})$ by shifting factors accordingly.[2] Conditions (ii) and (iii), are minor modifications of the commonly assumed identification restrictions. Without this restriction, we cannot identify $\mathbf{G}_\beta(\mathbf{X})$ separately because of the rotational indeterminacy of latent factor models. That is, $\mathbf{G}_\beta(\mathbf{X})\mathbf{F}\mathbf{J}_T = \mathbf{G}_\beta(\mathbf{X})\mathbf{H}^{-1}\mathbf{H}\mathbf{F}\mathbf{J}_T$ for any invertible matrix $\mathbf{H}$.

2.1 Methodology

Our Projected-PCA procedure first projects demeaned returns onto the cross-sectional firm-specific characteristics. The factor loading function is then estimated by applying standard a PCA procedure to the projected returns. Fan et al. (2016) show that the estimated factor loading function converges to the true factor loading function as the cross sectional sample increases, even for small time-series samples. This allows us to implement the procedure using rolling blocks of data to estimate portfolio weights for the next month. It also allows for time variation in factor risk premia and the extent to which any given characteristic can predict abnormal returns. We extend the PPCA

[2]For a similar restriction in literature, see equation (6) of Connor et al. (2012), who assume the cross-sectional orthogonality between alpha and beta for identification.

estimator to not only estimate factors, but also the mispricing function, which is not part of Fan et al. (2016).

We achieve the goal of constructing an arbitrage portfolio in three steps. In the first step, we demean returns and obtain an estimator of $\mathbf{G}_\beta(\mathbf{X})$ from applying the APC to demeaned projected returns. By demeaning the returns, we focus purely on the systematic risk not on the expected returns or realized premiums. In the second step, we estimate $\mathbf{G}_\alpha(\mathbf{X})$ by regressing the average returns on the characteristic space orthogonal to the estimated $\mathbf{G}_\beta(\mathbf{X})$ from the first step. Although the average returns contain both mispricing and risk premium from systematic risks, we extract the information on the mispricing by imposing the orthogonality to the systematic risks. In the third step, we use the estimated $\mathbf{G}_\alpha(\mathbf{X})$ to construct an arbitrage portfolio. Details of the three steps follow.

The first step of our procedure is the estimation of $\mathbf{G}_\beta(\mathbf{X})$. Recall that the observed returns in (2.3) are driven by not only $\mathbf{G}_\beta(\mathbf{X})$ but $\mathbf{G}_\alpha(\mathbf{X})$. We eliminate the effect of $\mathbf{G}_\alpha(\mathbf{X})$, by demeaning the observed returns:

$$\begin{aligned}\mathbf{RJ}_T &= \left(\mathbf{G}_\alpha(\mathbf{X}) + \Gamma_\alpha\right)\mathbf{1}_T'\mathbf{J}_T + \left(\mathbf{G}_\beta(\mathbf{X}) + \Gamma_\beta\right)\mathbf{F}'\mathbf{J}_T + \mathbf{EJ}_T \\ &= \left(\mathbf{G}_\beta(\mathbf{X}) + \Gamma_\beta\right)\mathbf{F}'\mathbf{J}_T + \mathbf{EJ}_T, \end{aligned} \tag{2.4}$$

where the last equality is from the property of $\mathbf{1}_T'\mathbf{J}_T = \mathbf{1}_T'\left(\mathbf{I}_T - \frac{1}{T}\mathbf{1}_T\mathbf{1}_T'\right) = \mathbf{1}_T' - \frac{T}{T}\mathbf{1}_T' = \mathbf{0}_T'$. For further isolation of $\mathbf{G}_\beta(\mathbf{X})$, we project the demeaned returns of (2.4) on the (linear) span of $\mathbf{X}$ by premultiplying the projection matrix $\mathbf{P} = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}$. Then, we get

$$\widehat{\mathbf{R}} \equiv \mathbf{PRJ}_T = \mathbf{PG}_\beta(\mathbf{X})\mathbf{F}'\mathbf{J}_T + \mathbf{P}\Gamma_\beta\mathbf{F}'\mathbf{J}_T + \mathbf{PEJ}_T. \tag{2.5}$$

Note that $\mathbf{PG}_\beta(\mathbf{X}) = \mathbf{G}_\beta(\mathbf{X})$ since $\mathbf{G}_\beta(\mathbf{X})$ is already in the linear span of $\mathbf{X}$. Also, the orthogonality ofΓ_β and $\mathbf{X}$ and the limits in Assumption 2(ii) make $\mathbf{P}\Gamma_\beta$ and $\mathbf{PE}$ negligible for large N. Hence, it holds that $\widehat{\mathbf{R}} = \mathbf{PRJ}_T \approx \mathbf{G}_\beta(\mathbf{X})\mathbf{F}'\mathbf{J}_T$ with large N. Finally, the following theorem shows that we can estimate $\mathbf{G}_\beta(\mathbf{X})$ by applying the standard principal component analysis to $\widehat{\mathbf{R}}$. We define the convergence of large dimensional random matrices as follows: For two $(N \times m)$ matrices $\mathbf{A}$ and $\mathbf{B}$ with a fixed m, we say that as N increases, $\mathbf{A} \xrightarrow{p} \mathbf{B}$ if, as N increases, $\frac{1}{N}(\mathbf{A} - \mathbf{B})'(\mathbf{A} - \mathbf{B}) \xrightarrow{p} \mathbf{0}_{m\times m}$.

Theorem 2.1. *Let* $\widehat{\mathbf{G}}_\beta(\mathbf{X})$ *denote the* $(N \times K)$ *matrix, the k-th column of which is* $\sqrt{N}$ *times the eigenvector of* $\frac{\widehat{\mathbf{R}}\widehat{\mathbf{R}}'}{N}$ *corresponding to the k-th largest eigenvalue of* $\frac{\widehat{\mathbf{R}}\widehat{\mathbf{R}}'}{N}$*, where* $\widehat{\mathbf{R}}$ *is given by (2.5). Under Assumptions 2 and 3, it holds that as* N *increases,* $\widehat{\mathbf{G}}_\beta(\mathbf{X}) \xrightarrow{p} \mathbf{G}_\beta(\mathbf{X})$. *All proofs are in the Appendix.*

Next, we proceed to estimate $\mathbf{G}_\alpha(\mathbf{X})$. Rather than demeaning $\mathbf{R}$, as we did for the estimation of $\mathbf{G}_\beta(\mathbf{X})$, we take the (weighted) mean of $\mathbf{R}$ by postmultiplying by the $(T \times 1)$ vector $\frac{1}{T}\mathbf{1}_T$.[3] From (2.3), the $(N \times 1)$ vector of average returns, $\frac{1}{T}\mathbf{R}\mathbf{1}_T = \overline{\mathbf{R}}$, has the following expression:

$$\begin{aligned}\overline{\mathbf{R}} &= (\mathbf{G}_\alpha(\mathbf{X}) + \Gamma_\alpha)\frac{1}{T}\mathbf{1}_T'\mathbf{1}_T + (\mathbf{G}_\beta(\mathbf{X}) + \Gamma_\beta)\frac{1}{T}\mathbf{F}'\mathbf{1}_T + \frac{1}{T}\mathbf{E}'\mathbf{1}_T \\ &= \mathbf{G}_\alpha(\mathbf{X}) + \Gamma_\alpha + (\mathbf{G}_\beta(\mathbf{X}) + \Gamma_\beta)\overline{\mathbf{F}} + \overline{\mathbf{E}}. \end{aligned} \tag{2.6}$$

Our objective is extracting $\mathbf{G}_\alpha(\mathbf{X})$ from $\overline{\mathbf{R}}$. Note that simply projecting $\overline{\mathbf{R}}$ to the linear span of $\mathbf{X}$ does not work because $\overline{\mathbf{R}}$ contains not only $\mathbf{G}_\alpha(\mathbf{X})$ but $\overline{\mathbf{F}}$. That is, projecting $\overline{\mathbf{R}}$ to the linear span of $\mathbf{X}$ confounds the cross-sectional predictability of returns due to mispricing with the predictability of returns due to factor risk premia. Hence, we project $\overline{\mathbf{R}}$ to the linear space of $\mathbf{X}$, orthogonal to $\widehat{\mathbf{G}}_\beta(\mathbf{X})$. The following theorem establishes that we can recover $\mathbf{G}_\alpha(\mathbf{X})$ with this approach.

Theorem 2.2. *Define* $\widehat{\mathbf{G}}_\alpha(\mathbf{X}) = \mathbf{X}\widehat{\boldsymbol{\theta}}$, *where the* $(L \times 1)$ *vector of* $\widehat{\boldsymbol{\theta}}$ *is given by the solution of the following constrained optimization problem:*

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \left(\overline{\mathbf{R}} - \mathbf{X}\boldsymbol{\theta}\right)'\left(\overline{\mathbf{R}} - \mathbf{X}\boldsymbol{\theta}\right) \quad \textit{subject to} \quad \widehat{\mathbf{G}}_\beta(\mathbf{X})'\mathbf{X}\boldsymbol{\theta} = \mathbf{0}_K,$$

where $\widehat{\mathbf{G}}_\beta(\mathbf{X})$ *is given by Theorem 2.1. Then, under Assumptions 2 and 3, as* N *increases, it holds that* $\widehat{\mathbf{G}}_\alpha(\mathbf{X}) \xrightarrow{p} \mathbf{G}_\alpha(\mathbf{X})$.

The problem in the above theorem is a conventional ordinary least square problem with linear equality constraints and the closed form solution is easily obtained.

Finally, we construct an arbitrage portfolio which optimally exploits the mispricing information in characteristics. Consider first an infeasible portfolio $\mathbf{w} = \frac{1}{N}\mathbf{G}_\alpha(\mathbf{X})$.

[3] We can weight the time series mean by post-multiplying any positive $(T \times 1)$ vector, $\mathbf{i}$, such that $\mathbf{1}_T'\mathbf{i} = 1$.

Then, from (2.3), we find that the return of this infeasible portfolio is given by

$$\mathbf{wR} = \left(\frac{1}{N}\mathbf{G}_{\alpha}\left(\mathbf{X}\right)'\mathbf{G}_{\alpha}\left(\mathbf{X}\right) + \frac{1}{N}\mathbf{G}_{\alpha}\left(\mathbf{X}\right)'\Gamma_{\alpha}\right)\mathbf{1}_T' + \left(\frac{1}{N}\mathbf{G}_{\alpha}\left(\mathbf{X}\right)'\mathbf{G}_{\beta}\left(\mathbf{X}\right) + \frac{1}{N}\mathbf{G}_{\alpha}\left(\mathbf{X}\right)'\Gamma_{\beta}\right)\mathbf{F}' + \frac{1}{N}\mathbf{G}_{\alpha}\left(\mathbf{X}\right)'\mathbf{E}.$$

From Assumptions 1-3, it is easy to verify that as N increases, only $\frac{1}{N}\mathbf{G}_{\alpha}\left(\mathbf{X}\right)'\mathbf{G}_{\alpha}\left(\mathbf{X}\right)$ converges to $\delta \geq 0$ and all other elements converge to zero such that $\mathbf{wR} \xrightarrow{p} \delta\mathbf{1}_T'$. The following theorem states that the feasible portfolio, $\widehat{\mathbf{w}} = \frac{1}{N}\widehat{\mathbf{G}}_{\alpha}\left(\mathbf{X}\right)$, achieves the same asymptotic property.

Theorem 2.3. *Define* $\widehat{\mathbf{w}} = \frac{1}{N}\widehat{\mathbf{G}}_{\alpha}\left(\mathbf{X}\right)$, *where the* $(N \times 1)$ *vector of* $\widehat{\mathbf{G}}_{\alpha}\left(\mathbf{X}\right)$ *is given in Theorem 2.2. Then, under Assumptions 1, 2 and 3 as* N *increases, it holds that* $\widehat{\mathbf{w}}\mathbf{R} \xrightarrow{p} \delta\mathbf{1}_T'$.

The above theorem is the punchline of this paper: An investor can consistently recover the arbitrage profits, should they exist, as the number of securities in the cross-section grows large.

We note that some readers may doubt the feasibility of the arbitrage portfolio weights of $\widehat{\mathbf{w}}$ in Theorem 2.3 because the estimated $\widehat{\mathbf{w}}$ uses the returns over the full sample period $t = 1, \cdots, T$. However, recall that the theory does not require large T. Hence, we can estimate $\mathbf{w}$ over one sample and calculate out-of-sample returns over a subsequent sample, as illustrated in Figure 1.

3 Simulation Evidence

In this section, we provide the simulation evidence that that the arbitrage portfolio proposed in the previous section yields the desired properties of generating arbitrage in an economy mimicking U.S. stock markets.

First, we describe the set of characteristics used for simulation. For the matrix of $\mathbf{X}$, we consider 62 characteristics, which are available at the end of 2010, the beginning of calibration period. The set of characteristics include past return such as momentum (returns from $t-12$ to $t-2$) and short-term reversal (returns from $t-2$ to $t-1$), the annual percentage change in total assets, return on operating assets, operating accruals and many others (the full list in given in Table 1 and described in more detail in Freyberger et al. (2017)).

Next, we take a stance on the return generating process of (2.1) by considering four popular asset pricing models of CAPM, Fama-French three factor model (FF3), Hou, Xue and Zhang four factor model (HXZ4), and Fama and French five factor model (FF5). However, we depart from those models by not restricting α to be zero. Initially we estimate $\mathbf{G}_\beta(\mathbf{X})$ using the value of K corresponding to the respective asset pricing model, i.e., $K = 1$ for the CAPM, $K = 3$ for the FF3, etc. Below we also report results using larger values of K .

More specifically, we calibrate α_i, $\boldsymbol{\beta}_i$, and the variance of residual returns, $\sigma^2_{i,\varepsilon} = \mathbb{E}\left[\varepsilon^2_{i,t}\right]$, of individual stocks for each of the four models from time series regression of excess returns of individual stocks on a constant and the factor realizations over the 36-month period from January 2011 to December 2013.[4] For the ease of interpretation, we normalize the cross-sectional variation of α_i so that the quantity of δ in Assumption 1 corresponds to 1 basis point per month, as follows. We estimate $\widehat{\alpha}_i$ from time series regression and fit the cross-sectional relation $\widehat{\alpha}_i = \mathbf{x}_i\widehat{\boldsymbol{\theta}} + e_i$. We rescale $\widetilde{\alpha}_i = k\widehat{\alpha}_i$ where $k = \frac{0.01}{\sqrt{\frac{\widehat{\boldsymbol{\theta}}'\mathbf{X}'\mathbf{X}\widehat{\boldsymbol{\theta}}}{N}}}$ and use the rescaled $\widetilde{\alpha}_i$ for (3.1). The choice of 36-month period is to follow our empirical application, discussed below, in which we set $T_0 = 36$ to estimate the arbitrage portfolios and hold the portfolio for the following periods in an out-of-sample manner as illustrated in Figure 1. There are 2,458 individual stocks with full time series over the calibration sample period. Because the consistency of our arbitrage portfolios is achieved with large cross-section of stocks, we consider $N = 1,000$ and $N = 2,000$, which are sampled from the 2,458 individual stocks.[5]

In each repetition, we simulate returns from

$$\mathbf{R} = \boldsymbol{\alpha}\mathbf{1}_T'\sqrt{\phi} + \mathbf{B}\mathbf{F}' + \mathbf{E}, \tag{3.1}$$

[4]We find that the simulation results are robust to other specifications. We incorporate the cross-sectional correlation among residual shocks by allowing 0.1 correlation among firms within each of 50 or 25 industries. We also explored other choices of $T_0 = 12,\ 24$ over different calibration calendar periods. The results are qualitatively and quantitatively similar.

[5]In the real trading environment, we face the random shock to the returns through systematic factors and residuals for a given set of large cross section of individual stocks. To reflect this, we fix the calibrated alpha, betas, and the variance of residuals of the given cross-section of individual stocks across 10,000 repetitions. Alternatively, we can draw individual stocks with replacement in each iteration. When N is either $N = 1,000$ or $N = 2,000$, simulation results confirm that this modification yields qualitatively and quantitatively similar results.

where $\boldsymbol{\alpha}$ and $\mathbf{B}$ are calibrated as in the above paragraph, $\mathbf{F}$ are bootstrapped from the realized factors over the 600-month sample from January 1967 to December 2016, $\mathbf{E}$ are drawn from normal distribution with the calibrated parameters as in the above paragraph. We consider $\sqrt{\phi} = 0,\ 1,$ and 2.

Figure 2 shows the results of using the Capital Asset Pricing Model (CAPM) as the base factor model in the return generating process. Note that $\widehat{\mathbf{w}}$ is estimated with the returns over $t = 1, \cdots$,36 and the return of the arbitrage portfolio is measured in the following twelve months over $t = 37, \cdots, 48$. That is, we use $T_0 = 36$ and $T = 48$ in the setup of Figure 1. We report the mean (solid line) and the 95% confidence intervals (dotted lines) of the time series of returns from our arbitrage portfolio $\widehat{\mathbf{w}}'\mathbf{R}_{37}, \cdots,\ \widehat{\mathbf{w}}'\mathbf{R}_{48}$ over the 12 month holding periods for each level of $\phi = 0,\ ,1$ and 2 and $N = 1,000$ and $N = 2,000$ from 10,000 repetitions. Note that the confidence intervals are narrower with $N = 2,000$ than those with $N = 1,000$. This result is empirically relevant because we can reasonably obtain a cross-section of this size in the U.S. stock market. As expected, when $\phi = 0$, or there do not exist any arbitrage opportunities, our arbitrage portfolio yields zero returns on average. Recall that α_i is rescaled so that $\frac{\mathbf{G}_\alpha(\mathbf{X})'\mathbf{G}_\alpha(\mathbf{X})}{N} \to$ 1b.p./month. Hence, the arbitrage portfolio is suppose to generate $\phi = \lim_{N\to\infty}\left(\frac{\left(\mathbf{G}_\alpha(\mathbf{X})\sqrt{\phi}\right)'\left(\mathbf{G}_\alpha(\mathbf{X})\sqrt{\phi}\right)}{N}\right)$. In fact, we observe that, when $\phi = 1$ or 2, the average of arbitrage portfolio returns corresponds to the target size of ϕ b.p./month, suggesting that our arbitrage portfolio actually generates arbitrage profits. We find similar results for other asset pricing models of FF3, HXZ4 and FF5, which are reported in Figures 3, 4, and 5, respectively.

So far, in extracting factor loadings from the projected returns, we use the true number of factors for each of the asset pricing models under consideration. However, because we do not know the true model in practice, we need to confront the issue on the number of factors. One feasible approach is using well-known tests on the number of factors using panel data such as the tests proposed by Bai and Ng (2002) or Ahn and Horenstein (2013). However, because our arbitrage portfolio needs to be orthogonal to the factor loadings on all factors, we take a safer strategy of using a sufficiently large number of factors. In fact, we find that the behavior of arbitrage portfolios is robust to using a larger number of factors than that of the true model. To see this, we repeat the analysis by using six factors and report the results in Figures 6-9, corresponding to

Figures 2-5, respectively.[6] The performance of arbitrage portfolio when over-fitting the number of factors is qualitatively and quantitatively similar to using the true number of factors.

4 Empirical Application

In this section we discuss the set of characteristics and the application of our methodology to U.S. stock market data.

4.1 Data

The data are the same as in Freyberger et al. (2017), we use stock return data from the Center for Research in Security Prices (CRSP) monthly file. As common in the literature, we limit the analysis to U.S. firms' common equity, which is trading on NYSE, Amex or Nasdaq. Accounting data are obtained from Compustat. As in Freyberger et al., we use accounting data from the fiscal year ending in calendar year $t-1$ for estimation starting from the end of June of year t until the end of May of year $t+1$ predicting returns from the beginning of July of year t until the end June of year $t+1$. Table 1 provides an overview of the characteristics used for estimation of mispricing function and factor loading function.[7] Our sample period is from 1965 through 2014.

4.2 Estimation

In the spirit of Ferson and Harvey (1994), we initially assume that the factor loading function and the mispricing function are linear in the characteristics, note that the methodology allows for nonlinear (but finite) expansions, which we explore below.

We estimate $\widehat{\mathbf{w}}$ with the returns over $t = 1, \cdots, 36$ and the return of the arbitrage portfolio is measured in the following one month $t = 37$. In the setup of Figure 1 we use $T_0 = 36$ and $T = 37$. For example, we first obtain projected and demeaned return of $\widehat{\mathbf{R}}$ corresponding to $\mathbf{PRJ}_{T_0}$ in (2.5) from a panel regression using 36 months from

[6] For the case of underestimated number of factors (using K less than the true number of factors) we find that arbitrage returns becomes much more volatile due to the unhedged exposure to systematic risks.

[7] The appendix of Freyberger et al. contains a detailed description about the construction of the data.

January 1965 to December 1967 as an estimation window. The t-th column of the $(N \times 36)$ matrix $\widehat{\mathbf{R}}$ is the demeaned projected return for the t-th month. Then we compute the $N \times N$ matrix $\widehat{\mathbf{R}}\widehat{\mathbf{R}}'$ and compute the first K eigenvectors. Note that the number of factors is in principle a free parameter in the estimation. It can be chosen as in Connor and Korajczyk (1993) or Bai and Ng (2002). Rather than relying on the correct estimation of the true number of factors, we present results for different values of K. Recall that the eigenvectors of $\widehat{\mathbf{R}}\widehat{\mathbf{R}}'$ represent the estimated factor loadings from Theorem 2.1. We then project the characteristics onto the estimated factor loadings with the constraints as in Theorem 2.2 and compute the arbitrage portfolio weights as in Theorem 2.3 for the following month of January 1968. We repeat this process till June 2014. For the out-of-sample implementation, we update the portfolio weights with characteristics as of the end of estimation period. In order to make the results comparable in scale to common equity factors, we scale the portfolio weights so that the in-sample standard deviation is 20% per year.

4.3 Empirical Results

Table 2 shows summary statistics for returns, standard deviation of the arbitrage portfolio for different numbers of eigenvectors. The average returns of the arbitrage portfolio and Sharpe ratios are increasing in the number of eigenvectors which are estimated. Note that this increase is driven by higher means rather than by lower standard deviations because the standard deviation is always normalized to 20% annually, over the estimation period. The out-of-sample standard deviation is close to the in-sample standard deviation. Moreover, the table also shows skewness, kurtosis and the best and the worst month for the arbitrage portfolio.

Next we run time-series regression of the arbitrage portfolio's returns onto common risk factors. In particular, we aim to explain the returns of the arbitrage portfolio with CAPM, the Fama-French three factor model, the Fama-French three factor model augmented with momentum, the Fama-French five factor model and the Fama-French five factor model augmented with momentum. Tables 3 through 11 show the results for estimating the arbitrage portfolio with different numbers of eigenvectors. Column (1) show the results for the CAPM. In all specifications the arbitrage portfolio earns a large positive and strongly significant alpha (shown in percent per month), column

(2) and column (3) show the results for the Fama-French three factor model and the Fama-French three factor model augmented with the Carhart (1997) momentum factor, column (4) shows the results for the Fama-French five factor model and column (6) shows the results for the Fama-French five factor model augmented with momentum. In Table 3 with one eigenvector, we can see that the alpha becomes larger for the Fama-French model augmented with the momentum factor because of the arbitrage portfolio's negative exposure to momentum. Including the momentum factor leads to a large increase in the adjusted R^2 of the regression. As we increase the number of eigenvectors, the alphas become more similar across the alternative benchmarks. For ten eigenvectors only the exposure to the size factor, smb, remains significant. Moreover, as we increase the number of eigenvectors, the R^2 of the factor models decreases. This seems to indicate that the returns generating process is driven by multiple factors. We illustrate the relation between out-of-sample alpha and alternative numbers of eigenvectors used in the estimator in Figure 10. We illustrate the relation between R^2 and the number of eigenvectors in Figure 11. In both figures we can see that the alpha and the R^2 "flatten out" after approximately six (alpha) and four (R^2) eigenvectors. Figure 12 summarizes the correlation of the arbitrage portfolio (using 10 eigenvectors) with common risk factors. We can see that the correlation is relatively small for all factors except size.

McLean and Pontiff (2016) and Linnainmaa and Roberts (2018) document that many anomalies have become significantly weaker post publication. In Figure 13 we plot the monthly excess returns of the arbitrage portfolio (using ten eigenvectors) and a linear time trend. The slope is slightly negative, but not in any economically meaningful way. While it is possible that data snooping will lead to reduced future performance of the arbitrage portfolio, many of the predictive characteristics are the result of research done decades ago. We conclude that the significant average excess returns are at least partially due to mispricing of assets.

5 Conclusion

We propose a new method to simultaneously recover conditional factor realizations (returns on "smart-beta" portfolios), estimate conditional factor loadings, and construct arbitrage portfolios, by extending the Projected Principal Component Analysis method of Fan et al. (2016). We show that our methodology works well in simulations of factor economies. Moreover, it works reliably on a large panel of U.S. stocks from 1968 to 2014. Although we give maximal power to a risk based explanation, we find that characteristics carry significant information about mispricing. We find that alphas against popular factor models of 1% to almost 2% per month.

Much research remains to be done. A natural next step is to investigate the properties of the estimated factors more closely and apply them in more traditional settings such as mutual fund performance evaluation. Moreover, our results about higher order expansions of the factor loading function (in the appendix) indicate that exploring other specifiations of the factor loading is also a promising direction for future work.

A Incorporating Nonlinearities

In Section 2, we have not taken a parametric stand on the functional form of $\mathbf{G}_\beta(\mathbf{X})$. In the application, we have assumed that $\mathbf{G}_\beta(\mathbf{X})$ is a linear function. In this section, we briefly outline one possible way to incorporate nonlinearities into $\mathbf{G}_\beta(\mathbf{X})$. In Fan et al., $\mathbf{G}_\beta(\mathbf{X})$, is approximated as by a series expansion in a nonparametric additive setting. The assumption of additivity ($\mathbf{G}_\beta(\mathbf{X}) = \sum g(x_1) + g(x_2) + ... + g(x_L)$) has the appealing property that, $\mathbf{G}_\beta(\mathbf{X})$, can be estimated without the so-called "curse of dimensionality" because the rate of convergence does not depend on the dimension of X, so that it can be estimated with many characteristics. However, it introduces a complication in the asymptotic theory, namely that the series expansion also grows with the cross-sectional sample size. Since our interest is primarily applied and to avoid these technicalities, we assume that $\mathbf{G}_\beta(\mathbf{X})$ can be well approximated by a fixed order polynomial expansion. In the application we will use Legendre polynomials to incorporate nonlinearities in the estimation of $\mathbf{G}_\beta(\mathbf{X})$.[8]

In Tables 12, 13, and 14 we show alphas of the arbitrage portfolio against various factor models, when we use second, third and fourth order Legendre polynomials in the estimation of $\mathbf{G}_\beta(\mathbf{X})$. The alphas are slightly smaller than in the linear specification but mostly still in excess of one percent per month and strongly statistically significant. Tables 15, 16, and 17 show the corresponding R^2's for higher order expansions of $\mathbf{G}_\beta(\mathbf{X})$. Interestingly, the R^2's are slightly lower than in the linear specification. This suggests that nonlinearities are helpful for estimating factors. Overall however, the results of the higher order expansions are consistent with the the linear specification and do not errode the arbitrage profits. However, they leave interesting avenues for future research.

B Proofs

Lemma B.1. *Let $\mathbf{Y}$ be a $(N \times T)$ matrix. Assume that the rank of $\mathbf{Y}$ is at least $K > 0$. Then, the k-th largest eigenvalue of $\mathbf{YY}'$ is identical to the k-th largest eigenvalue of $\mathbf{Y}'\mathbf{Y}$ for $k = 1, \cdots, K$.*

[8]Legendre polynomials are frequently used in econometrics to approximate unknown functions and fall into the more general class of "orthogonal polynomials". We refer to Bierens (2014)'s handbook chapter for a deep theoretical treatment of orthogonal polynomials.

Proof It suffices to show that λ is a strictly positive eigenvalue of $\mathbf{YY}'$ if and only if λ is a strictly positive eigenvalue of $\mathbf{Y}'\mathbf{Y}$. We show this in the following two steps.

Step 1: $\Rightarrow$) Let the $(N \times 1)$ vector $\mathbf{v}_0$ denote the eigenvector of $\mathbf{YY}'$ corresponding to the strictly positive eigenvalue of λ_0. Then, it holds that

$$\mathbf{YY}'\mathbf{v}_0 = \lambda_0 \mathbf{v}_0, \tag{B.1}$$

which implies

$$\mathbf{Y}'\mathbf{Y}\left(\mathbf{Y}'\mathbf{v}_0\right) = \lambda_0 \left(\mathbf{Y}'\mathbf{v}_0\right). \tag{B.2}$$

If $\mathbf{Y}'\mathbf{v}_0$ is a vector of zeros, $\lambda_0 = 0$ from (B.1) because $\mathbf{v}_0 \neq \mathbf{0}_N$, contradicting that λ_0 is strictly positive. Hence, $\mathbf{Y}'\mathbf{v}_0$ is a non-zero vector. From (B.2), λ_0 is the eigenvalue of $\mathbf{Y}'\mathbf{Y}$ with the corresponding eigenvector of $\mathbf{Y}'\mathbf{v}_0$.

Step 2: $\Leftarrow$) Replace $\mathbf{Y}$ with $\mathbf{Y}'$ in Step 1. $\square$

For a $(m \times m)$ diagonal matrix $\mathbf{D}$ with positive diagonal elements, we define $\mathbf{D}^a$ as the $(m \times m)$ diagonal matrix such that the k-th diagonal element of $\mathbf{D}^a$ is the k-th diagonal element of $\mathbf{D}$ to the power of a.

Lemma B.2. *Let $\mathbf{Y}$ be a $(N \times T)$ matrix. Assume that the first K eigenvalues of $\mathbf{Y}'\mathbf{Y}$ are distinct and strictly positive. Define the $(N \times K)$ matrix $\widehat{\mathbf{\Lambda}}$ such that the k-th column of $\widehat{\mathbf{\Lambda}}$ is the eigenvector of $\mathbf{YY}'$ corresponding to the k-th largest eigenvalue of $\mathbf{YY}'$. Let $\widetilde{\mathbf{\Lambda}} = \mathbf{Y}\widetilde{\mathbf{F}}\left(\widetilde{\mathbf{F}}'\widetilde{\mathbf{F}}\right)^{-1}$, where $\widetilde{\mathbf{F}} = \widehat{\mathbf{F}}\mathbf{D}^{1/2}$, the k-th column of the $(N \times K)$ matrix $\widehat{\mathbf{F}}$ is the eigenvector of $\mathbf{Y}'\mathbf{Y}$ corresponding to the k-th largest eigenvalue of $\mathbf{Y}'\mathbf{Y}$, and the k-th diagonal element of the $(K \times K)$ diagonal matrix $\mathbf{D}$ is the k-th largest eigenvalue of $\mathbf{Y}'\mathbf{Y}$. Then, it holds that*

$$\widehat{\mathbf{\Lambda}} = \widetilde{\mathbf{\Lambda}}.$$

Proof From Lemma B.1, the k-th largest eigenvalue of $\mathbf{Y}'\mathbf{Y}$ is the k-th largest eigenvalue of $\mathbf{YY}'$. Hence, $\widehat{\mathbf{\Lambda}}$ is identified by the following two conditions:

$$\text{i) } \widehat{\mathbf{\Lambda}}'\widehat{\mathbf{\Lambda}} = \mathbf{I}_K$$
$$\text{ii) } \widehat{\mathbf{\Lambda}}'\mathbf{YY}'\widehat{\mathbf{\Lambda}} = \mathbf{D}.$$

Using eigen-decomposition, we express the $(T \times T)$ matrix of $\mathbf{Y}'\mathbf{Y}$ as $\mathbf{QVQ}'$:

$$\mathbf{Y}'\mathbf{Y} = \mathbf{QVQ}'. \tag{B.3}$$

Note that the $(T \times K)$ matrix made out of the first K columns of $\mathbf{Q}$ is $\widehat{\mathbf{F}}$ and that the first

K diagonal elements of $\mathbf{V}$ correspond to the diagonal elements of $\mathbf{D}$:

$$\widehat{\mathbf{F}} = \mathbf{Q}\left[\mathbf{I}_K\ \mathbf{0}_{K\times(T-K)}\right]' \text{ and } \mathbf{D} = \left[\mathbf{I}_K\ \mathbf{0}_{K\times(T-K)}\right]\mathbf{V}\left[\mathbf{I}_K\ \mathbf{0}_{K\times(T-K)}\right]'. \tag{B.4}$$

We prove the lemma by showing that $\widetilde{\mathbf{\Lambda}}$ satisfies the two conditions of i) and ii) in the above when we set $\widehat{\mathbf{\Lambda}} = \widetilde{\mathbf{\Lambda}}$. Because $\widetilde{\mathbf{\Lambda}} = \mathbf{Y}\widetilde{\mathbf{F}}\left(\widetilde{\mathbf{F}}'\widetilde{\mathbf{F}}\right)^{-1} = \mathbf{Y}\widehat{\mathbf{F}}\left(\widehat{\mathbf{F}}'\widehat{\mathbf{F}}\right)^{-1}\mathbf{D}^{-0.5} = \mathbf{Y}\widehat{\mathbf{F}}\mathbf{D}^{-0.5}$, it follows that

$$\begin{aligned}\widetilde{\mathbf{\Lambda}}'\widetilde{\mathbf{\Lambda}} &= \mathbf{D}^{-0.5}\widehat{\mathbf{F}}'\mathbf{Y}'\mathbf{Y}\widehat{\mathbf{F}}\mathbf{D}^{-0.5} = \mathbf{D}^{-0.5}\left[\mathbf{I}_K\ \mathbf{0}_{K\times(T-K)}\right]\mathbf{Q}'\mathbf{Q}\mathbf{V}\mathbf{Q}'\mathbf{Q}\left[\mathbf{I}_K\ \mathbf{0}_{K\times(T-K)}\right]'\mathbf{D}^{-0.5}\\ &= \mathbf{D}^{-0.5}\left[\mathbf{I}_K\ \mathbf{0}_{K\times(T-K)}\right]\mathbf{V}\left[\mathbf{I}_K\ \mathbf{0}_{K\times(T-K)}\right]'\mathbf{D}^{-0.5} = \mathbf{D}^{-0.5}\mathbf{D}\mathbf{D}^{-0.5} = \mathbf{I}_K,\end{aligned} \tag{B.5}$$

where the second and fourth equalities are from (B.3) and (B.4), and that

$$\begin{aligned}\widetilde{\mathbf{\Lambda}}'\mathbf{Y}\mathbf{Y}'\widetilde{\mathbf{\Lambda}} &= \mathbf{D}^{-0.5}\widehat{\mathbf{F}}'\mathbf{Y}'\mathbf{Y}\mathbf{Y}'\mathbf{Y}\widehat{\mathbf{F}}\mathbf{D}^{-0.5} = \mathbf{D}^{-0.5}\widehat{\mathbf{F}}'\mathbf{Q}\mathbf{V}^2\mathbf{Q}'\widehat{\mathbf{F}}\mathbf{D}^{-0.5}\\ &= \mathbf{D}^{-0.5}\left[\mathbf{I}_K\ \mathbf{0}_{K\times(T-K)}\right]\mathbf{Q}'\mathbf{Q}\mathbf{V}^2\mathbf{Q}'\mathbf{Q}\left[\mathbf{I}_K\ \mathbf{0}_{K\times(T-K)}\right]'\mathbf{D}^{-0.5}\\ &= \mathbf{D}^{-0.5}\left[\mathbf{I}_K\ \mathbf{0}_{K\times(T-K)}\right]\mathbf{V}^2\left[\mathbf{I}_K\ \mathbf{0}_{K\times(T-K)}\right]'\mathbf{D}^{-0.5}\\ &= \mathbf{D}^{-0.5}\left(\left[\mathbf{I}_K\ \mathbf{0}_{K\times(T-K)}\right]\mathbf{V}\left[\mathbf{I}_K\ \mathbf{0}_{K\times(T-K)}\right]'\right)^2\mathbf{D}^{-0.5}\\ &= \mathbf{D}^{-0.5}\mathbf{D}^2\mathbf{D}^{-0.5} = \mathbf{D},\end{aligned} \tag{B.6}$$

where the second equality is from (B.3) and the third and sixth equalities are from (B.4). Finally, the two equalities of (B.5) and (B.6) prove the lemma. □

Lemma B.3. *Let* $\widehat{\mathbf{G}}_\beta(\mathbf{X})$ *denote the* $(N\times K)$ *matrix, the* k*-th column of which is* $\sqrt{N}$ *times the eigenvector of* $\frac{\widehat{\mathbf{R}}\widehat{\mathbf{R}}'}{N}$ *corresponding to the first* k*-th eigenvalue of* $\frac{\widehat{\mathbf{R}}\widehat{\mathbf{R}}'}{N}$*, where* $\widehat{\mathbf{R}}$ *is given by (2.5) as in Theorem 2.1. Define* $\widetilde{\mathbf{G}}_\beta(\mathbf{X}) = \widehat{\mathbf{R}}\widetilde{\mathbf{F}}\left(\widetilde{\mathbf{F}}'\widetilde{\mathbf{F}}\right)^{-1}$*, where* $\widetilde{\mathbf{F}} = \widehat{\mathbf{F}}\mathbf{D}^{1/2}$*, the* k*-th column of the* $(T\times K)$ *matrix* $\widehat{\mathbf{F}}$ *is the eigenvector of* $\frac{\widehat{\mathbf{R}}'\widehat{\mathbf{R}}}{N}$ *corresponding to the* k*-th largest eigenvalue of* $\frac{\widehat{\mathbf{R}}'\widehat{\mathbf{R}}}{N}$*, and the* k*-th element of the* $(K\times K)$ *diagonal matrix* $\mathbf{D}$ *is the* k*-th largest eigenvalue of* $\frac{\widehat{\mathbf{R}}'\widehat{\mathbf{R}}}{N}$*. Then, it holds that*
(i) $\widehat{\mathbf{G}}_\beta(\mathbf{X}) = \widetilde{\mathbf{G}}_\beta(\mathbf{X})$
(ii) $\mathbf{P}\widehat{\mathbf{G}}_\beta(\mathbf{X}) = \widehat{\mathbf{G}}_\beta(\mathbf{X})$.

Proof Note that $\frac{\widehat{\mathbf{R}}\widehat{\mathbf{R}}'}{N} = \left(\frac{\widehat{\mathbf{R}}}{\sqrt{N}}\right)'\left(\frac{\widehat{\mathbf{R}}}{\sqrt{N}}\right)$ and $\frac{\widehat{\mathbf{R}}'\widehat{\mathbf{R}}}{N} = \left(\frac{\widehat{\mathbf{R}}}{\sqrt{N}}\right)\left(\frac{\widehat{\mathbf{R}}}{\sqrt{N}}\right)'$ and that $\widetilde{\mathbf{G}}_\beta(\mathbf{X}) = \sqrt{N}\frac{\widehat{\mathbf{R}}}{\sqrt{N}}\widetilde{\mathbf{F}}\left(\widetilde{\mathbf{F}}'\widetilde{\mathbf{F}}\right)^{-1}$. Hence, (i) directly follows from Lemma B.2.

We turn to (ii). Because $\mathbf{P}\widetilde{\mathbf{G}}_\beta(\mathbf{X}) = \mathbf{P}\mathbf{P}\mathbf{R}\mathbf{J}_T\widetilde{\mathbf{F}}\left(\widetilde{\mathbf{F}}'\widetilde{\mathbf{F}}\right)^{-1} = \mathbf{P}\mathbf{R}\mathbf{J}_T\widetilde{\mathbf{F}}\left(\widetilde{\mathbf{F}}'\widetilde{\mathbf{F}}\right)^{-1} = \widetilde{\mathbf{G}}_\beta(\mathbf{X})$, (ii) is true from (i). This completes the proof of the lemma. □

Lemma B.4. *Under Assumptions 2 and 3(ii), it holds that as N increases,* $\frac{\widehat{\mathbf{R}}'\widehat{\mathbf{R}}}{N} \xrightarrow{p} \mathbf{J}_T\mathbf{F}\mathbf{F}'\mathbf{J}_T$.

Proof From (2.5), we have that

$$\widehat{\mathbf{R}} = l_{(1)} + l_{(2)} + l_{(3)},$$

where $l_{(1)} = \mathbf{P}\mathbf{G}_\beta\left(\mathbf{X}\right)\mathbf{F}'\mathbf{J}_T$, $l_{(2)} = \mathbf{P}\Gamma_\beta\mathbf{F}'\mathbf{J}_T$ and $l_{(3)} = \mathbf{P}\mathbf{E}\mathbf{J}_T$. Hence,

$$\frac{\widehat{\mathbf{R}}'\widehat{\mathbf{R}}}{N} = \sum_{i,j=1}^{3} \frac{1}{N} l'_{(i)} l_{(j)}. \tag{B.7}$$

Note that

$$\frac{1}{N} l'_{(1)} l_{(1)} = \mathbf{J}_T\mathbf{F}\left(\frac{\mathbf{G}_\beta\left(\mathbf{X}\right)'\mathbf{G}_\beta\left(\mathbf{X}\right)}{N}\right)\mathbf{F}'\mathbf{J}_T = \mathbf{J}_T\mathbf{F}\mathbf{F}'\mathbf{J}_T \tag{B.8}$$

from Assumption 3(ii) and that

$$\frac{1}{N} l'_{(1)} l_{(2)} = \mathbf{J}_T\mathbf{F}\left(\frac{\mathbf{G}_\beta\left(\mathbf{X}\right)'\Gamma_\beta}{N}\right)\mathbf{F}'\mathbf{J}_T \xrightarrow{p} \mathbf{J}_T\mathbf{F}\mathbf{0}_{K\times K}\mathbf{F}'\mathbf{J}_T = \mathbf{0}_{T\times T} \tag{B.9}$$

from Assumption 2(ii) and that

$$\frac{1}{N} l'_{(1)} l_{(3)} = \mathbf{J}_T\mathbf{F}\left(\frac{\mathbf{G}_\beta\left(\mathbf{X}\right)'\mathbf{E}}{N}\right)\mathbf{J}_T \xrightarrow{p} \mathbf{J}_T\mathbf{F}\mathbf{0}_{K\times T}\mathbf{J}_T = \mathbf{0}_{T\times T} \tag{B.10}$$

from Assumption 2(ii) and that

$$\frac{1}{N} l'_{(2)} l_{(2)} = \mathbf{J}_T\mathbf{F}\left(\frac{\Gamma'_\beta\mathbf{X}}{N}\right)\left(\frac{\mathbf{X}'\mathbf{X}}{N}\right)^{-1}\left(\frac{\mathbf{X}'\Gamma_\beta}{N}\right)\mathbf{F}'\mathbf{J}_T \xrightarrow{p} \mathbf{J}_T\mathbf{F}\mathbf{0}_{K\times L}\mathbf{V}_X^{-1}\mathbf{0}_{L\times K}\mathbf{F}'\mathbf{J}_T = \mathbf{0}_{T\times T} \tag{B.11}$$

from Assumptions 2(i) and 2(ii) and that

$$\frac{1}{N} l'_{(2)} l_{(3)} = \mathbf{J}_T\mathbf{F}\left(\frac{\Gamma'_\beta\mathbf{X}}{N}\right)\left(\frac{\mathbf{X}'\mathbf{X}}{N}\right)^{-1}\left(\frac{\mathbf{X}'\mathbf{E}}{N}\right)\mathbf{J}_T \xrightarrow{p} \mathbf{J}_T\mathbf{0}_{T\times L}\mathbf{V}_X^{-1}\mathbf{0}_{L\times T}\mathbf{J}_T = \mathbf{0}_{T\times T} \tag{B.12}$$

from Assumptions 2(i) and 2(ii) and that

$$\frac{1}{N} l'_{(3)} l_{(3)} = \mathbf{J}_T\left(\frac{\mathbf{E}'\mathbf{X}}{N}\right)\left(\frac{\mathbf{X}'\mathbf{X}}{N}\right)^{-1}\left(\frac{\mathbf{X}'\mathbf{E}}{N}\right)\mathbf{J}_T \xrightarrow{p} \mathbf{J}_T\mathbf{0}_{T\times L}\mathbf{V}_X^{-1}\mathbf{0}_{L\times T}\mathbf{J}_T = \mathbf{0}_{T\times T} \tag{B.13}$$

from Assumptions 2(i) and 2(ii).

Finally, plugging the results of (B.8)-(B.13) into (B.7), we have that $\frac{\widehat{\mathbf{R}}'\widehat{\mathbf{R}}}{N} \xrightarrow{p} \mathbf{J}_T\mathbf{F}\mathbf{F}'\mathbf{J}_T$, completing the proof of the lemma. □

Proof of Theorem 2.1 The following seven steps complete the proof of $\widehat{\mathbf{G}}_\beta(\mathbf{X}) \xrightarrow{p} \mathbf{G}_\beta(\mathbf{X})$.

Step 1. $\widehat{\mathbf{F}} \xrightarrow{p} \mathbf{J}_T\mathbf{F}(\mathbf{F}\mathbf{J}_T\mathbf{F}')^{-0.5}$: Recall that $\frac{\widehat{\mathbf{R}}'\widehat{\mathbf{R}}}{N} \xrightarrow{p} \mathbf{J}_T\mathbf{F}\mathbf{F}'\mathbf{J}_T$ from Lemma B.4 and $\widehat{\mathbf{F}}$ is the $(T \times K)$ matrix, each column of which is an eigenvector of $\frac{\widehat{\mathbf{R}}'\widehat{\mathbf{R}}}{N}$. Noting that $\left(\mathbf{J}_T\mathbf{F}(\mathbf{F}\mathbf{J}_T\mathbf{F}')^{-0.5}\right)'\left(\mathbf{J}_T\mathbf{F}(\mathbf{F}\mathbf{J}_T\mathbf{F}')^{-0.5}\right) = \mathbf{I}_K$ and that

$$\left(\mathbf{J}_T\mathbf{F}\left(\mathbf{F}\mathbf{J}_T\mathbf{F}'\right)^{-0.5}\right)'\mathbf{J}_T\mathbf{F}\mathbf{F}'\mathbf{J}_T\left(\mathbf{J}_T\mathbf{F}\left(\mathbf{F}\mathbf{J}_T\mathbf{F}'\right)^{-0.5}\right) = \mathbf{F}\mathbf{J}_T\mathbf{F}',$$

which is a diagonal matrix from Assumption 3(iii). Thus, $\mathbf{J}_T\mathbf{F}(\mathbf{F}\mathbf{J}_T\mathbf{F}')^{-0.5}$ is the $(T \times K)$ matrix, each column of which is an eigenvector of $\mathbf{J}_T\mathbf{F}\mathbf{F}'\mathbf{J}_T$. Due to the continuity of eigen-decomposition, it follows that $\widehat{\mathbf{F}} \xrightarrow{p} \mathbf{J}_T\mathbf{F}(\mathbf{F}\mathbf{J}_T\mathbf{F}')^{-0.5}$.

Step 2. $\mathbf{D} \xrightarrow{p} \mathbf{F}\mathbf{J}_T\mathbf{F}'$: In Step 1, we show that $\mathbf{F}\mathbf{J}_T\mathbf{F}'$ is the diagonal matrix whose diagonal elements are eigenvalues of $\mathbf{J}_T\mathbf{F}\mathbf{F}'\mathbf{J}_T$. Due to the continuity of eigendecomposition, it follows that $\mathbf{D} \xrightarrow{p} \mathbf{F}\mathbf{J}_T\mathbf{F}'$.

Step 3. $\widetilde{\mathbf{F}} \xrightarrow{p} \mathbf{J}_T\mathbf{F}$: From Steps 1 and 2, it holds that $\widetilde{\mathbf{F}} = \widehat{\mathbf{F}}\mathbf{D}^{0.5} \xrightarrow{p} \mathbf{J}_T\mathbf{F}(\mathbf{F}\mathbf{J}_T\mathbf{F}')^{-0.5}(\mathbf{F}\mathbf{J}_T\mathbf{F}')^{0.5} = \mathbf{J}_T\mathbf{F}$.

Step 4. $\mathbf{F}'\mathbf{J}_T\widetilde{\mathbf{F}}\left(\widetilde{\mathbf{F}}'\widetilde{\mathbf{F}}\right)^{-1} \xrightarrow{p} \mathbf{I}_K$: From Step 3, it holds that $\mathbf{F}'\mathbf{J}_T\widetilde{\mathbf{F}}\left(\widetilde{\mathbf{F}}'\widetilde{\mathbf{F}}\right)^{-1} \xrightarrow{p} \mathbf{F}'\mathbf{J}_T\mathbf{F}(\mathbf{F}\mathbf{J}_T\mathbf{F}')^{-1} = \mathbf{I}_K$.

Step 5 $\widetilde{\mathbf{G}}_\beta(\mathbf{X}) = \mathbf{P}\mathbf{R}\mathbf{J}_T\widetilde{\mathbf{F}}'\left(\widetilde{\mathbf{F}}'\widetilde{\mathbf{F}}\right)^{-1} \xrightarrow{p} \mathbf{G}_\beta(\mathbf{X})$: Using the expression of $\mathbf{P}\mathbf{R}\mathbf{J}_T$ in (2.5), we find that

$$\widetilde{\mathbf{G}}_\beta(\mathbf{X}) = \mathbf{G}_\beta(\mathbf{X})\mathbf{F}\mathbf{J}_T\widetilde{\mathbf{F}}'\left(\widetilde{\mathbf{F}}'\widetilde{\mathbf{F}}\right)^{-1} + \mathbf{P}\Gamma_\beta\mathbf{F}\mathbf{J}_T\widetilde{\mathbf{F}}'\left(\widetilde{\mathbf{F}}'\widetilde{\mathbf{F}}\right)^{-1} + \mathbf{P}\mathbf{E}\mathbf{J}_T\widetilde{\mathbf{F}}'\left(\widetilde{\mathbf{F}}'\widetilde{\mathbf{F}}\right)^{-1},$$

which gives

$$\widetilde{\mathbf{G}}_\beta(\mathbf{X}) - \mathbf{G}_\beta(\mathbf{X}) = m_{(1)} + m_{(2)} + m_{(3)},$$

where $m_{(1)} = \mathbf{G}_\beta(\mathbf{X})\left(\mathbf{F}\mathbf{J}_T\widetilde{\mathbf{F}}'\left(\widetilde{\mathbf{F}}'\widetilde{\mathbf{F}}\right)^{-1} - \mathbf{I}_K\right)$, $m_{(2)} = \mathbf{P}\Gamma_\beta\mathbf{F}\mathbf{J}_T\widetilde{\mathbf{F}}'\left(\widetilde{\mathbf{F}}'\widetilde{\mathbf{F}}\right)^{-1}$ and $m_{(3)} = \mathbf{P}\mathbf{E}\mathbf{J}_T\widetilde{\mathbf{F}}'\left(\widetilde{\mathbf{F}}'\widetilde{\mathbf{F}}\right)^{-1}$. Hence,

$$\frac{1}{N}\left(\widetilde{\mathbf{G}}_\beta(\mathbf{X}) - \mathbf{G}_\beta(\mathbf{X})\right)'\left(\widetilde{\mathbf{G}}_\beta(\mathbf{X}) - \mathbf{G}_\beta(\mathbf{X})\right) = \sum_{i,j=1}^{3}\frac{1}{N}m_{(i)}'m_{(j)}. \tag{B.14}$$

Note that

$$\frac{1}{N}m_{(1)}'m_{(1)} = \left(\mathbf{F}'\mathbf{J}_T\widetilde{\mathbf{F}}\left(\widetilde{\mathbf{F}}'\widetilde{\mathbf{F}}\right)^{-1} - \mathbf{I}_K\right)' \frac{\mathbf{G}_\beta\left(\mathbf{X}\right)'\mathbf{G}_\beta\left(\mathbf{X}\right)}{N} \left(\mathbf{F}'\mathbf{J}_T\widetilde{\mathbf{F}}\left(\widetilde{\mathbf{F}}'\widetilde{\mathbf{F}}\right)^{-1} - \mathbf{I}_K\right)$$
$$\xrightarrow{p} \left(\mathbf{I}_K - \mathbf{I}_K\right)'\mathbf{I}_K\left(\mathbf{I}_K - \mathbf{I}_K\right) = \mathbf{0}_{K\times K} \tag{B.15}$$

from Step 4 and Assumption 3(ii) and that

$$\frac{1}{N}m_{(1)}'m_{(2)} = \left(\mathbf{F}'\mathbf{J}_T\widetilde{\mathbf{F}}\left(\widetilde{\mathbf{F}}'\widetilde{\mathbf{F}}\right)^{-1} - \mathbf{I}_K\right)' \frac{\mathbf{G}_\beta\left(\mathbf{X}\right)'\Gamma_\beta}{N}\mathbf{F}'\mathbf{J}_T\widetilde{\mathbf{F}}\left(\widetilde{\mathbf{F}}'\widetilde{\mathbf{F}}\right)^{-1}$$
$$\xrightarrow{p} \left(\mathbf{I}_K - \mathbf{I}_K\right)'\mathbf{0}_{K\times K}\mathbf{I}_K = \mathbf{0}_{K\times K} \tag{B.16}$$

from Step 4 and Assumption 2(ii) and that

$$\frac{1}{N}m_{(1)}'m_{(3)} = \left(\mathbf{F}'\mathbf{J}_T\widetilde{\mathbf{F}}\left(\widetilde{\mathbf{F}}'\widetilde{\mathbf{F}}\right)^{-1} - \mathbf{I}_K\right)' \frac{\mathbf{G}_\beta\left(\mathbf{X}\right)'\mathbf{E}}{N}\mathbf{J}_T\widetilde{\mathbf{F}}\left(\widetilde{\mathbf{F}}'\widetilde{\mathbf{F}}\right)^{-1}$$
$$\xrightarrow{p} \left(\mathbf{I}_K - \mathbf{I}_K\right)'\mathbf{0}_{K\times T}\mathbf{J}_T\mathbf{F}\left(\mathbf{F}'\mathbf{J}_T\mathbf{F}\right)^{-1} = \mathbf{0}_{K\times K} \tag{B.17}$$

from Step 4 and Assumption 2(ii) and that

$$\frac{1}{N}m_{(2)}'m_{(2)} = \left(\widetilde{\mathbf{F}}'\widetilde{\mathbf{F}}\right)^{-1}\widetilde{\mathbf{F}}'\mathbf{J}_T\mathbf{F}\left(\frac{\Gamma_\beta'\mathbf{X}}{N}\right)\left(\frac{\mathbf{X}'\mathbf{X}}{N}\right)^{-1}\left(\frac{\mathbf{X}'\Gamma_\beta}{N}\right)\mathbf{F}'\mathbf{J}_T\widetilde{\mathbf{F}}\left(\widetilde{\mathbf{F}}'\widetilde{\mathbf{F}}\right)^{-1}$$
$$\xrightarrow{p} \mathbf{I}_K\mathbf{0}_{K\times L}\mathbf{V}_X^{-1}\mathbf{0}_{L\times K}\mathbf{I}_K = \mathbf{0}_{K\times K} \tag{B.18}$$

from Step 4 and Assumptions 2(i) and 2(ii) and that

$$\frac{1}{N}m_{(2)}'m_{(3)} = \left(\widetilde{\mathbf{F}}'\widetilde{\mathbf{F}}\right)^{-1}\widetilde{\mathbf{F}}'\mathbf{J}_T\mathbf{F}\left(\frac{\Gamma_\beta'\mathbf{X}}{N}\right)\left(\frac{\mathbf{X}'\mathbf{X}}{N}\right)^{-1}\left(\frac{\mathbf{X}'\mathbf{E}}{N}\right)\mathbf{J}_T\widetilde{\mathbf{F}}\left(\widetilde{\mathbf{F}}'\widetilde{\mathbf{F}}\right)^{-1}$$
$$\xrightarrow{p} \mathbf{I}_K\mathbf{0}_{K\times L}\mathbf{V}_X^{-1}\mathbf{0}_{L\times T}\mathbf{J}_T\mathbf{F}\left(\mathbf{F}'\mathbf{J}_T\mathbf{F}\right)^{-1} = \mathbf{0}_{K\times K} \tag{B.19}$$

from Step 4 and Assumption 2(i) and 2(iii) and that

$$\frac{1}{N}m_{(3)}'m_{(3)} = \left(\widetilde{\mathbf{F}}'\widetilde{\mathbf{F}}\right)^{-1}\widetilde{\mathbf{F}}'\mathbf{J}_T\left(\frac{\mathbf{E}'\mathbf{X}}{N}\right)\left(\frac{\mathbf{X}'\mathbf{X}}{N}\right)^{-1}\left(\frac{\mathbf{X}'\mathbf{E}}{N}\right)\mathbf{J}_T\widetilde{\mathbf{F}}\left(\widetilde{\mathbf{F}}'\widetilde{\mathbf{F}}\right)^{-1}$$
$$\xrightarrow{p} \left(\mathbf{F}'\mathbf{J}_T\mathbf{F}\right)^{-1}\mathbf{F}'\mathbf{J}_T\mathbf{0}_{T\times L}\mathbf{V}_X^{-1}\mathbf{0}_{L\times T}\mathbf{J}_T\mathbf{F}\left(\mathbf{F}'\mathbf{J}_T\mathbf{F}\right)^{-1} = \mathbf{0}_{K\times K} \tag{B.20}$$

Finally, plugging the results of (B.15)-(B.20) into (B.14), we have that

$$\frac{1}{N}\left(\widetilde{\mathbf{G}}_\beta\left(\mathbf{X}\right) - \mathbf{G}_\beta\left(\mathbf{X}\right)\right)'\left(\widetilde{\mathbf{G}}_\beta\left(\mathbf{X}\right) - \mathbf{G}_\beta\left(\mathbf{X}\right)\right) \xrightarrow{p} \mathbf{0}_{K\times K}.$$

Step 6: $\widehat{\mathbf{G}}_\beta(\mathbf{X}) = \widetilde{\mathbf{G}}_\beta(\mathbf{X})$: See Lemma B.3(i).

Step 7: $\widehat{\mathbf{G}}_\beta(\mathbf{X}) \xrightarrow{p} \mathbf{G}_\beta(\mathbf{X})$: This follows from Steps 5 and 6. □

Lemma B.5. *Consider $\widehat{\mathbf{G}}_\beta(\mathbf{X})$ defined in Theorem 2.1. Let $\mathbf{Y}$ be a $(N \times m)$ matrix. If $\frac{1}{N}\mathbf{Y}'\mathbf{Y} \xrightarrow{p} \mathbf{V}_Y$, a positive definite matrix, then the probability limit of $\frac{1}{N}\widehat{\mathbf{G}}_\beta(\mathbf{X})'\mathbf{Y}$ is identical to the $\frac{1}{N}\mathbf{G}_\beta(\mathbf{X})'\mathbf{Y}$.*

Proof It suffices to show that $\frac{1}{N}\mathbf{G}_\beta(\mathbf{X})'\mathbf{Y} - \frac{1}{N}\widehat{\mathbf{G}}_\beta(\mathbf{X})'\mathbf{Y} \xrightarrow{p} \mathbf{0}_{K\times m}$. Let $\mathbf{G}_\beta(\mathbf{X})_i$, $\widehat{\mathbf{G}}_\beta(\mathbf{X})_j$ and $\mathbf{Y}_j$ denote the i-th column of $\mathbf{G}_\beta(\mathbf{X})$, the i-th column of $\widehat{\mathbf{G}}_\beta(\mathbf{X})$, the j-th column of $\mathbf{Y}$. Then, the (i,j) element of $\frac{1}{N}\mathbf{G}_\beta(\mathbf{X})'\mathbf{Y} - \frac{1}{N}\widehat{\mathbf{G}}_\beta(\mathbf{X})'\mathbf{Y}$ has the following expression

$$\frac{1}{N}\mathbf{G}_\beta(\mathbf{X})_i'\mathbf{Y}_j - \frac{1}{N}\widehat{\mathbf{G}}_\beta(\mathbf{X})_i'\mathbf{Y}_j = \frac{1}{N}\left(\mathbf{G}_\beta(\mathbf{X})_i - \widehat{\mathbf{G}}_\beta(\mathbf{X})_i\right)'\mathbf{Y}_j.$$

Using the Cauchy–Schwarz inequality, we have that

$$\left(\frac{1}{N}\left(\mathbf{G}_\beta(\mathbf{X})_i - \widehat{\mathbf{G}}_\beta(\mathbf{X})_i\right)'\mathbf{Y}_j\right)^2 \leq \frac{1}{N}\left(\mathbf{G}_\beta(\mathbf{X})_i - \widehat{\mathbf{G}}_\beta(\mathbf{X})_i\right)'\left(\mathbf{G}_\beta(\mathbf{X})_i - \widehat{\mathbf{G}}_\beta(\mathbf{X})_i\right)\left(\frac{1}{N}\mathbf{Y}_j'\mathbf{Y}_j\right).$$

Because $\frac{1}{N}\mathbf{Y}'\mathbf{Y} \xrightarrow{p} \mathbf{V}_Y$, a positive definite matrix, by assumption and Theorem 2.1 says that $\frac{1}{N}\left(\mathbf{G}_\beta(\mathbf{X})_i - \widehat{\mathbf{G}}_\beta(\mathbf{X})_i\right)'\left(\mathbf{G}_\beta(\mathbf{X})_i - \widehat{\mathbf{G}}_\beta(\mathbf{X})_i\right) \xrightarrow{p} 0$, it holds that $\frac{1}{N}\left(\mathbf{G}_\beta(\mathbf{X})_i - \widehat{\mathbf{G}}_\beta(\mathbf{X})_i\right)'\mathbf{Y}_j \xrightarrow{p} 0$. Hence, $\frac{1}{N}\mathbf{G}_\beta(\mathbf{X})'\mathbf{Y} - \frac{1}{N}\widehat{\mathbf{G}}_\beta(\mathbf{X})'\mathbf{Y} \xrightarrow{p} \mathbf{0}_{K\times m}$, completing the proof of the lemma. □

Lemma B.6. *Consider $\widehat{\mathbf{G}}_\beta(\mathbf{X})$ defined in Theorem 2.1. Then, as N increases, $\frac{1}{N}\widehat{\mathbf{G}}_\beta(\mathbf{X})'\mathbf{R} \xrightarrow{p} \mathbf{F}$.*

Proof From Lemma B.5 and Assumption 2(i), it suffices to show that $\frac{1}{N}\mathbf{G}_\beta(\mathbf{X})'\mathbf{R} \xrightarrow{p} \mathbf{F}$. From the expression of $\mathbf{R}$ in (2.3),

$$\frac{\mathbf{G}_\beta(\mathbf{X})'\mathbf{R}}{N} = \left(\frac{\mathbf{G}_\beta(\mathbf{X})'\mathbf{G}_\alpha(\mathbf{X})}{N} + \frac{\mathbf{G}_\beta(\mathbf{X})'\Gamma_\alpha}{N}\right)\mathbf{1}_T' + \left(\frac{\mathbf{G}_\beta(\mathbf{X})'\mathbf{G}_\beta(\mathbf{X})}{N} + \frac{\mathbf{G}_\beta(\mathbf{X})'\Gamma_\beta}{N}\right)\mathbf{F}' + \frac{\mathbf{G}_\beta(\mathbf{X})'\mathbf{E}}{N}.$$

Then, from Assumptions 2(ii), 3(i), and 3(ii), it follows that $\frac{1}{N}\mathbf{G}_\beta(\mathbf{X})'\mathbf{R} \xrightarrow{p} \mathbf{F}$, which in conjunction with Assumption 2(i) and Lemma B.5 completes the proof of the lemma. □

Lemma B.7. *The minimization problem in Theorem 2.2 has the following closed form expression:*

$$\widehat{\boldsymbol{\theta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\overline{\mathbf{R}} - \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\widehat{\mathbf{G}}_\beta(\mathbf{X})\left(\widehat{\mathbf{G}}_\beta(\mathbf{X})'\widehat{\mathbf{G}}_\beta(\mathbf{X})\right)^{-1}\widehat{\mathbf{G}}_\beta(\mathbf{X})'\overline{\mathbf{R}}.$$

Proof We use the following Lagrangian to solve the constrained minimization problem:

$$\min_{\boldsymbol{\theta},\boldsymbol{\lambda}} \left(\overline{\mathbf{R}} - \mathbf{X}\boldsymbol{\theta}\right)' \left(\overline{\mathbf{R}} - \mathbf{X}\boldsymbol{\theta}\right) + \boldsymbol{\lambda}\widehat{\mathbf{G}}_{\beta}\left(\mathbf{X}\right)' \mathbf{X}\boldsymbol{\theta}.$$

The first order conditions give

$$\begin{bmatrix} 2\mathbf{X}'\mathbf{X} & \mathbf{X}'\widehat{\mathbf{G}}_{\beta}\left(\mathbf{X}\right) \\ \widehat{\mathbf{G}}_{\beta}\left(\mathbf{X}\right)'\mathbf{X} & 0 \end{bmatrix} \begin{bmatrix} \widehat{\boldsymbol{\theta}} \\ \lambda \end{bmatrix} = \begin{bmatrix} 2\mathbf{X}'\overline{\mathbf{R}} \\ 0 \end{bmatrix},$$

which yields

$$\begin{bmatrix} \widehat{\boldsymbol{\theta}} \\ \lambda \end{bmatrix} = \begin{bmatrix} 2\mathbf{X}'\mathbf{X} & \mathbf{X}'\widehat{\mathbf{G}}_{\beta}\left(\mathbf{X}\right) \\ \widehat{\mathbf{G}}_{\beta}\left(\mathbf{X}\right)'\mathbf{X} & 0 \end{bmatrix}^{-1} \begin{bmatrix} 2\mathbf{X}'\overline{\mathbf{R}} \\ 0 \end{bmatrix},$$

where the invertibility is guaranteed by Assumption 2(i) and the property of $\mathbf{P}\widehat{\mathbf{G}}_{\beta}\left(\mathbf{X}\right) = \widehat{\mathbf{G}}_{\beta}\left(\mathbf{X}\right)$ in Lemma B.3(ii). Then, standard block matrix inversion gives

$$\widehat{\boldsymbol{\theta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\overline{\mathbf{R}} - \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\widehat{\mathbf{G}}_{\beta}\left(\mathbf{X}\right)\left(\widehat{\mathbf{G}}_{\beta}\left(\mathbf{X}\right)'\widehat{\mathbf{G}}_{\beta}\left(\mathbf{X}\right)\right)^{-1}\widehat{\mathbf{G}}_{\beta}\left(\mathbf{X}\right)'\overline{\mathbf{R}},$$

which completes the proof of the lemma. □

Proof of Theorems 2.2 and 2.3 From Lemmas B.7 and B.3(ii), we have that

$$\widehat{\mathbf{G}}_{\alpha}\left(\mathbf{X}\right) = \mathbf{X}\widehat{\boldsymbol{\theta}} = \mathbf{P}\overline{\mathbf{R}} - \widehat{\mathbf{G}}_{\beta}\left(\mathbf{X}\right)\left(\widehat{\mathbf{G}}_{\beta}\left(\mathbf{X}\right)'\widehat{\mathbf{G}}_{\beta}\left(\mathbf{X}\right)\right)^{-1}\widehat{\mathbf{G}}_{\beta}\left(\mathbf{X}\right)'\overline{\mathbf{R}},$$

which in conjunction with the expression of $\overline{\mathbf{R}}$ in (2.6) yields

$$\widehat{\mathbf{G}}_{\alpha}\left(\mathbf{X}\right) - \mathbf{G}_{\alpha}\left(\mathbf{X}\right) = n_{(1)} + n_{(2)} + n_{(3)},$$

with $n_{(i)}$ for $i = 1, 2, 3$ are given by $n_{(1)} = \mathbf{P}\left(\Gamma_{\alpha} + \Gamma_{\beta}\overline{\mathbf{F}} + \overline{\mathbf{E}}\right)$, $n_{(2)} = \mathbf{G}_{\beta}\left(\mathbf{X}\right)\overline{\mathbf{F}}$ and $n_{(3)} = -\widehat{\mathbf{G}}_{\beta}\left(\mathbf{X}\right)\left(\widehat{\mathbf{G}}_{\beta}\left(\mathbf{X}\right)'\widehat{\mathbf{G}}_{\beta}\left(\mathbf{X}\right)\right)^{-1}\widehat{\mathbf{G}}_{\beta}\left(\mathbf{X}\right)'\overline{\mathbf{R}}$. Then,

$$\frac{1}{N}\left(\widehat{\mathbf{G}}_{\alpha}\left(\mathbf{X}\right) - \mathbf{G}_{\alpha}\left(\mathbf{X}\right)\right)'\left(\widehat{\mathbf{G}}_{\alpha}\left(\mathbf{X}\right) - \mathbf{G}_{\alpha}\left(\mathbf{X}\right)\right) = \sum_{i,j=1}^{3}\frac{1}{N}n_{(i)}'n_{(j)}. \tag{B.21}$$

Note that

$$\frac{1}{N}n_{(1)}'n_{(1)} = \left(\frac{\mathbf{X}'\Gamma_{\alpha}}{N} + \frac{\mathbf{X}'\Gamma_{\beta}}{N}\overline{\mathbf{F}} + \frac{\mathbf{X}'\mathbf{E}}{N}\frac{\mathbf{1}_T}{T}\right)'\left(\frac{\mathbf{X}'\mathbf{X}}{N}\right)^{-1}\left(\frac{\mathbf{X}'\Gamma_{\alpha}}{N} + \frac{\mathbf{X}'\Gamma_{\beta}}{N}\overline{\mathbf{F}} + \frac{\mathbf{X}'\mathbf{E}}{N}\frac{\mathbf{1}_T}{T}\right)$$
$$\xrightarrow{p} \left(\mathbf{0}_L + \mathbf{0}_{L\times K}'\overline{\mathbf{F}} + \mathbf{0}_{L\times T}\frac{\mathbf{1}_T}{T}\right)'\mathbf{V}_X^{-1}\left(\mathbf{0}_L + \mathbf{0}_{L\times K}'\overline{\mathbf{F}} + \mathbf{0}_{L\times T}\frac{\mathbf{1}_T}{T}\right) = 0 \tag{B.22}$$

from Assumptions 2(i) and 2(ii) and that

$$\frac{1}{N}n_{(1)}'n_{(2)} = \left(\frac{\mathbf{G}_{\beta}\left(\mathbf{X}\right)'\Gamma_{\alpha}}{N} + \frac{\mathbf{G}_{\beta}\left(\mathbf{X}\right)'\Gamma_{\beta}}{N}\overline{\mathbf{F}} + \frac{\mathbf{G}_{\beta}\left(\mathbf{X}\right)'\mathbf{E}}{N}\frac{\mathbf{1}_T}{T}\right)'\overline{\mathbf{F}}$$
$$\xrightarrow{p} \left(\mathbf{0}_K + \mathbf{0}_{K\times K}'\overline{\mathbf{F}} + \mathbf{0}_{K\times T}\mathbf{i}\right)'\overline{\mathbf{F}} = 0 \tag{B.23}$$

from Assumption 2(ii) and that

$$\frac{1}{N}n_{(1)}'n_{(3)} = -\left(\frac{\widehat{\mathbf{G}}_{\beta}\left(\mathbf{X}\right)'\Gamma_{\alpha}}{N} + \frac{\widehat{\mathbf{G}}_{\beta}\left(\mathbf{X}\right)'\Gamma_{\beta}}{N}\overline{\mathbf{F}} + \frac{\widehat{\mathbf{G}}_{\beta}\left(\mathbf{X}\right)'\mathbf{E}}{N}\frac{\mathbf{1}_T}{T}\right)'\frac{\widehat{\mathbf{G}}_{\beta}\left(\mathbf{X}\right)'\mathbf{R}}{N}\frac{\mathbf{1}_T}{T}$$
$$\xrightarrow{p} -\left(\mathbf{0}_K + \mathbf{0}_{K\times K}'\overline{\mathbf{F}} + \mathbf{0}_{K\times T}\frac{\mathbf{1}_T}{T}\right)'\overline{\mathbf{F}} = 0 \tag{B.24}$$

from Lemmas B.5 and B.6, and Assumption 2(ii) and that

$$\frac{1}{N}n_{(2)}'n_{(2)} = \overline{\mathbf{F}}'\left(\frac{\mathbf{G}_{\beta}\left(\mathbf{X}\right)'\mathbf{G}_{\beta}\left(\mathbf{X}\right)}{N}\right)\overline{\mathbf{F}} \xrightarrow{p} \overline{\mathbf{F}}'\overline{\mathbf{F}}. \tag{B.25}$$

from Assumption 3(ii) and that

$$\frac{1}{N}n_{(2)}'n_{(3)} = -\overline{\mathbf{F}}'\left(\frac{\mathbf{G}_{\beta}\left(\mathbf{X}\right)'\widehat{\mathbf{G}}_{\beta}\left(\mathbf{X}\right)}{N}\right)\frac{\widehat{\mathbf{G}}_{\beta}\left(\mathbf{X}\right)'\mathbf{R}}{N}\frac{\mathbf{1}_T}{T} \xrightarrow{p} -\overline{\mathbf{F}}'\overline{\mathbf{F}} \tag{B.26}$$

from Lemmas B.5 and B.6, and Assumption 3(ii) and that

$$\frac{1}{N}n_{(3)}'n_{(3)} = \frac{\mathbf{1}_T'}{T}\frac{\mathbf{R}'\widehat{\mathbf{G}}_{\beta}\left(\mathbf{X}\right)}{N}\frac{\widehat{\mathbf{G}}_{\beta}\left(\mathbf{X}\right)'\mathbf{R}}{N}\frac{\mathbf{1}_T}{T} \xrightarrow{p} \overline{\mathbf{F}}'\overline{\mathbf{F}} \tag{B.27}$$

from Lemma B.6. Finally, plugging the results of (B.22)-(B.27) into (B.21), we have that

$$\frac{1}{N}\left(\widehat{\mathbf{G}}_{\alpha}\left(\mathbf{X}\right) - \mathbf{G}_{\alpha}\left(\mathbf{X}\right)\right)'\left(\widehat{\mathbf{G}}_{\alpha}\left(\mathbf{X}\right) - \mathbf{G}_{\alpha}\left(\mathbf{X}\right)\right) \xrightarrow{p} 0, \tag{B.28}$$

which proves Theorem 2.2.

Next, we turn to Theorem 2.3.

$$\widehat{\mathbf{w}}'\mathbf{R} = \mathbf{w}'\mathbf{R} + (\widehat{\mathbf{w}} - \mathbf{w})'\,\mathbf{R}$$

We explain that $\mathbf{wR} \xrightarrow{p} \delta\mathbf{1}_T'$ in the text. Hence, it suffices to show that $(\widehat{\mathbf{w}} - \mathbf{w})'\,\mathbf{R}$ shrinks to zeros. Let $\mathbf{R}_t$ denote the t-th column of $\mathbf{R}$. Using the Cauchy–Schwarz inequality, we have that

$$\begin{aligned}
\left((\widehat{\mathbf{w}} - \mathbf{w})'\,\mathbf{R}_t\right)^2 &\leq (\widehat{\mathbf{w}} - \mathbf{w})'\,(\widehat{\mathbf{w}} - \mathbf{w})\left(\mathbf{R}_t'\mathbf{R}_t\right) \\
&= \frac{\left(\widehat{\mathbf{G}}_\alpha\left(\mathbf{X}\right) - \mathbf{G}_\alpha\left(\mathbf{X}\right)\right)'\left(\widehat{\mathbf{G}}_\alpha\left(\mathbf{X}\right) - \mathbf{G}_\alpha\left(\mathbf{X}\right)\right)}{N} \cdot \frac{\mathbf{R}_t'\mathbf{R}_t}{N} \xrightarrow{p} 0,
\end{aligned}$$

where the last limit is from (B.28) and Assumption 2(i). This completes the proof of Theorem 2.3. □

References

Ahn, Seung C, and Alex R Horenstein, 2013, Eigenvalue ratio test for the number of factors, *Econometrica* 81, 1203–1227.

Bai, Jushan, and Serena Ng, 2002, Determining the number of factors in approximate factor models, *Econometrica* 70, 191–221.

Banz, Rolf W, 1981, The relationship between return and market value of common stocks, *Journal of Financial Economics* 9, 3–18.

Bierens, Herman, 2014, *The Oxford handbook of applied nonparametric and semiparametric econometrics and statistics*, chapter The Hilbert Space Theoretical Foundation of Semi-Nonparametric Modeling, 3–37 (Oxford University Press).

Carhart, Mark M, 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57–82.

Connor, Gregory, Matthias Hagmann, and Oliver Linton, 2012, Efficient semiparametric estimation of the Fama–French model and extensions, *Econometrica* 80, 713–754.

Connor, Gregory, and Robert A Korajczyk, 1993, A test for the number of factors in an approximate factor model, *Journal of Finance* 48, 1263–1291.

Connor, Gregory, and Oliver Linton, 2007, Semiparametric estimation of a characteristic-based factor model of common stock returns, *Journal of Empirical Finance* 14, 694 – 717.

Daniel, Kent, and Sheridan Titman, 1997, Evidence on the characteristics of cross sectional variation in stock returns, *Journal of Finance* 52, 1–33.

Fama, Eugene F, and Kenneth R French, 1992, The cross-section of expected stock returns, *Journal of Finance* 47, 427–465.

Fama, Eugene F, and Kenneth R French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Fama, Eugene F, and Kenneth R French, 1996, Multifactor explanations of asset pricing anomalies, *Journal of Finance* 51, 55–84.

Fama, Eugene F, and Kenneth R French, 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1–22.

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

Ferson, Wayne E., and Campbell R. Harvey, 1994, Sources of risk and expected returns in global equity markets, *Journal of Banking & Finance* 18, 775–803.

Ferson, Wayne E, and Campbell R Harvey, 1997, Fundamental determinants of national equity market returns: A perspective on conditional asset pricing, *Journal of Banking & Finance* 21, 1625 – 1665.

Freyberger, Joachim, Andreas Neuhierl, and Michael Weber, 2017, Dissecting characteristics nonparametrically, *Working Paper* .

Gagliardini, Patrick, Elisa Ossola, and Olivier Scaillet, 2016, Time-varying risk premium in large cross-sectional equity data sets, *Econometrica* 84, 985–1046.

Harvey, Campbell R., Yan Liu, and Heqing Zhu, 2016, ... and the cross-section of expected returns, *Review of Financial Studies* 29, 5–68.

Hou, Kewei, Chen Xue, and Lu Zhang, 2015, Digesting anomalies: An investment approach, *Review of Financial Studies* 28, 660–705.

Kelly, Bryan, Seth Pruitt, and Yinan Su, 2017, Instrumented principal component analysis, *Unpublished Manuscript, University of Chicago* .

Kelly, Bryan, Seth Pruitt, and Yinan Su, 2018, Characteristics are covariances: A unified model of risk and return, *Unpublished Manuscript, University of Chicago* .

Kim, Soohun, and Georgios Skoulakis, 2018a, Ex-post risk premia estimation and asset pricing tests using large cross sections: The regression-calibration approach, *Journal of Econometrics* 204, 159–188.

Kim, Soohun, and Georgios Skoulakis, 2018b, Testing ex-post implications of asset pricing models using individual stocks, *Working Paper* .

Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh, 2018, Interpreting factor models, *The Journal of Finance* 73, 1183–1223.

Linnainmaa, Juhani T, and Michael R Roberts, 2018, The history of the cross-section of stock returns, *The Review of Financial Studies* 31, 2606–2649.

Lintner, John, 1965, The valuation of risky assets and the selection of risky investments in stock portfolios and capital budgets, *The Review of Economics and Statistics* 47, 13–37.

McLean, David R., and Jeffrey Pontiff, 2016, Does academic research destroy return predictability, *Journal of Finance* 71, 5–32.

Merton, Robert C., 1973, An intertemporal capital asset pricing model, *Econometrica* 41, 867–887.

Mossin, Jan, 1966, Equilibrium in a capital asset market, *Econometrica* 34, 768–783.

Newey, Whitney K, and Kenneth D West, 1987, A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica* 55, 703–708.

Rosenberg, Barr, Kenneth Reid, and Ronald Lanstein, 1985, Persuasive evidence of market inefficiency, *The Journal of Portfolio Management* 11, 9–16.

Ross, Stephen A, 1976, The arbitrage theory of capital asset pricing, *Journal of Economic Theory* 13, 341–360.

Sharpe, William F, 1964, Capital asset prices: A theory of market equilibrium under conditions of risk, *The Journal of Finance* 19, 425–442.

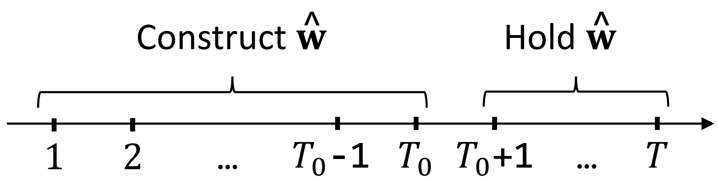
Stattman, Dennis, 1980, Book values and stock returns, *The Chicago MBA: A Journal of Selected Papers* 4, 25–45.

Treynor, Jack, 1962, Toward a theory of market value of risky assets.

Treynor, Jack, 1999, Toward a theory of market value of risky assets, in Robert A. Korajczyk, ed., *Asset Pricing and Portfolio Performance: Models, Strategy and Performance Metrics* (Risk Books).

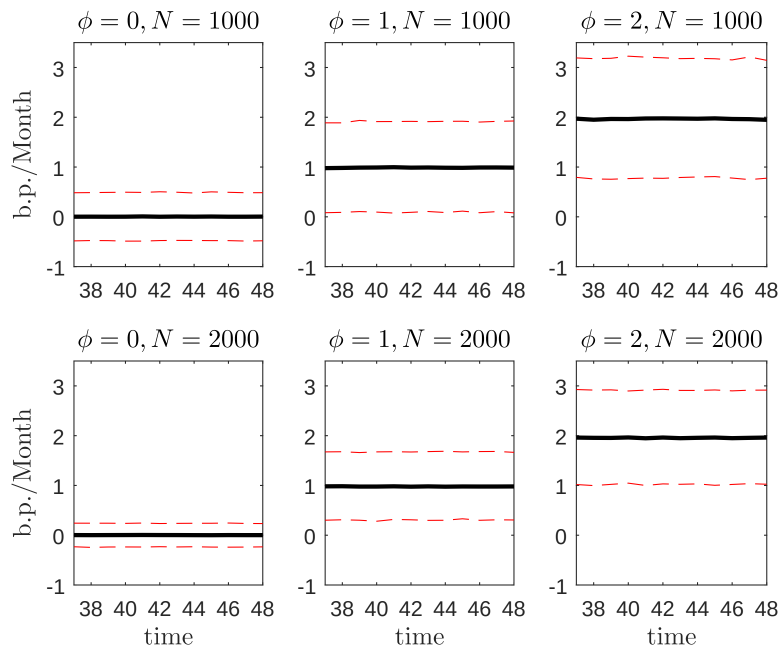
Figures and Tables

Figure 1: Out-of-sample Implementation of the Arbitrage Portfolio



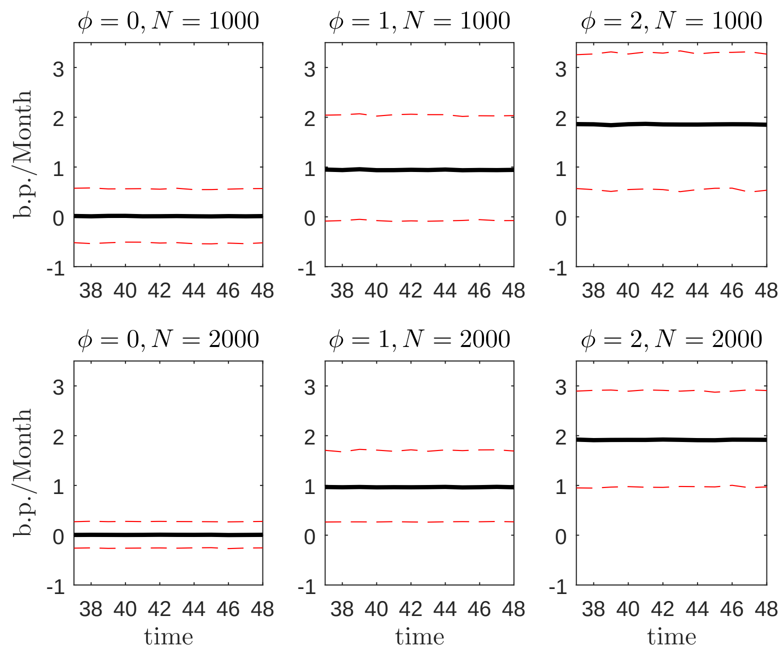
This figure illustrates how to implement the arbitrage portfolio in an out-of-sample manner. We construct $\widehat{\mathbf{w}}$ with the first set of data $t = 1, \cdots, T_0$ and hold the constructed portfolio of $\widehat{\mathbf{w}}$ over the second set of data $t = T_0 + 1, \cdots, T$ in an out-of-sample manner.

Figure 2: Simulated Arbitrage Portfolio Returns in CAPM



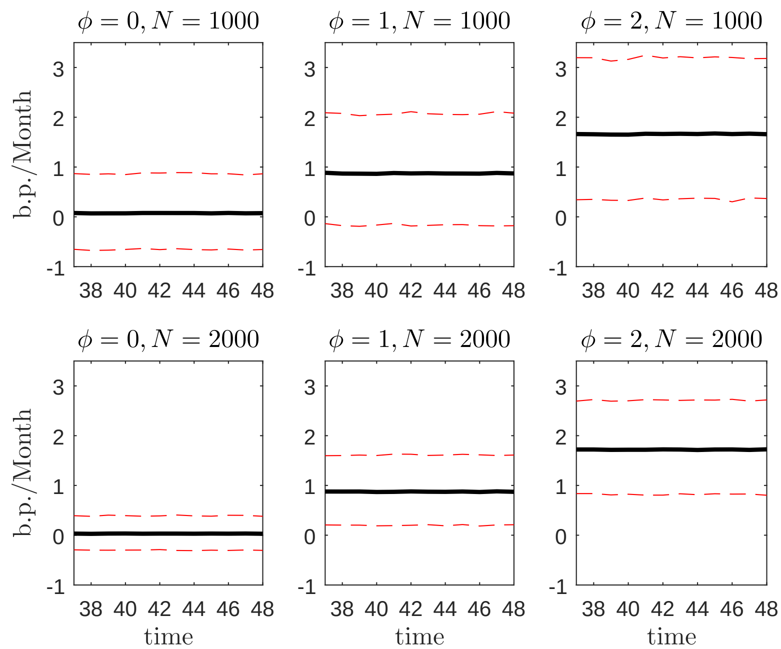
This figure plots the simulation results of the suggested arbitrage portfolio when the return generating process follows CAPM with potential mispricing. The arbitrage portfolio $\widehat{\mathbf{w}}$ is constructed with the returns from $t = 1$ to $t = 36$, and it generates returns from $t = 37$ to $t = 48$. The solid line represents the mean of the arbitrage portfolio returns over the holding periods of $t = 37$ to $t = 48$ from 10,000 repetitions. The dotted lines provide the 95% confidence intervals constructed from the same simulations. We use $K = 1$ for the construction of arbitrage portfolio.

Figure 3: Simulated Arbitrage Portfolio Returns in FF3



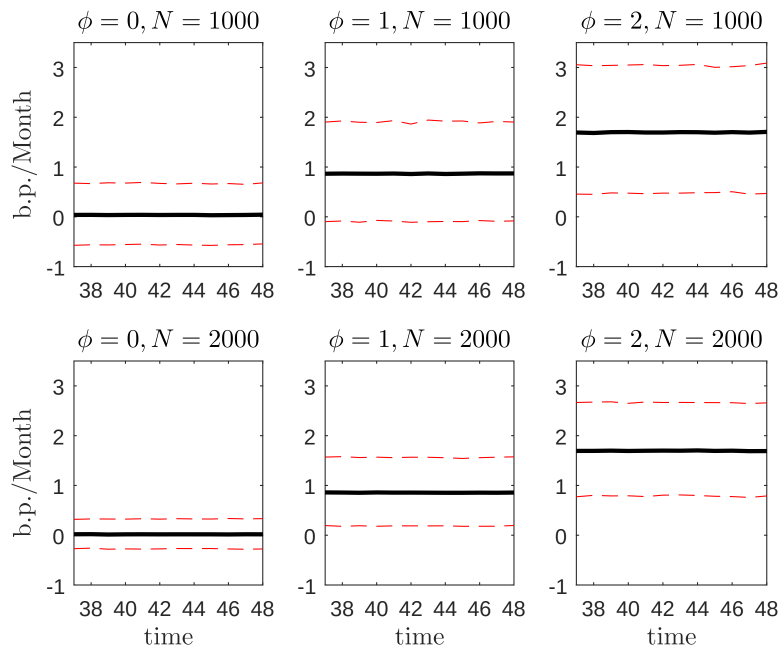
This figure plots the simulation results of the suggested arbitrage portfolio when the return generating process follows FF3 with potential mispricing. The arbitrage portfolio $\widehat{\mathbf{w}}$ is constructed with the returns from $t = 1$ to $t = 36$, and it generates returns from $t = 37$ to $t = 48$. The solid line represents the mean of the arbitrage portfolio returns over the holding periods of $t = 37$ to $t = 48$ from 10,000 repetitions. The dotted lines provide the 95% confidence intervals constructed from the same simulations. We use $K = 3$ for the construction of arbitrage portfolio.

Figure 4: Simulated Arbitrage Portfolio Returns in HXZ4



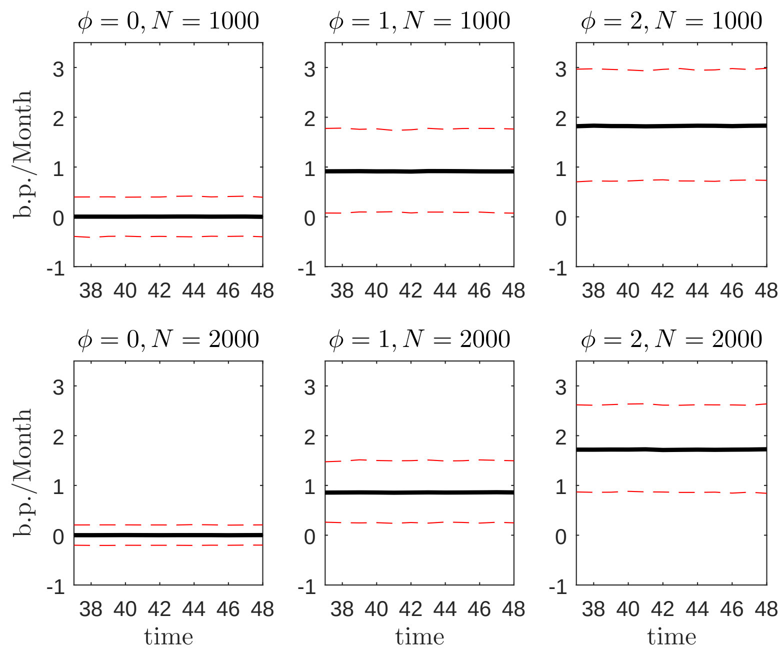
This figure plots the simulation results of the suggested arbitrage portfolio when the return generating process follows HXZ4 with potential mispricing. The arbitrage portfolio $\widehat{\mathbf{w}}$ is constructed with the returns from $t = 1$ to $t = 36$, and it generates returns from $t = 37$ to $t = 48$. The solid line represents the mean of the arbitrage portfolio returns over the holding periods of $t = 37$ to $t = 48$ from 10,000 repetitions. The dotted lines provide the 95% confidence intervals constructed from the same simulations. We use $K = 4$ for the construction of arbitrage portfolio.

Figure 5: Simulated Arbitrage Portfolio Returns in FF5



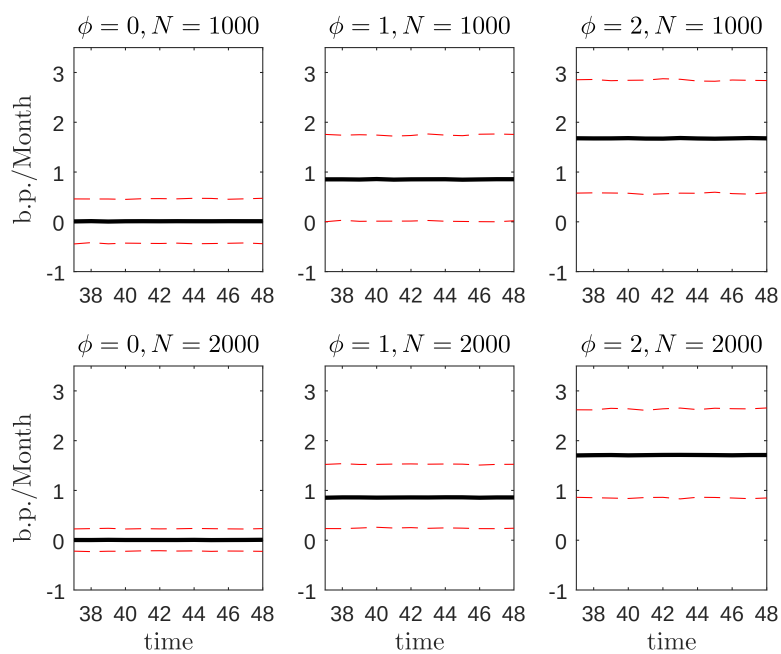
This figure plots the simulation results of the suggested arbitrage portfolio when the return generating process follows FF5 with potential mispricing. The arbitrage portfolio $\widehat{\mathbf{w}}$ is constructed with the returns from $t = 1$ to $t = 36$, and it generates returns from $t = 37$ to $t = 48$. The solid line represents the mean of the arbitrage portfolio returns over the holding periods of $t = 37$ to $t = 48$ from 10,000 repetitions. The dotted lines provide the 95% confidence intervals constructed from the same simulations. We use $K = 5$ for the construction of arbitrage portfolio.

Figure 6: Simulated Arbitrage Portfolio Returns in CAPM with $K = 6$



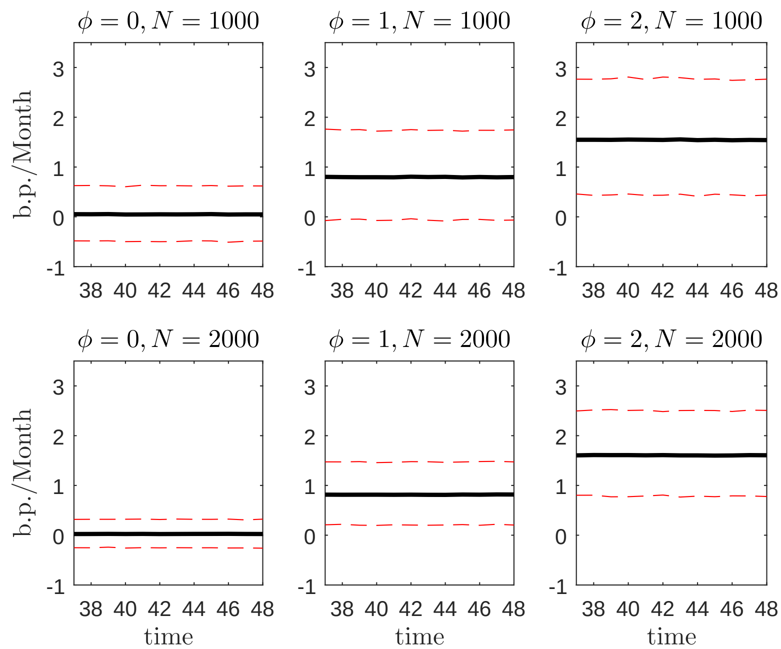
This figure plots the simulation results of the suggested arbitrage portfolio when the return generating process follows CAPM with potential mispricing. The arbitrage portfolio $\widehat{\mathbf{w}}$ is constructed with the returns from $t = 1$ to $t = 36$, and it generates returns from $t = 37$ to $t = 48$. The solid line represents the mean of the arbitrage portfolio returns over the holding periods of $t = 37$ to $t = 48$ from 10,000 repetitions. The dotted lines provide the 95% confidence intervals constructed from the same simulations. We use $K = 6$ for the construction of arbitrage portfolio.

Figure 7: Simulated Arbitrage Portfolio Returns in FF3 with $K = 6$



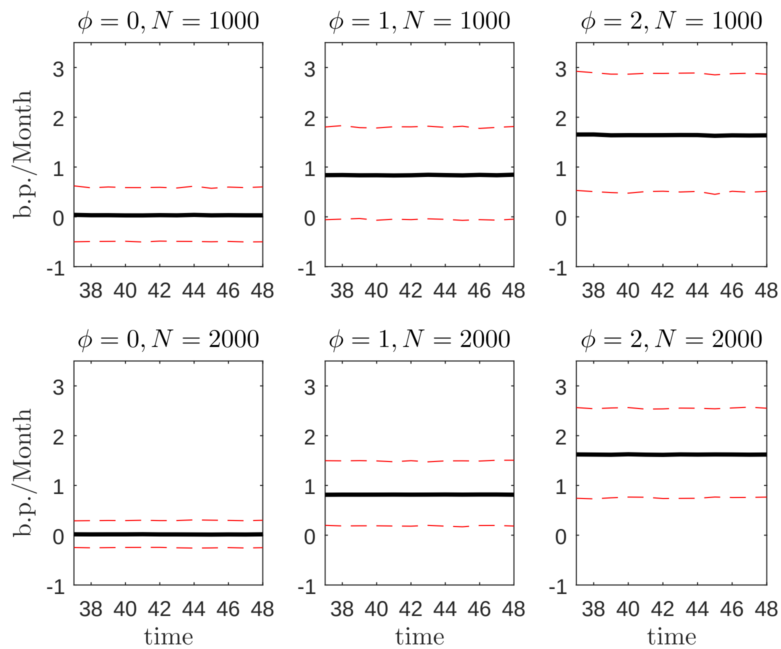
This figure plots the simulation results of the suggested arbitrage portfolio when the return generating process follows FF3 with potential mispricing. The arbitrage portfolio $\widehat{\mathbf{w}}$ is constructed with the returns from $t = 1$ to $t = 36$, and it generates returns from $t = 37$ to $t = 48$. The solid line represents the mean of the arbitrage portfolio returns over the holding periods of $t = 37$ to $t = 48$ from 10,000 repetitions. The dotted lines provide the 95% confidence intervals constructed from the same simulations. We use $K = 6$ for the construction of arbitrage portfolio.

Figure 8: Simulated Arbitrage Portfolio Returns in HXZ4 with $K = 6$



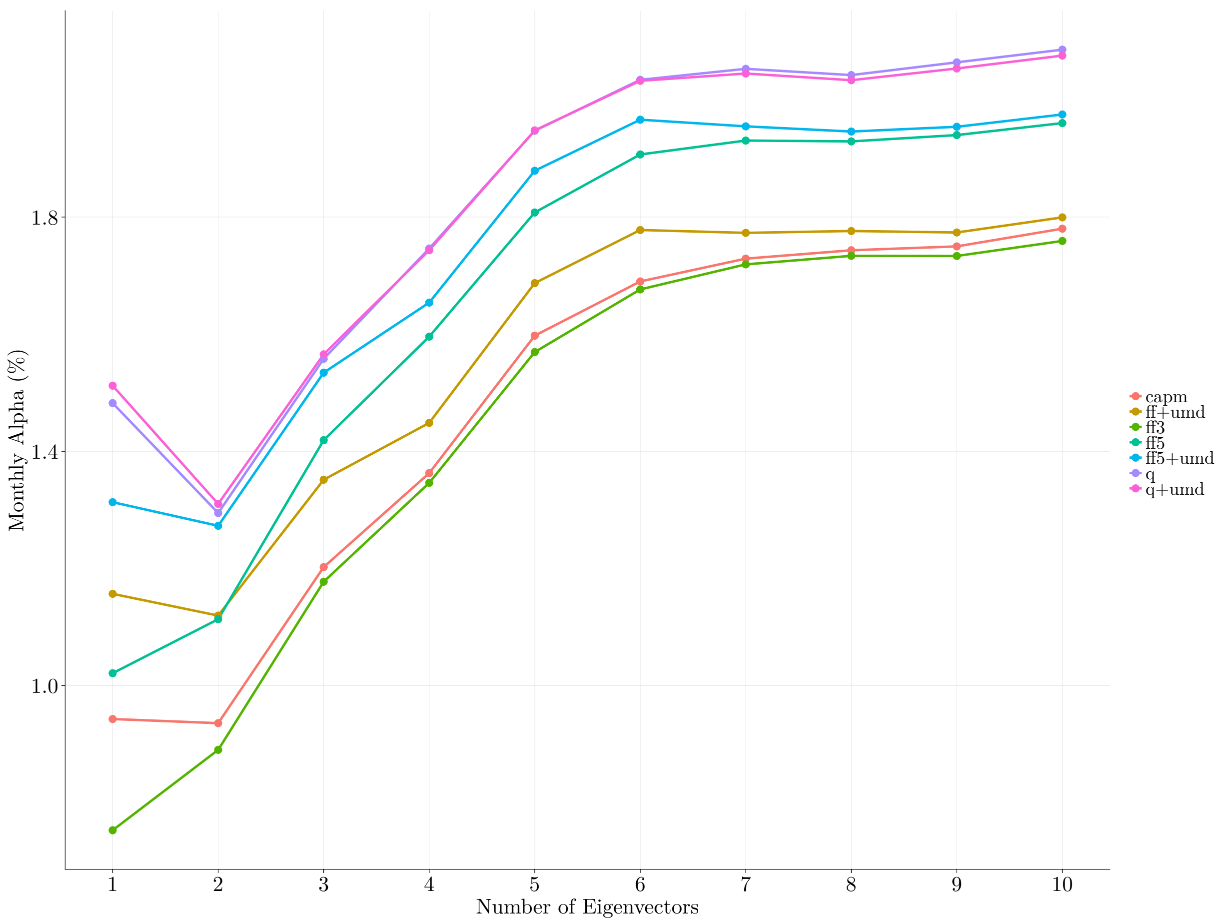
This figure plots the simulation results of the suggested arbitrage portfolio when the return generating process follows HXZ4 with potential mispricing. The arbitrage portfolio $\widehat{\mathbf{w}}$ is constructed with the returns from $t = 1$ to $t = 36$, and it generates returns from $t = 37$ to $t = 48$. The solid line represents the mean of the arbitrage portfolio returns over the holding periods of $t = 37$ to $t = 48$ from 10,000 repetitions. The dotted lines provide the 95% confidence intervals constructed from the same simulations. We use $K = 6$ for the construction of arbitrage portfolio.

Figure 9: Simulated Arbitrage Portfolio Returns in FF5 with $K = 6$



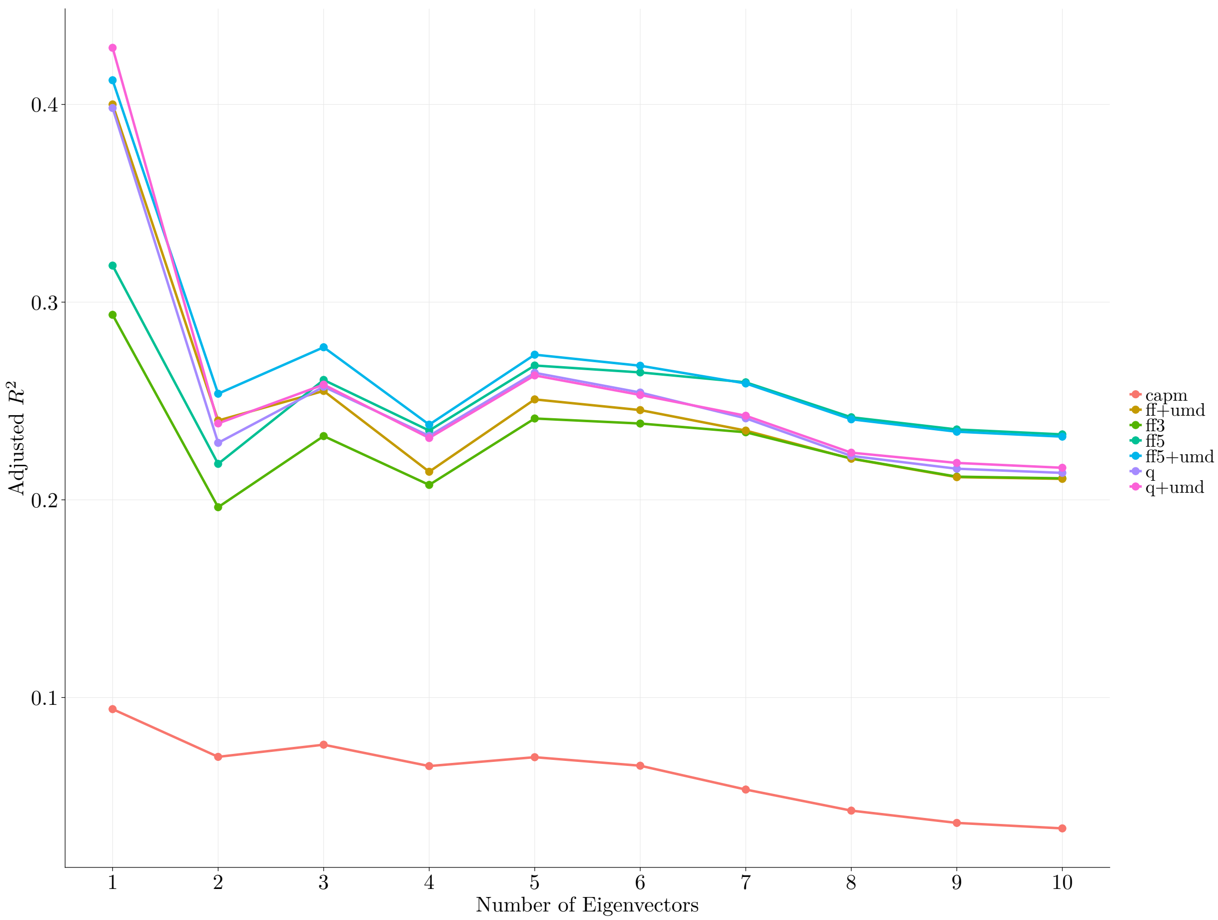
This figure plots the simulation results of the suggested arbitrage portfolio when the return generating process follows FF5 with potential mispricing. The arbitrage portfolio $\widehat{\mathbf{w}}$ is constructed with the returns from $t = 1$ to $t = 36$, and it generates returns from $t = 37$ to $t = 48$. The solid line represents the mean of the arbitrage portfolio returns over the holding periods of $t = 37$ to $t = 48$ from 10,000 repetitions. The dotted lines provide the 95% confidence intervals constructed from the same simulations. We use $K = 6$ for the construction of arbitrage portfolio.

Figure 10: Alpha for Varying the Number of Eigenvectors



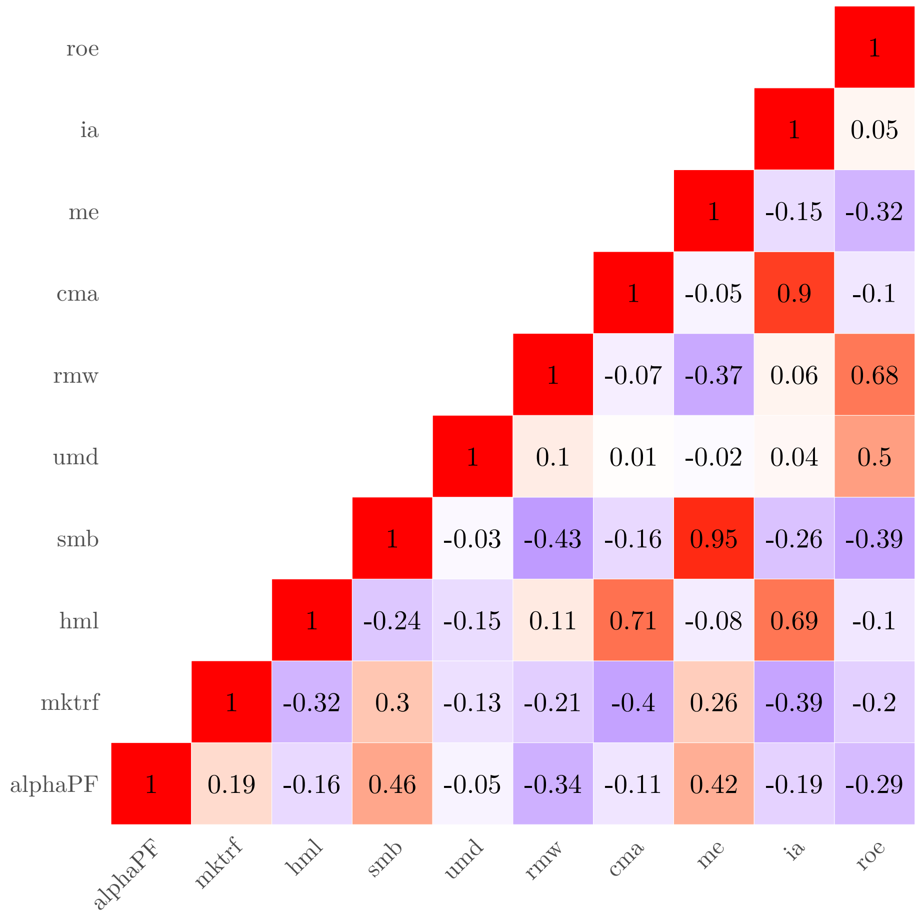
This figure shows the monthly alpha of the mispricing portfolio against the CAPM, the Fama-French three and five factor model, and their "momentum augmented" versions for one through ten eigenvectors. The sample period is from January 1968 to June 2014.

Figure 11: Adjusted R^2 for Varying the Number of Eigenvectors



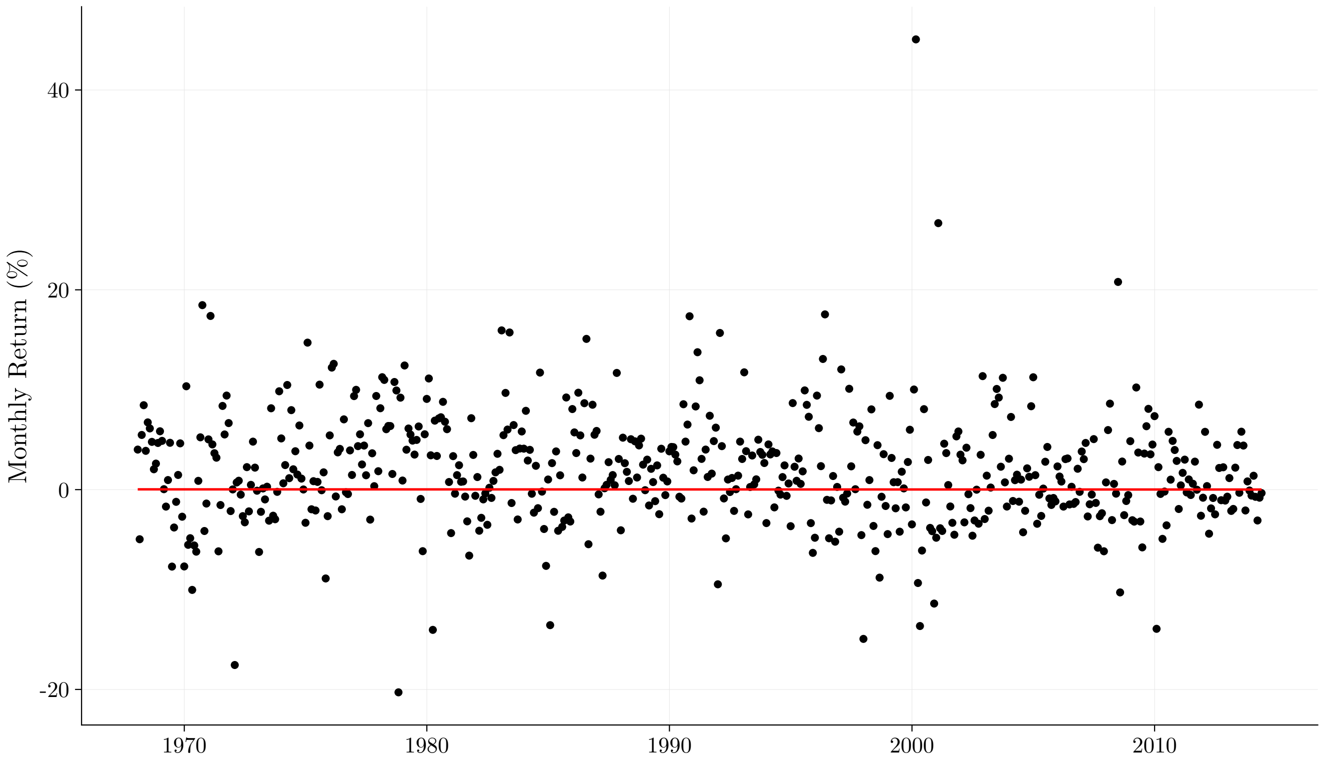
This figure shows the adjusted $R^2$of time series regression of the mispricing portfolio's returns onto common risk factors for one through ten eigenvectors. The sample period is January 1968 to June 2014.

Figure 12: Correlation Matrix with Common Factors



This figure shows the correlation matrix between the mispricing portfolio (alphaPF) and the Fama-French three and five factors as well as the momentum factor. The sample period is January 1968 to June 2014.

Figure 13: Monthly Returns of the Arbitrage Portfolio 1968 - 2014



This figure shows the monthy excess returns of the mispricing portfolio from 1968 through 2014 and a linear time trend (red). It is apparent that there is no economically meaningful decline in monthly excess returns.

Table 1: Firm Characteristics by Category

	Past-returns:			**Value:**	
(1)	r_{2-1}	Return 1 month before prediction	(33)	A2ME	Total assets to Size
(2)	r_{6-2}	Return from 6 to 2 months before prediction	(34)	BEME	Book to market ratio
(3)	r_{12-2}	Return from 12 to 2 months before prediction	(35)	$BEME_{adj}$	BEME - mean BEME in Fama-French 48 industry
(4)	r_{12-7}	Return from 12 to 7 months before prediction	(36)	C	Cash to AT
(5)	r_{36-13}	Return from 36 to 13 months before prediction	(37)	C2D	Cash flow to total liabilities
			(38)	ΔSO	Log change in split-adjusted shares outstanding
	Investment:		(39)	Debt2P	Total debt to Size
(6)	Investment	% change in AT	(40)	E2P	Income before extraordinary items to Size
(7)	ΔCEQ	% change in BE	(41)	Free CF	Free cash flow to BE
(8)	ΔPI2A	Change in PP&E and inventory over lagged AT	(42)	LDP	Trailing 12-months dividends to price
(9)	ΔShrout	% change in shares outstanding	(43)	NOP	Net payouts to Size
(10)	IVC	Change in inventory over average AT	(44)	O2P	Operating payouts to market cap
(11)	NOA	Net-operating assets over lagged AT	(45)	Q	Tobin's Q
			(46)	S2P	Sales to price
	Profitability:		(47)	Sales_g	Sales growth
(12)	ATO	Sales to lagged net operating assets			
(13)	CTO	Sales to lagged total assets		**Trading frictions:**	
(14)	Δ(ΔGM-ΔSales)	Δ(% change in gross margin and % change in sales)	(48)	AT	Total assets
(15)	EPS	Earnings per share	(49)	Beta	Correlation $\times$ ratio of vols
(16)	IPM	Pre-tax income over sales	(50)	Beta daily	CAPM beta using daily returns
(17)	PCM	Sales minus costs of goods sold to sales	(51)	DTO	De-trended Turnover - market Turnover
(18)	PM	OI after depreciation over sales	(52)	Idio vol	Idio vol of Fama-French 3 factor model
(19)	PM_adj	Profit margin - mean PM in Fama-French 48 industry	(53)	LME	Price times shares outstanding
(20)	Prof	Gross profitability over BE	(54)	LME_adj	Size - mean size in Fama-French 48 industry
(21)	RNA	OI after depreciation to lagged net operating assets	(55)	Lturnover	Last month's volume to shares outstanding
(22)	ROA	Income before extraordinary items to lagged AT	(56)	Rel_to_high_price	Price to 52 week high price
(23)	ROC	Size + longterm debt - total assets to cash	(57)	Ret_max	Maximum daily return
(24)	ROE	Income before extraordinary items to lagged BE	(58)	Spread	Average daily bid-ask spread
(25)	ROIC	Return on invested capital	(59)	Std turnover	Standard deviation of daily turnover
(26)	S2C	Sales to cash	(60)	Std volume	Standard deviation of daily volume
(27)	SAT	Sales to total assets	(61)	SUV	Standard unexplained volume
(28)	SAT_adj	SAT - mean SAT in Fama-French 48 industry	(62)	Total vol	Standard deviation of daily returns
	Intangibles:				
(29)	AOA	Absolute value of operating accruals			
(30)	OL	Costs of goods solds + SG&A to total assets			
(31)	Tan	Tangibility			
(32)	OA	Operating accruals			

This is a reproduction of Table 1 in Freyberger et al. (2017). It lists the characteristics we consider in our empirical analysis by category. We refer to their online appendix for a precise definition of these variables and their construction in conventional dataset (CRSP, Computstat). The sample period is January 1965 to June 2014.

Table 2: Portfolio Performance Statistics

# Eigenvectors	Mean (%)	Standard Deviation (%)	Sharpe Ratio	Skewness	Kurtosis	Worst Month (%)	Best Month (%)
1	13.61	20.47	0.67	1.50	8.50	-16.09	34.50
2	12.97	17.93	0.72	0.73	6.24	-18.17	25.97
3	16.31	18.62	0.88	1.00	8.50	-20.60	37.60
4	18.13	18.94	0.96	0.78	8.07	-19.93	38.39
5	20.98	18.65	1.12	0.99	9.29	-20.53	41.23
6	22.04	18.66	1.18	0.87	8.68	-21.11	39.83
7	22.36	18.86	1.19	0.98	10.46	-22.20	43.75
8	22.40	19.37	1.16	0.93	10.07	-21.87	43.74
9	22.38	19.43	1.15	0.88	10.05	-22.05	44.25
10	22.70	19.55	1.16	0.94	10.17	-20.29	45.07

This table reports annualized percentage means, annualized percentage standard deviations, annualized Sharpe Ratios, skewness, kurtosis, the best month and worst month return. The mispricing portfolio with one through ten eigenvectors is estimated every month using the steps outlined in Section 4. The sample period is January 1968 to June 2014.

Table 3: Risk Adjusted Returns with One Eigenvector

	CAPM	FF3	FF3+UMD	FF5	FF5+UMD	q	q+UMD
(Intercept)	0.94^{***}	0.75^{***}	1.16^{***}	1.02^{***}	1.31^{***}	1.48^{***}	1.51^{***}
	(0.24)	(0.18)	(0.19)	(0.22)	(0.22)	(0.25)	(0.22)
mktrf	0.40^{***}	0.26^{***}	0.18^{***}	0.20^{***}	0.15^{**}		
	(0.08)	(0.08)	(0.06)	(0.07)	(0.07)		
smb		0.89^{***}	0.88^{***}	0.77^{***}	0.78^{***}		
		(0.11)	(0.13)	(0.11)	(0.12)		
hml		0.25^{*}	0.10	0.43^{***}	0.20^{*}		
		(0.13)	(0.11)	(0.17)	(0.11)		
umd			-0.46^{***}		-0.43^{***}		-0.28^{***}
			(0.10)		(0.09)		(0.09)
rmw				-0.48^{***}	-0.37^{***}		
				(0.12)	(0.13)		
cma				-0.42^{*}	-0.20		
				(0.25)	(0.18)		
mkt						0.17^{**}	0.16^{**}
						(0.07)	(0.07)
me						0.56^{***}	0.63^{***}
						(0.11)	(0.12)
ia						-0.17	-0.16
						(0.15)	(0.13)
roe						-0.93^{***}	-0.68^{***}
						(0.14)	(0.12)
Adj. R^2	0.09	0.29	0.40	0.32	0.41	0.40	0.43
Num. obs.	557	557	557	557	557	557	557

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

This table reports alphas (%/month) and factor loadings on the Fama and French (1993), Carhart (1997), Fama and French (2015) and the *q*-factor model by Hou et al. (2015). The mispricing portfolio with one eigenvector is estimated every month using the steps outlined in Section 4. Newey and West (1987) standard errors are given in parentheses. The sample period is January 1968 to June 2014.

Table 4: Risk Adjusted Returns with Two Eigenvectors

	CAPM	FF3	FF3+UMD	FF5	FF5+UMD	q	q+UMD
(Intercept)	0.94^{***}	0.89^{***}	1.12^{***}	1.11^{***}	1.27^{***}	1.29^{***}	1.31^{***}
	(0.22)	(0.18)	(0.19)	(0.21)	(0.20)	(0.23)	(0.21)
mktrf	0.30^{***}	0.17^{**}	0.12^{*}	0.12	0.09		
	(0.08)	(0.08)	(0.07)	(0.08)	(0.07)		
smb		0.62^{***}	0.61^{***}	0.52^{***}	0.53^{***}		
		(0.13)	(0.14)	(0.13)	(0.13)		
hml		0.00	−0.08	0.15	0.02		
		(0.12)	(0.10)	(0.14)	(0.11)		
umd			-0.26^{***}		-0.24^{***}		−0.15
			(0.08)		(0.08)		(0.09)
rmw				-0.40^{***}	-0.34^{***}		
				(0.11)	(0.11)		
cma				-0.34^{*}	−0.23		
				(0.19)	(0.16)		
mkt						0.13^{**}	0.12^{*}
						(0.07)	(0.06)
me						0.42^{***}	0.46^{***}
						(0.13)	(0.14)
ia						-0.27^{*}	-0.26^{*}
						(0.15)	(0.14)
roe						-0.50^{***}	-0.36^{***}
						(0.13)	(0.14)
Adj. R^2	0.07	0.20	0.24	0.22	0.25	0.23	0.24
Num. obs.	557	557	557	557	557	557	557

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

This table reports alphas (%/month) and factor loadings on the Fama and French (1993), Carhart (1997), Fama and French (2015) and the *q*-factor model by Hou et al. (2015). The mispricing portfolio with two eigenvectors is estimated every month using the steps outlined in Section 4. Newey and West (1987) standard errors are given in parentheses. The sample period is January 1968 to June 2014.

Table 5: Risk Adjusted Returns with Three Eigenvectors

	CAPM	FF3	FF3+UMD	FF5	FF5+UMD	q	q+UMD
(Intercept)	1.20^{***}	1.18^{***}	1.35^{***}	1.42^{***}	1.53^{***}	1.56^{***}	1.57^{***}
	(0.23)	(0.20)	(0.21)	(0.22)	(0.22)	(0.23)	(0.23)
mktrf	0.33^{***}	0.17^{**}	0.13^{*}	0.12	0.10		
	(0.08)	(0.08)	(0.07)	(0.08)	(0.07)		
smb		0.71^{***}	0.70^{***}	0.58^{***}	0.59^{***}		
		(0.15)	(0.16)	(0.13)	(0.14)		
hml		−0.05	−0.12	0.08	−0.02		
		(0.12)	(0.11)	(0.14)	(0.11)		
umd			-0.20^{**}		-0.17^{**}		−0.07
			(0.09)		(0.08)		(0.10)
rmw				-0.49^{***}	-0.44^{***}		
				(0.11)	(0.13)		
cma				−0.30	−0.22		
				(0.19)	(0.16)		
mkt						0.13^{*}	0.13^{*}
						(0.07)	(0.07)
me						0.53^{***}	0.54^{***}
						(0.15)	(0.15)
ia						-0.34^{**}	-0.33^{**}
						(0.15)	(0.15)
roe						-0.46^{***}	-0.40^{***}
						(0.12)	(0.14)
Adj. R^2	0.08	0.23	0.26	0.26	0.28	0.26	0.26
Num. obs.	557	557	557	557	557	557	557

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

This table reports alphas (%/month) and factor loadings on the Fama and French (1993), Carhart (1997), Fama and French (2015) and the *q*-factor model by Hou et al. (2015). The mispricing portfolio with three eigenvectors is estimated every month using the steps outlined in Section 4. Newey and West (1987) standard errors are given in parentheses. The sample period is January 1968 to June 2014.

Table 6: Risk Adjusted Returns with Four Eigenvectors

	CAPM	FF3	FF3+UMD	FF5	FF5+UMD	q	q+UMD
(Intercept)	1.36^{***}	1.35^{***}	1.45^{***}	1.60^{***}	1.65^{***}	1.75^{***}	1.74^{***}
	(0.24)	(0.21)	(0.22)	(0.22)	(0.23)	(0.24)	(0.25)
mktrf	0.31^{***}	0.15^{*}	0.13^{*}	0.10	0.09		
	(0.07)	(0.08)	(0.07)	(0.07)	(0.07)		
smb		0.68^{***}	0.68^{***}	0.56^{***}	0.56^{***}		
		(0.15)	(0.16)	(0.14)	(0.14)		
hml		−0.07	−0.11	0.08	0.03		
		(0.13)	(0.12)	(0.14)	(0.12)		
umd			−0.12		−0.09		0.03
			(0.09)		(0.09)		(0.10)
rmw				-0.48^{***}	-0.46^{***}		
				(0.12)	(0.13)		
cma				-0.34^{*}	-0.30^{*}		
				(0.19)	(0.17)		
mkt						0.11	0.11
						(0.07)	(0.07)
me						0.51^{***}	0.50^{***}
						(0.16)	(0.15)
ia						-0.41^{***}	-0.41^{***}
						(0.15)	(0.15)
roe						-0.44^{***}	-0.46^{***}
						(0.12)	(0.13)
Adj. R^2	0.07	0.21	0.21	0.24	0.24	0.23	0.23
Num. obs.	557	557	557	557	557	557	557

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

This table reports alphas (%/month) and factor loadings on the Fama and French (1993), Carhart (1997), Fama and French (2015) and the *q*-factor model by Hou et al. (2015). The mispricing portfolio with four eigenvectors is estimated every month using the steps outlined in Section 4. Newey and West (1987) standard errors are given in parentheses. The sample period is January 1968 to June 2014.

Table 7: Risk Adjusted Returns with Six Eigenvectors

	CAPM	FF3	FF3+UMD	FF5	FF5+UMD	q	q+UMD
(Intercept)	1.69^{***}	1.68^{***}	1.78^{***}	1.91^{***}	1.97^{***}	2.03^{***}	2.03^{***}
	(0.24)	(0.22)	(0.24)	(0.23)	(0.24)	(0.26)	(0.26)
mktrf	0.30^{***}	0.13^{*}	0.11	0.09	0.08		
	(0.08)	(0.07)	(0.07)	(0.07)	(0.07)		
smb		0.74^{***}	0.74^{***}	0.62^{***}	0.62^{***}		
		(0.13)	(0.13)	(0.11)	(0.12)		
hml		-0.08	-0.12	0.04	-0.01		
		(0.11)	(0.11)	(0.13)	(0.13)		
umd			-0.11		-0.09		0.02
			(0.09)		(0.08)		(0.09)
rmw				-0.47^{***}	-0.45^{***}		
				(0.12)	(0.14)		
cma				-0.28	-0.24		
				(0.19)	(0.17)		
mkt						0.10	0.10
						(0.07)	(0.07)
me						0.57^{***}	0.57^{***}
						(0.14)	(0.13)
ia						-0.40^{**}	-0.40^{**}
						(0.17)	(0.16)
roe						-0.40^{***}	-0.42^{***}
						(0.13)	(0.14)
Adj. R^2	0.07	0.24	0.25	0.26	0.27	0.25	0.25
Num. obs.	557	557	557	557	557	557	557

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

This table reports alphas (%/month) and factor loadings on the Fama and French (1993), Carhart (1997), Fama and French (2015) and the *q*-factor model by Hou et al. (2015). The mispricing portfolio with six eigenvectors is estimated every month using the steps outlined in Section 4. Newey and West (1987) standard errors are given in parentheses. The sample period is January 1968 to June 2014.

Table 8: Risk Adjusted Returns with Seven Eigenvectors

	CAPM	FF3	FF3+UMD	FF5	FF5+UMD	q	q+UMD
(Intercept)	1.73***	1.72***	1.77***	1.93***	1.95***	2.05***	2.05***
	(0.25)	(0.22)	(0.23)	(0.24)	(0.25)	(0.27)	(0.27)
mktrf	0.28***	0.10	0.09	0.06	0.06		
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)		
smb		0.76***	0.76***	0.64***	0.64***		
		(0.13)	(0.14)	(0.11)	(0.12)		
hml		−0.09	−0.11	−0.01	−0.03		
		(0.11)	(0.11)	(0.12)	(0.12)		
umd			−0.06		−0.04		0.08
			(0.09)		(0.08)		(0.10)
rmw				−0.47***	−0.46***		
				(0.14)	(0.14)		
cma				−0.20	−0.18		
				(0.18)	(0.17)		
mkt						0.07	0.08
						(0.07)	(0.07)
me						0.60***	0.58***
						(0.15)	(0.13)
ia						−0.41**	−0.41**
						(0.17)	(0.17)
roe						−0.37***	−0.44***
						(0.14)	(0.15)
Adj. R^2	0.05	0.23	0.24	0.26	0.26	0.24	0.24
Num. obs.	557	557	557	557	557	557	557

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

This table reports alphas (%/month) and factor loadings on the Fama and French (1993), Carhart (1997), Fama and French (2015) and the *q*-factor model by Hou et al. (2015). The mispricing portfolio with seven eigenvectors is estimated every month using the steps outlined in Section 4. Newey and West (1987) standard errors are given in parentheses. The sample period is January 1968 to June 2014.

Table 9: Risk Adjusted Returns with Eight Eigenvectors

	CAPM	FF3	FF3+UMD	FF5	FF5+UMD	q	q+UMD
(Intercept)	1.74***	1.73***	1.78***	1.93***	1.95***	2.04***	2.03***
	(0.25)	(0.23)	(0.23)	(0.25)	(0.25)	(0.27)	(0.28)
mktrf	0.26***	0.08	0.07	0.04	0.04		
	(0.08)	(0.07)	(0.07)	(0.07)	(0.07)		
smb		0.78***	0.77***	0.66***	0.66***		
		(0.13)	(0.14)	(0.11)	(0.12)		
hml		−0.10	−0.11	−0.03	−0.04		
		(0.11)	(0.12)	(0.12)	(0.12)		
umd			−0.05		−0.03		0.08
			(0.09)		(0.09)		(0.10)
rmw				−0.45***	−0.44***		
				(0.14)	(0.14)		
cma				−0.17	−0.16		
				(0.18)	(0.17)		
mkt						0.05	0.06
						(0.07)	(0.07)
me						0.62***	0.60***
						(0.15)	(0.13)
ia						−0.40**	−0.40**
						(0.18)	(0.18)
roe						−0.34**	−0.42***
						(0.15)	(0.15)
Adj. R^2	0.04	0.22	0.22	0.24	0.24	0.22	0.22
Num. obs.	557	557	557	557	557	557	557

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

This table reports alphas (%/month) and factor loadings on the Fama and French (1993), Carhart (1997), Fama and French (2015) and the *q*-factor model by Hou et al. (2015). The mispricing portfolio with eight eigenvectors is estimated every month using the steps outlined in Section 4. Newey and West (1987) standard errors are given in parentheses. The sample period is January 1968 to June 2014.

Table 10: Risk Adjusted Returns with Nine Eigenvectors

	CAPM	FF3	FF3+UMD	FF5	FF5+UMD	q	q+UMD
(Intercept)	1.75^{***}	1.73^{***}	1.77^{***}	1.94^{***}	1.95^{***}	2.06^{***}	2.05^{***}
	(0.25)	(0.23)	(0.24)	(0.25)	(0.26)	(0.27)	(0.28)
mktrf	0.24^{***}	0.06	0.05	0.02	0.02		
	(0.08)	(0.07)	(0.07)	(0.07)	(0.07)		
smb		0.78^{***}	0.77^{***}	0.65^{***}	0.65^{***}		
		(0.14)	(0.14)	(0.12)	(0.12)		
hml		-0.08	-0.10	-0.01	-0.02		
		(0.11)	(0.12)	(0.12)	(0.12)		
umd			-0.05		-0.02		0.10
			(0.10)		(0.09)		(0.10)
rmw				-0.48^{***}	-0.47^{***}		
				(0.14)	(0.14)		
cma				-0.17	-0.16		
				(0.18)	(0.17)		
mkt						0.04	0.04
						(0.07)	(0.07)
me						0.60^{***}	0.58^{***}
						(0.15)	(0.13)
ia						-0.38^{**}	-0.38^{**}
						(0.19)	(0.19)
roe						-0.38^{**}	-0.48^{***}
						(0.15)	(0.15)
Adj. R^2	0.04	0.21	0.21	0.24	0.23	0.22	0.22
Num. obs.	557	557	557	557	557	557	557

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

This table reports alphas (%/month) and factor loadings on the Fama and French (1993), Carhart (1997), Fama and French (2015) and the *q*-factor model by Hou et al. (2015). The mispricing portfolio with nine eigenvectors is estimated every month using the steps outlined in Section 4. Newey and West (1987) standard errors are given in parentheses. The sample period is January 1968 to June 2014.

Table 11: Risk Adjusted Returns with Ten Eigenvectors

	CAPM	FF3	FF3+UMD	FF5	FF5+UMD	q	q+UMD
(Intercept)	1.78^{***}	1.76^{***}	1.80^{***}	1.96^{***}	1.98^{***}	2.09^{***}	2.08^{***}
	(0.25)	(0.23)	(0.24)	(0.25)	(0.26)	(0.27)	(0.28)
mktrf	0.23^{***}	0.05	0.05	0.02	0.01		
	(0.08)	(0.07)	(0.07)	(0.07)	(0.07)		
smb		0.79^{***}	0.79^{***}	0.66^{***}	0.66^{***}		
		(0.14)	(0.14)	(0.12)	(0.12)		
hml		-0.07	-0.09	-0.00	-0.02		
		(0.11)	(0.12)	(0.13)	(0.13)		
umd			-0.05		-0.02		0.10
			(0.10)		(0.09)		(0.10)
rmw				-0.46^{***}	-0.46^{***}		
				(0.14)	(0.14)		
cma				-0.17	-0.16		
				(0.19)	(0.17)		
mkt						0.03	0.03
						(0.07)	(0.07)
me						0.62^{***}	0.59^{***}
						(0.15)	(0.13)
ia						-0.37^{**}	-0.37^{**}
						(0.19)	(0.19)
roe						-0.38^{**}	-0.47^{***}
						(0.15)	(0.16)
Adj. R^2	0.03	0.21	0.21	0.23	0.23	0.21	0.22
Num. obs.	557	557	557	557	557	557	557

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

This table reports alphas (%/month) and factor loadings on the Fama and French (1993), Carhart (1997), Fama and French (2015) and the *q*-factor model by Hou et al. (2015). The mispricing portfolio with ten eigenvectors is estimated every month using the steps outlined in Section 4. Newey and West (1987) standard errors are given in parentheses. The sample period is January 1968 to June 2014.

Table 12: Alphas For Second Order Legendre Polynomials

# Eigenvectors	CAPM	FF3	FF3+UMD	FF5	FF5+UMD	q	q+UMD
1	0.92	0.73	1.08	1.00	1.26	1.43	1.45
2	0.83	0.79	0.94	1.01	1.11	1.19	1.19
3	1.08	1.04	1.16	1.29	1.36	1.45	1.45
4	1.12	1.10	1.08	1.33	1.30	1.47	1.45
5	1.30	1.28	1.29	1.51	1.49	1.64	1.62
6	1.42	1.40	1.40	1.60	1.58	1.74	1.72
7	1.41	1.40	1.34	1.58	1.52	1.70	1.68
8	1.37	1.37	1.27	1.53	1.44	1.63	1.60
9	1.32	1.33	1.22	1.49	1.38	1.58	1.55
10	1.24	1.25	1.12	1.40	1.29	1.49	1.45

This table reports alphas (%/month) against Fama and French (1993), Carhart (1997), Fama and French (2015) and the q-factor model by Hou et al. (2015). The mispricing is constructed using one through ten eigenvectors. It is estimated every month using the steps outlined in Section 2. The sample period is January 1968 to June 2014.

Table 13: Alphas For Third Order Legendre Polynomials

# Eigenvectors	CAPM	FF3	FF3+UMD	FF5	FF5+UMD	q	q+UMD
1	0.86	0.69	0.98	0.93	1.15	1.32	1.34
2	0.81	0.80	0.87	0.99	1.03	1.14	1.13
3	0.99	0.97	1.01	1.18	1.19	1.31	1.29
4	1.02	1.04	0.92	1.25	1.14	1.30	1.27
5	1.14	1.16	1.04	1.35	1.24	1.41	1.37
6	1.30	1.31	1.19	1.49	1.38	1.56	1.52
7	1.27	1.27	1.13	1.45	1.32	1.52	1.48
8	1.20	1.24	1.06	1.41	1.26	1.44	1.40
9	1.17	1.20	1.01	1.38	1.21	1.40	1.36
10	1.12	1.14	0.94	1.30	1.13	1.33	1.29

This table reports alphas (%/month) against Fama and French (1993), Carhart (1997), Fama and French (2015) and the q-factor model by Hou et al. (2015). The mispricing is constructed using one through ten eigenvectors. It is estimated every month using the steps outlined in Section 2. The sample period is January 1968 to June 2014.

Table 14: Alphas For Fourth Order Legendre Polynomials

# Eigenvectors	CAPM	FF3	FF3+UMD	FF5	FF5+UMD	q	q+UMD
1	0.86	0.69	0.95	0.94	1.12	1.29	1.30
2	0.80	0.77	0.83	0.96	0.99	1.10	1.08
3	0.95	0.91	0.95	1.11	1.12	1.25	1.23
4	0.98	1.00	0.85	1.19	1.06	1.22	1.19
5	1.07	1.08	0.95	1.26	1.14	1.31	1.28
6	1.21	1.21	1.09	1.38	1.27	1.45	1.41
7	1.17	1.17	1.02	1.33	1.20	1.40	1.36
8	1.13	1.15	0.98	1.30	1.16	1.35	1.31
9	1.08	1.08	0.89	1.24	1.08	1.28	1.24
10	1.06	1.06	0.87	1.21	1.04	1.25	1.21

This table reports alphas (%/month) against Fama and French (1993), Carhart (1997), Fama and French (2015) and the q-factor model by Hou et al. (2015). The mispricing is constructed using one through ten eigenvectors. It is estimated every month using the steps outlined in Section 2. The sample period is January 1968 to June 2014.

Table 15: Adjusted R^2 For Second Order Legendre Polynomials

# Eigenvectors	CAPM	FF3	FF3+UMD	FF5	FF5+UMD	q	q+UMD
1	0.06	0.27	0.35	0.30	0.36	0.37	0.38
2	0.04	0.16	0.17	0.18	0.19	0.19	0.18
3	0.05	0.18	0.19	0.21	0.21	0.21	0.21
4	0.03	0.16	0.16	0.18	0.18	0.17	0.18
5	0.04	0.17	0.17	0.19	0.19	0.18	0.19
6	0.03	0.18	0.18	0.19	0.19	0.18	0.19
7	0.02	0.18	0.18	0.19	0.19	0.17	0.19
8	0.02	0.17	0.17	0.18	0.19	0.15	0.18
9	0.01	0.17	0.17	0.19	0.19	0.15	0.18
10	0.01	0.17	0.18	0.19	0.20	0.15	0.18

This table reports R^2 in the regression of our arbitrage portfolio on the factors by Fama and French (1993), Carhart (1997), Fama and French (2015) and the q-factor model by Hou et al. (2015). The mispricing is constructed using one through ten eigenvectors. It is estimated every month using the steps outlined in Section 2. The sample period is January 1968 to June 2014.

Table 16: Adjusted R^2 For Third Order Legendre Polynomials

# Eigenvectors	CAPM	FF3	FF3+UMD	FF5	FF5+UMD	q	q+UMD
1	0.04	0.24	0.29	0.27	0.31	0.32	0.32
2	0.02	0.13	0.13	0.15	0.15	0.14	0.15
3	0.03	0.15	0.15	0.17	0.17	0.16	0.17
4	0.01	0.13	0.13	0.15	0.16	0.12	0.16
5	0.02	0.13	0.14	0.15	0.16	0.13	0.16
6	0.01	0.14	0.15	0.16	0.17	0.13	0.16
7	0.01	0.15	0.16	0.17	0.18	0.13	0.17
8	0.01	0.15	0.16	0.17	0.19	0.13	0.18
9	0.01	0.15	0.16	0.17	0.19	0.13	0.17
10	0.00	0.15	0.16	0.16	0.18	0.13	0.17

This table reports R^2 in the regression of our arbitrage portfolio on the factors by Fama and French (1993), Carhart (1997), Fama and French (2015) and the q-factor model by Hou et al. (2015). The mispricing is constructed using one through ten eigenvectors. It is estimated every month using the steps outlined in Section 2. The sample period is January 1968 to June 2014.

Table 17: Adjusted R^2 For Fourth Order Legendre Polynomials

# Eigenvectors	CAPM	FF3	FF3+UMD	FF5	FF5+UMD	q	q+UMD
1	0.03	0.22	0.26	0.25	0.28	0.29	0.29
2	0.02	0.12	0.12	0.14	0.14	0.14	0.14
3	0.02	0.14	0.14	0.16	0.16	0.15	0.16
4	0.01	0.12	0.13	0.14	0.15	0.10	0.15
5	0.01	0.12	0.13	0.14	0.15	0.12	0.15
6	0.01	0.14	0.15	0.16	0.16	0.13	0.16
7	0.01	0.15	0.16	0.17	0.18	0.13	0.17
8	0.01	0.15	0.16	0.17	0.18	0.14	0.17
9	0.01	0.15	0.16	0.16	0.18	0.13	0.17
10	0.00	0.15	0.16	0.16	0.18	0.13	0.17

This table reports R^2 in the regression of our arbitrage portfolio on the factors by Fama and French (1993), Carhart (1997), Fama and French (2015) and the q-factor model by Hou et al. (2015). The mispricing is constructed using one through ten eigenvectors. It is estimated every month using the steps outlined in Section 2. The sample period is January 1968 to June 2014.

Table 18: Risk Adjusted Returns with Five Eigenvectors over the subsample period from January 1968 to December 1990

	CAPM	FF3	FF3+UMD	FF5	FF5+UMD	q	q+UMD
(Intercept)	2.20***	2.10***	2.01***	2.20***	2.12***	2.19***	2.20***
	(0.35)	(0.32)	(0.34)	(0.34)	(0.36)	(0.41)	(0.42)
mktrf	0.19*	0.05	0.05	0.03	0.02		
	(0.10)	(0.08)	(0.08)	(0.09)	(0.09)		
smb		0.75***	0.78***	0.74***	0.76***		
		(0.20)	(0.19)	(0.19)	(0.18)		
hml		0.14	0.17	0.37**	0.40**		
		(0.13)	(0.14)	(0.16)	(0.17)		
umd			0.09		0.10		0.09
			(0.14)		(0.14)		(0.17)
rmw				0.01	−0.02		
				(0.29)	(0.28)		
cma				−0.44	−0.48		
				(0.33)	(0.31)		
mkt						0.03	0.03
						(0.09)	(0.09)
me						0.71***	0.71***
						(0.20)	(0.19)
ia						−0.15	−0.19
						(0.26)	(0.26)
roe						−0.01	−0.09
						(0.18)	(0.22)
Adj. R^2	0.03	0.18	0.18	0.18	0.18	0.16	0.16
Num. obs.	276	276	276	276	276	276	276

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

This table reports alphas (%/month) and factor loadings on the Fama and French (1993), Carhart (1997), Fama and French (2015) and the q-factor model by Hou et al. (2015). The mispricing portfolio with ten eigenvectors is estimated every month using the steps outlined in Section 4. Newey and West (1987) standard errors are given in parentheses. The sample period is January 1968 to December 1990.

Table 19: Risk Adjusted Returns with Five Eigenvectors over the subsample period from January 1991 to June 2014

	CAPM	FF3	FF3+UMD	FF5	FF5+UMD	q	q+UMD
(Intercept)	0.92***	0.93***	1.11***	1.17***	1.29***	1.40***	1.41***
	(0.27)	(0.25)	(0.23)	(0.29)	(0.23)	(0.29)	(0.29)
mktrf	0.48***	0.31***	0.23***	0.23***	0.18**		
	(0.07)	(0.08)	(0.08)	(0.08)	(0.08)		
smb		0.70***	0.72***	0.54***	0.58***		
		(0.17)	(0.18)	(0.12)	(0.12)		
hml		−0.22	−0.28**	−0.10	−0.21		
		(0.14)	(0.12)	(0.16)	(0.14)		
umd			−0.22***		−0.21***		−0.06
			(0.07)		(0.07)		(0.10)
rmw				−0.45**	−0.39**		
				(0.17)	(0.18)		
cma				−0.07	0.02		
				(0.20)	(0.16)		
mkt						0.14*	0.14
						(0.09)	(0.09)
me						0.53***	0.56***
						(0.16)	(0.16)
ia						−0.49***	−0.51***
						(0.17)	(0.16)
roe						−0.61***	−0.54***
						(0.14)	(0.17)
Adj. R^2	0.14	0.36	0.39	0.38	0.41	0.41	0.41
Num. obs.	281	281	281	281	281	281	281

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

This table reports alphas (%/month) and factor loadings on the Fama and French (1993), Carhart (1997), Fama and French (2015) and the *q*-factor model by Hou et al. (2015). The mispricing portfolio with ten eigenvectors is estimated every month using the steps outlined in Section 4. Newey and West (1987) standard errors are given in parentheses. The sample period is January 1991 to June 2014.