Arbitrage Portfolios*

Soohun Kim† Robert A. Korajczyk‡ Andreas Neuhierl§

April 15, 2019

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 characteristics' predictive power before any attribution to abnormal returns. We apply the methodology in simulated factor economies and to a large panel of U.S. stock returns from 1965–2014. The methodology works well in simulation and when applied to U.S. stocks. Empirically, we find the arbitrage portfolio has (statistically and economically) significant alphas relative to several popular asset pricing models and annualized Sharpe ratios ranging from 1.35 to 1.75.

*We are thankful to Dong-Hyun Ahn, Demir Bektic, Sylvain Champonnois, Zhuo Chen, Ian Dew-Becker, Alex Horenstein, Ravi Jagannathan, Andrew Karolyi, Yuan Liao, Hongyi Liu, Seth Pruitt, Andrea Tamoni, Olivier Scaillet, Gustavo Schwenkler, Michael Weber, Dacheng Xiu, Chu Zhang, Guofu Zhou, conference and seminar participants at Boston University, the European Winter Finance Summit, Northwestern University, Rutgers University, University of Chicago, University of Notre Dame, the Consortium on Factor Investing, the New Developments in Factor Investing Conference at Imperial College, the UW Madison Junior Finance Conference, and the 2018 New Zealand Finance Meeting. 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, Mendoza College of Business, Notre Dame, IN 46556. E-Mail: aneuhier@nd.edu.

1 Introduction

Many variables have shown some ability to predict the cross section of asset returns. This predictive power could be due to their ability to predict the cross section of systematic risk (beta); their ability to predict asset mispricing (alpha); and spurious cross-sectional relations due to overfitting (data snooping). Rosenberg and McKibben (1973) use 32 stock-level characteristics to predict the cross-section of systematic risk and find a significant relation with a number of characteristics common in the subsequent literature, such as asset size, book-to-market equity, share turnover, and a measure of quality. Betas on portfolios formed using firm-level characteristics have had much success in explaining the cross-section of returns (e.g., Fama and French (1993, 2015); Hou et al. (2015)). Daniel and Titman (1997) argue that it is difficult to disentangle a purely characteristics-based model (in which characteristics only predict alpha) from a risk-based model because the characteristics and factor loadings in the characteristic-sorted portfolios are collinear. In their influential approach to disentangling the beta vs. alpha explanations, assets are sorted into portfolios based on lagged beta estimates and firm characteristics. Returns on long–short portfolios made of long and short legs with similar beta exposure but different levels of the characteristics are designed to measure the pure returns to the characteristics. Similarly, returns on long–short portfolios made of portfolios with similar levels of the characteristic but different levels of beta exposure are designed to measure the pure risk premium. They find significant characteristic-based returns, controlling for betas, but not for beta-based returns, controlling for characteristics. These results rekindled the beta vs. alpha debate.

An issue with the double–sorting procedure arises when the true risk measures are related to firm characteristics. Regression-based estimates of systematic risk are often very noisy, and potentially stale, estimates of the true systematic risk. This may lead to the characteristic predicting returns, holding estimated betas constant, not because the characteristics predict abnormal returns, but because the characteristics are better predictors of beta (Ferson and Harvey (1997) and Berk (2000)). Regression estimates of systematic risks are known to be relatively imprecise. Furthermore, the issue of staleness of the estimates is somewhat inescapable because the estimates are usually backward-looking functions of unconditional covariances and variances. For example, 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 all correlated with leverage changes, commonly used characteristics (such as market capitalization, book-to-market equity ratios, and momentum) might help predict conditional betas, over and above the predictive power of unconditional betas. In addition to (a) the issue of staleness, double sorting has the disadvantage that (b) the approach handles one characteristic at a time and, hence, is unable to analyze many characteristics simultaneously and (c) sorting into portfolios may mask important variation in returns relative to using individual assets.

We propose a new methodology, which is an extension of the projected principal components procedure (PPCA) of Fan et al. (2016). The estimator can accommodate many characteristics simultaneously; can use individual assets, rather than portfolios; and conditions systematic risk estimates on current values of firm characteristics. Thus, the method addresses all three issues raised above. Our procedure gives characteristics maximal explanatory power for risk premia before we attribute any explanatory power to alphas.[1] We project time-series demeaned asset returns (which eliminates alpha) onto the characteristics (or, potentially, onto the series expansions of the characteristics). We then estimate the relation between factor betas and characteristics by applying principal components (PCA) to the projected returns. Given the estimated systematic factor loading function, we extract the relation between alpha and the characteristics that has maximal cross-sectional explanatory power, conditional on being orthogonal to the systematic factor loadings.

To illustrate the issue of characteristics versus noisy/stale estimates of beta and highlight the advantage of our approach over the double-sorting method, we simulate a simple economy in which the Capital Asset Pricing Model (CAPM) holds. Alpha, or abnormal returns, are identically zero, but the true underlying betas are functions, cross-sectionally, of a firm characteristic. The economy is simulated for 2,000 firms and 2,000 months. We perform month-by-month rolling sorts of assets based on OLS estimates of market betas (estimated over the previous sixty months) and the characteristic. We report average returns of double-sorted (first on characteristic and then on the estimated beta) portfolios in Table 1 (full details about the simulation are in the

[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 survive arbitrage activities by investors.

table legend). Although the true return generating process is the CAPM, the return differences of the high-minus-low characteristic portfolios (reported in the last row) are statistically significant while the return differences of the high-minus-low estimated beta portfolios (reported in the last column) are insignificant. Thus, the table seems to be indicating a strong relation between the characteristic and abnormal returns in an economy in which no abnormal returns exist. In contrast, when we apply our procedure (described fully below) to this economy, we find that the relation between abnormal returns and the characteristic is insignificantly different from zero (p-value of 0.82).

We also show that when there exists any relation between alpha and characteristics, one can use our method to construct an arbitrage portfolio that exploits such a relation. Our arbitrage portfolio weights are proportional to the estimated alpha function. We first apply our estimator in simulation and explore its finite sample properties as well as robustness to model misspecification. The estimator performs well in simulated factor economies, which we calibrate to mimic the CRSP/Compustat panel.

We apply the procedure to U.S. stock return data using the characteristics data set of Freyberger et al. (2019). In the baseline implementation, we use 12 months of data to estimate the weights of the arbitrage portfolio and then hold the portfolio for one month.[2] We then roll the estimation forward by one period and repeat the process. Therefore, we obtain portfolio returns that are out-of-sample relative to the estimation period, in the sense that the arbitrage portfolio weights for period t only use information from periods prior to t. The arbitrage portfolio has (statistically and economically) significant alphas relative to several popular asset pricing models and annualized Sharpe ratios ranging from 1.35 to 1.75 (depending on the number of systematic factors we estimate).

One possible way that data snooping could creep into the analysis is through the selection of firm characteristics, which may be based on studies that use data over the same sample period used to estimate the portfolio weights. As a check for this, we test for a trend in alpha over our sample period. Data snooping would lead us to expect a trend toward zero. We do find a slight downward trend, but it is economically inconsequential.

Our approach allows us to make a number of contributions to empirical asset pric-

[2]We also provide the robustness of our results when we use 24 or 36 months to estimate the weights of the arbitrage portfolio. We also show that the results are robust to alternative rebalancing periods. See Tables A.9-A.13 in the Appendix.

ing. First, we provide useful guidance in portfolio construction for investors who want to eliminate exposure to the common risks and focus on exploiting the mispricing 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)).[3] Our approach incorporates the cross-sectional predictive power of asset characteristics for factor 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. The "arbitrage" notion in our arbitrage portfolios is that we are constructing portfolios that hedge out the systematic risk associated with firm characteristics. In the limit, as the number of assets approaches infinity, the risk of the portfolio should approach zero. We do not assume that there are necessarily arbitrage opportunities. That is an empirical question. In the simulated economy above, there are no arbitrage opportunities, and our procedure applied to those data correctly finds no evidence of arbitrage opportunities. We are, ex ante, agnostic about whether the data support a purely beta-based explanation, a purely alpha-based explanation, or a combination of both. Our goal is to develop a procedure that does a good job disentangling these two effects. Our simulation results suggest that it does. The the empirical results using U.S. stock return data imply that the cross-sectional predictability is due to both beta- and alpha-effects.

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). There is a large literature that relates observable firm characteristics to expected returns, over and above those implied by extant asset pricing models. Early contributions to this literature were made by Banz (1981) (market capitalization), Stattman (1980) and Rosenberg et al. (1985) (book-to-market equity ratio), and Fama and French (1992) who provide an early synthesis of findings across multiple characteristics. The explanatory

[3] See Chen et al. (2018) for the extension of Daniel and Titman (1997) on various characteristics.

power of firm characteristics has led to alternative specifications of asset pricing models (e.g., Fama and French (1993, 1996)) and further testing of the ability of characteristics to explain the cross section of returns beyond that implied by the expanded set of asset pricing models. The recent meta study by Harvey et al. (2016) provides an extensive overview of many of the variables (coined the "zoo of new factors" by Cochrane (2011)) that the literature has produced and also raises important statistical concerns related to multiple hypothesis testing. After these influential papers, numerous efforts to systematically reduce the dimension of the cross sectional return predictors and have been undertaken, e.g. Freyberger et al. (2019), Feng et al. (2019) or Han et al. (2018).

A large portion of the earlier empirical literature works at the portfolio level. That is, rather than using individual assets to test models, researchers group assets into portfolios and conduct tests on these portfolios. Due to concerns about masking pricing errors by portfolio grouping, Connor and Korajczyk (1988) test the CAPM and a latent factor version of the APT using a large cross section of individual assets. Their tests assume that idiosyncratic correlations are non-zero only for firms in the same three-digit SIC code. Gagliardini et al. (2016) and Chaieb et al. (2018) also stress that the "pre-grouping" possibly masks important variation in alphas and betas 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 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 without the need to specify the factors, ex ante.

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 a similar methodology, instrumented principal component analysis (IPCA). Our work is closely related to that of 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, while we find that characteristics explain both risk and mispricing.

It is important to clarify the differences in economic questions between this paper and Kelly et al. (2018). Our focus is on identifying and utilizing both the cross-sectional and temporal relation of characteristics to risk or mispricing. Hence, we use the characteristics at the beginning of each estimation sub-interval of short horizon (of one year in our empirical work) to estimate the cross-sectional relation between alphas, betas, and characteristics but allow the cross-sectional relation to vary across sub-intervals. We apply the identified cross-sectional relation to the most recently observed characteristics to construct our portfolio weights. In contrast, Kelly et al. (2018) allow the characteristics to change period by period but hold the cross-sectional relation between characteristics and either risk or alpha *constant*. The dynamics in our procedure are primarily coming from changes in the cross-sectional relation between alphas, betas, and characteristics, along with updating 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. This would be the case if risk premia vary over time or if anomalies are arbitraged away after discovery. The IPCA 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 is stable over time. We also apply IPCA to form out-of-sample arbitrage portfolios using data over a short time interval in simulated economies and find the abnormal returns on the arbitrage portfolio to be noisier than those from our procedure.[4]

The rest of the paper is organized as follows. In Section 2, we describe our large cross-sectional economy and propose an estimator of arbitrage portfolio weights. 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 an 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 the U.S. equity market and provide evidence that our arbitrage portfolio indeed generates strong profitability

[4]This result does not mean that their method is deficient. Their asymptotic theory is based on large T. However, we intentionally design the simulation setup for small T to justify our theoretical results and empirical applications.

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 stable for short blocks of time (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 (plus risk premium) 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 $\mathbf{R}$ 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 $\mathbf{B}$ is $\boldsymbol{\beta}_i'$, the t-th row of the $(T \times K)$ matrix $\mathbf{F}$ is $\mathbf{f}_t'$=$[f_{1,t} \ \cdots \ f_{K,t}]$, and the (i, t) element of the $(N \times T)$ matrix $\mathbf{E}$ 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 nonparametric function of firm characteristics and estimate the model with the restriction that mispricing is zero, 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." Γ_α and Γ_β represent the sources of alpha and beta that are not related to the characteristics, $\mathbf{X}$. While the mispricing function, $\mathbf{G}_\alpha(\mathbf{X})$ and factor loading function, $\mathbf{G}_\beta(\mathbf{X})$, can be consistently estimated in the large N/small T setting used here, consistent estimates of Γ_α and Γ_β are not obtainable. Therefore, our procedure does not attempt to exploit the gammas, just their orthogonality to the characteristics. There are a number of ways in which one could incorporate non-linearity into the mispricing and factor loading functions. We chose $\mathbf{X}$ to be a large set of characteristics, possibly containing suitable polynomials of some 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:

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

First, we can learn about alpha and beta through $\mathbf{G}_\alpha(\mathbf{X})$ and $\mathbf{G}_\beta(\mathbf{X})$ even when data are relatively infrequently observed (such as monthly) over short horizon (such as a year) by instrumenting characteristics. This is a strong advantage over other factor extraction methods requiring large time series or high frequency observations. Second, because we set T as a short horizon, the process in (2.3) can be treated as a local approximation as an unconditional model of a conditional model over a long horizon model.[5] Third, our rolling estimation of (2.3) enables us to study the *temporal* relation

[5]We thank Yuan Liao for pointing out this. Our approach also works under smooth transition of $\mathbf{X}$ over a short horizon.

of characteristics to risk or mispricing. Many empirical approaches (e.g. Kelly et al. (2018), Ferson and Harvey (1999), Ghysels (1998)) construct conditional model by allowing the characteristics to change period-by-period but holding the cross-sectional relation between characteristics and either risk or alpha *constant*, which is not suitable for detecting anomalies that are arbitraged away after discovery.[6] By estimating (2.3) over rolling-windows, we can learn about the dynamics of $\mathbf{G}_{\alpha}(\mathbf{X})$ and $\mathbf{G}_{\beta}(\mathbf{X})$. Lastly, we do not need to necessarily have all important characteristics for risk and mispricing (2.3). Because any information in the missing characteristics is captured by Γ_{α} and Γ_{β}, our model already incorporates the possibility of misspecifying the set of characteristics. Hence, if some important characteristics are missing, we may lose some precision but will not generate spurious alpha.

Note that the Arbitrage Pricing Theory (APT, Ross (1976)) implies that the sum of squared pricing errors is finite, so 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 $0 \leq \frac{\mathbf{G}_{\alpha}(\mathbf{X})'\mathbf{G}_{\alpha}(\mathbf{X})}{N} \leq \frac{1}{N}\boldsymbol{\alpha}'\boldsymbol{\alpha}$, since $\frac{1}{N}\boldsymbol{\alpha}'\boldsymbol{\alpha}$ also involves $\frac{1}{N}\boldsymbol{\Gamma}_{\boldsymbol{\alpha}}'\boldsymbol{\Gamma}_{\boldsymbol{\alpha}}$. Allowing for significant mispricing of assets implies the cross-sectional average of the squared mispricing function $\mathbf{G}_{\alpha}(\mathbf{X})$ may be nonzero:

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

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

The above assumption specifies that the characteristics in $\mathbf{X}$ may contain information about nontrivial levels of asset mispricing, $\boldsymbol{\alpha}$. It is beyond the scope of this paper to examine the underlying cause of such a relation.[7] Assumption 1 does not imply that characteristics capture all potential mispricing. Mispricing orthogonal to the characteristics is reflected in Γ_{α}. The main objective of this paper is to provide 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 $\mathbf{B}$ allows us to form portfolios that yield abnormal returns (if $\delta > 0$) while hedging out the systematic risk associated with the firm characteristics.

The following are standard regularity conditions on the characteristics and residual

[6] Besides, Kothari and Shanken (1992) and Grundy and Martin (2001) theoretically verify that the relation between some firm characteristics and risk are guaranteed to be dynamic.

[7] See Jagannathan and Wang (2007), Baker and Wurgler (2006), Stambaugh and Yuan (2016), Frazzini and Pedersen (2014) among many for potential causes of mispricing.

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 define 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.[8] 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

[8] 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.

a standard 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 estimator to not only estimate factors, but also the mispricing function, a case not covered in 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 Asymptotic Principal Components (APC) to demeaned projected returns, (Connor and Korajczyk (1986)). By demeaning the returns, we focus purely on systematic risk not on expected returns or realized premiums. In the second step, we estimate $\mathbf{G}_\alpha(\mathbf{X})$ by regressing (in the cross-section) 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 premiums from systematic risks, we extract the information about the mispricing by imposing orthogonality to the systematic risks. In the third step, we use the estimated $\mathbf{G}_\alpha(\mathbf{X})$ to construct an arbitrage portfolio.

We define the convergence of large dimensional matrices as follows.

Definition. For two $(N \times m)$ random 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 $\frac{1}{N}(\mathbf{A}-\mathbf{B})'(\mathbf{A}-\mathbf{B}) \xrightarrow{p} \mathbf{0}_{m\times m}$.

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 both by $\mathbf{G}_\beta(\mathbf{X})$ and $\mathbf{G}_\alpha(\mathbf{X})$. We eliminate the effect of $\mathbf{G}_\alpha(\mathbf{X})$, by demeaning the observed returns:

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

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 by the projection matrix $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$.

Then, we get

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

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 $\mathbf{X}$. Due to the orthogonality of Γ_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, as in Fan et al. (2016), we estimate $\mathbf{G}_\beta(\mathbf{X})$ by applying standard principal component analysis to $\widehat{\mathbf{R}}$.

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, as N increases, $\widehat{\mathbf{G}}_\beta(\mathbf{X}) \xrightarrow{p} \mathbf{G}_\beta(\mathbf{X})$.*

To provide some intuition for the result, recall that $\widehat{\mathbf{R}}$ converges (as $N \to \infty$) to $\mathbf{G}_\beta(\mathbf{X})\mathbf{F}'\mathbf{J}_T$. Therefore, $\frac{\widehat{\mathbf{R}}\widehat{\mathbf{R}}'}{N}$ converges to $\frac{\mathbf{G}_\beta(\mathbf{X})}{\sqrt{N}}\mathbf{F}'\mathbf{J}_T\mathbf{F}\frac{\mathbf{G}_\beta(\mathbf{X})'}{\sqrt{N}}$. From Assumptions 3(ii), $\frac{\mathbf{G}_\beta(\mathbf{X})'\mathbf{G}_\beta(\mathbf{X})}{N} \to \mathbf{I}_K$, so each column of $\frac{\mathbf{G}_\beta(\mathbf{X})}{\sqrt{N}}$ can be treated as an eigenvector. Furthermore, $\mathbf{F}'\mathbf{J}_T\mathbf{F}$ is a diagonal matrix by Assumptions 3(iii), and hence, each diagonal element of $\mathbf{F}'\mathbf{J}_T\mathbf{F}$ can be interpreted as an eigenvalue. Resorting to these observations, we recover $\mathbf{G}_\beta(\mathbf{X})$ through the eigen-decomposition of $\frac{\widehat{\mathbf{R}}\widehat{\mathbf{R}}'}{N}$, as stated in 2.1.

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 mean of $\mathbf{R}$. by postmultiplying by the $(T \times 1)$ vector $\frac{1}{T}\mathbf{1}_T$.[9] From (2.3), the $(N \times 1)$ vector of average returns, $\frac{1}{T}\mathbf{R}\mathbf{1}_T = \overline{\mathbf{R}}$, is:

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

Our objective is to extract $\mathbf{G}_\alpha(\mathbf{X})$ 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 $\mathbf{G}_\beta(\mathbf{X})\overline{\mathbf{F}}$. 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 spanned by $\mathbf{X}$ that is orthogonal to $\widehat{\mathbf{G}}_\beta(\mathbf{X})$. The following theorem establishes that we can recover $\mathbf{G}_\alpha(\mathbf{X})$ with this approach.

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

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,* $\widehat{\mathbf{G}}_\alpha(\mathbf{X}) \xrightarrow{p} \mathbf{G}_\alpha(\mathbf{X})$.

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.[10]

Alternatively, the estimator in Theorem 2.2 can be derived within the conventional risk-adjusted approach as follows. Note that equation (2.7) can be rearranged as

$$\overline{\mathbf{R}} = \mathbf{G}_\beta(\mathbf{X})\overline{\mathbf{F}} + \left(\mathbf{G}_\alpha(\mathbf{X}) + \Gamma_\alpha + \Gamma_\beta + \overline{\mathbf{E}}\right) \tag{2.7}$$

and

$$\overline{\mathbf{R}} - \mathbf{G}_\beta(\mathbf{X})\overline{\mathbf{F}} = \mathbf{G}_\alpha(\mathbf{X}) + \left(\Gamma_\alpha + \Gamma_\beta\overline{\mathbf{F}} + \overline{\mathbf{E}}\right). \tag{2.8}$$

Recall that our objective is to estimate $\mathbf{G}_\alpha(\mathbf{X})$. Equation (2.8) shows that we can achieve this goal by regressing $\overline{\mathbf{R}} - \mathbf{G}_\beta(\mathbf{X})\overline{\mathbf{F}}$ on $\mathbf{X}$. Because we do not directly observe $\mathbf{G}_\beta(\mathbf{X})$ and $\overline{\mathbf{F}}$, we use $\widehat{\mathbf{G}}_\beta(\mathbf{X})$ from Theorem 2.1 and estimate $\overline{\mathbf{F}}$ by regressing $\overline{\mathbf{R}}$ on $\widehat{\mathbf{G}}_\beta(\mathbf{X})$, motivated by the expression (2.7). The two approaches yield identical results.

Finally, we construct an arbitrage portfolio that optimally exploits any mispricing information in characteristics. Consider first the true but unknown (and thus infeasible) arbitrage portfolio, $\mathbf{w} = \frac{1}{N}\mathbf{G}_\alpha(\mathbf{X})$. Then, from (2.3), we find that the return of this infeasible portfolio is given by

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

From Assumptions 1-3, it is easy to verify that as N increases, $\frac{1}{N}\mathbf{G}_\alpha(\mathbf{X})'\mathbf{G}_\alpha(\mathbf{X})$ con-

[10]The result in Theorem 2.2 can be extended to incorporate a weighting matrix to increase (or decrease) the importance of some stocks vs. others as follows. Consider a $(N \times N)$ diagonal matrix $\mathbf{W}$, the i-th diagonal element of which represents the weight for stock i. Then, we can estimate $\mathbf{G}_\alpha(\mathbf{X})$ by $\widehat{\mathbf{G}}_\alpha(\mathbf{X}) = \mathbf{X}\widehat{\boldsymbol{\theta}}_W$ such that $\widehat{\boldsymbol{\theta}}_W = \arg\min_{\boldsymbol{\theta}} \left(\overline{\mathbf{R}} - \mathbf{X}\boldsymbol{\theta}\right)' \mathbf{W} \left(\overline{\mathbf{R}} - \mathbf{X}\boldsymbol{\theta}\right)$ subject to $\widehat{\mathbf{G}}_\beta(\mathbf{X})' \mathbf{W}\mathbf{X}\boldsymbol{\theta} = \mathbf{0}_K$.

verges 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, $\widehat{\mathbf{w}} = \frac{1}{N}\widehat{\mathbf{G}}_\alpha(\mathbf{X})$, achieves the same asymptotic property.

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

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. Our estimator 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. The details of the out-of-sample applications are described in Section 4.

3 Simulation

In this section we analyze the properties of our estimator in simulations. The purpose of this exercise is three-fold. First, we illustrate the behavior of our arbitrage portfolio estimator in finite samples, similar in size, to those of the U.S. stock market.[11] Second, we explore the properties of the estimator if the number of factors is not known. Third, we document that our estimator is reasonably robust against model misspecification, in particular time-varying characteristics.

3.1 Setup

We first describe the set of characteristics used for simulation. For the matrix $\mathbf{X}$, we consider 61 characteristics, which are available at the end of 2010, the beginning of calibration period. The set of characteristics includes past returns 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, and operating accruals (the full list is given in Table 2).

We generate returns according to four popular asset pricing models, the CAPM, the Fama-French three-factor model (FF3), the Hou, Xue and Zhang four-factor model

[11] This section focuses on simulation evidence regarding our procedure's ability to accurately estimate the arbitrage profits (if any) as established in Theorem 2.3. We also confirm the results of Theorems 2.1 and 2.2. These additional results are available upon requests.

(HXZ4), and the Fama and French five-factor model (FF5). However, we depart from those models by not restricting α to be zero. The number of factors in our estimator, K, is set to the corresponding number in each asset pricing model, i.e., $K = 1$ for the CAPM, $K = 3$ for the FF3, etc. We explore the effects of selecting too few or too many factors in later sections.

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. For ease of interpretation, we normalize the cross-sectional variation of α_i so that the quantity 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 $\widehat{\alpha}_i = \mathbf{x}_i\boldsymbol{\theta}_\alpha + \gamma_{\alpha,i}$. We rescale $\widetilde{\alpha}_i = k\widehat{\alpha}_i$, where $k = \frac{0.01}{\sqrt{\frac{\boldsymbol{\theta}'_\alpha \mathbf{X}'\mathbf{X}\boldsymbol{\theta}_\alpha}{N}}}$, and use the rescaled $\widetilde{\alpha}_i$ in the simulated returns (3.1). Note that $\gamma_{\alpha,i}$, in the above cross-sectional relation, corresponds to the i-th element of $\boldsymbol{\Gamma}_\alpha$. Also, the calibrated betas are significantly correlated with characteristics.

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 a large cross section of stocks, we consider $N = 1,000$ and $N = 2,000$, which are sampled from the 2,458 individual stocks. In each repetition, we simulate returns from

$$\begin{aligned} \mathbf{R} &= \boldsymbol{\alpha}\mathbf{1}'_T\sqrt{\delta} + \mathbf{B}\mathbf{F}' + \mathbf{E} \\ &= \left(\mathbf{X}\boldsymbol{\theta}_\alpha\sqrt{\delta} + \Gamma_\alpha\sqrt{\delta}\right) + \left(\mathbf{X}\boldsymbol{\Theta}_\beta + \Gamma_\beta\right)\mathbf{F}' + \mathbf{E}, \end{aligned} \tag{3.1}$$

where $\boldsymbol{\alpha}$ and $\mathbf{B}$ are calibrated as in the above paragraph, $\mathbf{F}$ are resampled from the realized factors over the 600-month sample from January 1967 to December 2016, and $\mathbf{E}$ are drawn from a normal distribution with the calibrated $\sigma^2_{i,\varepsilon}$ parameters as in the above paragraph. We consider different cases of mispricing, i.e., $\delta = 0,\ 5,$ and 10.

3.2 Simulation Results

3.2.1 Correctly Specified Model

In our baseline scenario, we first investigate the performance of our estimator if we know the correct number of factors. Figure 2 shows the results for using the Capital Asset

Pricing Model (upper–left panel), the Fama-French three-factor model (upper–right panel), the Fama-French five-factor model (lower–left panel) and the Hou, Xue, and Zhang model (lower–right panel). Our findings are consistent across all models used for calibration. The weights of the arbitrage portfolio, $\widehat{\mathbf{w}}$, are estimated using the returns over $t = 1, \cdots, 12$, and the return of the arbitrage portfolio is computed in the following month, $t = 13$, as in our empirical application. That is, we use $T_0 = 12$ and $T = 13$ in the setup of Figure 1. We report the mean of the out-of-sample return as well as 95% confidence intervals for each level of $\delta = 0, 5,$ and 10 and $N = 1,000$ and $N = 2,000$ from 10,000 repetitions. The confidence intervals are considerably narrower with $N = 2,000$ than those with $N = 1,000$. This result is empirically relevant because we can obtain a cross section of this size in the U.S. stock market. As expected, when $\delta = 0$, or there do not exist any arbitrage opportunities, our arbitrage portfolio yields zero returns on average. Recall that α_i is rescaled so that $\frac{\mathbf{G}_\alpha(\mathbf{X})'\mathbf{G}_\alpha(\mathbf{X})}{N} \to 1$b.p./month. Hence, 3.1 implies that the arbitrage portfolio generates asymptotic arbitrage profits of $\delta = \lim_{N\to\infty} \left(\frac{\left(\mathbf{G}_\alpha(\mathbf{X})\sqrt{\delta}\right)'\left(\mathbf{G}_\alpha(\mathbf{X})\sqrt{\delta}\right)}{N} \right)$. In fact, we observe that, when the simulation parameters are $\delta = 0, 5,$ or 10, the average of arbitrage portfolio returns corresponds to the target size of δ b.p./month, suggesting that our arbitrage portfolio actually generates the expected level of arbitrage profits.

3.2.2 Unknown Number of Factors

In the previous section, we used the true number of factors in extracting factor loadings from the projected returns. In application, we do not know the correct number of factors. Estimating the number of factors is a long-standing problem in panel-data analysis for which many tests have been proposed, e.g., Connor and Korajczyk (1993), Bai and Ng (2002) or Ahn and Horenstein (2013), and is a nontrivial task as emphasized in Brown (1989). We therefore examine the effect of selecting one too few or one too many factors. Figure 3 reports the results when we set the number of extracted factors to be one more than the true number of factors. We find that the arbitrage portfolio's performance in Figure 3 is almost identical to those in Figure 2, where we set the number of extracted factors to be the number of true factors. Hence, we conclude that extracting one additional factor more than the true number does not seem to harm the performance of our arbitrage portfolios materially. This result is not too surprising because our arbitrage portfolio weights still achieve orthogonality to the systematic factors. Obviously, extracting many more extraneous factors will likely lead to imprecision in the estimates.

In contrast, if the number of extracted factors is less than the number of true factors, our methodology does not guarantee that the arbitrage portfolio weights are orthogonal to betas with respect to systematic factors. Figure 4 reports the performance of our arbitrage portfolios when we extract one less factor than the underlying model for the CAPM, FF3, HXZ4, and FF5. We find that the average returns are far off from the target level and the portfolio returns are much more volatile (presumably due to the exposure to systematic factors) relative to the case of overestimation (Figure 3 (too many) vs Figure 4 (too few)). As a a guideline for empirical analyses, we should therefore try to select slightly too many rather than too few factors, as the effects of selecting too few are far more severe than those of selecting too many. In the empirical analysis, we will explore the variation of the results as we change the number of factors.

3.2.3 Time-Varying Characteristics

The theory developed so far assumes that characteristics do not vary over time. In this section, we explore how our estimator will behave if this assumption is violated. We assume that each characteristic follows an AR(1) process. We find the AR(1) parameters of each characteristic as follows. For each characteristic and each firm, we have 36 observations of the characteristic over our calibration period. We estimate the AR(1) autoregressive coefficient over this time period and the variance of the residuals for each firm. We then determine the average AR(1) coefficient as the average across firms and also determine the variance of the residuals (for each characteristic) in the same way.

Across simulations, we fix the initial characteristic over the calibration period as $\mathbf{X}$. Let $x_{i,c}$ and $x_{i,c,t}$ denote the (i,c) element of $\mathbf{X}$ and $\mathbf{X}_t$, respectively. Then, we generate $\mathbf{X}_t$ with $x_{i,c,t} = x_{i,c} + \rho_c\left(x_{i,c,t} - x_{i,c}\right) + \sigma_c \varepsilon_{i,t}$, where ρ_c and σ_c^2 are the estimated AR(1) coefficient and variance of residuals of a certain characteristic c, and $\varepsilon_{i,t}$ is drawn from $N\left(0,1\right)$ as i.i.d over i and t. We then generate $\mathbf{R}_t$, the t-th column of $\mathbf{R}$, as follows:

$$\mathbf{R}_t = \boldsymbol{\alpha}_{t-1}\sqrt{\delta} + \mathbf{B}_{t-1}\mathbf{f}_t + \mathbf{E}_t,$$

where $\boldsymbol{\alpha}_{t-1} = \mathbf{X}_{t-1}\boldsymbol{\theta}_\alpha$, $\mathbf{B}_{t-1} = \mathbf{X}_{t-1}\boldsymbol{\Theta}_\beta + \Gamma_\beta$ and $\mathbf{E}_t$ is the t-th column of $\mathbf{E}$.[12]

Figure 5 reports the performance of our arbitrage portfolios when the returns are generated with the time-varying alpha $\boldsymbol{\alpha}_{t-1} = \mathbf{X}_{t-1}\boldsymbol{\theta}_\alpha$ and the time-varying beta $\mathbf{B}_{t-1} = \mathbf{X}_{t-1}\boldsymbol{\Theta}_\beta + \Gamma_\beta$, induced by time-varying characteristics. We find that our methodology is robust to the empirically relevant dynamics in the characteristics.

3.2.4 Further Robustness Checks

To further investigate the robustness of our estimator, we introduce correlated residuals. In each simulation, we randomly construct 50 clusters of equal numbers of stocks and generate the residual shocks so that the residual correlation between stocks in the same cluster is 0.1 and that between stocks in different clusters is zero. We calibrate the within-cluster residual correlation using the average correlation of residual shocks within a same industry relative to commonly used asset pricing models such as CAPM or FF3. The results are reported in Figure A.1 in the online appendix.

We also repeat the analysis using a different time period for calibration. In an alternative calibration, we use the data from the beginning of 2006 through 2008. This time period contains the extremely volatile second half of 2008. We report these results in the online Appendix, in Figure A.2. In addition, we provide simulation evidence of the robustness of our method to missing characteristics. To this end, in each repetition, we use 61 characteristics for simulating returns but drop randomly picked ten characteristics for computing $\widehat{\mathbf{w}}$. We plot the results in Figure A.3 of the online appendix. As an additional test, we also re-run the simulations and randomly select firms with replacement in each iteration, thereby illustrating the robustness to a slightly different composition of the panel. Overall, the performance of the estimator is very stable across all these modifications.

4 Empirical Application

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

[12]We obtain $\boldsymbol{\theta}_\alpha$ and $\boldsymbol{\Theta}_\beta$ by regressing the calibrated $\boldsymbol{\alpha}$ and $\mathbf{B}$ on $\mathbf{X}$. Also, we find Γ_β from $\Gamma_\beta = \mathbf{B} - \mathbf{X}\boldsymbol{\Theta}_\beta$.

4.1 Data

The data are the same as in Freyberger et al. (2019); we use stock return data from the Center for Research in Security Prices (CRSP) monthly file. As is 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. (2019), of year $t + 1$, predicting returns from the beginning of July of year t until the end June of year $t+1$. Table 2 provides an overview of the characteristics used for estimation of the mispricing function and the factor loading function.

To alleviate potential concerns about survivorship bias, which may arise because of backfilling, we require that a firm have a least two years of data in Compustat. Our sample period is from 1965 through 2014. For the full sample, we have approximately 1.6 million firm/month observations in our analysis.[13]

4.2 Estimation

We initially assume that the factor loading function and the mispricing function are linear in the characteristics.[14]

Figure 1 illustrates how we implement the arbitrage portfolio in an out-of-sample manner. We estimate $\widehat{\mathbf{w}}$ with the returns over $t = 1, \cdots, 12$, and the return of the arbitrage portfolio is measured in the following month, $t = 13$. We call the first period $t = 1, \cdots, 12$ the estimation period and the second period $t = 13$ the holding period (below we also report results for alternative lengths for both the estimation and the holding periods). Let $\mathbf{X}_0$ and $\mathbf{X}_{12}$ denote the characteristics at the beginning of estimation and holding periods, respectively. For example, we first use $\mathbf{X}_0$ to obtain the projected and demeaned return of $\widehat{\mathbf{R}}$ over the estimation period corresponding to $\mathbf{P}_{\mathbf{X}_0}\mathbf{R}\mathbf{J}_{12}$ in (2.5) (from a panel regression using 12 months from January 1967 to December 1967). The t-th column of the $(N \times 12)$ matrix $\widehat{\mathbf{R}}$ is the demeaned projected return for the t-th month. Then we compute the $N \times N$ matrix $\frac{\widehat{\mathbf{R}}\widehat{\mathbf{R}}'}{N}$ and the first K eigenvectors of the matrix. We then project the average returns onto characteristics

[13]The appendix in Freyberger et al. (2019) contains a detailed description of the construction of the data as well as numerous references to papers that have employed these characteristics in empirical applications.

[14]Note that our methodology allows for (parametric) nonlinearities, which we explore in Section 5.3. However, the results from employing these polynomial expansions are very similar to the linear case and are, therefore, relegated to the appendix.

subject to orthogonality to the estimated factor loadings as in Theorem 2.2 to obtain $\widehat{\boldsymbol{\theta}}$. In computing the arbitrage portfolio weights as in Theorem 2.3 for the following month of January 1968, we update characteristics with $\mathbf{X}_{12}$ in computing $\widehat{\mathbf{w}}$ such that $\widehat{\mathbf{w}} = \frac{1}{N}\mathbf{X}_{12}\widehat{\boldsymbol{\theta}}$. We repeat this process month by month until June 2014. 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 Performance of the Arbitrage Portfolio

In this section we document the out-of-sample performance of the arbitrage portfolio. Table 3 shows the summary statistics for returns of the arbitrage portfolio for different numbers of eigenvectors. From Table 3 we see that the returns and Sharpe ratios increase with the number of eigenvectors until about six eigenvectors. Employing more than six eigenvectors does not seem to materially harm the properties of the portfolio, but there also does not seem to be an improvement in any performance metric. Overall, the Sharpe ratios are very high, ranging from 1.35 to 1.75 . The increase in Sharpe ratios with increasing number of eigenvectors is driven by increasing means, not decreasing standard deviations, because the standard deviation is always normalized to be 20% in-sample. The out-of-sample standard deviation is close to the in-sample standard deviation. The table also displays the maximum drawdown, which ranges between 20.1% and 38.5%. These drawdown numbers are relatively moderate compared to the maximum drawdowns of common factors over the same time period. The four factors in Fama-French-Carhart model have maximum drawdowns of 55.71% (market factor), 52.78% (size factor), 44.68% (value factor) and 57.51% (momentum factor) over our sample period. In addition, skewness, kurtosis, and the best and worst month are also reported in Table 3.

The large Sharpe ratios of Table 3 could be driven by high exposures to common risk factors and therefore not be related to possible mispricing. Therefore, aiming to understand better the abnormal performance of the arbitrage portfolio, we run a time-series regression of the arbitrage portfolio's returns onto common risk factors.[15] In Tables 4 (one estimated factor) and 5 (six estimated factors), we report the risk-

[15]We are grateful to Kenneth French for making the factors involved in the CAPM, FF3, and FF5 models available on his website. We also thank Chen Xue for providing the data for the Hou et al. (2015) four-factor model.

adjusted returns of the arbitrage portfolio with respect to the CAPM (column 1), the Fama and French (1992) three-factor model (column 2), the Fama-French three-factor model augmented with the Carhart (1997) momentum factor (column 3), the Fama and French (2015) five-factor model (column 4), the Fama-French five-factor model augmented with the momentum factor (column 5), the Hou et al. (2015) four-factor model (column 6) and the HXZ model augmented with the momentum factor (column 7).

We limit our main discussion to the cases in which we extract one factor (one eigenvector) and six factors (six eigenvectors). The results for all other cases are contained in the online appendix. In Table 4 with one eigenvector, we can see that the alpha (or the intercept in the time-series regression) is fairly consistent across various asset pricing models. Although our arbitrage portfolio has significant exposures to some factors, the adjusted R^2 is fairly low with the minimum of 0.00 and the maximum of 0.22. We find consistent results when we increase the number of eigenvectors except that the alpha tends to increase. For example, the CAPM alpha of our arbitrage portfolio with six eigenvectors is 2.63 %/month (See Table 4), far higher than that with one eigenvector 1.79 %/month (See Table 5). We illustrate the relation between out-of-sample alpha and the number of eigenvectors used in the estimator in Figure 6. We can see that the alpha has a hump shape and decreases after approximately seveneigenvectors. We attribute the deterioration to the overfitting of systematic risks.

Figure 7 summarizes the correlation of the arbitrage portfolios (using 1 through 10 eigenvectors) with common risk factors. If we look at the correlation between the arbitrage portfolios, we see that the correlation between the arbitrage portfolio with one eigenvector and the other arbitrage portfolios drops as the number of eigenvectors increases, albeit it never drops below 0.8. If we compare the correlation of the arbitrage portfolios with five or more eigenvectors, we see that the correlation is consistently high, suggesting that the portfolio does not change very much after we extract five common factors. The correlation between the arbitrage portfolios and the common factors is relatively low except for the size factor, which again is consistent with the factor regressions in Tables 4 and 5 and the additional factor regressions in the online appendix.

4.4 Properties of the Arbitrage Portfolio

In this section, we explore the properties of the arbitrage portfolio more deeply. In particular, we open the "black box" and study the firm characteristics of the companies in the arbitrage portfolio. Furthermore, we discuss the time-series properties of the returns, the properties of the portfolio weights, as well as possible diminishing excess returns over time.

4.4.1 Time-Series Properties

To develop further intuition about the performance of the arbitrage portfolio, we explore its time-series properties more closely. In Figure 8 we plot the cumulative return. It is noteworthy, that the arbitrage portfolio did not have a negative return (for a full year) during the recent financial crises. Overall, the returns are positive in 42 out of 44 years. Also, the arbitrage portfolio does not have significantly different returns during NBER recessions versus other periods. With a simple regression of the portfolio return on a constant and an NBER recession indicator, i.e. $r_t = a + b \times \text{NBER}_t + \varepsilon_t$, we obtain point estimates of $\widehat{a} = 2.523$ (significant at the 1% level) and $\widehat{b} = 0.539$, with a p-value of 0.40. This strongly suggests that the portfolio returns are not systematically related to the business cycle.

In addition, we also explore whether the excess returns of the arbitrage portfolio diminish systematically over time. We test for a time trend, by estimating the following specification

$$r_t = a + b \times t^{\gamma} + \varepsilon_t. \tag{4.1}$$

We estimate the model using non-linear least squares, the point estimates are $\widehat{a} = 5.27$, $\widehat{b} = -0.127$, $\widehat{\gamma} = 0.5501$.[16] Only the intercept is significant at coventional levels, with a p-value of less than 0.01.[17] A possibly undesirable feature of this specification is that it does not rule out arbitrarily negative returns in the limit. However, it seems plausible to restrict the model to only allow returns to be zero in the limit. One easy way to achieve this is restrict the intercept to be zero and require a positive value for b in this case, we estimate $\widehat{b} = 10.18$ and $\widehat{\gamma} = -0.262$. This specification suggests a

[16] Note that this specification contains the linear time trend, $r_t = a + b \times t + \varepsilon_t$ as a special case.

[17] p-values of $\hat{b}$ and $\hat{\gamma}$ are 0.77 and 0.25, respectively.

mild decay in excess returns and predicts the returns to reach less than one percent per month in approximately 7000 periods. We plot the trend estimated from this specification in Figure 9. Both specifications confirm that the excess returns appear not to diminish systematically over time. This finding is important in the context of the work of McLean and Pontiff (2016) and Linnainmaa and Roberts (2018), who document that many anomalies have become significantly weaker post publication. 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.

4.4.2 Firm Characteristics

In Figure 10 we show a comparison of the long and short side for nine well-known characteristics for the arbitrage portfolio using six eigenvectors. All of the characteristics in Figure 10 are well-known cross-sectional return predictors: the book-to-market ratio (Fama and French (1992)), the debt-to-price ratio (Litzenberger and Ramaswamy (1979)), market equity (often referred to as "size," e.g., Banz (1981)), profitability (recently reexamined by Ball et al. (2015)), investment (Fama and French (2015)), operating accruals (Sloan (1996)), last month's turnover (Datar et al. (1998)), and short-term reversal as well as (standard) momentum, both of which are documented in Jegadeesh and Titman (1993)).

From Figure 10 we can see that the arbitrage portfolio is typically long smaller firms and short larger firms, which is consistent with the positive loading on the size factor in Table A.4. Another clear pattern emerging from the figure is that the arbitrage portfolio is typically long firms with low returns in the month preceding the portfolio formation. It is, however, very remarkable that there is no noticeable pattern for book-to-market, momentum, and investment, which is again consistent with small and insignificant loadings on the corresponding factors in Table A.4. Interestingly, the pattern for profitability is not very clear in the figure, but the portfolio has a significant negative loading on the "robust minus weak" factor in Table A.4. We show the cross-sectional comparison for all 61 characteristics in Figure A.4 in the online Appendix.

To gain more intuition about the relationship between characteristics and systematic risk on the one hand and mispricing on the other hand, we project the estimated factor loadings ($\widehat{\mathbf{G}}_{\beta}(\mathbf{X})$) and the estimated mispricing function ($\widehat{\mathbf{G}}_{\alpha}(\mathbf{X})$) onto the characteristic in each period. We normalize the coefficients cross-sectionally so that the highest coefficient always receives a value of one to ensure that we can compare the coefficients over time. Figure 11 shows the projection results for systematic risk and Figure 12 shows the corresponding results for mispricing. From Figure 11 we can see there appears to be a relatively stable relationship between some groups of characteristics (e.g. past returns, total volatility, idiosyncratic volatility and size related variables (LME and AT)) and factor loadings. However, from Figure 12 we can see that one few characteristics are consistently related to mispricing (size and total assets). Other characteristics are only related to mispricing for few periods "on-and-off". While there are clear limits to "eyeball econometrics", the results in Figure 12 underscore the importance of our time-varying approach.

4.4.3 Portfolio Weights

The theory does not impose any limits or discipline on the portfolio weights of the arbitrage portfolio. In the implementation, we scale the portfolio weights such that the in-sample standard deviation of the arbitrage portfolio is 20% annualized. In the implementation, we de-mean the characteristics so that the resulting portfolio weights of the arbitrage portfolio sum to zero, it therefore by construction a "zero-investment portfolio." However, we do not impose any constraints on the largest (smallest) position within the portfolio. It is therefore a potential concern that the portfolio allocates an unrealistically large amount into individual assets. In Figure 13, we plot the median, minimum, maximum as well as the 5% and 95% quantile of the weights in each month over the sample period from January 1968 to June 2014. The largest weight (in absolute value) over the entire sample is approximately 5.1%. In later parts of the sample, when the number of stocks is larger, the weights are considerably smaller, with the largest weights often being less than 1% in absolute value.

5 Robustness

The empirical implementation of the arbitrage portfolio in Section 4 naturally depends on several choices, such as the number of estimated factors (eigenvectors) or the length of the estimation window. It is therefore important to demonstrate that the results are robust to many of these choices, In the following, we relax many of these choices and show that our results do not depend crucially on these implementation choices.

5.1 Estimation Windows

In our main specification, we use 12 months and then roll the estimation window forward by one month. Since our theoretical results are derived in a *local in time* setting, i.e. T is fixed, the choice of T should not drive the result. To illustrate this, we re-estimate our main analysis with 24 and 36 months as the estimation window. Tables A.9 and A.10 show the general performance statistics and estimated alphas (against the standard factor models) for the arbitrage portfolio constructed using only 24 months as an estimation window. Overall, the results are quite similar to our baseline of a 12 months window, althoughslightly worse than the base case. If we use an estimation window of 36 months, the results are again similar to the baseline of 12, see Tables A.11 and A.12.

5.2 Holding Periods

The heatmap for the mispricing component in Figure 12 shows the loadings on the different characteristics change from period to period. It is therefore natural to investigate if the performance of the arbitrage portfolio deteriorates strongly if we do not rebalance the portfolio each month. In this robustness check we therefore hold the arbitrage portfolio for 2, 3, 6 and 12 months without rebalancing and analyze its performance.[18] The Sharpe ratios from this exercise are shown in Table A.13 . The results show that timely information is crucial for creating a arbitrage portfolio and that mispricing appears to be rather short lived. However, even if we only rebalance every second or every third month, the arbitrage portfolio still achieves annualized Sharpe ratios close to one. However, if we rebalance only once per year, the Sharpe ratio of the arbitrage portfolio is in the same order as the Sharpe ratio of the U.S. stock market overall and no abnormal performance can be obtained.

[18] Note that this approach does not yield same portfolio weights for each month in the holding periods. Although we do not trade over the holding period, the portfolio weights change as the value of securities evolve.

5.3 Nonlinear Estimation

In Section 2, we have not taken a parametric stand on the functional form of $\mathbf{G}_{\beta}(\mathbf{X})$. In the application, we estimate $\mathbf{G}_{\beta}(\mathbf{X})$ as 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. (2016), $\mathbf{G}_{\beta}(\mathbf{X})$ is approximated 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 $\mathbf{X}$, 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 $\mathbf{G}_{\beta}(\mathbf{X})$.[19]

In Table A.15 we show alphas of the arbitrage portfolio against various factor models when we use 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. This suggests that allowing nonlinearities enables our method to estimate systematic factors more effectively. Overall, however, the results of the higher-order expansions are consistent with the the linear specification and do not erode the arbitrage profits. However, they leave interesting avenues for future research.

5.4 Small Firms

The analysis in Section 4.4.2 suggests that the arbitrage portfolio tends to be long smaller firms and short larger firms. It is therefore important to understand if the results are materially driven by micro-cap stocks that are likely very illiquid. We therefore exclude all stocks below the 10% NYSE size quantile. Discarding firms below the 10% NYSE quantile eliminates much more than 10% of all firms, since the average NYSE firm is larger than the after firm listed on NASDAQ. Excluding these firms re-

[19]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's (2014) handbook chapter for a deep theoretical treatment of orthogonal polynomials.

duces the sample size on average by 38% per month, i.e. the total sample size shrinks from approximately 1.6 million observations to roughly 900,000 observations. We then re-compute the arbitrage portfolio using an estimation period of 12 months as in the baseline analysis. Tables A.16 and A.17 show portfolio performance measures and estimated alphas against various factor model. The performance does not weaken strongly relative to using all available firms and even excluding very small firms leads to portfolios with alphas in excess of 1% per month and annualized Sharpe ratios greater than 1, thereby reinforcing our earlier finding that characteristics contain information about mispricing.

5.5 Alternative Factor Models

In the previous sections, we relied on the "classic" risk factors suggested in the literature. While it is impossible to conduct an exhaustive analysis of all possible risk factors suggested throughout the empirical asset pricing literature, it is important to analyze the robustness of our results to "alternative" asset pricing factors, such as the liquidity factor of Pástor and Stambaugh (2003) or the betting-against-beta factor of Frazzini and Pedersen (2014).[20] Lastly, since we are dealing with an arbitrage or mispricing portfolio, we also employ the "mispricing factors" of Stambaugh and Yuan (2016).[21] Table A.18 shows the estimated alphas and factor exposures for these additional factor models for our baseline arbitrage portfolio (12 estimation months and six estimated factors). From Table A.18 we can see that the arbitrage portfolio still has high and strongly significant alpha's. Moreover, the portfolio is only marginally exposed to the "mispricing" factor of Stambaugh and Yuan (2016) . The exposure to the other "alternative" factors is insignificant.

6 Conclusion

We propose new methodology to simultaneously recover conditional factor realizations (returns on "smart-beta" portfolios), estimate conditional factor loadings, estimate conditional alphas using firm-level characteristics, and construct arbitrage portfolios. Our

[20] The betting-against-beta factor was obtained from the AQR factor database.

[21] We are grateful to Robert Stambaugh for making the illiquidity factor and the mispricing factors available on his website.

methodology extends the method of Projected Principal Components of Fan et al. (2016) to separately identify risk and mispricing. In an extensive simulation study, we show that our methodology works well in a finite sample and is also robust against various forms of misspecification, in particular, it does not break down with time-varying or some missing characteristics. The methodology only requires a large cross section and can accommodate a short time span.

In the empirical application in the CRSP/Compustat panel from 1968 to 2014, we find that characteristics carry significant information about mispricing despite giving maximal explanatory power to the statistical factor model. Alphas against popular factor models range between 1.5% and almost 3% per month.

While we do find significant abnormal returns for the arbitrage portfolio against popular existing factors, our main contribution is the development of new methodology that separately identifies alpha and beta and thereby correctly parses the ability of firm characteristics to explain the cross-section of returns into risk and mispricing components. This is important even if we had found no evidence of mispricing since some common techniques can lead to incorrect inferences.

A Proofs

Let $\mathbf{P}$ denote the projection matrix $\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}$.

Lemma A.1. *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 $\widehat{\mathbf{F}}$ and $\mathbf{D}$ such that 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 $\mathbf{Y}'\mathbf{Y}$. Define* the *$(N \times K)$ matrix $\widehat{\mathbf{\Lambda}}$ such that the k-th column of $\widehat{\mathbf{\Lambda}}$ is the eigenvector of $\mathbf{Y}\mathbf{Y}'$ corresponding to the k-th largest eigenvalue of $\mathbf{Y}\mathbf{Y}'$. Let $\widetilde{\mathbf{\Lambda}} = \mathbf{Y}\widetilde{\mathbf{F}}\left(\widetilde{\mathbf{F}}'\widetilde{\mathbf{F}}\right)^{-1}$, where $\widetilde{\mathbf{F}} = \widehat{\mathbf{F}}\mathbf{D}^{1/2}$. Then, it holds that*

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

Proof The k-th largest eigenvalue of $\mathbf{Y}'\mathbf{Y}$ is the k-th largest eigenvalue of $\mathbf{Y}\mathbf{Y}'$ (see Greene (2008) page 970). Hence, $\widehat{\mathbf{\Lambda}}$ is identified by the following two conditions:

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

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

$$\mathbf{Y}'\mathbf{Y} = \mathbf{Q}\mathbf{V}\mathbf{Q}'. \tag{A.1}$$

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 $\mathbf{D}$:

$$\widehat{\mathbf{F}} = \mathbf{Q}\left[\mathbf{I}_K \ \mathbf{0}_{K\times(T-K)}\right]' \text{ and } \mathbf{D} = \left[\mathbf{I}_K \ \mathbf{0}_{K\times(T-K)}\right]\mathbf{V}\left[\mathbf{I}_K \ \mathbf{0}_{K\times(T-K)}\right]'. \tag{A.2}$$

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

$$\begin{aligned}\widetilde{\mathbf{\Lambda}}'\widetilde{\mathbf{\Lambda}} &= \mathbf{D}^{-0.5}\widehat{\mathbf{F}}'\mathbf{Y}'\mathbf{Y}\widehat{\mathbf{F}}\mathbf{D}^{-0.5} = \mathbf{D}^{-0.5}\left[\mathbf{I}_K \ \mathbf{0}_{K\times(T-K)}\right]\mathbf{Q}'\mathbf{Q}\mathbf{V}\mathbf{Q}'\mathbf{Q}\left[\mathbf{I}_K \ \mathbf{0}_{K\times(T-K)}\right]'\mathbf{D}^{-0.5} \\ &= \mathbf{D}^{-0.5}\left[\mathbf{I}_K \ \mathbf{0}_{K\times(T-K)}\right]\mathbf{V}\left[\mathbf{I}_K \ \mathbf{0}_{K\times(T-K)}\right]'\mathbf{D}^{-0.5} = \mathbf{D}^{-0.5}\mathbf{D}\mathbf{D}^{-0.5} = \mathbf{I}_K,\end{aligned} \tag{A.3}$$

where the second and fourth equalities are from equation (A.1) and equation (A.2), and that

$$
\begin{aligned}
\widetilde{\mathbf{\Lambda}}'\mathbf{Y}\mathbf{Y}'\widetilde{\mathbf{\Lambda}} &= \mathbf{D}^{-0.5}\widehat{\mathbf{F}}'\mathbf{Y}'\mathbf{Y}\mathbf{Y}'\mathbf{Y}\widehat{\mathbf{F}}\mathbf{D}^{-0.5} = \mathbf{D}^{-0.5}\widehat{\mathbf{F}}'\mathbf{Q}\mathbf{V}^2\mathbf{Q}'\widehat{\mathbf{F}}\mathbf{D}^{-0.5} \\
&= \mathbf{D}^{-0.5}\left[\mathbf{I}_K \ \mathbf{0}_{K\times(T-K)}\right]\mathbf{Q}'\mathbf{Q}\mathbf{V}^2\mathbf{Q}'\mathbf{Q}\left[\mathbf{I}_K \ \mathbf{0}_{K\times(T-K)}\right]'\mathbf{D}^{-0.5} \\
&= \mathbf{D}^{-0.5}\left[\mathbf{I}_K \ \mathbf{0}_{K\times(T-K)}\right]\mathbf{V}^2\left[\mathbf{I}_K \ \mathbf{0}_{K\times(T-K)}\right]'\mathbf{D}^{-0.5} \\
&= \mathbf{D}^{-0.5}\left(\left[\mathbf{I}_K \ \mathbf{0}_{K\times(T-K)}\right]\mathbf{V}\left[\mathbf{I}_K \ \mathbf{0}_{K\times(T-K)}\right]'\right)^2\mathbf{D}^{-0.5} \\
&= \mathbf{D}^{-0.5}\mathbf{D}^2\mathbf{D}^{-0.5} = \mathbf{D}, \qquad (A.4)
\end{aligned}
$$

where the second equality is from equation (A.1) and the third and sixth equalities are from equation (A.2). Finally, the two equalities of equations (A.3) and (A.4) prove the lemma. □

Lemma A.2. *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 A.1.

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 A.2 shows there are two equivalent methods to estimate the factor loading matrix. A direct approach is to calculate $\widehat{\mathbf{G}}_\beta(\mathbf{X})$ by calculating the eigenvectors of the $N \times N$ matrix $\frac{\widehat{\mathbf{R}}\widehat{\mathbf{R}}'}{N}$ (which is not feasible for very large cross-sectional samples). The second approach is to first estimate the factors by asymptotic principal components (Connor and Korajczyk (1986)) using the eigenvectors of the much smaller $K \times K$ matrix $\frac{\widehat{\mathbf{R}}'\widehat{\mathbf{R}}}{N}$ and then to run regressions of returns on the factors to estimate the factor loadings $\widetilde{\mathbf{G}}_\beta(\mathbf{X})$.

Lemma A.3. *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,

$$\frac{\widehat{\mathbf{R}}'\widehat{\mathbf{R}}}{N} = \sum_{i=1}^{3}\sum_{j=1}^{3}\frac{1}{N}l_i'l_j. \tag{A.5}$$

Note that

$$\frac{1}{N}l_1'l_1 = \mathbf{J}_T\mathbf{F}\left(\frac{\mathbf{G}_\beta\left(\mathbf{X}\right)'\mathbf{G}_\beta\left(\mathbf{X}\right)}{N}\right)\mathbf{F}'\mathbf{J}_T = \mathbf{J}_T\mathbf{F}\mathbf{F}'\mathbf{J}_T \tag{A.6}$$

from Assumption 3(ii) and that

$$\frac{1}{N}l_1'l_2 = \mathbf{J}_T\mathbf{F}\left(\frac{\mathbf{G}_\beta\left(\mathbf{X}\right)'\Gamma_\beta}{N}\right)\mathbf{F}'\mathbf{J}_T \xrightarrow{p} \mathbf{J}_T\mathbf{F}\mathbf{0}_{K\times K}\mathbf{F}'\mathbf{J}_T = \mathbf{0}_{T\times T} \tag{A.7}$$

from Assumption 2(ii) and that

$$\frac{1}{N}l_1'l_3 = \mathbf{J}_T\mathbf{F}\left(\frac{\mathbf{G}_\beta\left(\mathbf{X}\right)'\mathbf{E}}{N}\right)\mathbf{J}_T \xrightarrow{p} \mathbf{J}_T\mathbf{F}\mathbf{0}_{K\times T}\mathbf{J}_T = \mathbf{0}_{T\times T} \tag{A.8}$$

from Assumption 2(ii) and that

$$\frac{1}{N}l_2'l_2 = \mathbf{J}_T\mathbf{F}\left(\frac{\Gamma_\beta'\mathbf{X}}{N}\right)\left(\frac{\mathbf{X}'\mathbf{X}}{N}\right)^{-1}\left(\frac{\mathbf{X}'\Gamma_\beta}{N}\right)\mathbf{F}'\mathbf{J}_T \xrightarrow{p} \mathbf{J}_T\mathbf{F}\mathbf{0}_{K\times L}\mathbf{V}_X^{-1}\mathbf{0}_{L\times K}\mathbf{F}'\mathbf{J}_T = \mathbf{0}_{T\times T} \tag{A.9}$$

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

$$\frac{1}{N}l_2'l_3 = \mathbf{J}_T\mathbf{F}\left(\frac{\Gamma_\beta'\mathbf{X}}{N}\right)\left(\frac{\mathbf{X}'\mathbf{X}}{N}\right)^{-1}\left(\frac{\mathbf{X}'\mathbf{E}}{N}\right)\mathbf{J}_T \xrightarrow{p} \mathbf{J}_T\mathbf{0}_{T\times L}\mathbf{V}_X^{-1}\mathbf{0}_{L\times T}\mathbf{J}_T = \mathbf{0}_{T\times T} \tag{A.10}$$

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

$$\frac{1}{N}l_3'l_3 = \mathbf{J}_T\left(\frac{\mathbf{E}'\mathbf{X}}{N}\right)\left(\frac{\mathbf{X}'\mathbf{X}}{N}\right)^{-1}\left(\frac{\mathbf{X}'\mathbf{E}}{N}\right)\mathbf{J}_T \xrightarrow{p} \mathbf{J}_T\mathbf{0}_{T\times L}\mathbf{V}_X^{-1}\mathbf{0}_{L\times T}\mathbf{J}_T = \mathbf{0}_{T\times T} \tag{A.11}$$

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

Finally, plugging the results of equations (A.6)-(A.11) into (A.5), 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. $\square$

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 A.3 and $\widehat{\mathbf{F}}$ is the $(T \times K)$ matrix, each column of which is an eigenvector of $\frac{\widehat{\mathbf{R}}'\widehat{\mathbf{R}}}{N}$.

Note 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 eigendecomposition, 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, $\widetilde{\mathbf{F}} = \widehat{\mathbf{F}}\mathbf{D}^{0.5} \xrightarrow{p} \mathbf{J}_T\mathbf{F}(\mathbf{F}'\mathbf{J}_T\mathbf{F})^{-0.5}(\mathbf{F}'\mathbf{J}_T\mathbf{F})^{0.5} = \mathbf{J}_T\mathbf{F}$.

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, $\mathbf{F}'\mathbf{J}_T\widetilde{\mathbf{F}}\left(\widetilde{\mathbf{F}}'\widetilde{\mathbf{F}}\right)^{-1} \xrightarrow{p} \mathbf{F}'\mathbf{J}_T\mathbf{F}(\mathbf{F}'\mathbf{J}_T\mathbf{F})^{-1} = \mathbf{I}_K$.

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

$$m_1 = \mathbf{G}_\beta(\mathbf{X})\left(\mathbf{F}'\mathbf{J}_T\widetilde{\mathbf{F}}\left(\widetilde{\mathbf{F}}'\widetilde{\mathbf{F}}\right)^{-1} - \mathbf{I}_K\right),$$
$$m_2 = \mathbf{P}\Gamma_\beta\mathbf{F}'\mathbf{J}_T\widetilde{\mathbf{F}}\left(\widetilde{\mathbf{F}}'\widetilde{\mathbf{F}}\right)^{-1},$$
$$m_3 = \mathbf{P}\mathbf{E}\mathbf{J}_T\widetilde{\mathbf{F}}\left(\widetilde{\mathbf{F}}'\widetilde{\mathbf{F}}\right)^{-1}.$$

Hence,

$$\frac{1}{N}\left(\widetilde{\mathbf{G}}_{\beta}\left(\mathbf{X}\right)-\mathbf{G}_{\beta}\left(\mathbf{X}\right)\right)'\left(\widetilde{\mathbf{G}}_{\beta}\left(\mathbf{X}\right)-\mathbf{G}_{\beta}\left(\mathbf{X}\right)\right)=\sum_{i=1}^{3}\sum_{j=1}^{3}\frac{1}{N}m_i'm_j. \tag{A.12}$$

Note that

$$\begin{aligned}\frac{1}{N}m_1'm_1 &= \left(\mathbf{F}'\mathbf{J}_T\widetilde{\mathbf{F}}\left(\widetilde{\mathbf{F}}'\widetilde{\mathbf{F}}\right)^{-1}-\mathbf{I}_K\right)'\frac{\mathbf{G}_{\beta}\left(\mathbf{X}\right)'\mathbf{G}_{\beta}\left(\mathbf{X}\right)}{N}\left(\mathbf{F}'\mathbf{J}_T\widetilde{\mathbf{F}}\left(\widetilde{\mathbf{F}}'\widetilde{\mathbf{F}}\right)^{-1}-\mathbf{I}_K\right)\\ &\xrightarrow{p}\left(\mathbf{I}_K-\mathbf{I}_K\right)'\mathbf{I}_K\left(\mathbf{I}_K-\mathbf{I}_K\right)=\mathbf{0}_{K\times K}\end{aligned} \tag{A.13}$$

from Step 4 and Assumption 3(ii) and that

$$\begin{aligned}\frac{1}{N}m_1'm_2 &= \left(\mathbf{F}'\mathbf{J}_T\widetilde{\mathbf{F}}\left(\widetilde{\mathbf{F}}'\widetilde{\mathbf{F}}\right)^{-1}-\mathbf{I}_K\right)'\frac{\mathbf{G}_{\beta}\left(\mathbf{X}\right)'\Gamma_{\beta}}{N}\mathbf{F}'\mathbf{J}_T\widetilde{\mathbf{F}}\left(\widetilde{\mathbf{F}}'\widetilde{\mathbf{F}}\right)^{-1}\\ &\xrightarrow{p}\left(\mathbf{I}_K-\mathbf{I}_K\right)'\mathbf{0}_{K\times K}\mathbf{I}_K=\mathbf{0}_{K\times K}\end{aligned} \tag{A.14}$$

from Step 4 and Assumption 2(ii) and that

$$\begin{aligned}\frac{1}{N}m_1'm_3 &= \left(\mathbf{F}'\mathbf{J}_T\widetilde{\mathbf{F}}\left(\widetilde{\mathbf{F}}'\widetilde{\mathbf{F}}\right)^{-1}-\mathbf{I}_K\right)'\frac{\mathbf{G}_{\beta}\left(\mathbf{X}\right)'\mathbf{E}}{N}\mathbf{J}_T\widetilde{\mathbf{F}}\left(\widetilde{\mathbf{F}}'\widetilde{\mathbf{F}}\right)^{-1}\\ &\xrightarrow{p}\left(\mathbf{I}_K-\mathbf{I}_K\right)'\mathbf{0}_{K\times T}\mathbf{J}_T\mathbf{F}\left(\mathbf{F}'\mathbf{J}_T\mathbf{F}\right)^{-1}=\mathbf{0}_{K\times K}\end{aligned} \tag{A.15}$$

from Step 4 and Assumption 2(ii) and that

$$\begin{aligned}\frac{1}{N}m_2'm_2 &= \left(\widetilde{\mathbf{F}}'\widetilde{\mathbf{F}}\right)^{-1}\widetilde{\mathbf{F}}'\mathbf{J}_T\mathbf{F}\left(\frac{\Gamma_{\beta}'\mathbf{X}}{N}\right)\left(\frac{\mathbf{X}'\mathbf{X}}{N}\right)^{-1}\left(\frac{\mathbf{X}'\Gamma_{\beta}}{N}\right)\mathbf{F}'\mathbf{J}_T\widetilde{\mathbf{F}}\left(\widetilde{\mathbf{F}}'\widetilde{\mathbf{F}}\right)^{-1}\\ &\xrightarrow{p}\mathbf{I}_K\mathbf{0}_{K\times L}\mathbf{V}_X^{-1}\mathbf{0}_{L\times K}\mathbf{I}_K=\mathbf{0}_{K\times K}\end{aligned} \tag{A.16}$$

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

$$\begin{aligned}\frac{1}{N}m_2'm_3 &= \left(\widetilde{\mathbf{F}}'\widetilde{\mathbf{F}}\right)^{-1}\widetilde{\mathbf{F}}'\mathbf{J}_T\mathbf{F}\left(\frac{\Gamma_{\beta}'\mathbf{X}}{N}\right)\left(\frac{\mathbf{X}'\mathbf{X}}{N}\right)^{-1}\left(\frac{\mathbf{X}'\mathbf{E}}{N}\right)\mathbf{J}_T\widetilde{\mathbf{F}}\left(\widetilde{\mathbf{F}}'\widetilde{\mathbf{F}}\right)^{-1}\\ &\xrightarrow{p}\mathbf{I}_K\mathbf{0}_{K\times L}\mathbf{V}_X^{-1}\mathbf{0}_{L\times T}\mathbf{J}_T\mathbf{F}\left(\mathbf{F}'\mathbf{J}_T\mathbf{F}\right)^{-1}=\mathbf{0}_{K\times K}\end{aligned} \tag{A.17}$$

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

$$\begin{aligned}\frac{1}{N}m_3'm_3 &= \left(\widetilde{\mathbf{F}}'\widetilde{\mathbf{F}}\right)^{-1}\widetilde{\mathbf{F}}'\mathbf{J}_T\left(\frac{\mathbf{E}'\mathbf{X}}{N}\right)\left(\frac{\mathbf{X}'\mathbf{X}}{N}\right)^{-1}\left(\frac{\mathbf{X}'\mathbf{E}}{N}\right)\mathbf{J}_T\widetilde{\mathbf{F}}\left(\widetilde{\mathbf{F}}'\widetilde{\mathbf{F}}\right)^{-1}\\ &\xrightarrow{p}\left(\mathbf{F}'\mathbf{J}_T\mathbf{F}\right)^{-1}\mathbf{F}'\mathbf{J}_T\mathbf{0}_{T\times L}\mathbf{V}_X^{-1}\mathbf{0}_{L\times T}\mathbf{J}_T\mathbf{F}\left(\mathbf{F}'\mathbf{J}_T\mathbf{F}\right)^{-1}=\mathbf{0}_{K\times K}.\end{aligned} \tag{A.18}$$

Finally, plugging the results of equations (A.13)-(A.18) into equation (A.12), 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: $\widehat{\mathbf{G}}_{\beta}\left(\mathbf{X}\right)=\widetilde{\mathbf{G}}_{\beta}\left(\mathbf{X}\right)$: See Lemma A.2(i).

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

Lemma A.4. *Consider $\widehat{\mathbf{G}}_{\beta}\left(\mathbf{X}\right)$ 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 $\frac{1}{N}\widehat{\mathbf{G}}_{\beta}\left(\mathbf{X}\right)'\mathbf{Y}$ is identical to the limit of $\frac{1}{N}\mathbf{G}_{\beta}\left(\mathbf{X}\right)'\mathbf{Y}$.*

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

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

From the Cauchy–Schwarz inequality, we have that

$$\begin{aligned}&\left(\frac{1}{N}\left(\mathbf{G}_{\beta}\left(\mathbf{X}\right)_{i}-\widehat{\mathbf{G}}_{\beta}\left(\mathbf{X}\right)_{i}\right)'\mathbf{Y}_{j}\right)^{2}\\ \leq&\frac{1}{N}\left(\mathbf{G}_{\beta}\left(\mathbf{X}\right)_{i}-\widehat{\mathbf{G}}_{\beta}\left(\mathbf{X}\right)_{i}\right)'\left(\mathbf{G}_{\beta}\left(\mathbf{X}\right)_{i}-\widehat{\mathbf{G}}_{\beta}\left(\mathbf{X}\right)_{i}\right)\left(\frac{1}{N}\mathbf{Y}_{j}'\mathbf{Y}_{j}\right)\end{aligned}$$

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 $\frac{1}{N}\left(\mathbf{G}_{\beta}\left(\mathbf{X}\right)_{i}-\widehat{\mathbf{G}}_{\beta}\left(\mathbf{X}\right)_{i}\right)'\left(\mathbf{G}_{\beta}\left(\mathbf{X}\right)_{i}-\widehat{\mathbf{G}}_{\beta}\left(\mathbf{X}\right)_{i}\right)\xrightarrow{p}0$, the above inequality implies that $\frac{1}{N}\left(\mathbf{G}_{\beta}\left(\mathbf{X}\right)_{i}-\widehat{\mathbf{G}}_{\beta}\left(\mathbf{X}\right)_{i}\right)'\mathbf{Y}_{j}\xrightarrow{p}0$. Hence, $\frac{1}{N}\mathbf{G}_{\beta}\left(\mathbf{X}\right)'\mathbf{Y}-\frac{1}{N}\widehat{\mathbf{G}}_{\beta}\left(\mathbf{X}\right)'\mathbf{Y}\xrightarrow{p}\mathbf{0}_{K\times m}$, completing the proof of the lemma. □

Lemma A.5. *Consider $\widehat{\mathbf{G}}_{\beta}\left(\mathbf{X}\right)$ in Theorem 2.1. Then, as N increases, $\frac{1}{N}\widehat{\mathbf{G}}_{\beta}\left(\mathbf{X}\right)'\mathbf{R}\xrightarrow{p}\mathbf{F}$.*

Proof From Lemma A.4 and Assumption 2(i), it suffices to show that $\frac{1}{N}\mathbf{G}_{\beta}\left(\mathbf{X}\right)'\mathbf{R}\xrightarrow{p}\mathbf{F}$. From the expression of $\mathbf{R}$ in (2.3),

$$\begin{aligned}\frac{\mathbf{G}_{\beta}\left(\mathbf{X}\right)'\mathbf{R}}{N}=&\left(\frac{\mathbf{G}_{\beta}\left(\mathbf{X}\right)'\mathbf{G}_{\alpha}\left(\mathbf{X}\right)}{N}+\frac{\mathbf{G}_{\beta}\left(\mathbf{X}\right)'\Gamma_{\alpha}}{N}\right)\mathbf{1}_{T}'\\&+\left(\frac{\mathbf{G}_{\beta}\left(\mathbf{X}\right)'\mathbf{G}_{\beta}\left(\mathbf{X}\right)}{N}+\frac{\mathbf{G}_{\beta}\left(\mathbf{X}\right)'\Gamma_{\beta}}{N}\right)\mathbf{F}'+\frac{\mathbf{G}_{\beta}\left(\mathbf{X}\right)'\mathbf{E}}{N}.\end{aligned}$$

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 A.4 completes the proof of the lemma. □

Lemma A.6. *The minimization problem in Theorem 2.2 has the following closed form solution:*

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

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

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

The first order conditions are

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

which yields

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

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

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

which completes the proof of the lemma. □

Proof of Theorems 2.2 and 2.3 Recall that $\mathbf{P} = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'$. From Lemmas A.2(ii) and A.6, we have that

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

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

$$\widehat{\mathbf{G}}_\alpha(\mathbf{X}) - \mathbf{G}_\alpha(\mathbf{X}) = n_1 + n_2 + n_3,$$

with n_i for $i = 1, 2, 3$ are given by $n_1 = \mathbf{P}\left(\Gamma_\alpha + \Gamma_\beta\overline{\mathbf{F}} + \overline{\mathbf{E}}\right)$, $n_2 = \mathbf{G}_\beta(\mathbf{X})\overline{\mathbf{F}}$, 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,

$$\frac{1}{N}\left(\widehat{\mathbf{G}}_{\alpha}\left(\mathbf{X}\right)-\mathbf{G}_{\alpha}\left(\mathbf{X}\right)\right)'\left(\widehat{\mathbf{G}}_{\alpha}\left(\mathbf{X}\right)-\mathbf{G}_{\alpha}\left(\mathbf{X}\right)\right)=\sum_{i,j=1}^{3}\frac{1}{N}n_i'n_j. \tag{A.19}$$

Note that

$$\begin{aligned}\frac{1}{N}n_1'n_1 &= \left(\frac{\mathbf{X}'\Gamma_{\alpha}}{N}+\frac{\mathbf{X}'\Gamma_{\beta}}{N}\overline{\mathbf{F}}+\frac{\mathbf{X}'\mathbf{E}}{N}\frac{\mathbf{1}_T}{T}\right)'\left(\frac{\mathbf{X}'\mathbf{X}}{N}\right)^{-1}\left(\frac{\mathbf{X}'\Gamma_{\alpha}}{N}+\frac{\mathbf{X}'\Gamma_{\beta}}{N}\overline{\mathbf{F}}+\frac{\mathbf{X}'\mathbf{E}}{N}\frac{\mathbf{1}_T}{T}\right)\\ &\xrightarrow{p}\left(\mathbf{0}_L+\mathbf{0}_{L\times K}\overline{\mathbf{F}}+\mathbf{0}_{L\times T}\frac{\mathbf{1}_T}{T}\right)'\mathbf{V}_X^{-1}\left(\mathbf{0}_L+\mathbf{0}_{L\times K}\overline{\mathbf{F}}+\mathbf{0}_{L\times T}\frac{\mathbf{1}_T}{T}\right)=0\end{aligned} \tag{A.20}$$

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

$$\begin{aligned}\frac{1}{N}n_1'n_2 &= \left(\frac{\mathbf{G}_{\beta}\left(\mathbf{X}\right)'\Gamma_{\alpha}}{N}+\frac{\mathbf{G}_{\beta}\left(\mathbf{X}\right)'\Gamma_{\beta}}{N}\overline{\mathbf{F}}+\frac{\mathbf{G}_{\beta}\left(\mathbf{X}\right)'\mathbf{E}}{N}\frac{\mathbf{1}_T}{T}\right)'\overline{\mathbf{F}}\\ &\xrightarrow{p}\left(\mathbf{0}_K+\mathbf{0}_{K\times K}\overline{\mathbf{F}}+\mathbf{0}_{K\times T}\mathbf{i}\right)'\overline{\mathbf{F}}=0\end{aligned} \tag{A.21}$$

from Assumption 2(ii) and that

$$\begin{aligned}\frac{1}{N}n_1'n_3 &= -\left(\frac{\widehat{\mathbf{G}}_{\beta}\left(\mathbf{X}\right)'\Gamma_{\alpha}}{N}+\frac{\widehat{\mathbf{G}}_{\beta}\left(\mathbf{X}\right)'\Gamma_{\beta}}{N}\overline{\mathbf{F}}+\frac{\widehat{\mathbf{G}}_{\beta}\left(\mathbf{X}\right)'\mathbf{E}}{N}\frac{\mathbf{1}_T}{T}\right)'\frac{\widehat{\mathbf{G}}_{\beta}\left(\mathbf{X}\right)'\mathbf{R}}{N}\frac{\mathbf{1}_T}{T}\\ &\xrightarrow{p}-\left(\mathbf{0}_K+\mathbf{0}_{K\times K}\overline{\mathbf{F}}+\mathbf{0}_{K\times T}\frac{\mathbf{1}_T}{T}\right)'\overline{\mathbf{F}}=0\end{aligned} \tag{A.22}$$

from Lemmas A.4 and A.5 and Assumption 2(ii) and that

$$\frac{1}{N}n_2'n_2=\overline{\mathbf{F}}'\left(\frac{\mathbf{G}_{\beta}\left(\mathbf{X}\right)'\mathbf{G}_{\beta}\left(\mathbf{X}\right)}{N}\right)\overline{\mathbf{F}}\xrightarrow{p}\overline{\mathbf{F}}'\overline{\mathbf{F}}. \tag{A.23}$$

from Assumption 3(ii) and that

$$\frac{1}{N}n_2'n_3=-\overline{\mathbf{F}}'\left(\frac{\mathbf{G}_{\beta}\left(\mathbf{X}\right)'\widehat{\mathbf{G}}_{\beta}\left(\mathbf{X}\right)}{N}\right)\frac{\widehat{\mathbf{G}}_{\beta}\left(\mathbf{X}\right)'\mathbf{R}}{N}\frac{\mathbf{1}_T}{T}\xrightarrow{p}-\overline{\mathbf{F}}'\overline{\mathbf{F}} \tag{A.24}$$

from Lemmas A.4 and A.5 and Assumption 3(ii) and that

$$\frac{1}{N}n_3'n_3=\frac{\mathbf{1}_T'}{T}\frac{\mathbf{R}'\widehat{\mathbf{G}}_{\beta}\left(\mathbf{X}\right)}{N}\frac{\widehat{\mathbf{G}}_{\beta}\left(\mathbf{X}\right)'\mathbf{R}}{N}\frac{\mathbf{1}_T}{T}\xrightarrow{p}\overline{\mathbf{F}}'\overline{\mathbf{F}} \tag{A.25}$$

from Lemma A.5. Finally, plugging the results of equations (A.20)-(A.25) into equation

(A.19), we have that

$$\frac{1}{N}\left(\widehat{\mathbf{G}}_{\alpha}(\mathbf{X})-\mathbf{G}_{\alpha}(\mathbf{X})\right)'\left(\widehat{\mathbf{G}}_{\alpha}(\mathbf{X})-\mathbf{G}_{\alpha}(\mathbf{X})\right)\xrightarrow{p}0, \tag{A.26}$$

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 zero. Let $\mathbf{R}_t$ denote the t-th column of $\mathbf{R}$. Using the Cauchy–Schwarz inequality, we have that

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

where the last limit is from (A.26) 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.

Baker, Malcolm, and Jeffrey Wurgler, 2006, Investor sentiment and the cross-section of stock returns, *The journal of Finance* 61, 1645–1680.

Ball, Ray, Joseph Gerakos, Juhani T Linnainmaa, and Valeri V Nikolaev, 2015, Deflating profitability, *Journal of Financial Economics* 117, 225–248.

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

Berk, Jonathan B, 2000, Sorting out sorts, *The Journal of Finance* 55, 407–427.

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).

Brown, Stephen J, 1989, The number of factors in security returns, *the Journal of Finance* 44, 1247–1262.

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

Chaieb, Ines, Hugues Langlois, and Olivier Scaillet, 2018, Time-varying risk premia in large international equity markets, *Swiss Finance Institute Research Paper* .

Chen, Zhuo, Bibo Liu, Huijun Wang, Zhengwei Wang, and Jianfeng Yu, 2018, Characteristics-based factors, Tsinghua University, Working Paper.

Cochrane, John H., 2011, Presidential address: Discount rates, *Journal of Finance* 66, 1047–1108.

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, 1986, Performance measurement with the arbitrage pricing theory: A new framework for analysis, *Journal of Financial Economics* 15, 373–394.

Connor, Gregory, and Robert A Korajczyk, 1988, Risk and return in an equilibrium apt: Application of a new test methodology, *Journal of Financial Economics* 21, 255–289.

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.

Datar, Vinay T, Narayan Y Naik, and Robert Radcliffe, 1998, Liquidity and stock returns: An alternative test, *Journal of Financial Markets* 1, 203–219.

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.

Feng, Guanhao, Stefano Giglio, and Dacheng Xiu, 2019, Taming the factor zoo: A test of new factors, *Fama-Miller Working Paper* 17–04.

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.

Ferson, Wayne E, and Campbell R Harvey, 1999, Conditioning variables and the cross section of stock returns, *The Journal of Finance* 54, 1325–1360.

Frazzini, Andrea, and Lasse Heje Pedersen, 2014, Betting against beta, *Journal of Financial Economics* 111, 1–25.

Freyberger, Joachim, Andreas Neuhierl, and Michael Weber, 2019, 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.

Ghysels, Eric, 1998, On stable factor structures in the pricing of risk: do time-varying betas help or hurt?, *The Journal of Finance* 53, 549–573.

Greene, William H., 2008, *Econometric Analysis*, sixth edition (Pearson Education, Inc.).

Grundy, Bruce D, and J Spencer Martin Martin, 2001, Understanding the nature of the risks and the source of the rewards to momentum investing, *The Review of Financial Studies* 14, 29–78.

Han, Yufeng, Ai He, David Rapach, and Guofu Zhou, 2018, How many firm characteristics drive us stock returns?, *Working Paper* .

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.

Jagannathan, Ravi, and Yong Wang, 2007, Lazy investors, discretionary consumption, and the cross-section of stock returns, *The Journal of Finance* 62, 1623–1661.

Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, *Journal of Finance* 48, 65–91.

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, *Journal of Financial Economics* forthcoming.

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 over short horizons, *Working Paper* .

Kothari, Shanker P, and Jay Shanken, 1992, Stock return variation and expected dividends: A time-series and cross-sectional analysis, *Journal of Financial Economics* 31, 177–210.

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.

Litzenberger, Robert H, and Krishna Ramaswamy, 1979, The effect of personal taxes and dividends on capital asset prices: Theory and empirical evidence, *Journal of Financial Economics* 7, 163–195.

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.

Pástor, L'uboš, and Robert F Stambaugh, 2003, Liquidity risk and expected stock returns, *Journal of Political economy* 111, 642–685.

Rosenberg, Barr, and Walt McKibben, 1973, The prediction of systematic and specific risk in common stocks, *The Journal of Financial and Quantitative Analysis* 8, 317–333.

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.

Sloan, Richard, 1996, Do stock prices fully reflect information in accruals and cash flows about future earnings?, *Accounting Review* 71, 289–315.

Stambaugh, Robert F., and Yu Yuan, 2016, Mispricing factors, *Review of Financial Studies* 30, 1270–1315.

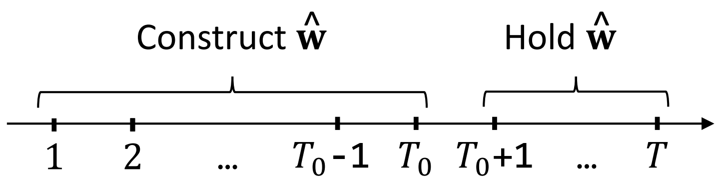
Stattman, Dennis, 1980, Book values and stock returns, *The Chicago MBA: A Journal of Selected Papers* 4, 25–45.

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).

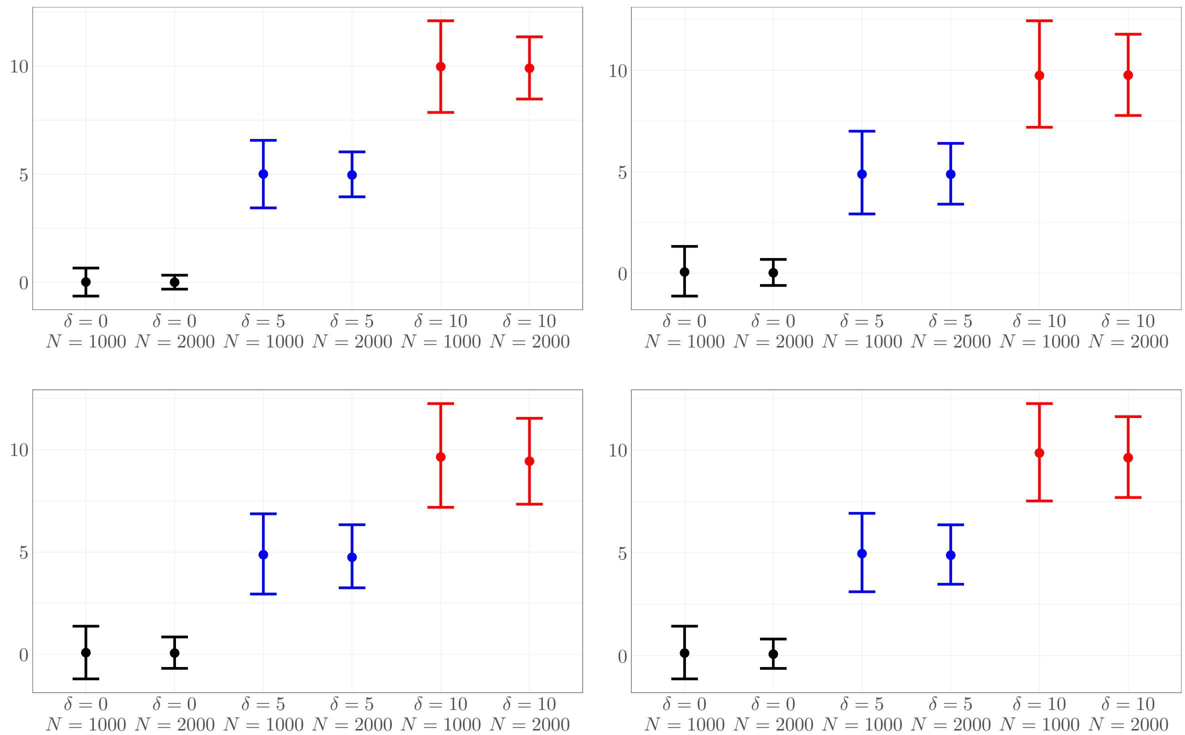
Figures and Tables

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



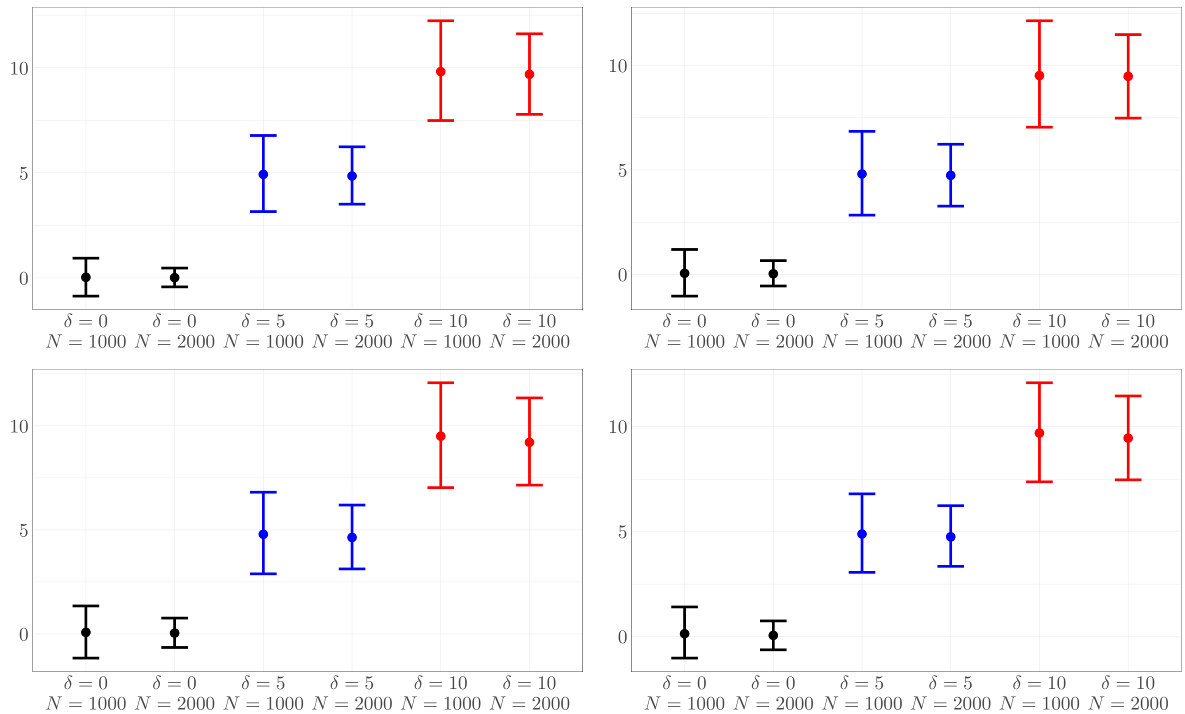
This figure illustrates how to implement the arbitrage portfolio in an out-of-sample manner. We construct $\widehat{\mathbf{w}}$ with the first set of data $t = 1, \cdots, T_0$ and hold the constructed portfolio of $\widehat{\mathbf{w}}$ over the second set of data $t = T_0 + 1, \cdots, T$ in an out-of-sample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

Figure 2: Simulated Arbitrage Portfolio Returns in the CAPM, FF 3, FF5, and HXZ4 Models (correctly specified model)



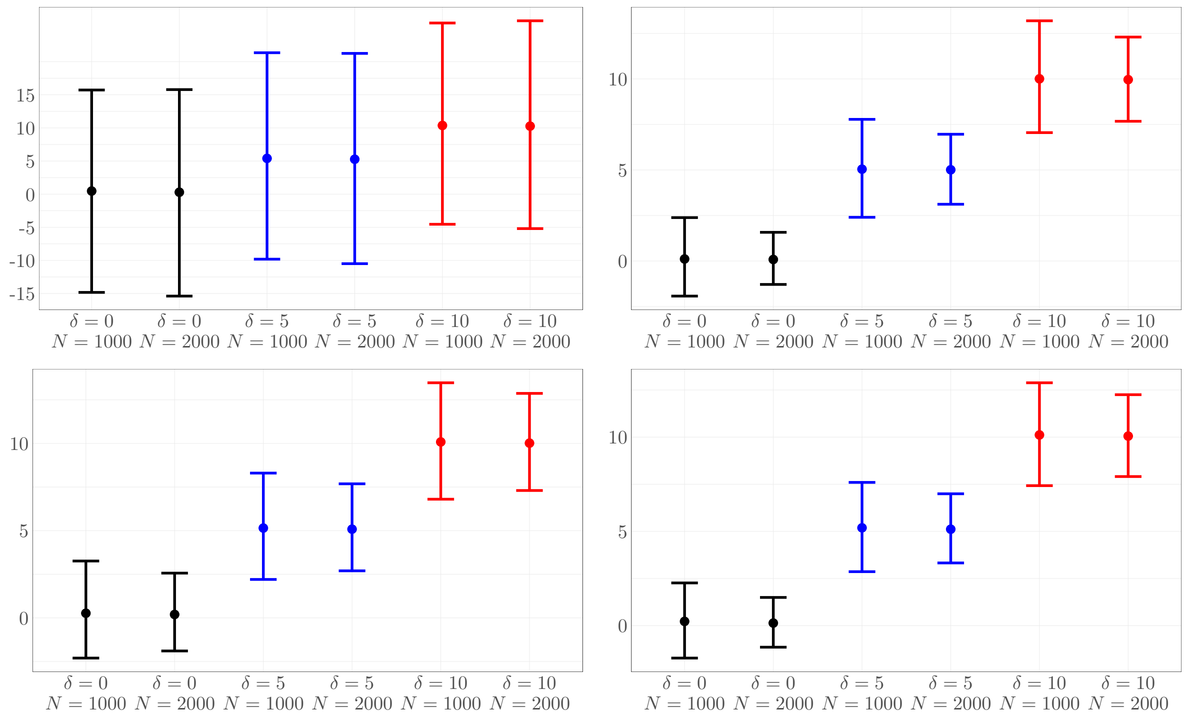
This figure shows the simulation results of the arbitrage portfolio when the return generating process is calibrated to the CAPM (upper–left panel), the Fama-French three-factor model (upper–right panel), the Fama-French five-factor model (lower–left panel) and the Hou-Xue-Zhang four-factor model (lower–right panel). The arbitrage portfolio $\widehat{\mathbf{w}}$ is constructed with the returns from $t = 1$ to $t = 12$, and it generates returns for one month out-of-sample (at a time). The solid dot represents the mean of the arbitrage portfolio in the out-of-sample period over 10,000 simulations. The error bars provide a 95% confidence interval. In this simulation, we use the correct number of factors in constructing the arbitrage portfolio, i.e. $K = 1$ for the CAPM, $K = 3$ for the Fama-French three-factor model, $K = 5$ for the Fama-French five-factor model, and $K = 4$ for the Hou-Xue-Zhang four-factor model.

Figure 3: Simulated Arbitrage Portfolio Returns in the CAPM, FF 3, FF5, and HXZ4 Models with $K_{\text{wrong}} = K_{\text{true}} + 1$ (selecting too many factors)



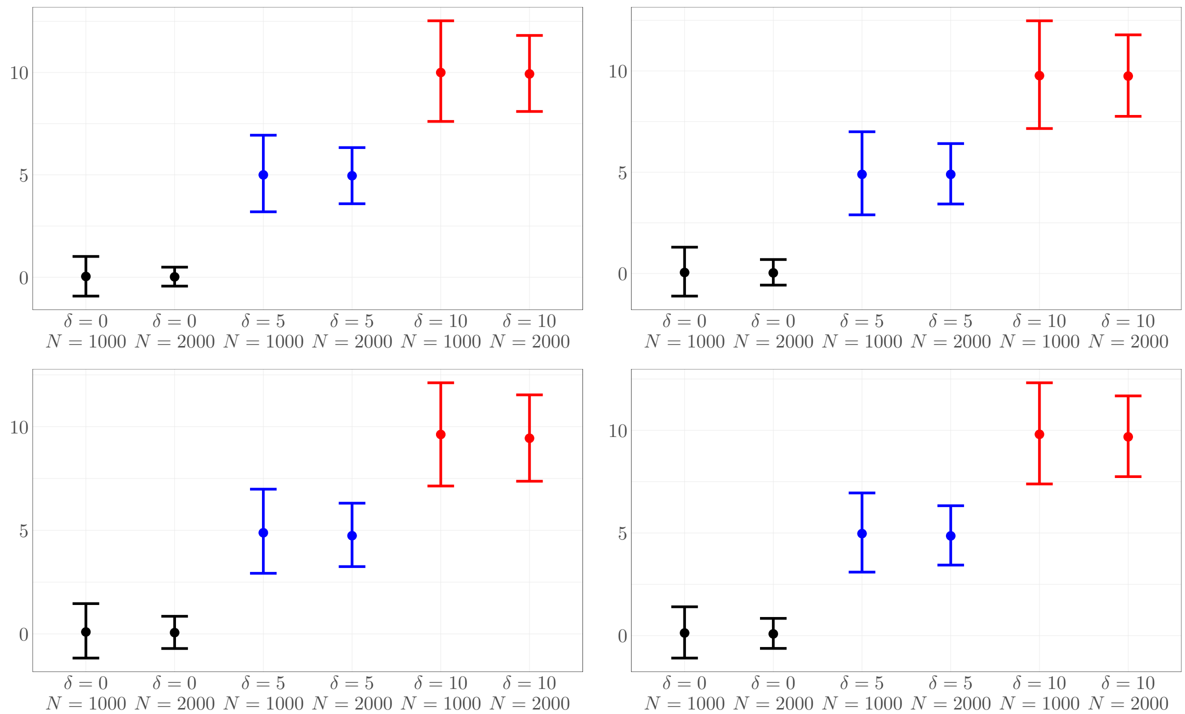
This figure shows the simulation results of the arbitrage portfolio when the return generating process is calibrated to the CAPM (upper–left panel), the Fama-French three-factor model (upper–right panel), the Fama-French five-factor model (lower–left panel) and the Hou-Xue-Zhang four-factor model (lower–right panel). The arbitrage portfolio $\widehat{\mathbf{w}}$ is constructed with the returns from $t = 1$ to $t = 12$, and it generates returns for one month out-of-sample (at a time). The solid dot represents the mean of the arbitrage portfolio in the out-of-sample period over 10,000 simulations. The error bars provide a 95% confidence interval. In this simulation, we use too many factors in constructing the arbitrage portfolio, i.e. $K_{\text{wrong}} = 2$ for the CAPM, $K_{\text{wrong}} = 4$ for the Fama-French three-factor model, $K_{\text{wrong}} = 6$ for the Fama-French five-factor model, and $K_{\text{wrong}} = 5$ for the Hou-Xue-Zhang four-factor model.

Figure 4: Simulated Arbitrage Portfolio Returns in the CAPM, FF 3, FF5, and HXZ4 Models with $K_{\text{wrong}} = K_{\text{true}} - 1$ (selecting too few factors)



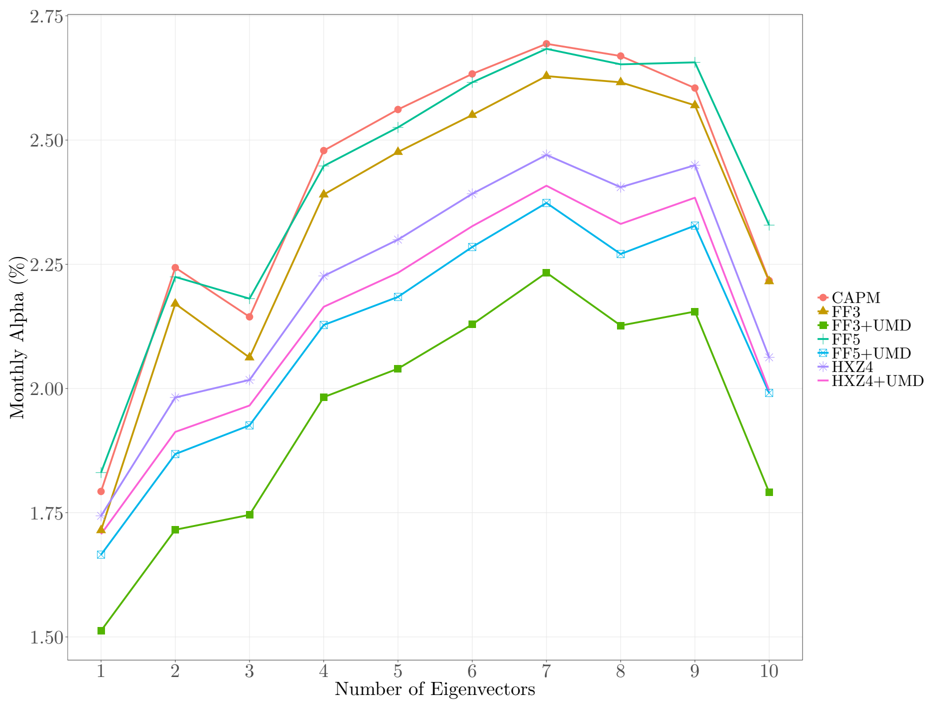
This figure shows the simulation results of the arbitrage portfolio when the return generating process is calibrated to the CAPM (upper–left panel), the Fama-French three-factor model (upper–right panel), the Fama-French five-factor model (lower–left panel) and the Hou-Xue-Zhang four-factor model (lower–right panel). The arbitrage portfolio $\widehat{\mathbf{w}}$ is constructed with the returns from $t = 1$ to $t = 12$, and it generates returns for one month out-of-sample (at a time). The solid dot represents the mean of the arbitrage portfolio in the out-of-sample period over 10,000 simulations. The error bars provide a 95% confidence interval. In this simulation, we use too few factors in constructing the arbitrage portfolio, i.e. $K_{\text{wrong}} = 0$ for the CAPM, $K_{\text{wrong}} = 2$ for the Fama-French three-factor model, $K_{\text{wrong}} = 4$ for the Fama-French five-factor model, and $K_{\text{wrong}} = 3$ for the Hou-Xue-Zhang four-factor model.

Figure 5: Simulated Arbitrage Portfolio Returns in the CAPM, FF 3, FF5, and HXZ4 Models with Time-Varying Characteristics



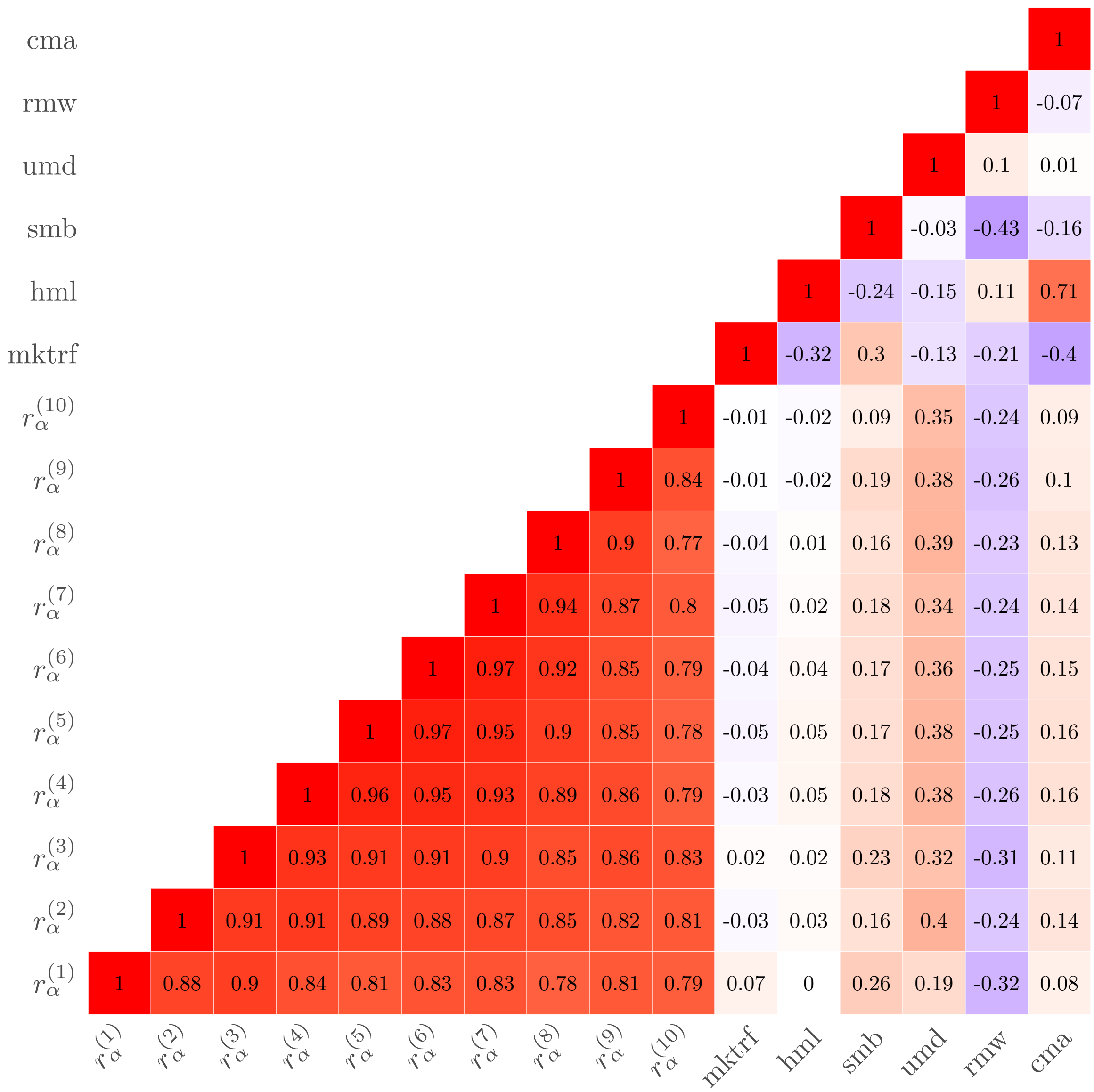
This figure shows the simulation results of the arbitrage portfolio when the return generating process is calibrated to the CAPM (upper–left panel), the Fama-French three-factor model (upper–right panel), the Fama-French five-factor model (lower–left panel) and the Hou-Xue-Zhang four-factor model (lower–right panel). The arbitrage portfolio $\widehat{\mathbf{w}}$ is constructed with the returns from $t = 1$ to $t = 12$, and it generates returns for one month out-of-sample (at a time). The solid dot represents the mean of the arbitrage portfolio in the out-of-sample period over 10,000 simulations. The error bars provide a 95% confidence interval. In this simulation, we use the correct number of factors in constructing the arbitrage portfolio, i.e. $K = 1$ for the CAPM, $K = 3$ for the Fama-French three-factor model, $K = 5$ for the Fama-French five-factor model, and $K = 4$ for the Hou-Xue-Zhang four-factor model. Time-varying characteristics are generated by fitting an AR(1) process to the empirically observed characteristics. The construction is detailed in Section 3.2.4.

Figure 6: Alpha for Varying the Number of Eigenvectors



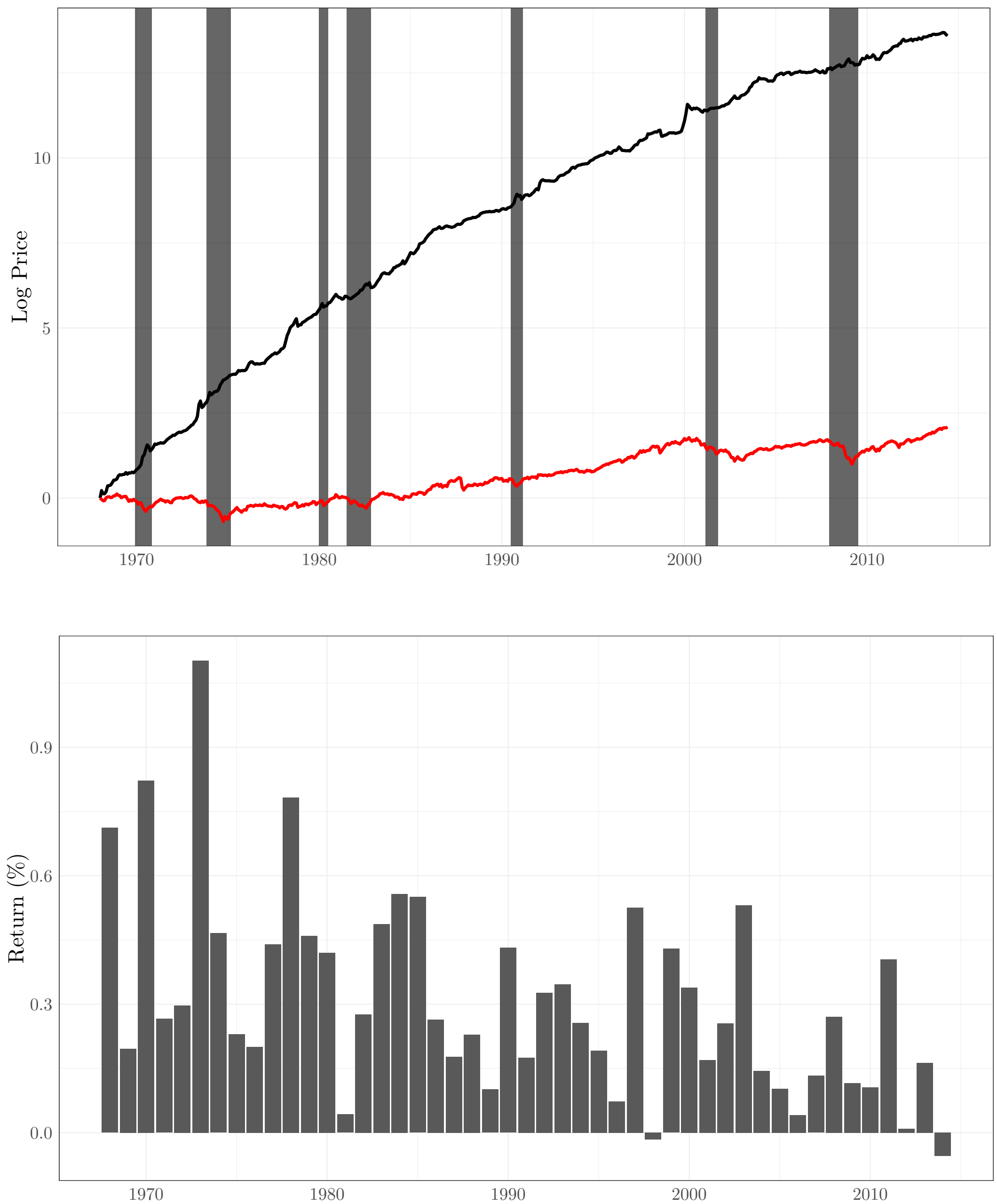
This figure shows the monthly alpha of the arbitrage portfolio against the CAPM, the Fama-French three- and five-factor model, and their "momentum augmented" versions for one through ten eigenvectors. The sample period is from January 1968 to June 2014.

Figure 7: Correlation Matrix with Common Factors



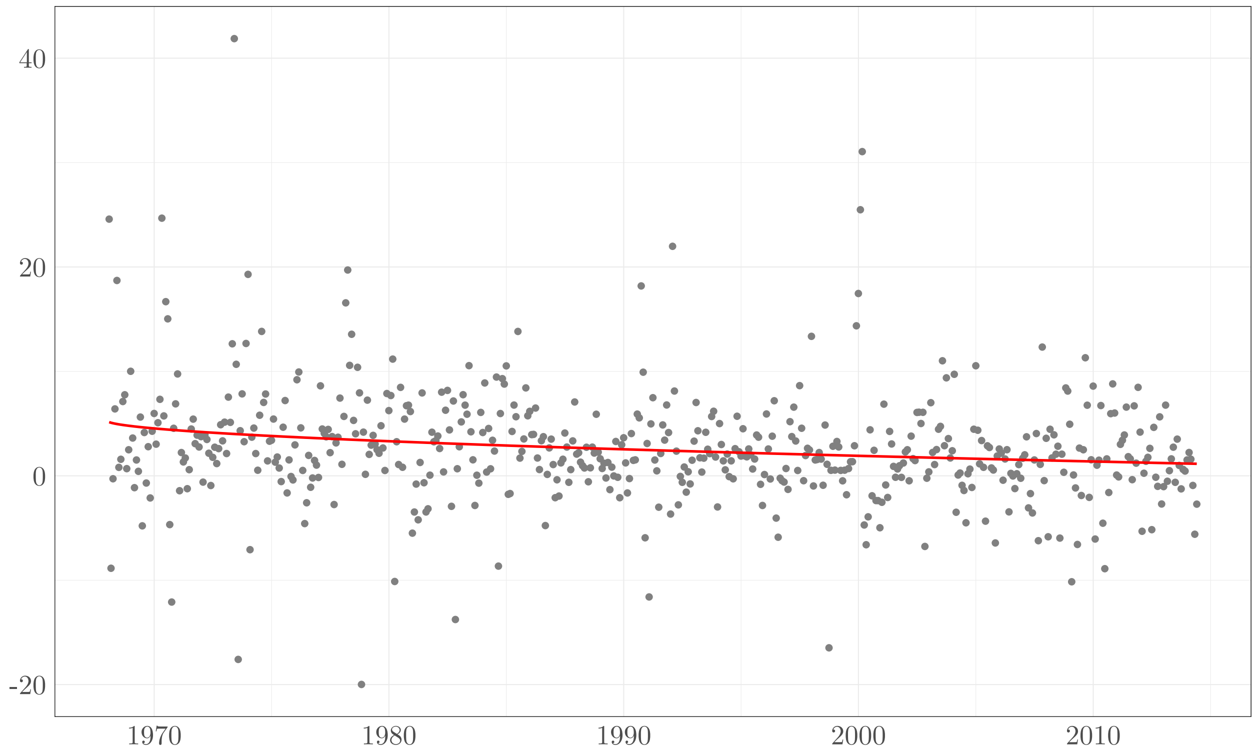
This figure shows the correlation matrix between the arbitrage portfolios with 1 through 10 eigenvectors, $r_\alpha^{(1)}$, $r_\alpha^{(2)}, \ldots, r_\alpha^{(10)}$, and the Fama-French three and five factors as well as the momentum factor. The sample period is January 1968 to June 2014.

Figure 8: Price Path and Yearly Returns of the Arbitrage Portfolio



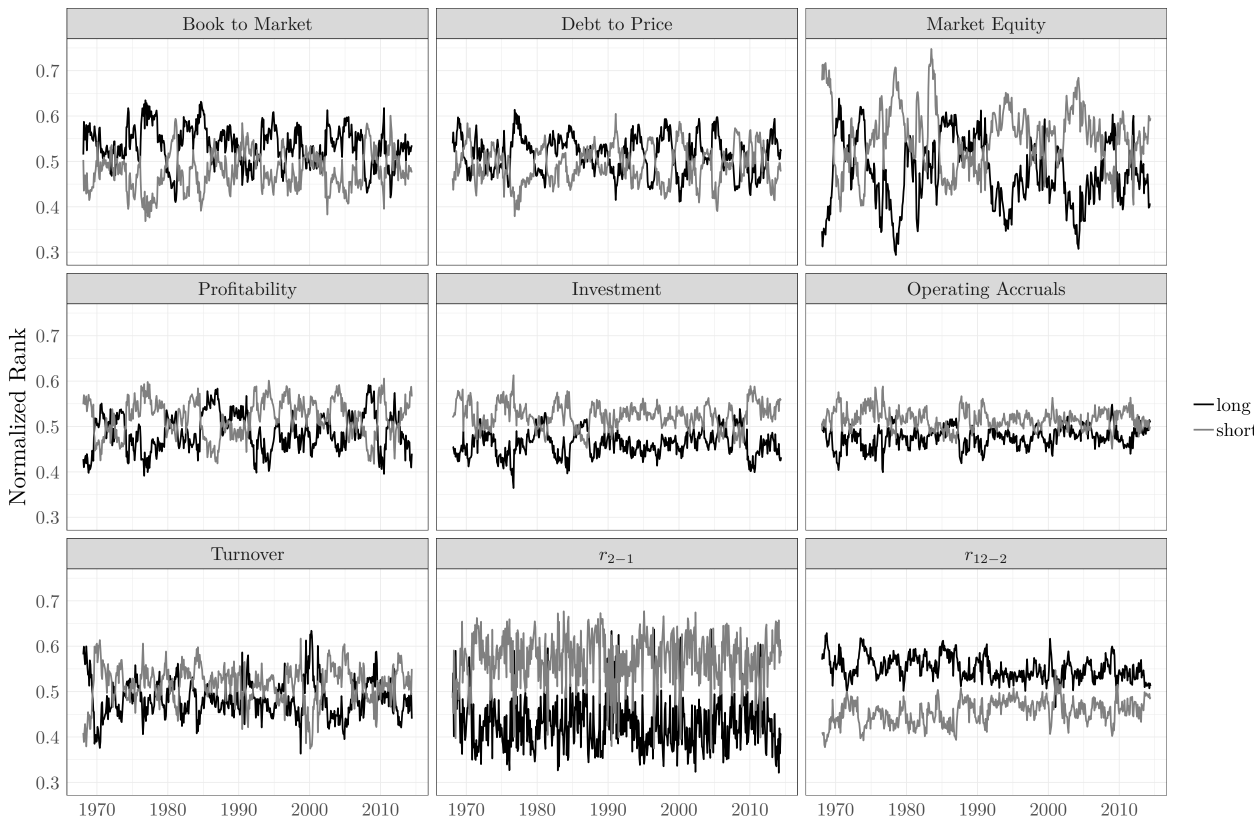
The top panel of the figure shows the logarithmic price path (i.e., the cumulative returns) of the arbitrage portfolio (using six eigenvectors) in black line and the market portfolio in red line. The areas shaded in gray depict NBER recessions. The lower panel shows the yearly returns of the arbitrage portfolio (with six eigenvectors). The sample period is January 1968 to June 2014.

Figure 9: Monthly Returns of the Arbitrage Portfolio 1968–2014



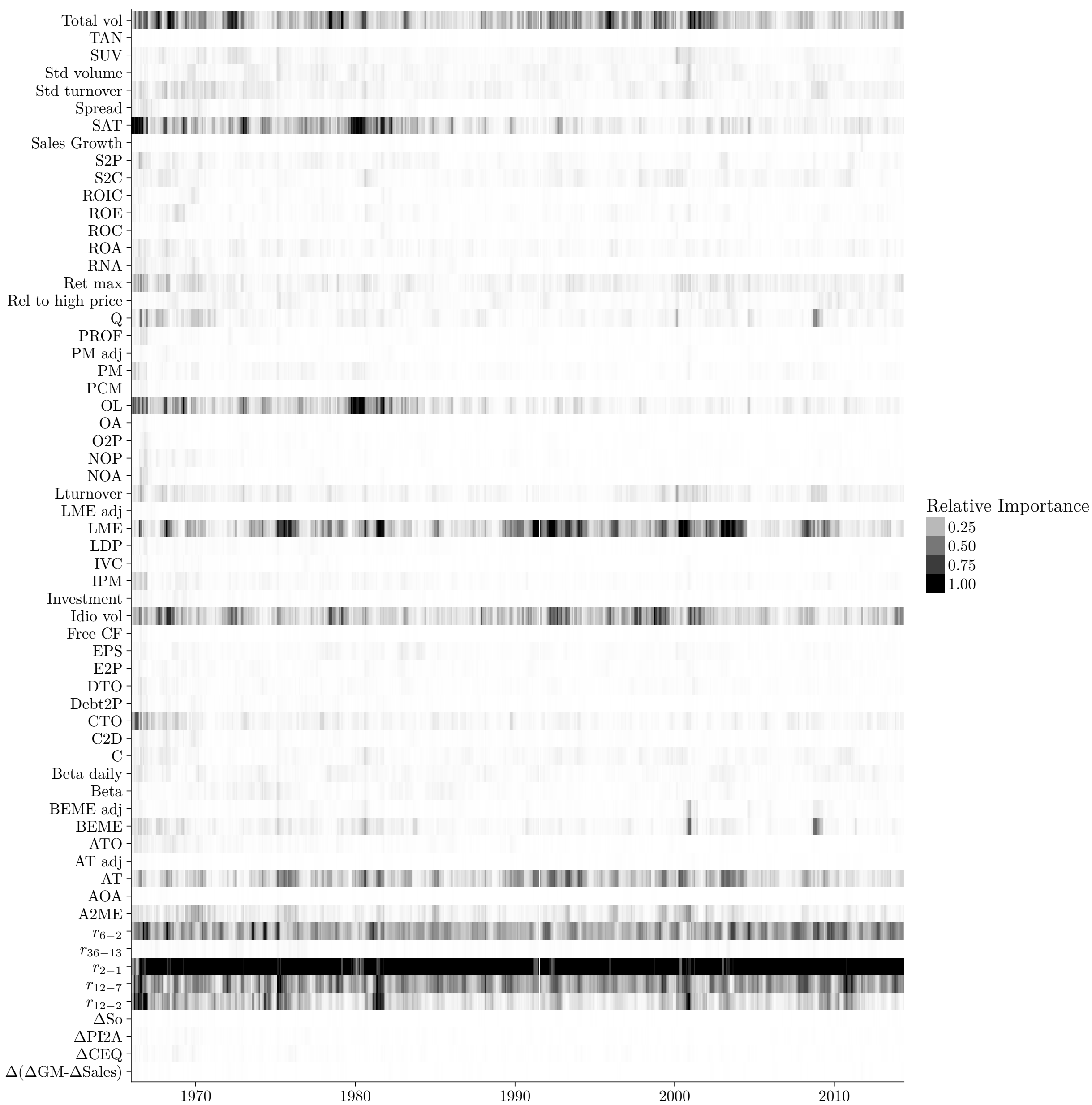
This figure shows the monthly excess returns of the arbitrage portfolio (six eigenvectors) from January 1968 through June 2014 and a time trend (red). The time trend is estimated by $r_t = a + b \times t^{\gamma} + \varepsilon_t$ with $\widehat{a} = 5.27$, $\widehat{b} = -0.127$, $\widehat{\gamma} = 0.5501$.

Figure 10: Firm Characteristics of the Long and Short Leg of the Arbitrage Portfolio



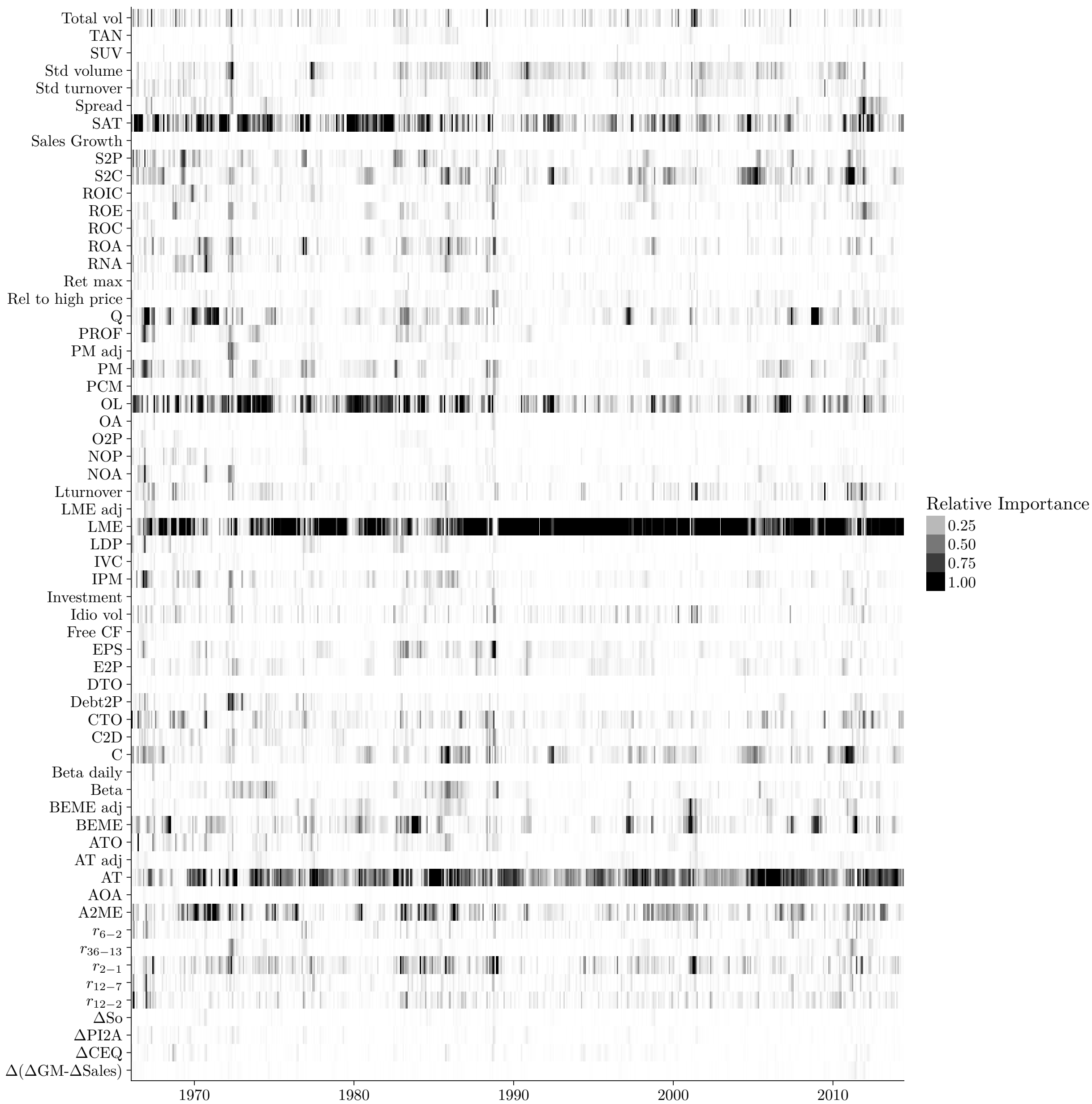
This figure shows the normalized rank of nine cross-sectional return characteristics for the long and short leg of the arbitrage portfolio. The firm characteristics are the book-to-market ratio, the debt-to-price ratio, market equity (size), profitability, investment, operating accruals, last month's volume, the return one month before portfolio formation (r_{2-1}) and the return from 12 to 2 month before portfolio formation (r_{12-2}). Each month, the characteristics are normalized to be in the unit interval, i.e., the normalized characteristics is computed as $\tilde{c}_{i,t} = \frac{\text{rank}(c_{it})}{N_t+1}$, where c_{it} denotes the "raw" characteristic value and N_t denotes the number of firms in month t. The rank normalization facilitates an easy comparison cross-sectionally and over time. The sample period is January 1968 to June 2014.

Figure 11: Beta Heatmap



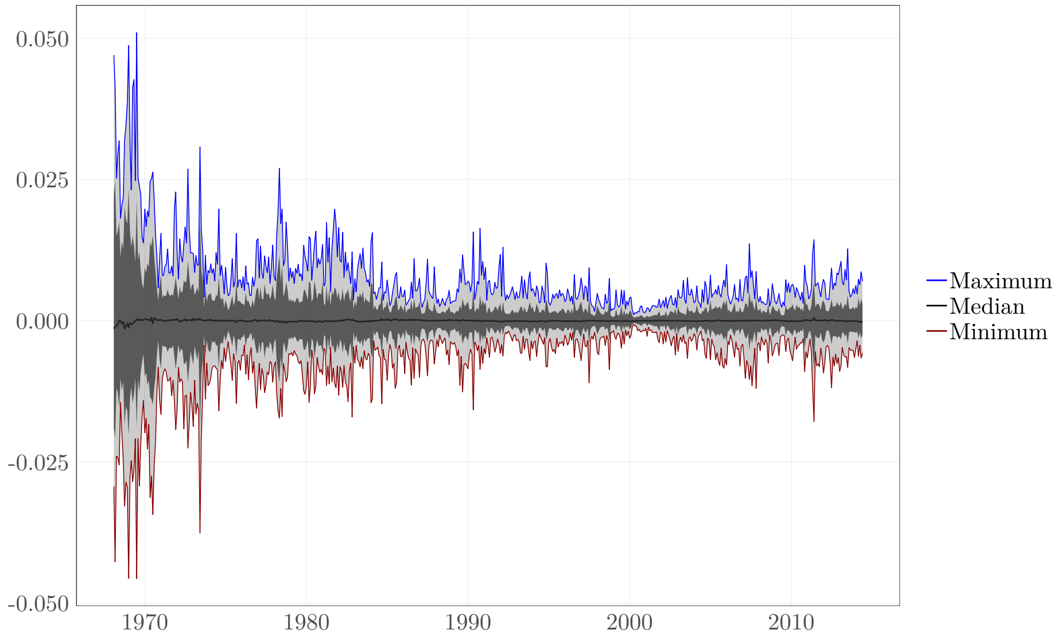
This figure plots beta-heatmap $\widehat{\beta}^{\texttt{norm}}_{(l)}$ for each characteristic l. We compute $\widehat{\beta}^{\texttt{norm}}_{(l)}$ as follows. We project the k-th column of $\widehat{\mathbf{G}}_{\beta}(\mathbf{X})$ onto the characteristics at each month: $\widehat{\mathbf{G}}_{\beta}(\mathbf{X})_k = \beta_{0,k} + \mathbf{X}\boldsymbol{\beta}_k + \varepsilon$. We then take absolute value of $\widehat{\boldsymbol{\beta}}_k$ for each characteristic and compute $\widehat{\beta}_{(l)} = \sum_{k=1}^{K} |\widehat{\boldsymbol{\beta}}_{k,l}|$. We then normalize cross-sectionally to have the normalized sum of absolute coefficients $\widehat{\beta}^{\texttt{norm}}_{(j)} = \frac{\widehat{\beta}_{(j)}}{\max_j \widehat{\beta}_{(j)}}$. This way, the characteristic with the largest (absolute) sum of coefficients gets a 1. We then repeat this process each month, sliding the estimation window forward. The sample period is January 1968 to June 2014.

Figure 12: Alpha Heatmap



This figure plots beta-heatmap $\widehat{\alpha}_{(l)}^{\texttt{norm}}$ for each characteristic l. We compute $\widehat{\alpha}_{(l)}^{\texttt{norm}}$ as follows. We project $\widehat{\mathbf{G}}_{\alpha}(\mathbf{X})$ onto the characteristics at each month: $\widehat{\mathbf{G}}_{\alpha}(\mathbf{X}) = \alpha_0 + \mathbf{X}\boldsymbol{\alpha} + \varepsilon$. We then take absolute value of $\widehat{\boldsymbol{\alpha}}$ for each characteristic and compute $\widehat{\alpha}_{(l)} = |\widehat{\boldsymbol{\alpha}}_l|$. We then normalize cross-sectionally to have the normalized sum of absolute coefficients $\widehat{\alpha}_{(j)}^{\texttt{norm}} = \frac{\widehat{\alpha}_{(j)}}{\max_j \widehat{\alpha}_{(j)}}$. This way, the characteristic with the largest (absolute) sum of coefficients gets a 1. We then repeat this process each month, sliding the estimation window forward. The sample period is January 1968 to June 2014.

Figure 13: Portfolio Weights



This figure shows the median, minimum, maximum, and the 5% and 95% quantiles of the portfolio weights of the arbitrage portfolio (with five eigenvectors). The solid black line is the median portfolio weight in a given month, the dark-gray area depicts the 5% and 95% quantiles of the weights in a month and the light-gray area depicts the monthly minimum and maximum. The sample period is January 1968 to June 2014.

Table 1: Average Returns on Double-Sorted Portfolio in a Simulated CAPM Economy

			Past Beta										
	Characteristic		Low 1	2	3	4	5	6	7	8	9	High 10	10-1
Low	1	0.24	0.25	0.29	0.34	0.26	0.15	0.34	0.25	0.18	0.14	0.23	-0.03
	2	0.36	0.40	0.32	0.39	0.36	0.37	0.39	0.29	0.32	0.25	0.47	0.06
	3	0.42	0.37	0.41	0.42	0.41	0.46	0.28	0.47	0.46	0.48	0.46	0.09
	4	0.45	0.46	0.45	0.45	0.36	0.42	0.44	0.39	0.45	0.58	0.47	0.02
	5	0.47	0.48	0.37	0.48	0.45	0.52	0.51	0.47	0.48	0.44	0.47	-0.01
	6	0.53	0.56	0.51	0.61	0.43	0.47	0.56	0.59	0.47	0.56	0.55	-0.01
	7	0.58	0.52	0.54	0.61	0.58	0.60	0.60	0.51	0.56	0.71	0.59	0.07
	8	0.59	0.56	0.54	0.56	0.63	0.60	0.46	0.66	0.70	0.62	0.56	0.01
	9	0.67	0.67	0.68	0.59	0.71	0.66	0.64	0.71	0.69	0.67	0.70	0.03
High	10	0.78	0.78	0.74	0.68	0.83	0.75	0.74	0.85	0.78	0.80	0.86	0.08
	10-1	0.54***	0.52***	0.45***	0.33**	0.57***	0.61***	0.40***	0.60***	0.60***	0.66***	0.63***	

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

This table reports average returns of double-sorted (first on characteristic and then on the estimated beta using past 60 month returns) portfolios. We simulate excess returns $R_{i,t}$ for $i = 1, \cdots, 2000$ and $t = 1, \cdots, 2000$ with the following calibration: $f_{M,t} \sim \mathcal{N}\left(\mu_M, \sigma_M^2\right)$, $\beta_i \sim \mathcal{N}\left(1, \sigma_\beta^2\right)$, $\varepsilon_{i,t} \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right)$, where $\mu_M = 5\%/12$, $\sigma_M = \sqrt{(20\%)^2/12}$, $\sigma_\beta = 0.4$, $\sigma_\varepsilon = 2\sigma_M$. Reported numbers are the averages over $t = 61, \cdots, 2000$.

Table 2: Firm Characteristics by Category

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

This is a reproduction of Table 1 in Freyberger et al. (2019). 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 data set (CRSP, Compustat). The sample period is January 1965 to June 2014.

Table 3: Portfolio Performance Statistics

# Eigenvectors	Mean (%)	Standard Deviation (%)	Sharpe Ratio	Skewness	Kurtosis	Maximum Drawdown	Worst Month (%)	Best Month (%)
1	21.91	15.19	1.44	1.15	7.69	22.37	-18.77	29.46
2	26.70	17.97	1.49	0.48	5.86	23.16	-22.44	30.06
3	25.86	15.12	1.71	1.20	9.66	20.91	-19.89	34.05
4	29.54	16.90	1.75	1.07	6.57	22.21	-19.61	30.66
5	30.38	18.25	1.66	1.05	7.10	20.08	-20.08	36.16
6	31.30	18.72	1.67	1.26	8.71	20.84	-19.99	41.88
7	31.98	18.61	1.72	1.29	9.06	21.92	-20.21	42.82
8	31.72	20.10	1.58	1.24	10.49	27.92	-26.02	42.74
9	31.18	17.84	1.75	1.39	8.58	23.74	-20.43	36.56
10	26.51	19.58	1.35	0.18	15.17	38.52	-38.52	41.19

This table reports annualized percentage means, annualized percentage standard deviations, annualized Sharpe Ratios, skewness, kurtosis, the maximum drawdown, and the best and worst month returns. The arbitrage 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 4: Risk-Adjusted Returns with One Eigenvector

	CAPM	FF3	FF3+UMD	FF5	FF5+UMD	HXZ4	HXZ4+UMD
alpha	1.79***	1.71***	1.51***	1.83***	1.67***	1.74***	1.71***
	(0.22)	(0.22)	(0.21)	(0.25)	(0.25)	(0.27)	(0.28)
mktrf	0.07	0.01	0.05	0.01	0.04		
	(0.08)	(0.07)	(0.06)	(0.07)	(0.06)		
smb		0.38**	0.39***	0.24	0.24**		
		(0.18)	(0.15)	(0.15)	(0.12)		
hml		0.10	0.18	0.01	0.14		
		(0.14)	(0.15)	(0.15)	(0.14)		
umd			0.23*		0.25**		0.35***
			(0.12)		(0.11)		(0.11)
rmw				−0.48***	−0.54***		
				(0.18)	(0.14)		
cma				0.20	0.08		
				(0.23)	(0.19)		
mkt						0.03	0.05
						(0.07)	(0.06)
me						0.30*	0.22*
						(0.18)	(0.12)
ia						0.18	0.16
						(0.23)	(0.21)
roe						−0.18	−0.50***
						(0.17)	(0.14)
Adj. R^2	0.00	0.07	0.12	0.12	0.18	0.07	0.15
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 factors by Fama and French (1993), Carhart (1997), Fama and French (2015) and the q-factor model (HXZ4) by Hou et al. (2015). The arbitrage 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 5: Risk-Adjusted Returns with Six Eigenvectors

	CAPM	FF3	FF3+UMD	FF5	FF5+UMD	HXZ4	HXZ4+UMD
alpha	2.63***	2.55***	2.13***	2.62***	2.28***	2.39***	2.33***
	(0.29)	(0.30)	(0.25)	(0.33)	(0.28)	(0.35)	(0.30)
mktrf	−0.05	−0.10	−0.01	−0.09	−0.02		
	(0.08)	(0.07)	(0.06)	(0.07)	(0.06)		
smb		0.36*	0.38**	0.21	0.20		
		(0.20)	(0.15)	(0.15)	(0.13)		
hml		0.11	0.27*	−0.07	0.20		
		(0.14)	(0.16)	(0.17)	(0.15)		
umd			0.48***		0.49***		0.61***
			(0.12)		(0.10)		(0.11)
rmw				−0.48**	−0.61***		
				(0.20)	(0.16)		
cma				0.41	0.16		
				(0.26)	(0.23)		
mkt						−0.06	−0.02
						(0.08)	(0.06)
me						0.35*	0.19
						(0.21)	(0.14)
ia						0.37	0.35
						(0.27)	(0.24)
roe						−0.04	−0.61***
						(0.18)	(0.15)
Adj. R^2	0.00	0.04	0.18	0.08	0.23	0.05	0.22
Num. obs.	557	557	557	557	557	557	557

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

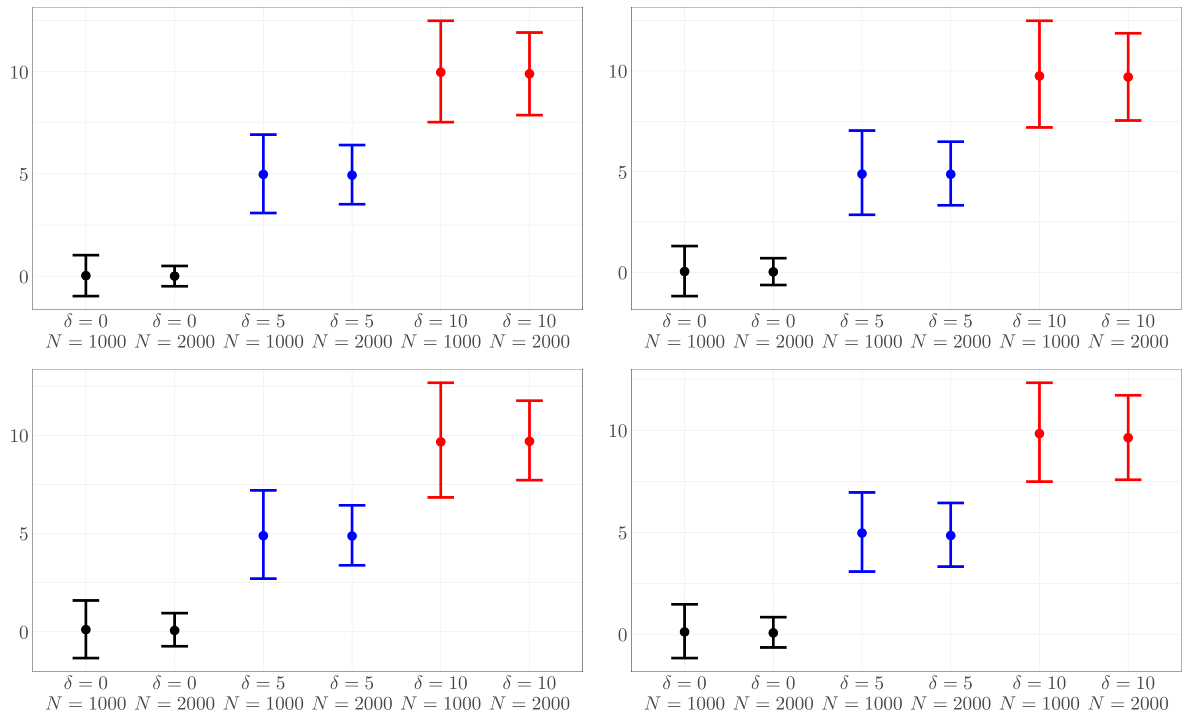
This table reports alphas (%/month) and factor loadings on the factors by the Fama and French (1993), Carhart (1997), Fama and French (2015) and the q-factor model (HXZ4) by Hou et al. (2015). The arbitrage portfolio with six eigenvectors is estimated every month using the steps outlined in Section 4. Newey and West (1987) standard errors are given in parentheses. The sample period is January 1968 to June 2014.

Arbitrage Portfolio in Large Panels
Online Appendix

Not for Publication

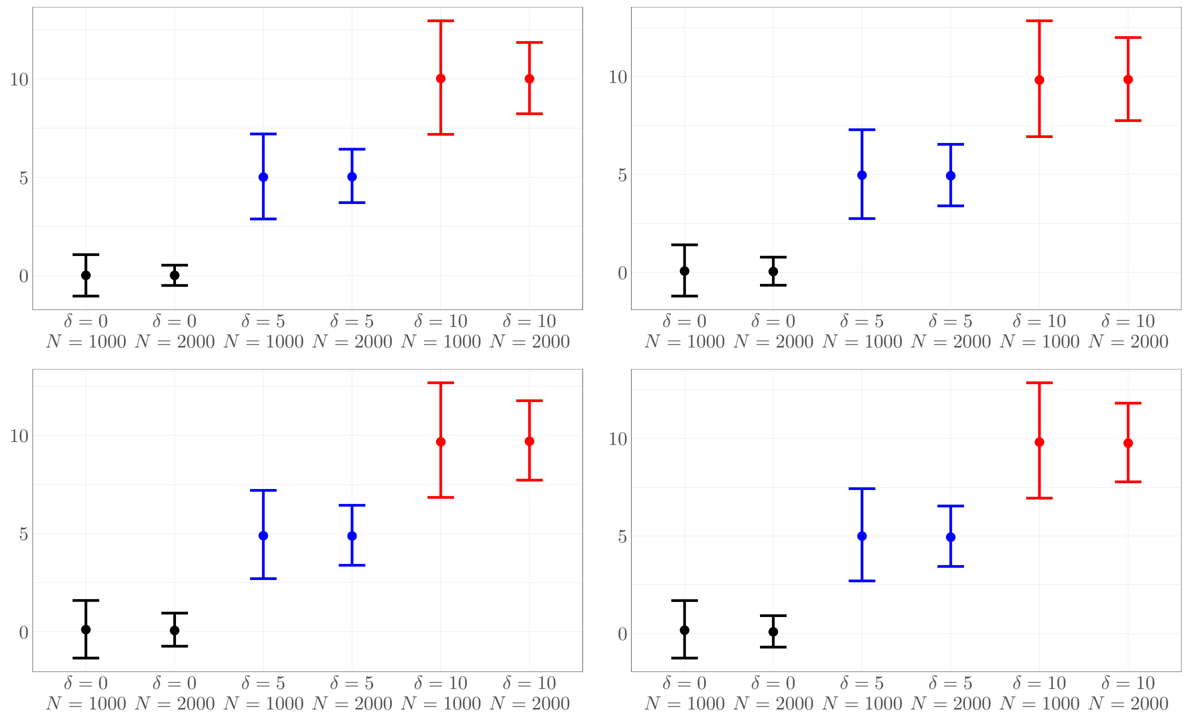
Additional Figures and Tables

Figure A.1: Simulated Arbitrage Portfolio Returns in the CAPM, FF 3, FF5, and HXZ4 Models (correlated errors)



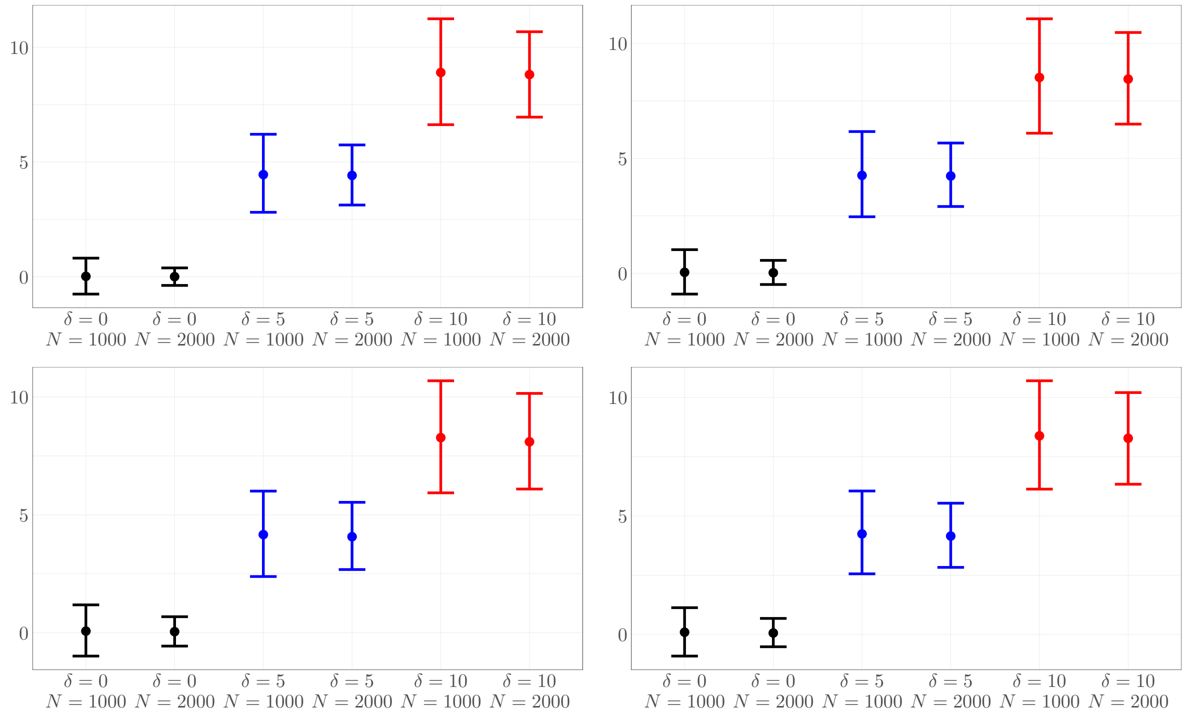
This figure shows the simulation results of the arbitrage portfolio when the return generating process is calibrated to the CAPM (upper–left panel), the Fama-French three-factor model (upper–right panel), the Fama-French five-factor model (lower–left panel) and the Hou-Xue-Zhang four-factor model (lower–right panel). The arbitrage portfolio $\widehat{\mathbf{w}}$ is constructed with the returns from $t = 1$ to $t = 12$, and it generates returns for one month out-of-sample (at a time). The solid dot represents the mean of the arbitrage portfolio in the out-of-sample period over 10,000 simulations. The error bars provide a 95% confidence interval. In this simulation, we use the correct number of factors in constructing the arbitrage portfolio, i.e. $K = 1$ for the CAPM, $K = 3$ for the Fama-French three-factor model, $K = 5$ for the Fama-French five-factor model, and $K = 4$ for the Hou-Xue-Zhang four-factor model. We generated correlated errors, by creating industry clusters, with "within-industry correlation". Details of the data-generation are given in Section 3.2.4.

Figure A.2: Simulated Arbitrage Portfolio Returns in the CAPM, FF 3, FF5, and HXZ4 Models (calibration period 2006 - 2008)



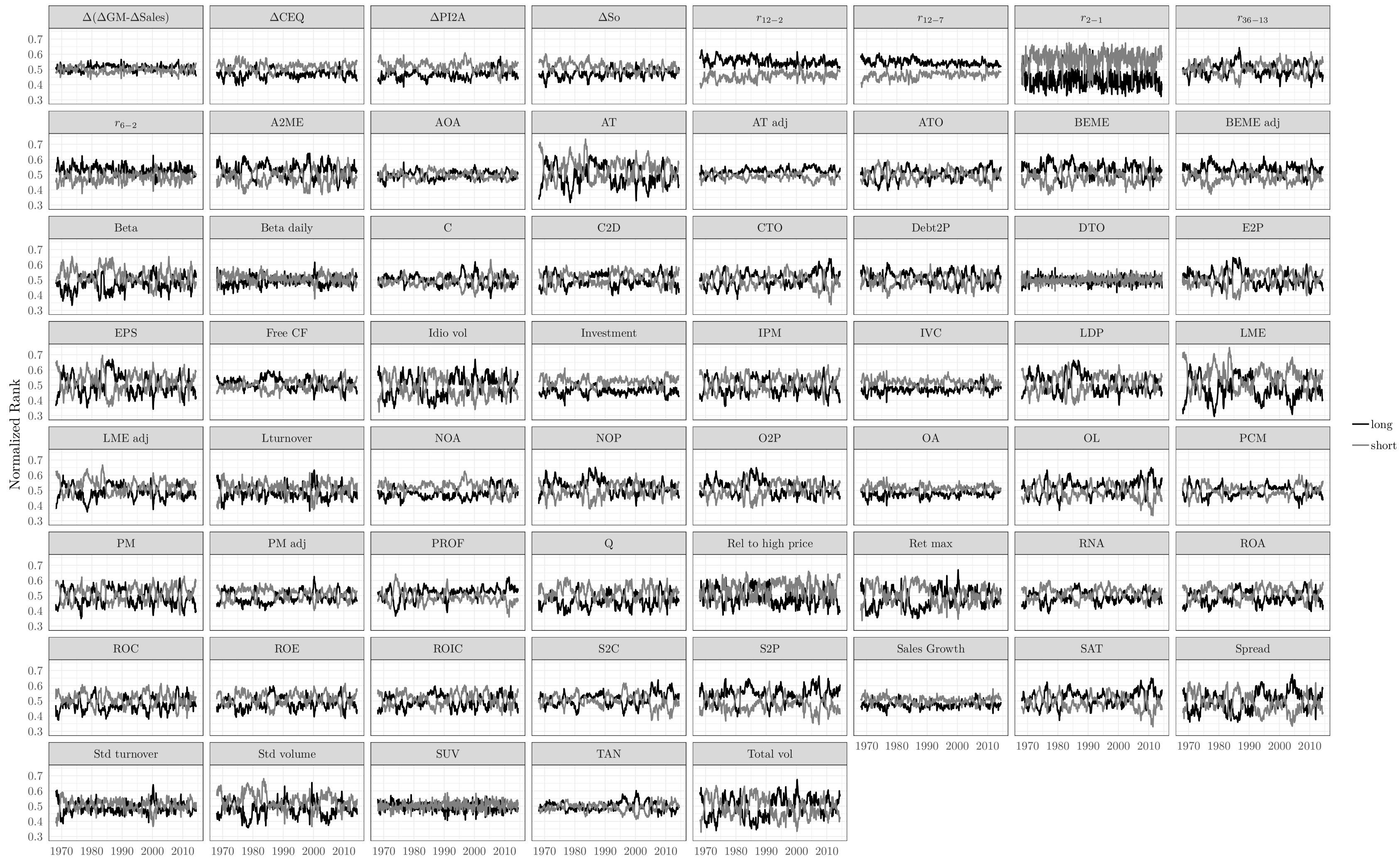
This figure shows the simulation results of the arbitrage portfolio when the return generating process is calibrated to the CAPM (upper–left panel), the Fama-French three-factor model (upper–right panel), the Fama-French five-factor model (lower–left panel) and the Hou-Xue-Zhang four-factor model (lower–right panel). The arbitrage portfolio $\widehat{\mathbf{w}}$ is constructed with the returns from $t = 1$ to $t = 12$, and it generates returns for one month out-of-sample (at a time). The solid dot represents the mean of the arbitrage portfolio in the out-of-sample period over 10,000 simulations. In this simulation, we use the correct number of factors in constructing the arbitrage portfolio, i.e. $K = 1$ for the CAPM, $K = 3$ for the Fama-French three-factor model, $K = 5$ for the Fama-French five-factor model, and $K = 4$ for the Hou-Xue-Zhang four-factor model. For this figure, we calibrate the parameters of the economy using the data from 2006 through 2008 to cover the parts of more volatile recent financial crisis.

Figure A.3: Simulated Arbitrage Portfolio Returns in the CAPM, FF 3, FF5, and HXZ4 Models with Missing Characteristics



This figure shows the simulation results of the arbitrage portfolio when the return generating process is calibrated to the CAPM (upper–left panel), the Fama-French three-factor model (upper–right panel), the Fama-French five-factor model (lower–left panel) and the Hou-Xue-Zhang four-factor model (lower–right panel). The arbitrage portfolio $\widehat{\mathbf{w}}$ is constructed with the returns from $t = 1$ to $t = 12$, and it generates returns for one month out-of-sample (at a time). The solid dot represents the mean of the arbitrage portfolio in the out-of-sample period over 10,000 simulations. The error bars provide a 95% confidence interval. In this simulation, we use the correct number of factors in constructing the arbitrage portfolio, i.e. $K = 1$ for the CAPM, $K = 3$ for the Fama-French three-factor model, $K = 5$ for the Fama-French five-factor model, and $K = 4$ for the Hou-Xue-Zhang four-factor model. For each repetition, we use 61 characteristics for simulating returns but drop randomly picked ten characteristics for computing $\widehat{\mathbf{w}}$. The construction is detailed in Section 3.2.4.

Figure A.4: Firm Characteristics of the Arbitrage Portfolio



The figure shows the normalized rank of the full set of cross-sectional return characteristics for the long and short leg of the arbitrage portfolio for all the characteristics used in the empirical analysis and described in Table 2. Each months, the characteristics are normalized to be in the unit interval, i.e. the normalized characteristics is computed as, $\widetilde{c}_{it} = \frac{\text{rank}(c_{it})}{N_t+1}$, where c_{it} denotes the "raw" characteristic value and N_t denotes the number of firms in month t. The rank normalization facilitates and easy comparison cross-sectionally and over time. The sample period is January 1968 to June 2014.

Table A.1: Risk-Adjusted Returns with Two Eigenvectors

	CAPM	FF3	FF3+UMD	FF5	FF5+UMD	HXZ4	HXZ4+UMD
alpha	2.24^{***}	2.17^{***}	1.72^{***}	2.22^{***}	1.87^{***}	1.98^{***}	1.91^{***}
	(0.26)	(0.26)	(0.25)	(0.28)	(0.27)	(0.31)	(0.28)
mktrf	−0.04	−0.09	0.01	−0.07	0.00		
	(0.08)	(0.07)	(0.06)	(0.08)	(0.06)		
smb		0.33	0.35^{**}	0.19	0.17		
		(0.21)	(0.14)	(0.18)	(0.12)		
hml		0.10	0.27^{*}	−0.09	0.20		
		(0.15)	(0.14)	(0.17)	(0.13)		
umd			0.51^{***}		0.53^{***}		0.65^{***}
			(0.13)		(0.12)		(0.12)
rmw				-0.45^{**}	-0.59^{***}		
				(0.20)	(0.14)		
cma				0.41^{*}	0.15		
				(0.24)	(0.18)		
mkt						−0.03	0.01
						(0.08)	(0.07)
me						0.31	0.14
						(0.22)	(0.13)
ia						0.38	0.35
						(0.26)	(0.23)
roe						−0.01	-0.61^{***}
						(0.18)	(0.14)
Adj. R^2	-0.00	0.03	0.21	0.08	0.26	0.04	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 factors by the Fama and French (1993), Carhart (1997), Fama and French (2015) and the q-factor model (HXZ4) by Hou et al. (2015). The arbitrage 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 A.2: Risk-Adjusted Returns with Three Eigenvectors

	CAPM	FF3	FF3+UMD	FF5	FF5+UMD	HXZ4	HXZ4+UMD
alpha	2.14^{***}	2.06^{***}	1.75^{***}	2.18^{***}	1.93^{***}	2.02^{***}	1.97^{***}
	(0.24)	(0.25)	(0.22)	(0.27)	(0.25)	(0.30)	(0.26)
mktrf	0.02	−0.03	0.04	−0.03	0.02		
	(0.07)	(0.06)	(0.05)	(0.06)	(0.05)		
smb		0.36^{*}	0.37^{**}	0.21	0.21^{*}		
		(0.20)	(0.15)	(0.15)	(0.12)		
hml		0.11	0.23^{*}	0.01	0.22^{*}		
		(0.14)	(0.14)	(0.15)	(0.13)		
umd			0.36^{***}		0.38^{***}		0.49^{***}
			(0.12)		(0.11)		(0.12)
rmw				-0.49^{**}	-0.59^{***}		
				(0.20)	(0.16)		
cma				0.21	0.02		
				(0.22)	(0.19)		
mkt						−0.01	0.02
						(0.06)	(0.05)
me						0.32^{*}	0.20
						(0.19)	(0.13)
ia						0.24	0.22
						(0.24)	(0.22)
roe						−0.11	-0.56^{***}
						(0.16)	(0.13)
Adj. R^2	-0.00	0.06	0.18	0.12	0.25	0.06	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 factors by the Fama and French (1993), Carhart (1997), Fama and French (2015) and the q-factor model (HXZ4) by Hou et al. (2015). The arbitrage 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 A.3: Risk-Adjusted Returns with Four Eigenvectors

	CAPM	FF3	FF3+UMD	FF5	FF5+UMD	HXZ4	HXZ4+UMD
alpha	2.48^{***}	2.39^{***}	1.98^{***}	2.45^{***}	2.13^{***}	2.23^{***}	2.16^{***}
	(0.26)	(0.27)	(0.23)	(0.29)	(0.24)	(0.30)	(0.25)
mktrf	-0.04	-0.08	0.01	-0.06	-0.00		
	(0.07)	(0.07)	(0.06)	(0.07)	(0.06)		
smb		0.35^{*}	0.36^{***}	0.20	0.19^{*}		
		(0.18)	(0.14)	(0.14)	(0.12)		
hml		0.13	0.28^{**}	-0.05	0.21		
		(0.13)	(0.14)	(0.16)	(0.13)		
umd			0.46^{***}		0.48^{***}		0.59^{***}
			(0.10)		(0.09)		(0.10)
rmw				-0.44^{**}	-0.57^{***}		
				(0.19)	(0.15)		
cma				0.39^{*}	0.16		
				(0.23)	(0.20)		
mkt						-0.04	0.00
						(0.07)	(0.06)
me						0.33^{*}	0.19
						(0.19)	(0.12)
ia						0.38	0.36^{*}
						(0.24)	(0.20)
roe						-0.03	-0.57^{***}
						(0.16)	(0.13)
Adj. R^2	-0.00	0.04	0.20	0.09	0.26	0.05	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 factors by the Fama and French (1993), Carhart (1997), Fama and French (2015) and the q-factor model (HXZ4) by Hou et al. (2015). The arbitrage 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 A.4: Risk-Adjusted Returns with Five Eigenvectors

	CAPM	FF3	FF3+UMD	FF5	FF5+UMD	HXZ4	HXZ4+UMD
alpha	2.56^{***}	2.48^{***}	2.04^{***}	2.53^{***}	2.18^{***}	2.30^{***}	2.23^{***}
	(0.28)	(0.29)	(0.24)	(0.31)	(0.27)	(0.33)	(0.28)
mktrf	−0.06	−0.11	−0.02	−0.09	−0.03		
	(0.08)	(0.07)	(0.06)	(0.07)	(0.06)		
smb		0.36^{*}	0.38^{***}	0.22	0.21		
		(0.20)	(0.14)	(0.16)	(0.13)		
hml		0.12	0.28^{*}	−0.08	0.20		
		(0.14)	(0.16)	(0.17)	(0.14)		
umd			0.49^{***}		0.51^{***}		0.63^{***}
			(0.12)		(0.11)		(0.11)
rmw				-0.45^{**}	-0.59^{***}		
				(0.21)	(0.16)		
cma				0.44^{*}	0.18		
				(0.25)	(0.22)		
mkt						−0.06	−0.02
						(0.08)	(0.06)
me						0.35^{*}	0.20
						(0.20)	(0.13)
ia						0.39	0.36
						(0.26)	(0.24)
roe						−0.03	-0.61^{***}
						(0.18)	(0.14)
Adj. R^2	0.00	0.04	0.20	0.09	0.25	0.05	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 factors by the Fama and French (1993), Carhart (1997), Fama and French (2015) and the q-factor model (HXZ4) by Hou et al. (2015). The arbitrage portfolio with five 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 A.5: Risk-Adjusted Returns with Seven Eigenvectors

	CAPM	FF3	FF3+UMD	FF5	FF5+UMD	HXZ4	HXZ4+UMD
alpha	2.69***	2.63***	2.23***	2.68***	2.37***	2.47***	2.41***
	(0.29)	(0.30)	(0.26)	(0.33)	(0.29)	(0.35)	(0.32)
mktrf	−0.06	−0.12*	−0.04	−0.11	−0.05		
	(0.08)	(0.07)	(0.06)	(0.07)	(0.06)		
smb		0.39**	0.40***	0.24	0.23*		
		(0.19)	(0.15)	(0.16)	(0.13)		
hml		0.08	0.22	−0.12	0.14		
		(0.14)	(0.15)	(0.17)	(0.15)		
umd			0.45***		0.46***		0.59***
			(0.12)		(0.11)		(0.11)
rmw				−0.45**	−0.57***		
				(0.20)	(0.15)		
cma				0.41	0.18		
				(0.26)	(0.22)		
mkt						−0.07	−0.04
						(0.08)	(0.06)
me						0.36*	0.22
						(0.20)	(0.14)
ia						0.33	0.31
						(0.27)	(0.25)
roe						−0.05	−0.59***
						(0.19)	(0.16)
Adj. R^2	0.00	0.04	0.17	0.09	0.22	0.05	0.21
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 factors by the Fama and French (1993), Carhart (1997), Fama and French (2015) and the q-factor model (HXZ4) by Hou et al. (2015). The arbitrage 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 A.6: Risk-Adjusted Returns with Eight Eigenvectors

	CAPM	FF3	FF3+UMD	FF5	FF5+UMD	HXZ4	HXZ4+UMD
alpha	2.67^{***}	2.62^{***}	2.13^{***}	2.65^{***}	2.27^{***}	2.41^{***}	2.33^{***}
	(0.29)	(0.30)	(0.26)	(0.33)	(0.28)	(0.36)	(0.31)
mktrf	−0.05	−0.12	−0.02	−0.09	−0.02		
	(0.08)	(0.07)	(0.06)	(0.08)	(0.07)		
smb		0.37^{**}	0.39^{***}	0.22	0.21		
		(0.18)	(0.15)	(0.18)	(0.14)		
hml		0.05	0.23	−0.17	0.14		
		(0.14)	(0.16)	(0.17)	(0.15)		
umd			0.55^{***}		0.57^{***}		0.70^{***}
			(0.12)		(0.11)		(0.12)
rmw				-0.44^{**}	-0.59^{***}		
				(0.20)	(0.15)		
cma				0.48^{**}	0.20		
				(0.24)	(0.22)		
mkt						−0.06	−0.01
						(0.08)	(0.06)
me						0.35^{*}	0.18
						(0.21)	(0.14)
ia						0.33	0.30
						(0.29)	(0.25)
roe						0.03	-0.61^{***}
						(0.21)	(0.16)
Adj. R^2	-0.00	0.03	0.19	0.07	0.24	0.03	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 factors by the Fama and French (1993), Carhart (1997), Fama and French (2015) and the q-factor model (HXZ4) by Hou et al. (2015). The arbitrage 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 A.7: Risk-Adjusted Returns with Nine Eigenvectors

	CAPM	FF3	FF3+UMD	FF5	FF5+UMD	HXZ4	HXZ4+UMD
alpha	2.60^{***}	2.57^{***}	2.15^{***}	2.66^{***}	2.33^{***}	2.45^{***}	2.38^{***}
	(0.29)	(0.30)	(0.25)	(0.33)	(0.28)	(0.36)	(0.30)
mktrf	−0.01	−0.08	0.01	−0.07	−0.01		
	(0.07)	(0.07)	(0.05)	(0.07)	(0.06)		
smb		0.35^{**}	0.37^{**}	0.20	0.19^{*}		
		(0.17)	(0.14)	(0.14)	(0.11)		
hml		0.02	0.17	−0.13	0.13		
		(0.13)	(0.15)	(0.17)	(0.13)		
umd			0.47^{***}		0.49^{***}		0.62^{***}
			(0.12)		(0.10)		(0.11)
rmw				-0.48^{**}	-0.61^{***}		
				(0.20)	(0.15)		
cma				0.33	0.09		
				(0.23)	(0.18)		
mkt						−0.04	−0.00
						(0.08)	(0.06)
me						0.31^{*}	0.16
						(0.19)	(0.12)
ia						0.21	0.19
						(0.25)	(0.23)
roe						−0.03	-0.60^{***}
						(0.19)	(0.15)
Adj. R^2	-0.00	0.04	0.18	0.08	0.24	0.03	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 factors by the Fama and French (1993), Carhart (1997), Fama and French (2015) and the *q*-factor model (HXZ4) by Hou et al. (2015). The arbitrage 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 A.8: Risk-Adjusted Returns with Ten Eigenvectors

	CAPM	FF3	FF3+UMD	FF5	FF5+UMD	HXZ4	HXZ4+UMD
alpha	2.22^{***}	2.22^{***}	1.79^{***}	2.33^{***}	1.99^{***}	2.06^{***}	2.00^{***}
	(0.28)	(0.30)	(0.27)	(0.33)	(0.30)	(0.37)	(0.32)
mktrf	−0.02	−0.06	0.03	−0.05	0.01		
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)		
smb		0.19	0.20	0.01	−0.01		
		(0.24)	(0.18)	(0.19)	(0.16)		
hml		−0.03	0.13	−0.20	0.08		
		(0.16)	(0.14)	(0.17)	(0.13)		
umd			0.48^{***}		0.50^{***}		0.63^{***}
			(0.13)		(0.12)		(0.14)
rmw				-0.59^{***}	-0.72^{***}		
				(0.19)	(0.16)		
cma				0.37	0.12		
				(0.26)	(0.22)		
mkt						−0.01	0.03
						(0.08)	(0.07)
me						0.18	0.02
						(0.24)	(0.17)
ia						0.22	0.20
						(0.28)	(0.23)
roe						−0.00	-0.58^{***}
						(0.18)	(0.17)
Adj. R^2	-0.00	0.01	0.13	0.06	0.20	0.01	0.18
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 factors by the Fama and French (1993), Carhart (1997), Fama and French (2015) and the q-factor model (HXZ4) by Hou et al. (2015). The arbitrage 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 A.9: Portfolio Performance Statistics for using 24-month Estimation Period

# Eigenvectors	Mean (%)	Standard Deviation (%)	Sharpe Ratio	Skewness	Kurtosis	Maximum Drawdown	Worst Month (%)	Best Month (%)
1	16.47	17.50	0.94	1.08	6.63	33.32	-24.59	31.17
2	19.12	18.36	1.04	0.95	9.34	33.15	-24.76	41.65
3	23.76	17.64	1.35	0.72	6.42	34.73	-23.04	32.91
4	27.39	19.20	1.43	0.83	5.86	34.88	-24.05	35.59
5	28.67	19.07	1.50	0.95	6.65	32.70	-24.99	39.83
6	29.95	20.38	1.47	1.70	12.59	32.36	-24.06	49.63
7	30.35	20.27	1.50	1.64	11.51	32.10	-23.72	47.76
8	30.51	20.22	1.51	1.35	9.20	32.34	-24.46	44.29
9	31.21	20.28	1.54	1.09	6.69	31.56	-24.89	39.56
10	31.47	20.43	1.54	1.03	6.03	30.21	-23.88	38.89

This table reports annualized percentage means, annualized percentage standard deviations, annualized Sharpe Ratios, skewness, kurtosis, and the best and worst month returns. The arbitrage 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. We use 24 months of data to estimate the weights of the arbitrage portfolio and then hold the portfolio for one month.

Table A.10: Alphas using 24-month Estimation Period

# Eigenvectors	CAPM	FF3	FF3+UMD	FF5	FF5+UMD	HXZ4	HXZ4+UMD
1	1.27	1.15	1.24	1.30	1.36	1.49	1.48
2	1.53	1.44	1.28	1.56	1.43	1.57	1.53
3	1.91	1.84	1.68	1.96	1.82	1.93	1.89
4	2.21	2.13	1.91	2.26	2.09	2.22	2.18
5	2.31	2.22	1.99	2.37	2.18	2.31	2.27
6	2.43	2.34	2.08	2.45	2.24	2.41	2.36
7	2.47	2.38	2.12	2.50	2.29	2.46	2.41
8	2.49	2.39	2.15	2.53	2.34	2.48	2.44
9	2.55	2.46	2.21	2.60	2.40	2.54	2.49
10	2.58	2.48	2.24	2.61	2.41	2.55	2.51

This table reports alphas (%/month) against Fama and French (1993), Carhart (1997), Fama and French (2015) and the q-factor model (HXZ4) by Hou et al. (2015). The arbitrage portfolio 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. We use 24 months of data to estimate the weights of the arbitrage portfolio and then hold the portfolio for one month.

Table A.11: Portfolio Performance Statistics for using 36-month Estimation Period

# Eigenvectors	Mean (%)	Standard Deviation (%)	Sharpe Ratio	Skewness	Kurtosis	Maximum Drawdown	Worst Month (%)	Best Month (%)
1	13.40	20.17	0.66	1.44	5.58	41.51	-14.03	35.82
2	12.49	18.41	0.68	0.93	4.54	44.63	-19.64	35.33
3	16.58	18.70	0.89	0.88	4.63	36.60	-25.17	31.03
4	18.46	19.34	0.95	0.69	4.13	41.67	-24.92	32.15
5	21.86	19.28	1.13	0.93	3.80	39.66	-16.51	34.54
6	23.17	19.28	1.20	0.82	3.32	36.10	-16.73	33.25
7	23.98	19.24	1.25	0.85	3.69	35.65	-17.05	36.13
8	23.99	19.62	1.22	0.79	3.42	36.53	-17.57	34.89
9	24.55	19.75	1.24	0.78	3.72	39.03	-17.98	36.35
10	25.05	19.72	1.27	0.82	3.63	37.79	-18.01	36.61

This table reports annualized percentage means, annualized percentage standard deviations, annualized Sharpe Ratios, skewness, kurtosis, and the best and worst month returns. The arbitrage 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. We use 36 months of data to estimate the weights of the arbitrage portfolio and then hold the portfolio for one month.

Table A.12: Alphas using 36-month Estimation Period

# Eigenvectors	CAPM	FF3	FF3+UMD	FF5	FF5+UMD	HXZ4	HXZ4+UMD
1	0.94	0.75	1.14	1.04	1.31	1.46	1.49
2	0.90	0.82	1.00	1.06	1.18	1.22	1.23
3	1.22	1.14	1.29	1.41	1.50	1.55	1.55
4	1.37	1.28	1.35	1.56	1.59	1.68	1.68
5	1.65	1.53	1.60	1.80	1.83	1.93	1.92
6	1.76	1.66	1.71	1.92	1.93	2.03	2.02
7	1.84	1.75	1.77	1.99	1.99	2.10	2.09
8	1.85	1.75	1.77	1.99	1.98	2.08	2.07
9	1.90	1.81	1.82	2.05	2.04	2.15	2.13
10	1.94	1.85	1.87	2.09	2.08	2.19	2.18

This table reports alphas (%/month) against Fama and French (1993), Carhart (1997), Fama and French (2015) and the *q*-factor model (HXZ4) by Hou et al. (2015). The arbitrage portfolio 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. We use 36 months of data to estimate the weights of the arbitrage portfolio and then hold the portfolio for one month.

Table A.13: Sharpe Ratios using various rebalancing frequencies

	Rebalancing Frequency per Year					
# Eigenvectors	1	2	3	4	6	12
1	0.35	0.43	0.48	0.72	0.96	1.44
2	0.44	0.58	0.68	0.87	1.06	1.49
3	0.42	0.71	0.74	0.91	1.25	1.71
4	0.47	0.70	0.73	0.92	1.26	1.75
5	0.47	0.67	0.58	0.81	1.10	1.66
6	0.47	0.69	0.62	0.80	1.13	1.67
7	0.45	0.74	0.69	0.80	1.22	1.72
8	0.41	0.68	0.69	0.77	1.13	1.58
9	0.44	0.68	0.73	0.93	1.22	1.75
10	0.31	0.55	0.57	0.72	1.10	1.35

This table reports annualized Sharpe Ratios of our arbitrage portfolios with various rebalancing frequencies. The last column of 12 rebalancing frequency corresponds to our baseline case. We use 12 months of data to estimate the weights of the arbitrage portfolio and then hold the portfolio over for the next 12/frequency months, at the end of which we rebalance the arbitrage portfolio using the previous 12 month data. The sample period is January 1968 to June 2014.

Table A.14: Portfolio Performance Statistics for Fourth Order Legendre Polynomials

# Eigenvectors	Mean (%)	Standard Deviation (%)	Sharpe Ratio	Skewness	Kurtosis	Maximum Drawdown	Worst Month (%)	Best Month (%)
1	30.22	17.81	1.70	0.37	6.70	26.78	-24.27	30.39
2	35.38	20.34	1.74	0.19	4.33	26.11	-26.11	29.04
3	32.85	16.51	1.99	0.34	6.22	25.94	-25.94	31.07
4	35.50	18.69	1.90	-0.10	4.32	26.37	-26.37	28.05
5	36.19	19.08	1.90	-0.10	3.94	28.48	-26.77	28.56
6	37.74	20.10	1.88	0.15	5.40	31.73	-31.73	34.20
7	38.69	19.85	1.95	0.20	5.54	32.29	-32.29	31.64
8	37.69	21.67	1.74	-0.76	11.96	47.21	-47.21	33.55
9	36.42	20.95	1.74	-0.56	9.65	41.64	-41.64	33.93
10	32.12	20.53	1.56	-0.14	11.69	41.15	-41.15	35.55

This table reports annualized percentage means, annualized percentage standard deviations, annualized Sharpe Ratios, skewness, kurtosis, and the best and worst month returns. The arbitrage 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 A.15: Alphas for Fourth Order Legendre Polynomials

# Eigenvectors	CAPM	FF3	FF3+UMD	FF5	FF5+UMD	HXZ4	HXZ4+UMD
1	2.52	2.49	2.10	2.57	2.26	2.34	2.28
2	2.99	2.97	2.37	2.99	2.52	2.66	2.58
3	2.75	2.70	2.26	2.76	2.40	2.51	2.45
4	2.98	2.92	2.36	2.94	2.50	2.60	2.53
5	3.04	2.98	2.39	3.00	2.54	2.62	2.53
6	3.17	3.14	2.53	3.18	2.70	2.80	2.71
7	3.25	3.23	2.63	3.22	2.76	2.85	2.77
8	3.16	3.17	2.52	3.16	2.66	2.73	2.64
9	3.05	3.05	2.41	3.04	2.55	2.61	2.53
10	2.70	2.69	2.11	2.64	2.20	2.19	2.12

This table reports alphas (%/month) against Fama and French (1993), Carhart (1997), Fama and French (2015) and the q-factor model (HXZ4) by Hou et al. (2015). The arbitrage portfolio 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 A.16: Portfolio Performance Statistics without Micro-Cap Stocks

# Eigenvectors	Mean (%)	Standard Deviation (%)	Sharpe Ratio	Skewness	Kurtosis	Maximum Drawdown	Worst Month (%)	Best Month (%)
1	17.58	15.55	1.13	0.59	6.98	22.28	-21.62	30.87
2	21.81	18.78	1.16	0.23	5.19	25.35	-25.35	30.87
3	21.06	15.34	1.37	0.78	6.82	22.02	-22.02	31.28
4	23.75	17.79	1.34	0.62	5.39	22.77	-22.24	34.13
5	24.24	19.05	1.27	0.61	5.68	31.86	-22.54	36.65
6	24.17	19.79	1.22	0.44	5.39	27.26	-26.07	36.78
7	25.44	20.45	1.24	0.75	6.80	28.18	-28.18	42.03
8	25.17	21.01	1.20	0.81	6.78	28.47	-28.47	43.24
9	21.60	19.99	1.08	-0.35	11.37	42.97	-36.85	38.29
10	18.51	20.66	0.90	0.28	6.48	30.43	-27.13	36.65

This table reports annualized percentage means, annualized percentage standard deviations, annualized Sharpe Ratios, skewness, kurtosis, and the best and worst month returns. The arbitrage 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. We exclude micro-cap stocks, smaller than 10% quantile of the market capitalization among NYSE traded stocks.

Table A.17: Alphas without Micro-Cap Stocks

# Eigenvectors	CAPM	FF3	FF3+UMD	FF5	FF5+UMD	HXZ4	HXZ4+UMD
1	1.42	1.37	1.09	1.47	1.25	1.30	1.26
2	1.82	1.78	1.24	1.88	1.45	1.53	1.46
3	1.74	1.68	1.29	1.79	1.47	1.54	1.48
4	1.99	1.96	1.45	2.04	1.64	1.73	1.66
5	2.04	2.03	1.46	2.10	1.65	1.73	1.65
6	2.04	2.02	1.43	2.11	1.64	1.74	1.65
7	2.15	2.16	1.57	2.26	1.80	1.89	1.81
8	2.13	2.14	1.56	2.18	1.72	1.85	1.77
9	1.81	1.85	1.28	1.90	1.45	1.58	1.50
10	1.53	1.54	0.96	1.58	1.12	1.20	1.12

This table reports alphas (%/month) against Fama and French (1993), Carhart (1997), Fama and French (2015) and the q-factor model (HXZ4) by Hou et al. (2015). The arbitrage portfolio 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. We exclude micro-cap stocks, smaller than 10% quantile of the market capitalization among NYSE traded stocks.

Table A.18: Risk-Adjusted Returns with respect to Alternative Factor Models

	SY MP	FF3+UMD+LIQ	FF3+UMD+BAB	FF3+UMD+STREV
(Intercept)	2.19***	2.12***	2.18***	2.18***
	(0.28)	(0.26)	(0.27)	(0.26)
mktrf	0.01	−0.01	−0.01	0.00
	(0.08)	(0.06)	(0.06)	(0.06)
smb	0.43**	0.38***	0.38***	0.39***
	(0.20)	(0.14)	(0.15)	(0.15)
mgmt	0.30			
	(0.21)			
perf	0.22*			
	(0.12)			
hml		0.27*	0.32**	0.28*
		(0.14)	(0.13)	(0.16)
umd		0.48***	0.50***	0.46***
		(0.11)	(0.11)	(0.11)
liqf		0.03		
		(0.06)		
bab			−0.10	
			(0.09)	
strev				−0.11
				(0.10)
Adj. R^2	0.06	0.17	0.18	0.18
Num. obs.	557	557	557	557

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

This table reports alphas (%/month) and factor loadings on the factors by the several alternative factor models. The arbitrage 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.