

Arbitrage Portfolios*

Soohun Kim

College of Business, Korea Advanced Institute of Science and Technology

Robert A. Korajczyk

Kellogg School of Management, Northwestern University

Andreas Neuhierl

Olin Business School, Washington University in St. Louis

August 6, 2020

Forthcoming, *Review of Financial Studies*.

*We are thankful to Dong-Hyun Ahn, Demir Bektic, Sylvain Champonnois, Wojciech Charemza, Zhuo Chen, Ian Dew-Becker, Philip Dybvig, Emmanuel Eyiah-Donkor, Ryan Garino, Guillaume Rousset, Alex Horenstein, Ravi Jagannathan, Wei Jiang (the editor), Chotibhak Jotikasthira, Andrew Karolyi, Yuan Liao, Gene Lu, Hongyi Liu, Yan Liu, Seth Pruitt, Oleg Rytchkov, Andrea Tamoni, Olivier Scaillet, Gustavo Schwenkler, Raman Uppal, Michael Weber, Dacheng Xiu, Nancy Xu, Chu Zhang, and Guofu Zhou; the anonymous referees; and conference and seminar participants at Boston University, Northeastern University, Northwestern University, Rutgers University, Temple University, Texas A&M, University of Chicago, the Ohio State University, University of Missouri, University of Georgia, University of Houston, Michigan State University, University of Notre Dame, Washington University in St. Louis, Wirtschaftsuniversität Wien, KAIST, the USC Conference on Panel Data Forecasting, the Consortium on Factor Investing, the New Developments in Factor Investing Conference at Imperial College, the Mutual Funds, Hedge Funds and Factor Investing Conference at Lancaster University, the Society for Financial Econometrics Annual meeting, the UW Madison Junior Finance Conference, the New Zealand Finance Meeting, the China International Conference in Finance, the Econometric Research in Finance Workshop, the European Finance Association annual meeting, the EFMA annual meeting, the European Winter Finance Summit, the UT Dallas Fall Finance Conference, the University of Oregon Summer Finance Conference, the Paris December Finance Meeting, the Midwest Econometrics Group annual meeting, and the World Symposium on Investment Research. This work was supported by the Unigestion Alternative Risk Premia Research Academy, Université Paris-Dauphine. Send correspondence to Soohun Kim, soohun@kaist.ac.kr.

Arbitrage Portfolios

Abstract

We propose a new methodology for forming arbitrage portfolios that utilizes the 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 is made to abnormal returns. We apply the methodology to simulated economies and to a large panel of U.S. stock returns. The methodology works well in our simulation and when applied to 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 1.31 to 1.66. (*JEL* G12, C58, C38)

As has been documented in the literature, many variables have demonstrated 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); or 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. Researchers have had much success in forming portfolios using firm-level characteristics and then analyzing the obtained betas to explain the cross-section of returns (e.g., Fama and French (1993, 2015); Hou, Xue, and Zhang (2015)). Daniel and Titman (1997) argue that disentangling a purely characteristics-based model (in which characteristics only predict alpha) from a risk-based model is difficult, because the characteristics and factor loadings in the characteristic-sorted portfolios are collinear. In their influential approach to disentangling the beta versus alpha explanations, the authors sort assets 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 versus 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 the issue of staleness, double sorting has the disadvantages that the approach handles one characteristic at a time and, hence, is unable to analyze many characteristics simultaneously and that 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, Liao, and Wang (2016). The estimator can accommodate many characteristics simultaneously; can use individual assets or 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 premiums 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 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 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. We perform month-by-month rolling sorts of assets based on ordinary least squares (OLS) estimates of market betas (estimated over the previous 60 months) and the characteristic. We report average returns of double-sorted portfolios in Table 1 (full details about the simulation are in the table legend). Although the true return generating process is the CAPM, the return differences of

¹Kozak, Nagel, and Santosh (2018) argue that the distinction between risk premiums 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.

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 indicate 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 to this economy, we find that the relation between abnormal returns and the characteristic is insignificantly different from zero (p -value of .82).

We also show that when a relation between alpha and characteristics exists, 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, Neuhierl, and Weber (2020), updated from June 2014 to December 2018. In the baseline implementation, we use 12 months of data to estimate the weights of the arbitrage portfolio and then hold the portfolio for 1 month. Our results are robust to using 24- or 36-month estimation periods. 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 1.31 to 1.66 (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.

Kelly, Pruitt, and Su (2017, 2019) develop and apply a similar methodology, called instrumented principal component analysis (IPCA). They 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 re-

turns is due to their ability to predict the cross-section of risk loadings rather than mispricing, while we find that mispricing is significant. It is important to clarify the differences in economic questions between this paper and Kelly, Pruitt, and Su (2019). We focus on identifying and utilizing both the cross-sectional and the temporal relations of characteristics to risk and mispricing. Hence, we use the characteristics at the beginning of each estimation subinterval of short horizon (1 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 subintervals. We apply the identified cross-sectional relation to the most recently observed characteristics to construct our portfolio weights.

In contrast, Kelly, Pruitt, and Su (2019) allow the characteristics to change period by period but hold the cross-sectional relation between characteristics and either risk or alpha *constant* over the full sample period. The dynamics in our procedure primarily come from changes in the cross-sectional relation between alphas, betas, and characteristics, along with updating characteristics across subintervals of time. The dynamics in Kelly, Pruitt, and Su (2019) only come from the time-series variation of characteristics, holding the cross-sectional relation constant. Our procedure will tend to perform better in situations in which characteristics are relatively stable over our short estimation period (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 the characteristic/beta relation varies over time or if anomalies change over time (e.g., exhibit momentum themselves or are arbitrated away after discovery). Alti and Titman (2019) provide a micro foundation of such time-varying relation and Chinco, Neuhierl, and Weber (2020) document time variation in the ex ante likelihood of the occurrence of anomalies. 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.

Our method performs well with large cross-sectional and short time-series data. In contrast, IPCA seems to have low power in situations in which the time-series sample, T , is small. This can be seen in table A.1 of the internet appendix of Kelly, Pruitt, and Su (2019) (see Corollary 1 of Kelly, Pruitt, and Su (2017) for the large T requirement for consistency). For comparison, we 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.² Also, our arbitrage portfolio has a significant alpha relative to the latent factors estimated by the IPCA procedures of Kelly, Pruitt, and Su (2019).

Our approach allows us to make a number of contributions to the empirical asset pricing literature. First, we provide useful guidance about portfolio construction to 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 versus 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, Hagmann, and Linton (2012) for prespecified factor models and Fan, Liao, and Wang (2016), Light, Maslov, and Rytchkov (2017), and Kelly, Pruitt, and Su (2019) 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. In the simulated economy above, there are no arbitrage opportunities, and our procedure applied to those data correctly finds no evidence of arbitrage opportunities. This is corroborated in more extensive simulation results, reported in Section 2. Our procedure can separately identify risk (beta) and mispricing (alpha). The empirical results using U.S. stock return data imply that the cross-sectional predictability is not only due to risk exposure effects (beta) but also due to mispricing effects (alpha).

1 The Model

We assume that there exists a large number of securities indexed by $i = 1, \dots, N$, and the return-generating processes for those individual securities are stable for short blocks of time (e.g., dozens of months) $t = 1, \dots, T$. We allow the return-generating process to change across time periods. The return-generating process of each security follows a

²This result does not mean that their method is deficient. Their asymptotic theory is based on a large T . However, we intentionally design the simulation setup for a small T to compare to our real-time estimated arbitrage portfolio returns.

³See Chen et al. (2018), Herskovic, Moreira, and Muir (2019), and Daniel et al. (2020), who extend the Daniel and Titman (1997) procedure to various characteristics.

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,

$$R_{i,t} = \alpha_i + \beta_i' \mathbf{f}_t + e_{i,t}, \quad i = 1, \dots, N \text{ and } t = 1, \dots, T, \quad (1)$$

where $\beta_i = [\beta_{i,1} \cdots \beta_{i,K}]'$ is the $(K \times 1)$ vector of 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 (1). Throughout, we use $\mathbf{0}_m$, $\mathbf{1}_m$, and $\mathbf{0}_{m \times l}$ to denote the $(m \times 1)$ vectors of zeros and ones and the $(m \times l)$ matrix of zeros, respectively. The return-generating process of (1) is compactly expressed in matrix form:

$$\mathbf{R} = \boldsymbol{\alpha} \mathbf{1}_T' + \mathbf{B} \mathbf{F}' + \mathbf{E}, \quad (2)$$

where the (i, t) element of the $(N \times T)$ matrix \mathbf{R} is $R_{i,t}$, $\boldsymbol{\alpha}$ 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 β_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 PPCA approach of Fan, Liao, and Wang (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} :

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

where $\mathbf{G}_\alpha(\mathbf{X}) : \mathbb{R}^{N \times L} \rightarrow \mathbb{R}^N$, $\mathbf{G}_\beta(\mathbf{X}) : \mathbb{R}^{N \times L} \rightarrow \mathbb{R}^{N \times K}$, 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 with small T . Therefore, our procedure does not attempt to exploit the gammas, just their orthogonality to the characteristics. One could incorporate nonlinearity into the mispricing and factor loading functions a number of ways. 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) as follows:

$$\mathbf{R} = (\mathbf{G}_\alpha(\mathbf{X}) + \Gamma_\alpha) \mathbf{1}'_T + (\mathbf{G}_\beta(\mathbf{X}) + \Gamma_\beta) \mathbf{F}' + \mathbf{E}. \quad (3)$$

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 (3) can be treated as a locally unconditional approximation of a conditional model.⁴ Third, our rolling estimation of (3) enables us to study the *temporal* relation of characteristics to risk or mispricing. Many empirical approaches (e.g., Kelly, Pruitt, and Su (2019); Ferson and Harvey (1999); Ghysels (1998)) construct conditional models 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 arbitrated away after discovery. By estimating (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 have all important characteristics for risk and mispricing (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 $\frac{1}{N} \boldsymbol{\alpha}' \boldsymbol{\alpha} \rightarrow 0$. Hence, in an economy governed by the APT, it follows that $\frac{\mathbf{G}_\alpha(\mathbf{X})' \mathbf{G}_\alpha(\mathbf{X})}{N} \rightarrow 0$, because $0 \leq \frac{\mathbf{G}_\alpha(\mathbf{X})' \mathbf{G}_\alpha(\mathbf{X})}{N} \leq \frac{1}{N} \boldsymbol{\alpha}' \boldsymbol{\alpha}$, since

⁴We thank Yuan Liao for pointing this out. Our approach also works under a smooth transition of \mathbf{X} over a short horizon.

$\frac{1}{N}\boldsymbol{\alpha}'\boldsymbol{\alpha}$ 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 $\mathbf{G}_{\boldsymbol{\alpha}}(\mathbf{X})$ may be nonzero:

Assumption 1. As $N \rightarrow \infty$,

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

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.⁵ Assumption 1 does not imply that characteristics capture all potential mispricing. Mispricing orthogonal to the characteristics is reflected in $\boldsymbol{\Gamma}_{\boldsymbol{\alpha}}$. 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 relations 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 returns.

Assumption 2. As $N \rightarrow \infty$, it holds that

- (i) $\frac{\mathbf{R}'\mathbf{R}}{N} \xrightarrow{p} \mathbf{V}_R$ and $\frac{\mathbf{X}'\mathbf{X}}{N} \rightarrow \mathbf{V}_X$, where \mathbf{V}_R and \mathbf{V}_X are positive definite matrices,
(ii) $\frac{\mathbf{G}_{\boldsymbol{\beta}}(\mathbf{X})'\boldsymbol{\Gamma}_{\boldsymbol{\alpha}}}{N} \xrightarrow{p} \mathbf{0}_K$, $\frac{\mathbf{G}_{\boldsymbol{\beta}}(\mathbf{X})'\boldsymbol{\Gamma}_{\boldsymbol{\beta}}}{N} \xrightarrow{p} \mathbf{0}_{K \times K}$, $\frac{\mathbf{X}'\boldsymbol{\Gamma}_{\boldsymbol{\alpha}}}{N} \xrightarrow{p} \mathbf{0}_L$, $\frac{\mathbf{X}'\boldsymbol{\Gamma}_{\boldsymbol{\beta}}}{N} \xrightarrow{p} \mathbf{0}_{L \times K}$, $\frac{\mathbf{G}_{\boldsymbol{\beta}}(\mathbf{X})'\mathbf{E}}{N} \xrightarrow{p} \mathbf{0}_{K \times T}$
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 $\mathbf{G}_{\boldsymbol{\alpha}}(\mathbf{X})$ and $\mathbf{G}_{\boldsymbol{\beta}}(\mathbf{X})$. 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.

⁵See Jagannathan and Wang (2007), Baker and Wurgler (2006), Stambaugh and Yuan (2016), and Frazzini and Pedersen (2014), among many others, for potential causes of mispricing.

Assumption 3. As $N \rightarrow \infty$, we assume

- (i) $\frac{\mathbf{G}_\beta(\mathbf{X})'\mathbf{G}_\alpha(\mathbf{X})}{N} \rightarrow \mathbf{0}_K$,
- (ii) $\frac{\mathbf{G}_\beta(\mathbf{X})'\mathbf{G}_\beta(\mathbf{X})}{N} \rightarrow \mathbf{I}_K$ 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 (Connor, Hagmann, and Linton (2012) utilize a similar orthogonality condition for identification). 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} .

1.1 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 a standard PCA procedure to the projected returns. Fan, Liao, and Wang (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 premiums 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, Liao, and Wang (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 PCA to demeaned projected returns. By demeaning the returns, we purely focus 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 infor-

mation 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 (3) are driven by both $\mathbf{G}_\beta(\mathbf{X})$ and $\mathbf{G}_\alpha(\mathbf{X})$. We eliminate the effect of $\mathbf{G}_\alpha(\mathbf{X})$, by demeaning the observed returns:

$$\begin{aligned} \mathbf{R}\mathbf{J}_T &= (\mathbf{G}_\alpha(\mathbf{X}) + \Gamma_\alpha) \mathbf{1}'_T \mathbf{J}_T + (\mathbf{G}_\beta(\mathbf{X}) + \Gamma_\beta) \mathbf{F}' \mathbf{J}_T + \mathbf{E} \mathbf{J}_T \\ &= (\mathbf{G}_\beta(\mathbf{X}) + \Gamma_\beta) \mathbf{F}' \mathbf{J}_T + \mathbf{E} \mathbf{J}_T, \end{aligned} \quad (4)$$

where the last equality is from the property of $\mathbf{1}'_T \mathbf{J}_T = \mathbf{1}'_T (\mathbf{I}_T - \frac{1}{T} \mathbf{1}_T \mathbf{1}'_T) = \mathbf{1}'_T - \frac{T}{T} \mathbf{1}'_T = \mathbf{0}'_T$. For further isolation of $\mathbf{G}_\beta(\mathbf{X})$, we project the demeaned returns of (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

$$\hat{\mathbf{R}} \equiv \mathbf{P}\mathbf{R}\mathbf{J}_T = \mathbf{P}\mathbf{G}_\beta(\mathbf{X}) \mathbf{F}' \mathbf{J}_T + \mathbf{P}\Gamma_\beta \mathbf{F}' \mathbf{J}_T + \mathbf{P}\mathbf{E} \mathbf{J}_T. \quad (5)$$

Note that $\mathbf{P}\mathbf{G}_\beta(\mathbf{X}) = \mathbf{G}_\beta(\mathbf{X})$, since $\mathbf{G}_\beta(\mathbf{X})$ is already in the linear span of \mathbf{X} . The orthogonality of Γ_β and \mathbf{X} and the limits in Assumption 2(ii) make $\mathbf{P}\Gamma_\beta$ and $\mathbf{P}\mathbf{E}$ negligible for large N . Hence, it holds that $\hat{\mathbf{R}} = \mathbf{P}\mathbf{R}\mathbf{J}_T \approx \mathbf{G}_\beta(\mathbf{X}) \mathbf{F}' \mathbf{J}_T$ with large N . Finally, as in Fan, Liao, and Wang (2016), we estimate $\mathbf{G}_\beta(\mathbf{X})$ by applying standard principal component analysis to $\hat{\mathbf{R}}$.

Theorem 1. Let $\hat{\mathbf{G}}_\beta(\mathbf{X})$ denote the $(N \times K)$ matrix, the k th column of which is \sqrt{N} times the eigenvector of $\frac{\hat{\mathbf{R}}\hat{\mathbf{R}}'}{N}$ corresponding to the k th largest eigenvalue of $\frac{\hat{\mathbf{R}}\hat{\mathbf{R}}'}{N}$, where $\hat{\mathbf{R}}$ is given by (5). Under Assumptions 2 and 3, as N increases, $\hat{\mathbf{G}}_\beta(\mathbf{X}) \xrightarrow{p} \mathbf{G}_\beta(\mathbf{X})$.

Proof All proofs are in the appendix. □

As some intuition for the result, recall that $\hat{\mathbf{R}}$ converges (as $N \rightarrow \infty$) to $\mathbf{G}_\beta(\mathbf{X}) \mathbf{F}' \mathbf{J}_T$. Therefore, $\frac{\hat{\mathbf{R}}\hat{\mathbf{R}}'}{N}$ 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), $\frac{\mathbf{G}_\beta(\mathbf{X})' \mathbf{G}_\beta(\mathbf{X})}{N} \rightarrow$

\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 $\frac{\hat{\mathbf{R}}\hat{\mathbf{R}}'}{N}$, as stated in Theorem 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 $\frac{1}{T}\mathbf{1}_T$ (we can weight the time-series mean by post-multiplying any $(T \times 1)$ vector \mathbf{i} such that $\mathbf{1}_T'\mathbf{i} = 1$). From (3), the $(N \times 1)$ vector of average returns, $\frac{1}{T}\mathbf{R}\mathbf{1}_T = \bar{\mathbf{R}}$, is

$$\begin{aligned}\bar{\mathbf{R}} &= (\mathbf{G}_\alpha(\mathbf{X}) + \Gamma_\alpha) \frac{1}{T}\mathbf{1}_T'\mathbf{1}_T + (\mathbf{G}_\beta(\mathbf{X}) + \Gamma_\beta) \frac{1}{T}\mathbf{F}'\mathbf{1}_T + \frac{1}{T}\mathbf{E}'\mathbf{1}_T \\ &= \mathbf{G}_\alpha(\mathbf{X}) + \Gamma_\alpha + (\mathbf{G}_\beta(\mathbf{X}) + \Gamma_\beta) \bar{\mathbf{F}} + \bar{\mathbf{E}},\end{aligned}\quad (6)$$

Our objective is to extract $\mathbf{G}_\alpha(\mathbf{X})$ from $\bar{\mathbf{R}}$. Note that simply projecting $\bar{\mathbf{R}}$ to the linear span of \mathbf{X} does not work because $\bar{\mathbf{R}}$ contains not only $\mathbf{G}_\alpha(\mathbf{X})$ but also $\mathbf{G}_\beta(\mathbf{X})\bar{\mathbf{F}}$. That is, projecting $\bar{\mathbf{R}}$ 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 premiums. Hence, we project $\bar{\mathbf{R}}$ to the linear space spanned by \mathbf{X} that is orthogonal to $\hat{\mathbf{G}}_\beta(\mathbf{X})$. The following theorem establishes that we can recover $\mathbf{G}_\alpha(\mathbf{X})$ with this approach.

Theorem 2. Define $\hat{\mathbf{G}}_\alpha(\mathbf{X}) = \mathbf{X}\hat{\boldsymbol{\theta}}$, where the $(L \times 1)$ vector of $\hat{\boldsymbol{\theta}}$ is given by the solution of the following constrained optimization problem:

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} (\bar{\mathbf{R}} - \mathbf{X}\boldsymbol{\theta})' (\bar{\mathbf{R}} - \mathbf{X}\boldsymbol{\theta}) \quad \text{subject to} \quad \hat{\mathbf{G}}_\beta(\mathbf{X})' \mathbf{X}\boldsymbol{\theta} = \mathbf{0}_K, \quad (7)$$

where $\hat{\mathbf{G}}_\beta(\mathbf{X})$ is given by Theorem 1. Then, under Assumptions 2 and 3, as N increases, $\hat{\mathbf{G}}_\alpha(\mathbf{X}) \xrightarrow{p} \mathbf{G}_\alpha(\mathbf{X})$.

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.⁶

Alternatively, the estimator in Theorem 2 can be derived within the conventional

⁶The result in Theorem 2 can be extended to incorporate a weighting matrix to increase (or decrease) the importance of some stocks versus 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 $\hat{\mathbf{G}}_\alpha(\mathbf{X}) = \mathbf{X}\hat{\boldsymbol{\theta}}_W$ such that $\hat{\boldsymbol{\theta}}_W = \arg \min_{\boldsymbol{\theta}} (\bar{\mathbf{R}} - \mathbf{X}\boldsymbol{\theta})' \mathbf{W} (\bar{\mathbf{R}} - \mathbf{X}\boldsymbol{\theta})$ subject to $\hat{\mathbf{G}}_\beta(\mathbf{X})' \mathbf{W}\mathbf{X}\boldsymbol{\theta} = \mathbf{0}_K$.

risk-adjusted approach as follows. Note that equation (8) can be rearranged as

$$\bar{\mathbf{R}} = \mathbf{G}_\beta(\mathbf{X}) \bar{\mathbf{F}} + (\mathbf{G}_\alpha(\mathbf{X}) + \Gamma_\alpha + \Gamma_\beta \bar{\mathbf{F}} + \bar{\mathbf{E}}) \quad (8)$$

and

$$\bar{\mathbf{R}} - \mathbf{G}_\beta(\mathbf{X}) \bar{\mathbf{F}} = \mathbf{G}_\alpha(\mathbf{X}) + (\Gamma_\alpha + \Gamma_\beta \bar{\mathbf{F}} + \bar{\mathbf{E}}). \quad (9)$$

Recall that our objective is to estimate $\mathbf{G}_\alpha(\mathbf{X})$. Equation (9) shows that we can achieve this goal by regressing $\bar{\mathbf{R}} - \mathbf{G}_\beta(\mathbf{X}) \bar{\mathbf{F}}$ on \mathbf{X} . Because we do not directly observe $\mathbf{G}_\beta(\mathbf{X})$ and $\bar{\mathbf{F}}$, we use $\hat{\mathbf{G}}_\beta(\mathbf{X})$ from Theorem 1 and estimate $\bar{\mathbf{F}}$ by regressing $\bar{\mathbf{R}}$ on $\hat{\mathbf{G}}_\beta(\mathbf{X})$, motivated by the expression (8). 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 (3), we find that the return of this infeasible portfolio is given by

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

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

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

The above theorem delivers 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. The details of the out-of-sample applications are described in Section 3.

The PPCA method does not require that all factors have betas that are explained by \mathbf{X} . That is, there could be factors for which the corresponding column of $\mathbf{G}_\beta(\mathbf{X})$ is equal to $\mathbf{0}$ and the betas are completely determined by the corresponding column of Γ_β .

In this case, we will not be able to identify those factors as N increases. The arbitrage portfolio weights are orthogonal to those factor loadings since $\mathbf{G}_\alpha(\mathbf{X})$ is orthogonal to Γ_β . In this case, K represents the number of factors explained by \mathbf{X} and additional K^* factors may not be explained by \mathbf{X} . PPCA also requires at least K linearly independent characteristics that explain the cross-section of betas in order to estimate the K latent factors whose betas are associated with \mathbf{X} (Fan, Liao, and Wang (2016) call this the genuine projection assumption; see their assumption 3.1). Below, we present a graphical analysis suggesting sufficient important characteristics for K in the range typically used in factor models. When choosing the number of factors to use in the analysis, one must bear in mind that too many factors may violate the genuine projection assumption. We address the issue of extracting too many factors through conducting simulations.

2 Simulation

In this section, we analyze the properties of our estimator in simulations. The purpose of this exercise is threefold. First, we illustrate the behavior of our arbitrage portfolio estimator in finite samples similar in size to those of the U.S. stock market.⁷ 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.

2.1 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 (see Table 2 for the full list).

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

⁷This section focuses on simulation evidence for our procedure's ability to accurately estimate the arbitrage profits (if any) as established in Theorem 3. We also confirm the results of Theorems 1 and 2. These additional results are available on 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, that is, $K = 1$ for the CAPM, $K = 3$ for the FF3, etc. In later sections, we will explore the effects of selecting too few or too many factors.

We calibrate α_i , β_i , and the variance of residual returns, $\sigma_{i,\varepsilon}^2 = \mathbb{E}[\varepsilon_{i,t}^2]$, 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 (bp) per month, as follows: we estimate $\hat{\alpha}_i$ from time-series regression and fit the cross-sectional relation $\hat{\alpha}_i = \mathbf{x}_i \boldsymbol{\theta}_\alpha + \gamma_{\alpha,i}$. We rescale $\tilde{\alpha}_i = k \hat{\alpha}_i$, where $k = \frac{0.01}{\sqrt{\frac{\boldsymbol{\theta}'_\alpha \mathbf{X}' \mathbf{X} \boldsymbol{\theta}_\alpha}{N}}}$, and use the rescaled $\tilde{\alpha}_i$ in the simulated returns (10). Note that $k \gamma_{\alpha,i}$, 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 a 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

$$\begin{aligned} \mathbf{R} &= \boldsymbol{\alpha} \mathbf{1}'_T \sqrt{\delta} + \mathbf{B} \mathbf{F}' + \mathbf{E} \\ &= \left(\mathbf{X} \boldsymbol{\theta}_\alpha \sqrt{\delta} + \boldsymbol{\Gamma}_\alpha \sqrt{\delta} \right) + (\mathbf{X} \boldsymbol{\Theta}_\beta + \boldsymbol{\Gamma}_\beta) \mathbf{F}' + \mathbf{E}, \end{aligned} \tag{10}$$

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 $\sigma_{i,\varepsilon}^2$ parameters as in the above paragraph. We consider different cases of mispricing, that is, $\delta = 0, 5$, and 10 .

2.2 Simulation results

2.2.1 Correctly specified model.

In our baseline scenario, we first investigate the performance of our estimator if we know the correct number of factors. Figure 1 shows the results for using 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, and Zhang model (lower-right panel). Our findings are consistent across all models used for calibration. The weights of the arbitrage portfolio, $\hat{\mathbf{w}}$, are estimated using the returns over $t = 1, \dots, 12$, 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 $T = 13$. 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 $\frac{\mathbf{G}_\alpha(\mathbf{X})'\mathbf{G}_\alpha(\mathbf{X})}{N} \rightarrow 1\text{bp/month}$. Hence, equation (10) implies that the arbitrage portfolio generates asymptotic arbitrage profits of $\delta = \lim_{N \rightarrow \infty} \left(\frac{(\mathbf{G}_\alpha(\mathbf{X})\sqrt{\delta})'(\mathbf{G}_\alpha(\mathbf{X})\sqrt{\delta})}{N} \right)$. 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 bp/month, suggesting that our arbitrage portfolio actually generates the expected level of arbitrage profits.

2.2.2 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 (see, e.g., Connor and Korajczyk (1993); Bai and Ng (2002); 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 2 reports the results when we set the number of extracted factors to be one more than the true number of factors. This is a case in which the genuine projection assumption of PPCA is violated since no characteristics explain the betas of factor $K + 1$. We find that the arbitrage portfolio's average performance in Figure 2 is almost identical to those in Figure 1, where we set the number of extracted factors to be the number of true factors. The confidence bands are slightly larger when extracting one too many factors. We conclude that extracting one factor more than

the true number does not seem to harm the performance of our arbitrage portfolios materially. 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 3 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, while the average returns are close to the true level, the portfolio returns are much more volatile (presumably because of the exposure to systematic factors) relative to the case of overestimation (Figure 2 (too many) versus Figure 3 (too few)). As 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.

2.2.3 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 the 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 $x_{i,c,t} = x_{i,c} + \rho_c(x_{i,c,t} - x_{i,c}) + \sigma_c \varepsilon_{i,t}$, 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 $N(0, 1)$ 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 $\alpha_{t-1} = \mathbf{X}_{t-1}\boldsymbol{\theta}_\alpha$, $\mathbf{B}_{t-1} = \mathbf{X}_{t-1}\boldsymbol{\Theta}_\beta + \Gamma_\beta$ and \mathbf{E}_t is the t th column of \mathbf{E} .⁸

Figure 4 reports the performance of our arbitrage portfolios when the returns are generated with the time-varying alpha $\alpha_{t-1} = \mathbf{X}_{t-1}\boldsymbol{\theta}_\alpha$ 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.

2.2.4 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. Figure A.1 in the Internet Appendix reports the results.

We also repeat the analysis using a different time period for the 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. Figure A.2 in the Internet Appendix reports these results. 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 $\hat{\mathbf{w}}$. Figure A.3 in the Internet Appendix plots the results. As an additional test, we also rerun 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.

3 Empirical Application

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

⁸We obtain $\boldsymbol{\theta}_\alpha$ and $\boldsymbol{\Theta}_\beta$ by regressing the calibrated α and \mathbf{B} on \mathbf{X} . Also, we find Γ_β from $\Gamma_\beta = \mathbf{B} - \mathbf{X}\boldsymbol{\Theta}_\beta$.

3.1 Data

The characteristic data are the same as in Freyberger, Neuhierl, and Weber (2020) but extended from June 2014 to December 2018. The assets studied are equities 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, Neuhierl, and Weber (2020), we use accounting data from the fiscal year ending in calendar year $t - 1$ for estimation starting from the end of June of year t until the end of May of year $t + 1$, predicting returns from the beginning of July of year t until the end June of year $t + 1$. Table 2 provides an overview of the characteristics used for estimating 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 at least 2 years of data in Compustat before inclusion in the sample. Our sample period is from 1965 through 2018. We use 12 to 36 months for the estimation period. For consistency, all versions of the arbitrage portfolio are run from January 1968 to December 2018. For the full sample, we have approximately 1.75 million firm-month observations in our analysis. The appendix in Freyberger, Neuhierl, and Weber (2020) describes the construction of the characteristic data in detail and cites numerous references to papers employing these characteristics in empirical applications.

3.2 Estimation

We initially assume that the factor loading function and the mispricing function are linear in the characteristics. Our methodology allows for (parametric) nonlinearities, which we explore in Section 4.3. We estimate $\hat{\mathbf{w}}$ with the returns over $t = 1, \dots, 12$, and the return of the arbitrage portfolio is measured in the following month, $t = 13$. We call the first period $t = 1, \dots, 12$ 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 $\hat{\mathbf{R}}$ over the estimation period corresponding to $\mathbf{P}_{\mathbf{X}_0} \mathbf{R} \mathbf{J}_{12}$ in (5) (from a panel regression using 12 months from

January 1967 to December 1967). The t th column of the $(N \times 12)$ matrix $\hat{\mathbf{R}}$ is the demeaned projected return for the t th month. Then we compute the $N \times N$ matrix $\frac{\hat{\mathbf{R}}\hat{\mathbf{R}}'}{N}$ and the first K eigenvectors of the matrix. We then project the average returns onto characteristics subject to orthogonality to the estimated factor loadings as in Theorem 2 to obtain $\hat{\boldsymbol{\theta}}$. In computing the arbitrage portfolio weights as in Theorem 3 for the following month of January 1968, we update characteristics with \mathbf{X}_{12} in computing $\hat{\mathbf{w}}$ such that $\hat{\mathbf{w}} = \frac{1}{N}\mathbf{X}_{12}\hat{\boldsymbol{\theta}}$. We repeat this process month by month until December 2018. To make the results comparable in scale to common equity factors, we scale the portfolio weights so that the in-sample standard deviation of the arbitrage portfolio returns is 20% per year.

3.3 Performance of the arbitrage portfolio

Table 3 shows the summary statistics for returns of the arbitrage portfolio for different numbers of eigenvectors (latent factors). 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 1.31 to 1.66. 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 55.68% (market factor), 55.04% (size factor), 40.92% (value factor) and 57.31% (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 should not be driven by high exposures to common risk factors. To test this we run a time-series regression of the arbitrage portfolio's returns onto common risk factors.⁹ In Tables 4 (one estimated factor) and 5 (six

⁹We 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, Xue,

estimated factors), we report the risk-adjusted returns of the arbitrage portfolio with respect to the CAPM (column 1), the Fama and French (1993) 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, Xue, and Zhang (2015) four-factor model (column 6), and the HXZ model augmented with the momentum factor (column 7).

We limit our main discussion to those cases in which we extract one factor (one eigenvector) and six factors (six eigenvectors). The Internet Appendix contains the results for all other cases. In Table 4 with one eigenvector, we can see that the estimated 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 0.17. 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 2.49%/month (see Table 5), which is far higher than that with one eigenvector at 1.64%/month. Figure 5 illustrates the relation between the out-of-sample alpha and the number of eigenvectors used in the estimator. The hump-shaped alpha decreases after approximately seven eigenvectors. Given that we use only 12 monthly observations in the estimation, we attribute the alpha's deterioration to overfitting during the estimation period. Tables A.1 through A.8 show the detailed results for estimating the arbitrage portfolio with the other numbers of eigenvectors through 10. Across all specifications and against all factor models, the arbitrage portfolio has a large positive and highly significant alpha. Strengthening the interpretation that characteristics contain information not only about factor loadings but also about mispricing. The alphas range between 1.38% and 2.59% per month.

Figure 6 summarizes the correlation of the arbitrage portfolios (using 1 through 10 eigenvectors) with each other and with common risk factors. 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 0.77. 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

and Zhang (2015) four-factor model.

much after we extract five common factors. The correlation between the arbitrage portfolios and the common factors is relatively low except for the size and momentum factors, which again is consistent with the factor regressions in Tables 4 and 5 and the additional factor regressions in the Internet Appendix.

3.4 Properties of the arbitrage portfolio

In this section, we more deeply explore the properties of the arbitrage portfolio. 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, possible diminishing excess returns over time, as well as the importance of controlling systematic risks.

3.4.1 Time-series properties.

To develop further intuition about the performance of the arbitrage portfolio, we explore its time-series properties more closely. Figure 7 plots the cumulative return. It is noteworthy that the arbitrage portfolio did not have a negative return (for a full calendar year) during the 2008 to 2010 financial crisis. Overall, the returns are positive in 49 of 51 years. Also, the arbitrage portfolio does not have significantly different returns during NBER recessions versus other periods. For a regression of the portfolio return on a constant and an NBER recession indicator, that is, $r_t = a + b \times \text{NBER}_t + \varepsilon_t$, we obtain point estimates of $\hat{a} = 2.36$ (significant at the 1% level) and $\hat{b} = 0.71$, with a p -value of .26. 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 systematically diminish over time. We test for a time trend by estimating the following specification:

$$r_t = a + b \times t^\gamma + \varepsilon_t. \quad (11)$$

We estimate the model using nonlinear least squares, and the point estimates are $\hat{a} = 5.22$, $\hat{b} = -0.11$, $\hat{\gamma} = 0.57$. This specification contains the linear time trend, $r_t = a + b \times t + \varepsilon_t$ as a special case. Only the intercept is significant at conventional levels, with a p -value of less than .01 (p -values of \hat{b} and $\hat{\gamma}$ are .75 and .18, respectively). A possibly undesirable feature of this specification is that it does not rule out arbitrar-

ily 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 $\hat{b} = 10.67$ and $\hat{\gamma} = -0.28$. This specification suggests a mild decay in excess returns and predicts the returns to reach less than 1% per month in approximately 5,000 months. Figure 8 plots the trend estimated from the specification estimated with the intercept. 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 data snooping could possibly lead to reduced future performance of an arbitrage portfolio, many of the predictive characteristics are the result of decades-old research. We conclude that the significant average excess returns are at least partially due to mispricing of assets.

3.4.2 Firm characteristics.

Figure 9 compares the long and short sides for nine well-known characteristics for the arbitrage portfolio using six eigenvectors. The second panel shows the time-series average of the normalized rank of the nine characteristics for the long and short sides of the arbitrage portfolio. All of the characteristics in Figure 9 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, Naik, and Radcliffe (1998)), and short-term reversal as well as (standard) momentum, both of which are documented in Jegadeesh and Titman (1993)).

From Figure 9, we can see that the biggest differences in average normalized ranks are for the two momentum-related characteristics, followed by size, investment, book-to-market, turnover, accruals, profitability, and debt-to-price. 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. Figure A.4 in the Internet Appendix shows the

normalized ranks of the long and short sides of all 61 characteristics.

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 ($\hat{\mathbf{G}}_{\beta}(\mathbf{X})$) and the estimated mispricing function ($\hat{\mathbf{G}}_{\alpha}(\mathbf{X})$) onto the characteristic vector 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 10 shows the projection results for systematic risk, and Figure 11 shows the corresponding results for mispricing. From Figure 10, 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 11, we can see that few characteristics are consistently related to mispricing (for example, size and total assets). Other characteristics are only related to mispricing for few periods “on-and-off.” While “eyeball econometrics” clearly has limits, the results in Figure 11 underscore the importance of our time-varying approach.

3.4.3 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 demean the characteristics so that the resultant portfolio weights of the arbitrage portfolio sum to zero. Therefore, by construction it is 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. Figure 12 plots the median, minimum, and maximum as well as the 5% and 95% quantiles of the weights in each month over the sample period from January 1968 to December 2018. 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.

3.4.4 Predicting raw mean returns.

We also estimate the relation between raw mean returns and the characteristics (using an OLS regression like equation (7), but without the orthogonality constraint) and form portfolios based on those predictions. This portfolio confounds alpha and risk premiums since, unlike our arbitrage portfolio, it is not constrained to have zero exposure to the latent factors. Table A.18 in the Internet Appendix reports the unconstrained portfolio's performance relative to standard asset pricing models. The alphas are significant, although they are about 0.7%–0.8% per month lower than the 6-eigenvector arbitrage portfolio results. The Sharpe ratio of the portfolio without the orthogonality constraint is 1.02 (vs. 1.6 for the 6-eigenvector arbitrage portfolio). Since this portfolio is formed without the orthogonality constraint, it will load on both alphas and risk premiums. Therefore, the portfolio tends to load more on some of the standard prespecified factors. This is particularly true for the market portfolio (actually a significantly negative loading), UMD (loadings around 0.9 rather than 0.5), and HML (when UMD is included). The lower alphas and Sharpe ratios are due, in part, to the scaling to a constant volatility of 20%. The unconstrained portfolio has more factor exposure, so it has to trim the alpha exposure to meet the total risk constraint, leading to lower alphas and Sharpe ratios. The correlation between the arbitrage portfolio and the unconstrained portfolio is 0.72.

4 Robustness

The empirical implementation of the arbitrage portfolio in Section 3 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 on these implementation choices.

4.1 Estimation windows

In our main specification, we use 12 months and then roll the estimation window forward by 1 month. Since our theoretical results are derived in a *local in time* setting, that is, T is fixed, the choice of T should not drive the result. To illustrate this, we reestimate

our main analysis with 24 and 36 months as the estimation window. Tables A.9 and A.10 (A.11 and A.12) in the Internet Appendix show the general performance statistics and estimated alphas (against the standard factor models) for the arbitrage portfolio constructed using 24 (36) months as the estimation window. Overall, the results are quite similar to our baseline of a 12-month window, although slightly worse than the base case.

4.2 Effect of staleness in $\mathbf{G}_\alpha(\mathbf{X})$ and \mathbf{X}

The heatmaps for the mispricing component and betas, in Figures 11 and 10, show the loadings on the different characteristics change from period to period. This suggests that an advantage of the rolling estimation of $\mathbf{G}_\alpha(\mathbf{X})$ is an important component in the performance of the arbitrage portfolio. To get a sense for the importance of changing mispricing, θ_α , relative to changing values of the characteristics, \mathbf{X} , in $\mathbf{G}_\alpha(\mathbf{X}) = \mathbf{X}\theta_\alpha$ we estimate the alphas and Sharpe ratios for arbitrage portfolios using different lags of θ_α and \mathbf{X} . These are reported in Tables 6 and 7. We calculate results where we lag the estimates of θ_α and the values of \mathbf{X} from 1 to 12 months. The (1, 1) entry corresponds to our main results for a 6-eigenvector model. Looking across a row increases the lag in $\hat{\theta}_\alpha$, holding the lag in \mathbf{X} constant. Looking down a column increases the lag in \mathbf{X} holding the lag in $\hat{\theta}_\alpha$ constant. Staleness in both thetas and characteristics is important for determining alphas. The average decline in alpha when moving from 1 to 12 lags of \mathbf{X} is 52%, and the average decline in alpha when moving from 1 to 12 lags of theta is 58%. Thus, updating both the characteristic and the mispricing relations seem to be important.

4.3 Nonlinear estimation

In Section 1, 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, Liao, and Wang (2016), $\mathbf{G}_\beta(\mathbf{X})$ is approximated by a series expansion in a nonparametric additive setting. The assumption of additivity ($\mathbf{G}_\beta(\mathbf{X}) = \sum g(x_1) + g(x_2) + \dots + g(x_L)$) 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} , hence 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. To avoid this issue we assume that $\mathbf{G}_\beta(\mathbf{X})$ can be well approximated by a fixed-order polynomial expansion. In the application, we use Legendre polynomials to incorporate nonlinearities in the estimation of $\mathbf{G}_\beta(\mathbf{X})$.¹⁰

In Table A.14, 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 universally higher in the polynomial specification. They range from 1.89% to 3.02% per month and are statistically significant uniformly at the 1% level. This suggests that allowing nonlinearities enables our method to estimate systematic factors and mispricing more effectively.

4.4 Small firms

The analysis in Section 3.4.2 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 average firm listed on NASDAQ. Excluding these firms reduces the sample size on average by 38% per month; that is, the total sample size shrinks from approximately 1.75 million observations to roughly 1.0 million observations. We then recompute the arbitrage portfolio using an estimation period of 12 months as in the baseline analysis. Tables A.15 and A.16 in the Internet Appendix show portfolio performance measures and estimated alphas against various factor model. Alphas decline in magnitude, but are still in the range of 1% to 2% per month and are statistically significant uniformly at the 1% level. Sharpe ratios actually improve for most specifications when the smallest firms are excluded. This is likely due to the higher idiosyncratic risk of smaller firms. The results reinforce our earlier finding that characteristics contain information about mispricing.

¹⁰Legendre polynomials are frequently used in econometrics to approximate unknown functions and fall into the more general class of “orthogonal polynomials.” We refer readers to Bierens’s (2014) handbook, which includes a deep theoretical treatment of orthogonal polynomials.

4.5 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).¹¹ Lastly, since we are dealing with an arbitrage or mispricing portfolio, we also employ the “mispricing factors” of Stambaugh and Yuan (2016).¹² Table A.17 in the Internet Appendix shows the estimated alphas and factor exposures for these additional factor models for our baseline arbitrage portfolio (12 estimation months and 6 estimated factors). From Table A.17, we can see that the arbitrage portfolio still has high and strongly significant alphas. The arbitrage portfolio has a small, but statistically significant, exposure to the “mispricing” factor of Stambaugh and Yuan (2016). The exposures to the other “alternative” factors are insignificant.

We also report the performance of the alphas of the arbitrage portfolio (with six factors) relative to several alternative latent-factor statistical benchmarks. In panel A of Table A.19 in the Internet Appendix, we examine three alternative statistical factor models, each for four different numbers of factors (1, 5, 10, and 20). Two of the benchmarks are formed using 610 decile portfolios formed by sorting by the lagged value of the characteristics. The first set of benchmarks is estimated using APC (Connor and Korajczyk (1986)) and is labeled “j PCs,” where j is the number of factors. The APC approach is similar to that in Kozak, Nagel, and Santosh (2018). The second set is estimated using the risk premium PCA method of Lettau and Pelger (2020b,a) and are labeled “j RP-PC.” The last set of benchmarks is estimated on individual equities using the IPCA method of Kelly, Pruitt, and Su (2019). This is not a rolling estimate over a short estimation window, but a full-period estimation using the algorithm of Kelly, Pruitt, and Su (2019) which accommodates missing data. This gives IPCA the large T that it apparently needs. Our arbitrage portfolio has statistically significant alphas (ranging from 1.12% to 2.45% per month) against all of the benchmarks.

¹¹The betting-against-beta factor was obtained from the AQR factor database: <https://www.aqr.com/Insights/Datasets>.

¹²We are grateful to Robert Stambaugh for making the illiquidity and mispricing factors available on his website: <http://finance.wharton.upenn.edu/~stambaugh/>.

4.6 Alternative aggregation of characteristics

Our approach of building the arbitrage portfolio has a clear interpretation and is solidly grounded in theory. However, it is not the only approach to aggregate the predictive information in many firm characteristics for building portfolios. Lewellen (2015) and Stambaugh and Yuan (2016) have introduced other prominent approaches. In this section, we apply their methods to our data and study the resultant portfolio. Lewellen's approach uses rolling Fama and MacBeth (1973) regression slopes to create return predictions at the firm level and then sorts stocks into portfolios based on the predicted return. Stambaugh and Yuan's procedure computes a composite rank by ordering individual firm characteristics and then constructs portfolios based on low versus high aggregate ranks.¹³ In both cases, we construct a monthly rebalanced "10-1" portfolio. Tables A.20 and A.22 show the alphas against common factors models, and Tables A.21 and A.23 show the alphas against alternative factor models. While the procedures are able to capture attractive returns and alphas against standard factor models, they also load more strongly on standard factors and the R^2 s are considerably higher. The discrepancy becomes even more apparent when we confront the portfolios with the higher bar of the statistical factors. Table A.19 in the Internet Appendix shows the alphas and R^2 s of the arbitrage, Lewellen, and Stambaugh-Yuan portfolios using several alternative statistical factor models. Panel A shows the results for the arbitrage portfolio for convenience; panel B shows the results for the portfolio constructed using Lewellen's approach, and panel C shows the results from applying the Stambaugh and Yuan aggregate rank procedure. Table A.19 shows that the R^2 s against statistical factors is much higher for the Lewellen and Stambaugh and Yuan portfolios relative to the arbitrage portfolio. It is also interesting that the alphas of the Lewellen and Stambaugh and Yuan approaches occasionally become insignificant when confronted with this "higher bar," particularly for the RP-PC factors. This does not happen for the arbitrage portfolio. It is therefore clear that the other approaches confound alpha and beta and their apparent success for predicting returns stems to a partially from their risk exposures.

¹³We reorder the characteristics so that low ranks correspond to low predicted returns, and high ranks correspond to high predicted returns.

5 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 methodology extends the method of PPCA of Fan, Liao, and Wang (2016) to separately identify risk and mispricing. It has several advantages, relative to double sorting on past estimated betas and lagged characteristics, which is frequently used in the literature. In simulations, we show that our methodology works well in a small- T sample and is robust to a variety of deviations from its underlying assumptions. The methodology only requires a large cross-section and can accommodate a short time span. This allows us to incorporate factor momentum (e.g., Arnott et al. (2019); Gupta and Kelly (2019)) or anomaly decay (e.g., because of attempts to trade on the anomaly) into a real-time portfolio construction process. Our approach correctly parses the ability of firm characteristics to explain the cross-section of returns into risk and mispricing components.

When applied to U.S. equities over the period from 1968 to 2018, we find that characteristics carry significant information about mispricing despite giving maximal explanatory power to the risk loadings of the statistical factor model. Alphas against popular factor models range between 1.4% and 2.6% per month (all of which are statistically significant at the 1% level). For some characteristics (e.g., intermediate momentum) the arbitrage portfolio has consistently signed weights, although the magnitudes of the exposure vary considerably over time (as shown in Figure 9). For other characteristics (e.g., profitability and turnover), the arbitrage portfolio has weights that vary in sign and magnitude. This highlights the importance of allowing the relation between mispricing and characteristics, that is, the cross-sectional pricing of characteristics, to vary over time.

Proofs

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

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

$$\hat{\mathbf{\Lambda}} = \tilde{\mathbf{\Lambda}}.$$

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), p. 970). Hence, $\hat{\mathbf{\Lambda}}$ is identified by the following two conditions:

$$\begin{aligned} \text{i) } & \hat{\mathbf{\Lambda}}'\hat{\mathbf{\Lambda}} = \mathbf{I}_K \\ \text{ii) } & \hat{\mathbf{\Lambda}}'\mathbf{Y}\mathbf{Y}'\hat{\mathbf{\Lambda}} = \mathbf{D}. \end{aligned}$$

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

$$\mathbf{Y}'\mathbf{Y} = \mathbf{Q}\mathbf{V}\mathbf{Q}'. \quad (\text{A.1})$$

Note that the $(T \times K)$ matrix composed of the first K columns of \mathbf{Q} is $\hat{\mathbf{F}}$ and that the first K -diagonal elements of \mathbf{V} correspond to the diagonal elements of \mathbf{D} :

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

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

$$\begin{aligned} \tilde{\mathbf{\Lambda}}'\tilde{\mathbf{\Lambda}} &= \mathbf{D}^{-0.5}\hat{\mathbf{F}}'\mathbf{Y}'\mathbf{Y}\hat{\mathbf{F}}\mathbf{D}^{-0.5} = \mathbf{D}^{-0.5} [\mathbf{I}_K \mathbf{0}_{K \times (T-K)}] \mathbf{Q}'\mathbf{Q}\mathbf{V}\mathbf{Q}'\mathbf{Q} [\mathbf{I}_K \mathbf{0}_{K \times (T-K)}]' \mathbf{D}^{-0.5} \\ &= \mathbf{D}^{-0.5} [\mathbf{I}_K \mathbf{0}_{K \times (T-K)}] \mathbf{V} [\mathbf{I}_K \mathbf{0}_{K \times (T-K)}]' \mathbf{D}^{-0.5} = \mathbf{D}^{-0.5}\mathbf{D}\mathbf{D}^{-0.5} = \mathbf{I}_K, \end{aligned} \quad (\text{A.3})$$

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

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

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. \square

Lemma 2. Let $\hat{\mathbf{G}}_\beta(\mathbf{X})$ denote the $(N \times K)$ matrix, the k th column of which is \sqrt{N} times the eigenvector of $\frac{\hat{\mathbf{R}}\hat{\mathbf{R}}'}{N}$ corresponding to the first k th eigenvalue of $\frac{\hat{\mathbf{R}}\hat{\mathbf{R}}'}{N}$, where $\hat{\mathbf{R}}$ is given by (5) as in Theorem 1. Define $\tilde{\mathbf{G}}_\beta(\mathbf{X}) = \hat{\mathbf{R}}\tilde{\mathbf{F}}(\tilde{\mathbf{F}}'\tilde{\mathbf{F}})^{-1}$, where $\tilde{\mathbf{F}} = \hat{\mathbf{F}}\mathbf{D}^{1/2}$; the k th column of the $(T \times K)$ matrix $\hat{\mathbf{F}}$ is the eigenvector of $\frac{\hat{\mathbf{R}}'\hat{\mathbf{R}}}{N}$ corresponding to the k th largest eigenvalue of $\frac{\hat{\mathbf{R}}'\hat{\mathbf{R}}}{N}$; and the k th element of the $(K \times K)$ -diagonal matrix \mathbf{D} is the k th largest eigenvalue of $\frac{\hat{\mathbf{R}}'\hat{\mathbf{R}}}{N}$. Then, it holds that

- (i) $\hat{\mathbf{G}}_\beta(\mathbf{X}) = \tilde{\mathbf{G}}_\beta(\mathbf{X})$
- (ii) $\mathbf{P}\hat{\mathbf{G}}_\beta(\mathbf{X}) = \hat{\mathbf{G}}_\beta(\mathbf{X})$.

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

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

Lemma 2 shows two equivalent methods for estimating the factor loading matrix. A direct approach is to calculate $\hat{\mathbf{G}}_\beta(\mathbf{X})$ by calculating the eigenvectors of the $N \times N$ matrix $\frac{\hat{\mathbf{R}}\hat{\mathbf{R}}'}{N}$ (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 $\frac{\hat{\mathbf{R}}'\hat{\mathbf{R}}}{N}$ and then to run regressions of returns on the factors to estimate the factor loadings $\tilde{\mathbf{G}}_\beta(\mathbf{X})$.

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

Proof From (5), we have that

$$\widehat{\mathbf{R}} = l_1 + l_2 + l_3,$$

where $l_1 = \mathbf{P}\mathbf{G}_\beta(\mathbf{X})\mathbf{F}'\mathbf{J}_T$, $l_2 = \mathbf{P}\Gamma_\beta\mathbf{F}'\mathbf{J}_T$ and $l_3 = \mathbf{P}\mathbf{E}\mathbf{J}_T$. Hence,

$$\frac{\widehat{\mathbf{R}}'\widehat{\mathbf{R}}}{N} = \sum_{i=1}^3 \sum_{j=1}^3 \frac{1}{N} l'_i l_j. \quad (\text{A.5})$$

Note that

$$\frac{1}{N} l'_1 l_1 = \mathbf{J}_T \mathbf{F} \left(\frac{\mathbf{G}_\beta(\mathbf{X})' \mathbf{G}_\beta(\mathbf{X})}{N} \right) \mathbf{F}' \mathbf{J}_T \xrightarrow{p} \mathbf{J}_T \mathbf{F} \mathbf{F}' \mathbf{J}_T \quad (\text{A.6})$$

from Assumption 3(ii) and that

$$\frac{1}{N} l'_1 l_2 = \mathbf{J}_T \mathbf{F} \left(\frac{\mathbf{G}_\beta(\mathbf{X})' \Gamma_\beta}{N} \right) \mathbf{F}' \mathbf{J}_T \xrightarrow{p} \mathbf{J}_T \mathbf{F} \mathbf{0}_{K \times K} \mathbf{F}' \mathbf{J}_T = \mathbf{0}_{T \times T} \quad (\text{A.7})$$

from Assumption 2(ii) and that

$$\frac{1}{N} l'_1 l_3 = \mathbf{J}_T \mathbf{F} \left(\frac{\mathbf{G}_\beta(\mathbf{X})' \mathbf{E}}{N} \right) \mathbf{J}_T \xrightarrow{p} \mathbf{J}_T \mathbf{F} \mathbf{0}_{K \times T} \mathbf{J}_T = \mathbf{0}_{T \times T} \quad (\text{A.8})$$

from Assumption 2(ii) and that

$$\frac{1}{N} l'_2 l_2 = \mathbf{J}_T \mathbf{F} \left(\frac{\Gamma'_\beta \mathbf{X}}{N} \right) \left(\frac{\mathbf{X}' \mathbf{X}}{N} \right)^{-1} \left(\frac{\mathbf{X}' \Gamma_\beta}{N} \right) \mathbf{F}' \mathbf{J}_T \xrightarrow{p} \mathbf{J}_T \mathbf{F} \mathbf{0}_{K \times L} \mathbf{V}_X^{-1} \mathbf{0}_{L \times K} \mathbf{F}' \mathbf{J}_T = \mathbf{0}_{T \times T} \quad (\text{A.9})$$

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

$$\frac{1}{N} l'_2 l_3 = \mathbf{J}_T \mathbf{F} \left(\frac{\Gamma'_\beta \mathbf{X}}{N} \right) \left(\frac{\mathbf{X}' \mathbf{X}}{N} \right)^{-1} \left(\frac{\mathbf{X}' \mathbf{E}}{N} \right) \mathbf{J}_T \xrightarrow{p} \mathbf{J}_T \mathbf{0}_{T \times L} \mathbf{V}_X^{-1} \mathbf{0}_{L \times T} \mathbf{J}_T = \mathbf{0}_{T \times T} \quad (\text{A.10})$$

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

$$\frac{1}{N} l'_3 l_3 = \mathbf{J}_T \left(\frac{\mathbf{E}' \mathbf{X}}{N} \right) \left(\frac{\mathbf{X}' \mathbf{X}}{N} \right)^{-1} \left(\frac{\mathbf{X}' \mathbf{E}}{N} \right) \mathbf{J}_T \xrightarrow{p} \mathbf{J}_T \mathbf{0}_{T \times L} \mathbf{V}_X^{-1} \mathbf{0}_{L \times T} \mathbf{J}_T = \mathbf{0}_{T \times T} \quad (\text{A.11})$$

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 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 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

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

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 $\mathbf{J}_T \mathbf{F} \mathbf{F}' \mathbf{J}_T$. Because of the continuity of eigen-decomposition, it follows that $\widehat{\mathbf{F}} \xrightarrow{p} \mathbf{J}_T \mathbf{F} (\mathbf{F}' \mathbf{J}_T \mathbf{F})^{-0.5}$.

Step 2. $\mathbf{D} \xrightarrow{p} \mathbf{F}' \mathbf{J}_T \mathbf{F}$: In Step 1, we show that $\mathbf{F}' \mathbf{J}_T \mathbf{F}$ is the diagonal matrix whose diagonal elements are eigenvalues of $\mathbf{J}_T \mathbf{F} \mathbf{F}' \mathbf{J}_T$. Because of the continuity of eigen-decomposition, 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. $\mathbf{F}' \mathbf{J}_T \widetilde{\mathbf{F}} (\widetilde{\mathbf{F}}' \widetilde{\mathbf{F}})^{-1} \xrightarrow{p} \mathbf{I}_K$: From Step 3, $\mathbf{F}' \mathbf{J}_T \widetilde{\mathbf{F}} (\widetilde{\mathbf{F}}' \widetilde{\mathbf{F}})^{-1} \xrightarrow{p} \mathbf{F}' \mathbf{J}_T \mathbf{F} (\mathbf{F}' \mathbf{J}_T \mathbf{F})^{-1} = \mathbf{I}_K$.

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

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

which gives

$$\widetilde{\mathbf{G}}_\beta(\mathbf{X}) - \mathbf{G}_\beta(\mathbf{X}) = m_1 + m_2 + m_3,$$

where

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

Hence,

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

Note that

$$\begin{aligned} \frac{1}{N}m'_1m_1 &= \left(\mathbf{F}'\mathbf{J}_T\tilde{\mathbf{F}} \left(\tilde{\mathbf{F}}'\tilde{\mathbf{F}} \right)^{-1} - \mathbf{I}_K \right)' \frac{\mathbf{G}_\beta(\mathbf{X})'\mathbf{G}_\beta(\mathbf{X})}{N} \left(\mathbf{F}'\mathbf{J}_T\tilde{\mathbf{F}} \left(\tilde{\mathbf{F}}'\tilde{\mathbf{F}} \right)^{-1} - \mathbf{I}_K \right) \\ &\xrightarrow{p} (\mathbf{I}_K - \mathbf{I}_K)' \mathbf{I}_K (\mathbf{I}_K - \mathbf{I}_K) = \mathbf{0}_{K \times K} \end{aligned} \quad (\text{A.13})$$

from Step 4 and Assumption 3(ii) and that

$$\begin{aligned} \frac{1}{N}m'_1m_2 &= \left(\mathbf{F}'\mathbf{J}_T\tilde{\mathbf{F}} \left(\tilde{\mathbf{F}}'\tilde{\mathbf{F}} \right)^{-1} - \mathbf{I}_K \right)' \frac{\mathbf{G}_\beta(\mathbf{X})'\Gamma_\beta}{N} \mathbf{F}'\mathbf{J}_T\tilde{\mathbf{F}} \left(\tilde{\mathbf{F}}'\tilde{\mathbf{F}} \right)^{-1} \\ &\xrightarrow{p} (\mathbf{I}_K - \mathbf{I}_K)' \mathbf{0}_{K \times K} \mathbf{I}_K = \mathbf{0}_{K \times K} \end{aligned} \quad (\text{A.14})$$

from Step 4 and Assumption 2(ii) and that

$$\begin{aligned} \frac{1}{N}m'_1m_3 &= \left(\mathbf{F}'\mathbf{J}_T\tilde{\mathbf{F}} \left(\tilde{\mathbf{F}}'\tilde{\mathbf{F}} \right)^{-1} - \mathbf{I}_K \right)' \frac{\mathbf{G}_\beta(\mathbf{X})'\mathbf{E}}{N} \mathbf{J}_T\tilde{\mathbf{F}} \left(\tilde{\mathbf{F}}'\tilde{\mathbf{F}} \right)^{-1} \\ &\xrightarrow{p} (\mathbf{I}_K - \mathbf{I}_K)' \mathbf{0}_{K \times T} \mathbf{J}_T \mathbf{F} (\mathbf{F}'\mathbf{J}_T\mathbf{F})^{-1} = \mathbf{0}_{K \times K} \end{aligned} \quad (\text{A.15})$$

from Step 4 and Assumption 2(ii) and that

$$\begin{aligned} \frac{1}{N}m'_2m_2 &= \left(\tilde{\mathbf{F}}'\tilde{\mathbf{F}} \right)^{-1} \tilde{\mathbf{F}}'\mathbf{J}_T\mathbf{F} \left(\frac{\Gamma'_\beta\mathbf{X}}{N} \right) \left(\frac{\mathbf{X}'\mathbf{X}}{N} \right)^{-1} \left(\frac{\mathbf{X}'\Gamma_\beta}{N} \right) \mathbf{F}'\mathbf{J}_T\tilde{\mathbf{F}} \left(\tilde{\mathbf{F}}'\tilde{\mathbf{F}} \right)^{-1} \\ &\xrightarrow{p} \mathbf{I}_K \mathbf{0}_{K \times L} \mathbf{V}_X^{-1} \mathbf{0}_{L \times K} \mathbf{I}_K = \mathbf{0}_{K \times K} \end{aligned} \quad (\text{A.16})$$

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

$$\begin{aligned} \frac{1}{N}m'_2m_3 &= \left(\tilde{\mathbf{F}}'\tilde{\mathbf{F}} \right)^{-1} \tilde{\mathbf{F}}'\mathbf{J}_T\mathbf{F} \left(\frac{\Gamma'_\beta\mathbf{X}}{N} \right) \left(\frac{\mathbf{X}'\mathbf{X}}{N} \right)^{-1} \left(\frac{\mathbf{X}'\mathbf{E}}{N} \right) \mathbf{J}_T\tilde{\mathbf{F}} \left(\tilde{\mathbf{F}}'\tilde{\mathbf{F}} \right)^{-1} \\ &\xrightarrow{p} \mathbf{I}_K \mathbf{0}_{K \times L} \mathbf{V}_X^{-1} \mathbf{0}_{L \times T} \mathbf{J}_T \mathbf{F} (\mathbf{F}'\mathbf{J}_T\mathbf{F})^{-1} = \mathbf{0}_{K \times K} \end{aligned} \quad (\text{A.17})$$

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

$$\begin{aligned} \frac{1}{N}m'_3m_3 &= \left(\tilde{\mathbf{F}}'\tilde{\mathbf{F}} \right)^{-1} \tilde{\mathbf{F}}'\mathbf{J}_T \left(\frac{\mathbf{E}'\mathbf{X}}{N} \right) \left(\frac{\mathbf{X}'\mathbf{X}}{N} \right)^{-1} \left(\frac{\mathbf{X}'\mathbf{E}}{N} \right) \mathbf{J}_T\tilde{\mathbf{F}} \left(\tilde{\mathbf{F}}'\tilde{\mathbf{F}} \right)^{-1} \\ &\xrightarrow{p} (\mathbf{F}'\mathbf{J}_T\mathbf{F})^{-1} \mathbf{F}'\mathbf{J}_T \mathbf{0}_{T \times L} \mathbf{V}_X^{-1} \mathbf{0}_{L \times T} \mathbf{J}_T \mathbf{F} (\mathbf{F}'\mathbf{J}_T\mathbf{F})^{-1} = \mathbf{0}_{K \times K}. \end{aligned} \quad (\text{A.18})$$

Finally, plugging the results of equations (A.13)-(A.18) into equation (A.12), we have that

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

Step 6. $\widehat{\mathbf{G}}_\beta(\mathbf{X}) = \widetilde{\mathbf{G}}_\beta(\mathbf{X})$: See Lemma 2(i).

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

Lemma 4. Consider $\widehat{\mathbf{G}}_\beta(\mathbf{X})$ defined in Theorem 1. Let \mathbf{Y} be a $(N \times m)$ matrix. If $\frac{1}{N}\mathbf{Y}'\mathbf{Y} \xrightarrow{p} \mathbf{V}_Y$, a positive definite matrix, then the probability limit of $\frac{1}{N}\widehat{\mathbf{G}}_\beta(\mathbf{X})'\mathbf{Y}$ is identical to the limit of $\frac{1}{N}\mathbf{G}_\beta(\mathbf{X})'\mathbf{Y}$.

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

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

From the Cauchy-Schwarz inequality, we have that

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

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

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

Proof From Lemma 4 and Assumption 2(i), it suffices to show that $\frac{1}{N}\mathbf{G}_\beta(\mathbf{X})'\mathbf{R} \xrightarrow{p} \mathbf{F}$. From the expression of \mathbf{R} in (3),

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

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 4 completes the proof of the lemma. \square

Lemma 6. *The minimization problem in Theorem 2 has the following closed form solution:*

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

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

$$\min_{\boldsymbol{\theta}, \lambda} (\bar{\mathbf{R}} - \mathbf{X}\boldsymbol{\theta})' (\bar{\mathbf{R}} - \mathbf{X}\boldsymbol{\theta}) + \lambda \hat{\mathbf{G}}_{\beta}(\mathbf{X})' \mathbf{X}\boldsymbol{\theta}.$$

The first order conditions are

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

which yields

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

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

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

which completes the proof of the lemma. \square

Proof of Theorems 2 and 3 Recall that $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. From Lemmas 2(ii) and 6, we have that

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

which in conjunction with the expression of $\bar{\mathbf{R}}$ in (6) yields

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

with n_i for $i = 1, 2, 3$ are given by $n_1 = \mathbf{P}(\Gamma_{\alpha} + \Gamma_{\beta}\bar{\mathbf{F}} + \bar{\mathbf{E}})$, $n_2 = \mathbf{G}_{\beta}(\mathbf{X})\bar{\mathbf{F}}$, and $n_3 = -\hat{\mathbf{G}}_{\beta}(\mathbf{X}) \left(\hat{\mathbf{G}}_{\beta}(\mathbf{X})' \hat{\mathbf{G}}_{\beta}(\mathbf{X}) \right)^{-1} \hat{\mathbf{G}}_{\beta}(\mathbf{X})' \bar{\mathbf{R}}$. Then,

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

Note that

$$\begin{aligned} \frac{1}{N}n'_1n_1 &= \left(\frac{\mathbf{X}'\Gamma_\alpha}{N} + \frac{\mathbf{X}'\Gamma_\beta}{N}\bar{\mathbf{F}} + \frac{\mathbf{X}'\mathbf{E}\mathbf{1}_T}{N\frac{1}{T}} \right)' \left(\frac{\mathbf{X}'\mathbf{X}}{N} \right)^{-1} \left(\frac{\mathbf{X}'\Gamma_\alpha}{N} + \frac{\mathbf{X}'\Gamma_\beta}{N}\bar{\mathbf{F}} + \frac{\mathbf{X}'\mathbf{E}\mathbf{1}_T}{N\frac{1}{T}} \right) \\ &\xrightarrow{p} \left(\mathbf{0}_L + \mathbf{0}_{L \times K}\bar{\mathbf{F}} + \mathbf{0}_{L \times T}\frac{\mathbf{1}_T}{T} \right)' \mathbf{V}_X^{-1} \left(\mathbf{0}_L + \mathbf{0}_{L \times K}\bar{\mathbf{F}} + \mathbf{0}_{L \times T}\frac{\mathbf{1}_T}{T} \right) = 0 \end{aligned} \quad (\text{A.20})$$

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

$$\frac{1}{N}n'_2n_2 = \bar{\mathbf{F}}' \left(\frac{\mathbf{G}_\beta(\mathbf{X})' \mathbf{G}_\beta(\mathbf{X})}{N} \right) \bar{\mathbf{F}} \xrightarrow{p} \bar{\mathbf{F}}' \bar{\mathbf{F}}. \quad (\text{A.21})$$

from Assumption 3(ii) and that

$$\frac{1}{N}n'_3n_3 = \frac{\mathbf{1}_T' \mathbf{R}' \hat{\mathbf{G}}_\beta(\mathbf{X})}{T} \frac{\hat{\mathbf{G}}_\beta(\mathbf{X})' \mathbf{R} \mathbf{1}_T}{N} \frac{1}{T} \xrightarrow{p} \bar{\mathbf{F}}' \bar{\mathbf{F}} \quad (\text{A.22})$$

from Lemma 5 and that

$$\frac{1}{N}n'_1n_2 \leq \sqrt{\frac{1}{N}n'_1n_1} \sqrt{\frac{1}{N}n'_2n_2} \xrightarrow{p} 0 \quad (\text{A.23})$$

from the Cauchy-Schwarz inequality and the limits of (A.20) and (A.21) and that

$$\frac{1}{N}n'_1n_3 \leq \sqrt{\frac{1}{N}n'_1n_1} \sqrt{\frac{1}{N}n'_3n_3} \xrightarrow{p} 0 \quad (\text{A.24})$$

from the Cauchy-Schwarz inequality and the limits of (A.20) and (A.22) and that

$$\frac{1}{N}n'_2n_3 = -\bar{\mathbf{F}}' \left(\frac{\mathbf{G}_\beta(\mathbf{X})' \hat{\mathbf{G}}_\beta(\mathbf{X})}{N} \right) \frac{\hat{\mathbf{G}}_\beta(\mathbf{X})' \mathbf{R} \mathbf{1}_T}{N} \frac{1}{T} \xrightarrow{p} -\bar{\mathbf{F}}' \bar{\mathbf{F}} \quad (\text{A.25})$$

from Lemmas 4 and 5 and Assumption 3(ii). Finally, plugging the results of equations (A.20)-(A.25) into equation (A.19), we have that

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

which proves Theorem 2.

Next, we turn to Theorem 3.

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

We explain that $\mathbf{w}' \mathbf{R} \xrightarrow{p} \delta \mathbf{1}_T'$ in the text. Hence, it suffices to show that $(\hat{\mathbf{w}} - \mathbf{w})' \mathbf{R}$ shrinks

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

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

where the last limit is from (A.26) and Assumption 2(i). This completes the proof of Theorem 3. □

References

- Ahn, S. C., and A. R. Horenstein. 2013. Eigenvalue ratio test for the number of factors. *Econometrica* 81:1203–27.
- Alti, A., and S. Titman. 2019. A dynamic model of characteristic-based return predictability. *Journal of Finance* 74:3187–216.
- Arnott, R. D., M. Clements, V. Kalesnik, and J. T. Linnainmaa. 2019. Factor momentum. *Available at SSRN 3116974* .
- Bai, J., and S. Ng. 2002. Determining the number of factors in approximate factor models. *Econometrica* 70:191–221.
- Baker, M., and J. Wurgler. 2006. Investor sentiment and the cross-section of stock returns. *Journal of Finance* 61:1645–80.
- Ball, R., J. Gerakos, J. T. Linnainmaa, and V. V. Nikolaev. 2015. Deflating profitability. *Journal of Financial Economics* 117:225–48.
- Banz, R. W. 1981. The relationship between return and market value of common stocks. *Journal of Financial Economics* 9:3–18.
- Berk, J. B. 2000. Sorting out sorts. *Journal of Finance* 55:407–27.
- Bierens, H. 2014. *The oxford handbook of applied nonparametric and semiparametric econometrics and statistics*, chap. The Hilbert Space Theoretical Foundation of Semi-Nonparametric Modeling, 3–37. Oxford University Press.
- Brown, S. J. 1989. The number of factors in security returns. *Journal of Finance* 44:1247–62.
- Carhart, M. M. 1997. On persistence in mutual fund performance. *Journal of Finance* 52:57–82.
- Chen, Z., B. Liu, H. Wang, Z. Wang, and J. Yu. 2018. Characteristics-based factors. Tsinghua University, Working Paper.
- Chinco, A., A. Neuhierl, and M. Weber. 2020. Estimating the anomaly base rate. *Journal of Financial Economics (forthcoming)* .

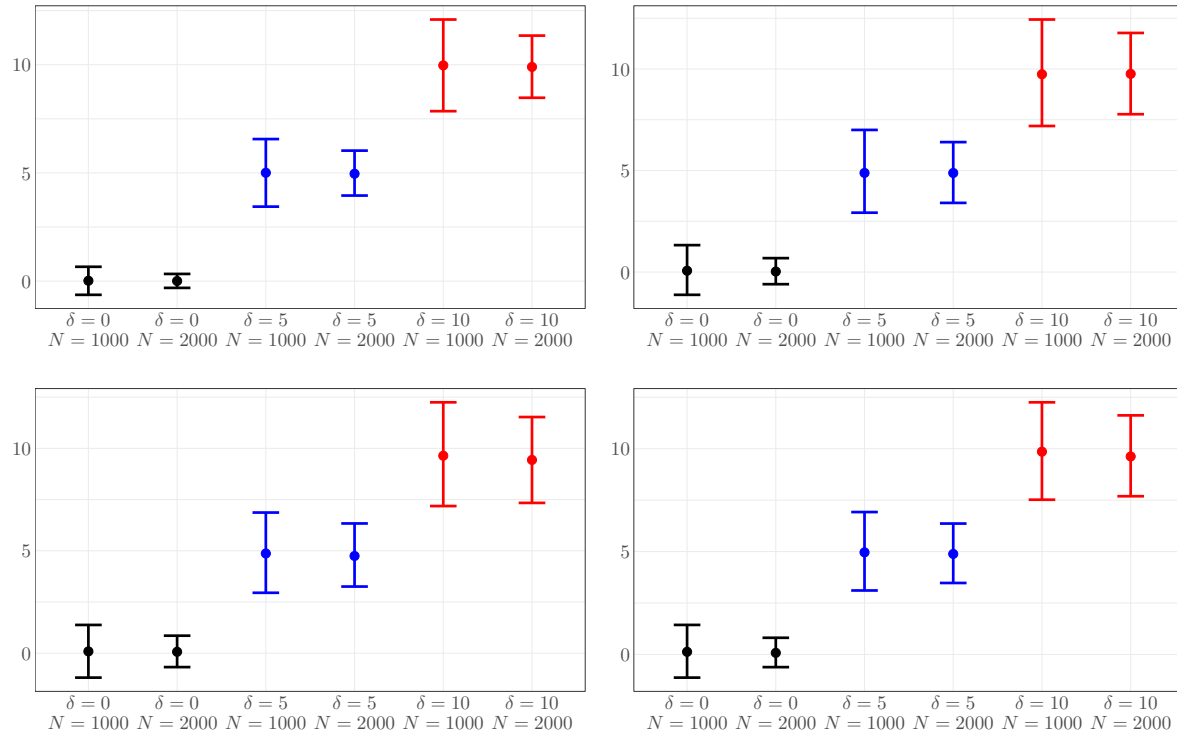
- Connor, G., M. Hagmann, and O. Linton. 2012. Efficient semiparametric estimation of the Fama–French model and extensions. *Econometrica* 80:713–54.
- Connor, G., and R. A. Korajczyk. 1986. Performance measurement with the arbitrage pricing theory: A new framework for analysis. *Journal of Financial Economics* 15:373–94.
- . 1993. A test for the number of factors in an approximate factor model. *Journal of Finance* 48:1263–91.
- Connor, G., and O. Linton. 2007. Semiparametric estimation of a characteristic-based factor model of common stock returns. *Journal of Empirical Finance* 14:694–717.
- Daniel, K., L. Mota, S. Rottke, and T. Santos. 2020. The cross-section of risk and returns. *Review of Financial Studies* 33:1927–79.
- Daniel, K., and S. Titman. 1997. Evidence on the characteristics of cross sectional variation in stock returns. *Journal of Finance* 52:1–33.
- Datar, V. T., N. Y. Naik, and R. Radcliffe. 1998. Liquidity and stock returns: An alternative test. *Journal of Financial Markets* 1:203–19.
- Fama, E. F., and K. R. French. 1992. The cross-section of expected stock returns. *Journal of Finance* 47:427–65.
- . 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33:3–56.
- . 2015. A five-factor asset pricing model. *Journal of Financial Economics* 116:1–22.
- Fama, E. F., and J. D. MacBeth. 1973. Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy* 81:607–36.
- Fan, J., Y. Liao, and W. Wang. 2016. Projected principal component analysis in factor models. *Annals of Statistics* 44:219–54.
- Ferson, W. E., and C. R. Harvey. 1997. Fundamental determinants of national equity market returns: A perspective on conditional asset pricing. *Journal of Banking & Finance* 21:1625 – 1665.

- . 1999. Conditioning variables and the cross section of stock returns. *Journal of Finance* 54:1325–60.
- Frazzini, A., and L. H. Pedersen. 2014. Betting against beta. *Journal of Financial Economics* 111:1–25.
- Freyberger, J., A. Neuhierl, and M. Weber. 2020. Dissecting characteristics nonparametrically. *Review of Financial Studies* 33:2326–77.
- Ghysels, E. 1998. On stable factor structures in the pricing of risk: do time-varying betas help or hurt? *Journal of Finance* 53:549–73.
- Greene, W. H. 2008. *Econometric analysis*. 6th ed. Pearson Education, Inc.
- Gupta, T., and B. Kelly. 2019. Factor momentum everywhere. *Journal of Portfolio Management* 45:13–36.
- Herskovic, B., A. Moreira, and T. Muir. 2019. Hedging risk factors. *Available at SSRN 3148693* .
- Hou, K., C. Xue, and L. Zhang. 2015. Digesting anomalies: An investment approach. *Review of Financial Studies* 28:660–705.
- Jagannathan, R., and Y. Wang. 2007. Lazy investors, discretionary consumption, and the cross-section of stock returns. *Journal of Finance* 62:1623–61.
- Jegadeesh, N., and S. Titman. 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance* 48:65–91.
- Kelly, B., S. Pruitt, and Y. Su. 2017. Instrumented principal component analysis. *Unpublished Manuscript, University of Chicago* .
- . 2019. Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics* 134:501–34.
- Kozak, S., S. Nagel, and S. Santosh. 2018. Interpreting factor models. *Journal of Finance* 73:1183–223.
- Lettau, M., and M. Pelger. 2020a. Estimating latent asset-pricing factors. *Journal of Econometrics* 218:1–31.

- . 2020b. Factors that fit the time series and cross-section of stock returns. *Review of Financial Studies* 33:2274–325.
- Lewellen, J. 2015. The cross section of expected stock returns. *Critical Finance Review* 4:1–44.
- Light, N., D. Maslov, and O. Rytchkov. 2017. Aggregation of information about the cross section of stock returns: A latent variable approach. *Review of Financial Studies* 30:1339–81.
- Linnainmaa, J. T., and M. R. Roberts. 2018. The history of the cross-section of stock returns. *Review of Financial Studies* 31:2606–49.
- Litzenberger, R. H., and K. Ramaswamy. 1979. The effect of personal taxes and dividends on capital asset prices: Theory and empirical evidence. *Journal of Financial Economics* 7:163–95.
- McLean, D. R., and J. Pontiff. 2016. Does academic research destroy return predictability. *Journal of Finance* 71:5–32.
- Newey, W. K., and K. D. West. 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55:703–8.
- Pástor, v., and R. F. Stambaugh. 2003. Liquidity risk and expected stock returns. *Journal of Political Economy* 111:642–85.
- Rosenberg, B., and W. McKibben. 1973. The prediction of systematic and specific risk in common stocks. *Journal of Financial and Quantitative Analysis* 8:317–33.
- Ross, S. A. 1976. The arbitrage theory of capital asset pricing. *Journal of Economic Theory* 13:341–60.
- Sloan, R. 1996. Do stock prices fully reflect information in accruals and cash flows about future earnings? *The Accounting Review* 71:289–315.
- Stambaugh, R. F., and Y. Yuan. 2016. Mispricing factors. *Review of Financial Studies* 30:1270–315.

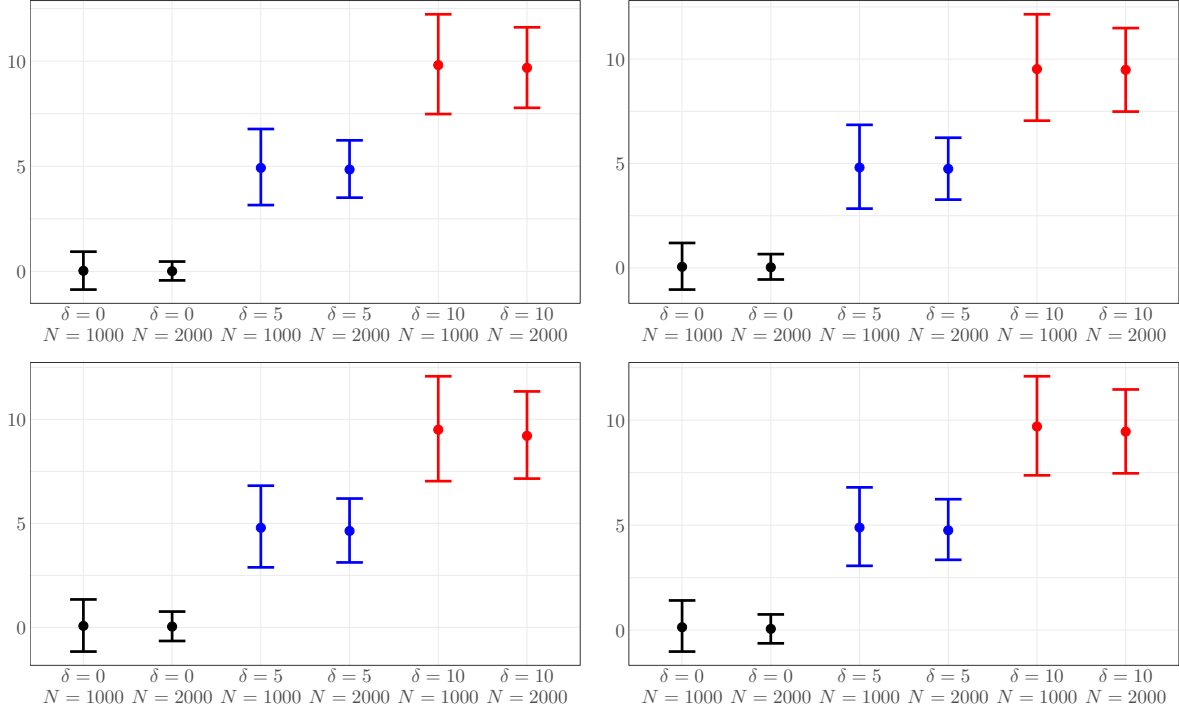
Figures and Tables

Figure 1: Simulated arbitrage portfolio returns in the CAPM, FF3, FF5, and HXZ4 models (correctly specified model)



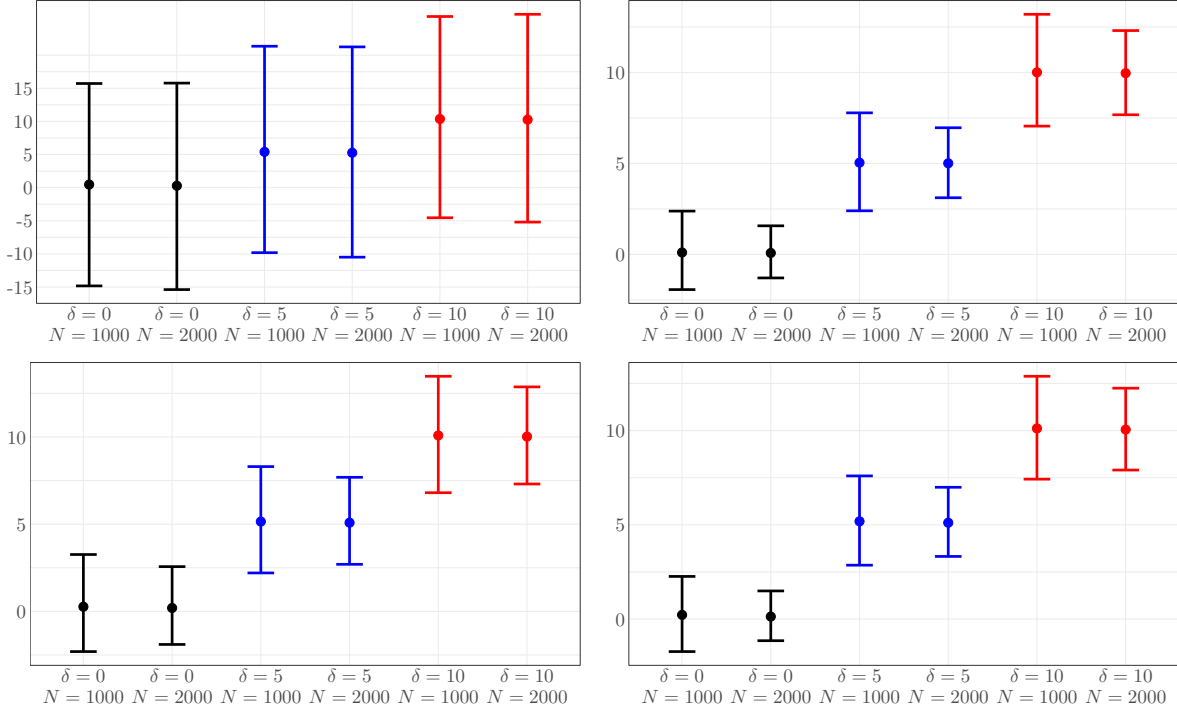
This figure shows the simulation results of the arbitrage portfolio when the return-generating process is calibrated to the CAPM (upper-left panel), the Fama-French three-factor model (upper-right panel), the Fama-French five-factor model (lower-left panel), and the Hou-Xue-Zhang four-factor model (lower-right panel). The arbitrage portfolio $\hat{\mathbf{w}}$ is constructed with the returns from $t = 1$ to $t = 12$, and it generates 1-month out-of-sample returns (at time 13). The solid dot represents the mean of the arbitrage portfolio in the out-of-sample period over 10,000 simulations. Error bars represent the 95% confidence interval. In this simulation, we use the correct number of factors to construct the arbitrage portfolio, that is, $K = 1$ 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 2: Simulated arbitrage portfolio returns in the CAPM, FF3, FF5, and HXZ4 models with $K_{\text{wrong}} = K_{\text{true}} + 1$ (selecting too many factors)



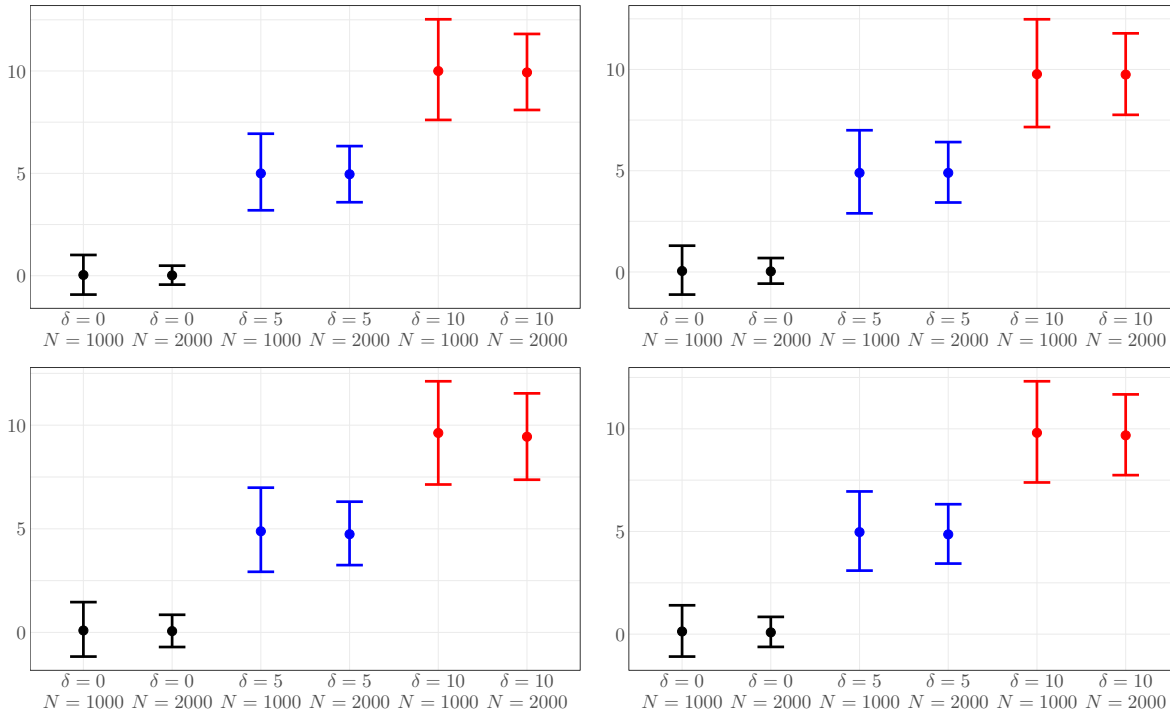
This figure shows the simulation results of the arbitrage portfolio when the return-generating process is calibrated to the CAPM (upper-left panel), the Fama-French three-factor model (upper-right panel), the Fama-French five-factor model (lower-left panel), and the Hou-Xue-Zhang four-factor model (lower-right panel). The arbitrage portfolio $\hat{\mathbf{w}}$ is constructed with the returns from $t = 1$ to $t = 12$, and it generates 1-month out-of-sample returns (at time 13). The solid dot represents the mean of the arbitrage portfolio in the out-of-sample period over 10,000 simulations. Error bars represent the 95% confidence interval. In this simulation, we use too many factors in constructing the arbitrage portfolio, that is, $K_{\text{wrong}} = 2$ 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 3: Simulated arbitrage portfolio returns in the CAPM, FF3, FF5, and HXZ4 models with $K_{\text{wrong}} = K_{\text{true}} - 1$ (selecting too few factors)



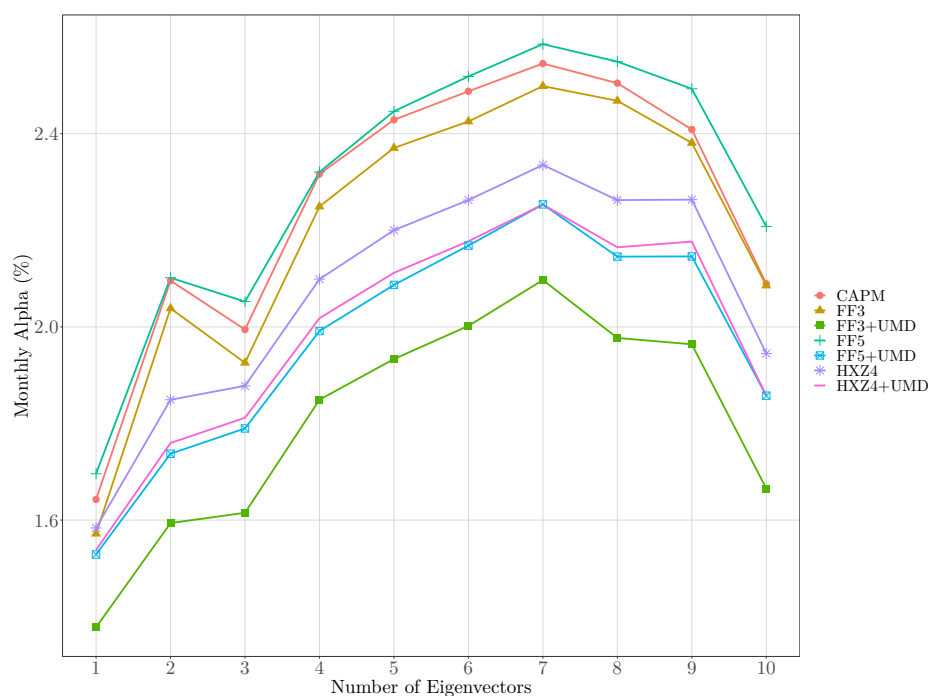
This figure shows the simulation results of the arbitrage portfolio when the return-generating process is calibrated to the CAPM (upper-left panel), the Fama-French three-factor model (upper-right panel), the Fama-French five-factor model (lower-left panel), and the Hou-Xue-Zhang four-factor model (lower-right panel). The arbitrage portfolio $\hat{\mathbf{w}}$ is constructed with the returns from $t = 1$ to $t = 12$, and it generates 1-month out-of-sample returns (at time 13). The solid dot represents the mean of the arbitrage portfolio in the out-of-sample period over 10,000 simulations. Error bars represent the 95% confidence interval. In this simulation, we use too few factors in constructing the arbitrage portfolio, that is, $K_{\text{wrong}} = 0$ 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 4: Simulated arbitrage portfolio returns in the CAPM, FF 3, FF5, and HXZ4 models with time-varying characteristics



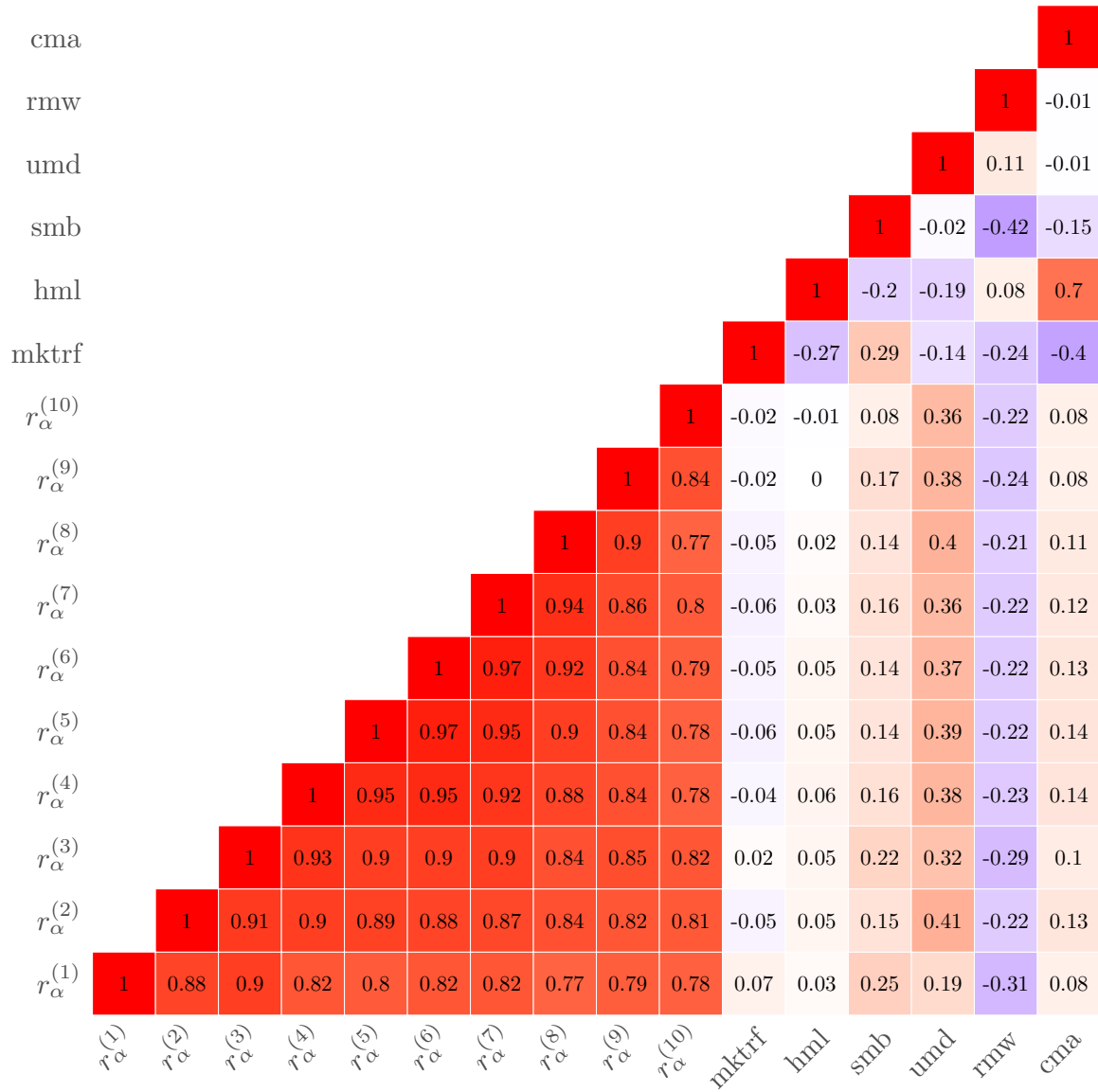
This figure shows the simulation results of the arbitrage portfolio when the return-generating process is calibrated to the CAPM (upper-left panel), the Fama-French three-factor model (upper-right panel), the Fama-French five-factor model (lower-left panel), and the Hou-Xue-Zhang four-factor model (lower-right panel). The arbitrage portfolio $\hat{\mathbf{w}}$ is constructed with the returns from $t = 1$ to $t = 12$, and it generates 1-month out-of-sample returns (at time 13). The solid dot represents the mean of the arbitrage portfolio in the out-of-sample period over 10,000 simulations. Error bars represent the 95% confidence interval. In this simulation, we use the correct number of factors in constructing the arbitrage portfolio, that is, $K = 1$ 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. Section 2.2.3 details the construction.

Figure 5: Alpha for varying the number of eigenvectors



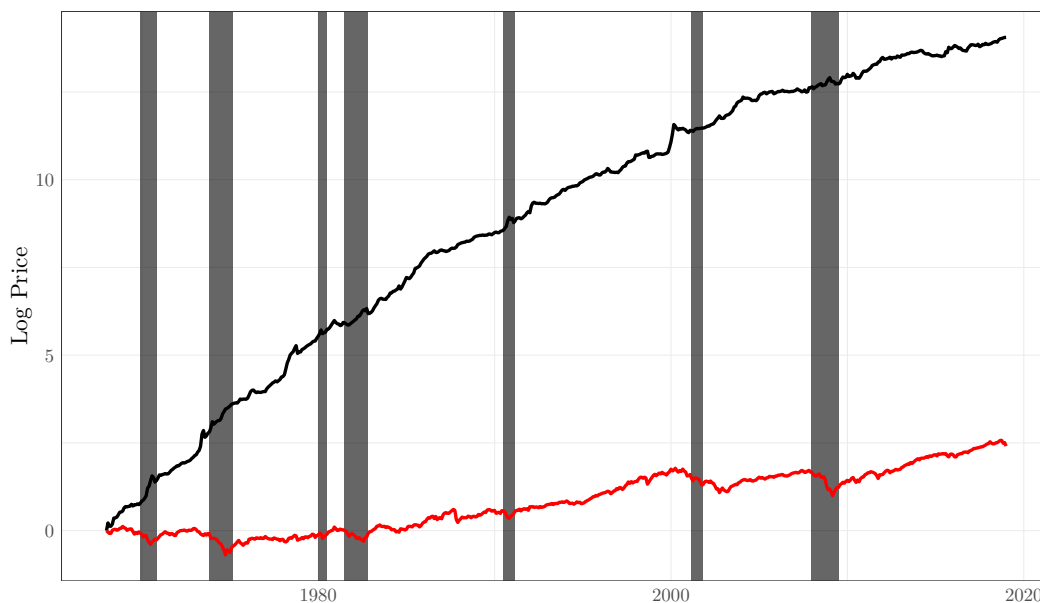
This figure shows the monthly alpha of the arbitrage portfolio against the CAPM, the Fama-French three-factor model, and the Fama-French five-factor model, as well as their “momentum-augmented” versions for 1 through 10 eigenvectors. The sample period is from January 1968 to December 2018.

Figure 6: Correlation matrix with common factors



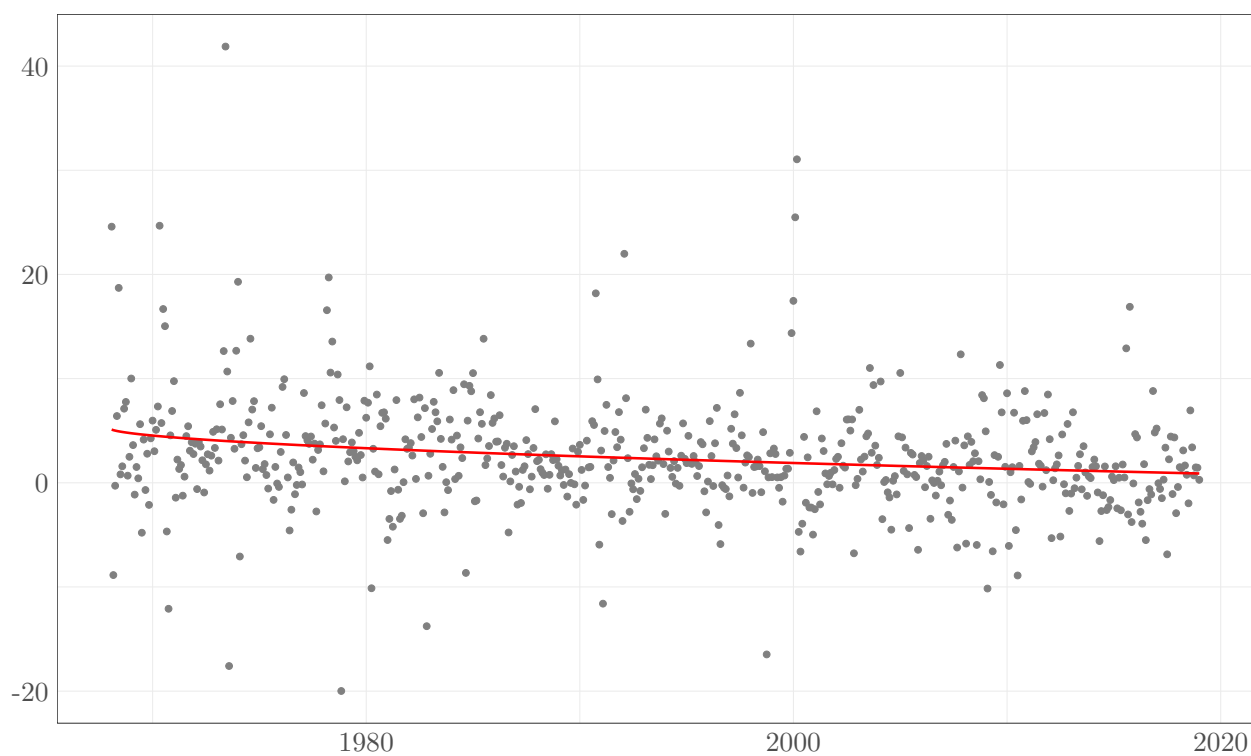
This figure shows the correlation matrix between the arbitrage portfolios with 1 through 10 eigenvectors, $r_{\alpha}^{(1)}, r_{\alpha}^{(2)}, \dots, r_{\alpha}^{(10)}$, the Fama-French five factors, and the momentum factor. The sample period is January 1968 to December 2018.

Figure 7: Price path and yearly returns of the arbitrage portfolio



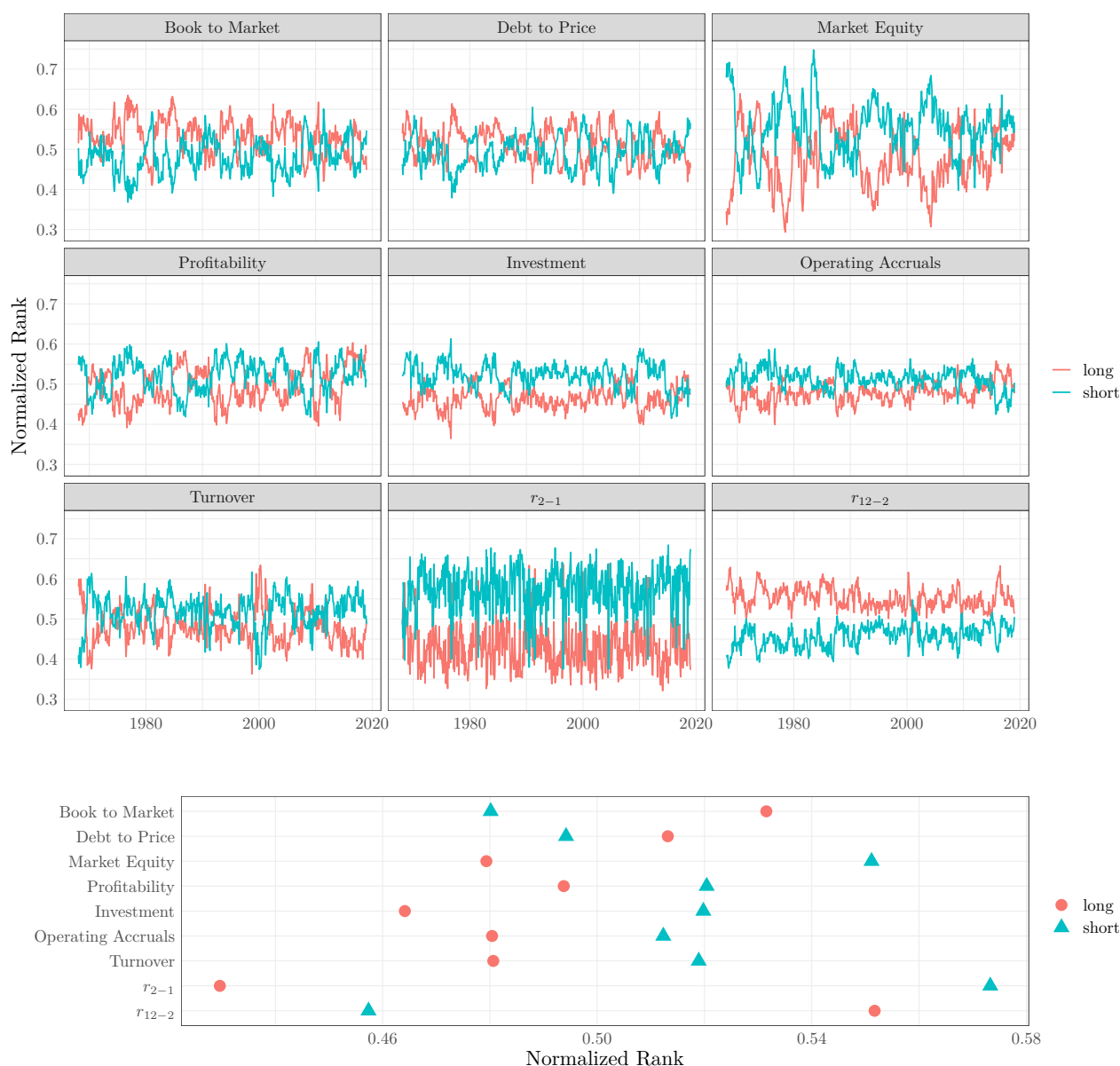
The black line represents the logarithmic price path (i.e., the cumulative returns) of the arbitrage portfolio (using six eigenvectors), and the red line represents the market portfolio. The gray-shaded areas represent NBER recessions. The sample period is January 1968 to December 2018.

Figure 8: Monthly returns of the arbitrage portfolio, 1968–2018



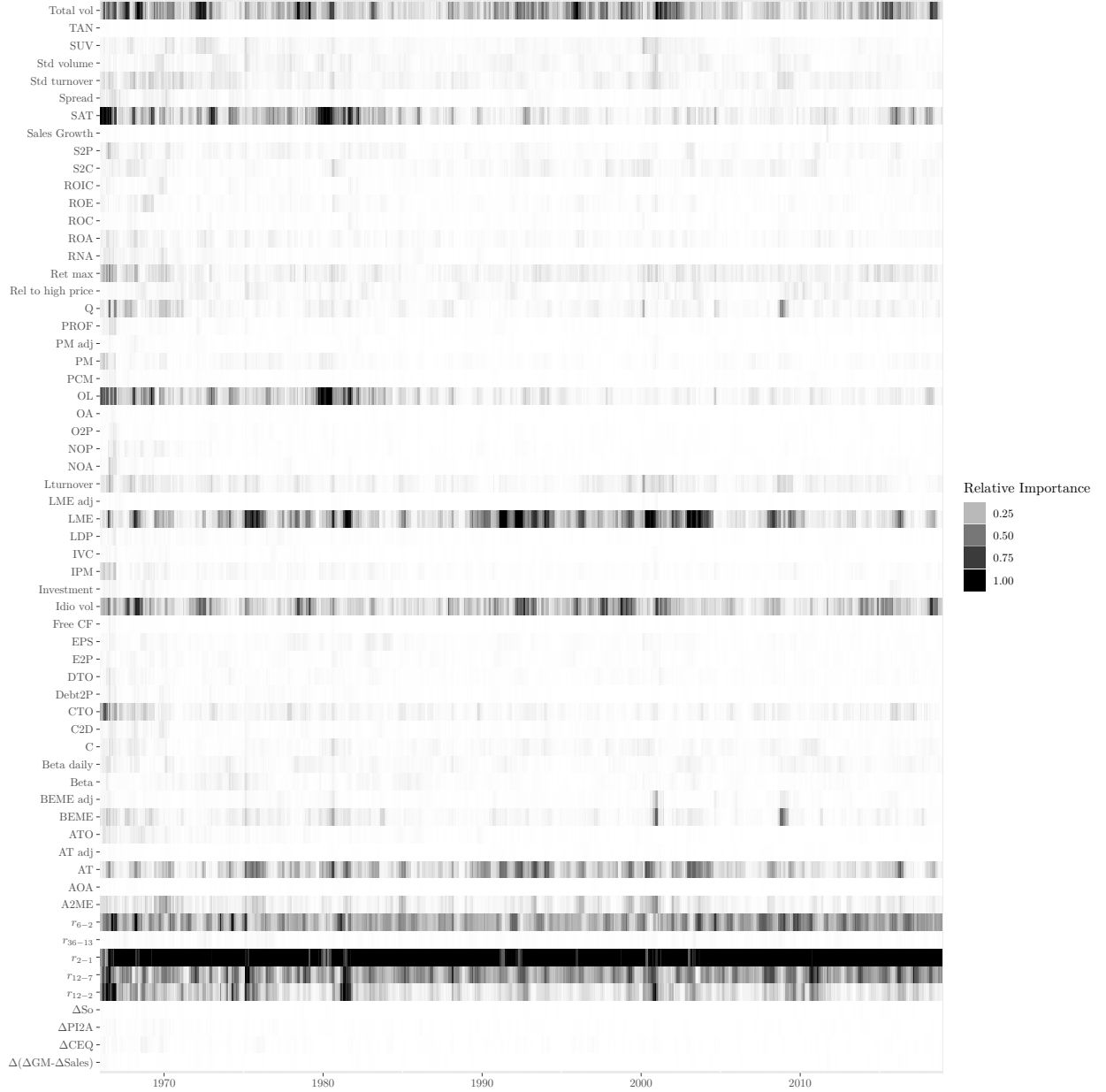
This figure shows the monthly excess returns of the arbitrage portfolio (six eigenvectors) from January 1968 through December 2018 and a time trend (red). The time trend is estimated by $r_t = a + b \times t^\gamma + \varepsilon_t$ with $\hat{a} = 5.22$, $\hat{b} = -0.11$, $\hat{\gamma} = 0.57$.

Figure 9: Firm characteristics of the long and short legs of the arbitrage portfolio



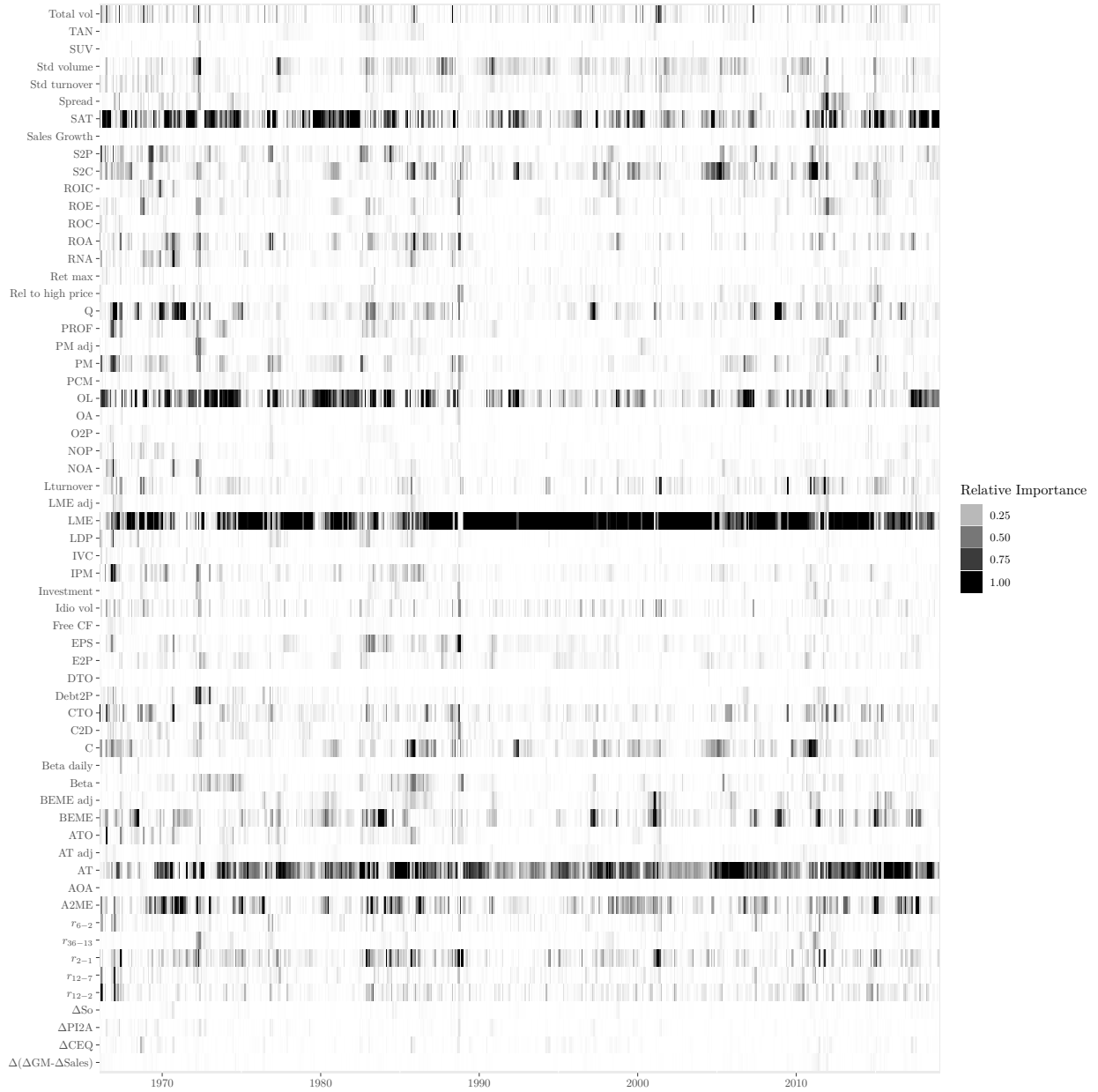
This figure shows the normalized rank of nine cross-sectional return characteristics for the long and short legs of the arbitrage portfolio. The firm characteristics are the book-to-market ratio, the debt-to-price ratio, market equity (size), profitability, investment, operating accruals, last month's volume, the return 1 month before portfolio formation (r_{2-1}) and the return from 12 to 2 months before portfolio formation (r_{12-2}). Each month, the characteristics are normalized to be in the unit interval; that is, 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 cross-sectional comparison over time. The sample period is January 1968 to December 2018.

Figure 10: Beta heatmap



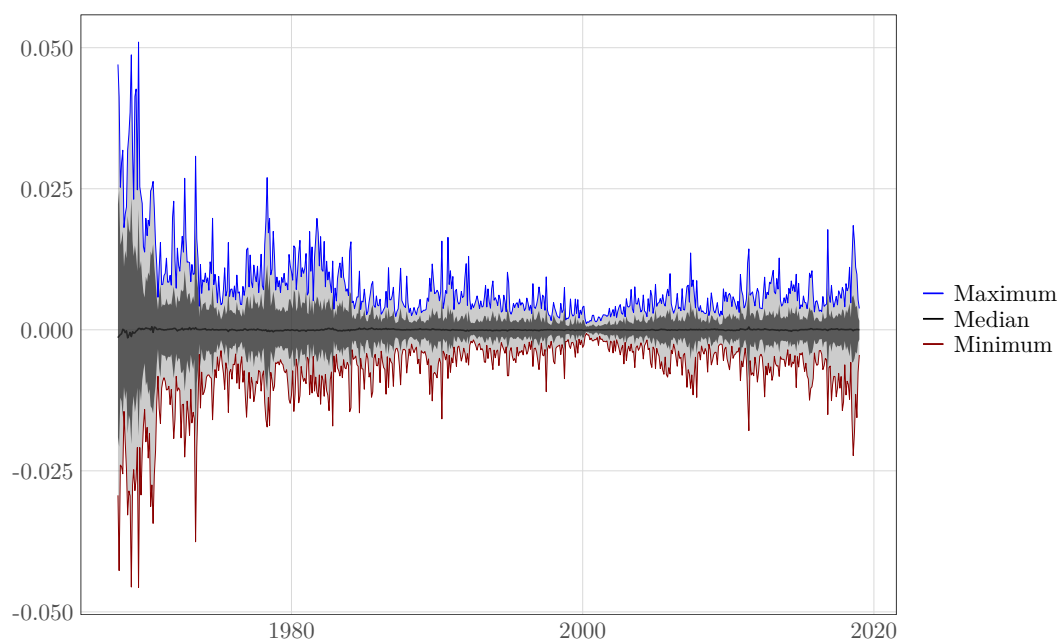
This figure plots a beta heatmap $\hat{\beta}_{(l)}^{\text{norm}}$ for each characteristic, l , computed as follows. We project the k th column of $\hat{\mathbf{G}}_{\beta}(\mathbf{X})$ onto the characteristics at each month: $\hat{\mathbf{G}}_{\beta}(\mathbf{X})_k = \beta_{0,k} + \mathbf{X}\beta_k + \varepsilon$. We then take the absolute value of $\hat{\beta}_k$ for each characteristic and compute $\hat{\beta}_{(l)} = \sum_{k=1}^K |\hat{\beta}_{k,l}|$. We then normalize cross-sectionally to obtain the normalized sum of absolute coefficients $\hat{\beta}_{(l)}^{\text{norm}} = \frac{\hat{\beta}_{(l)}}{\max_j \hat{\beta}_{(j)}}$. This way, the characteristic with the largest (absolute) sum of coefficients has $\hat{\beta}_{(l)}^{\text{norm}} = 1$. We then repeat this process each month, sliding the estimation window forward. The sample period is January 1968 to December 2018.

Figure 11: Alpha heatmap



This figure plots an alpha heatmap $\hat{\alpha}_{(l)}^{\text{norm}}$ for each characteristic, l , computed as follows. We project $\hat{\mathbf{G}}_{\alpha}(\mathbf{X})$ onto the characteristics at each month: $\hat{\mathbf{G}}_{\alpha}(\mathbf{X}) = \alpha_0 + \mathbf{X}\boldsymbol{\alpha} + \varepsilon$. We then take the absolute value of $\hat{\boldsymbol{\alpha}}$ for each characteristic and compute $\hat{\alpha}_{(l)} = |\hat{\alpha}_l|$. We then normalize cross-sectionally to obtain the normalized absolute coefficient $\hat{\alpha}_{(l)}^{\text{norm}} = \frac{\hat{\alpha}_{(l)}}{\max_j \hat{\alpha}_{(j)}}$. This way, the characteristic with the largest (absolute) coefficient has $\hat{\alpha}_{(l)}^{\text{norm}} = 1$. We then repeat this process each month, sliding the estimation window forward. The sample period is January 1968 to December 2018.

Figure 12: Portfolio weights



This figure shows the median, minimum, maximum, 5% quantile and 95% quantile of the portfolio weights of the arbitrage portfolio (with five eigenvectors). The solid-black line represents the median portfolio weight in a given month; the dark-gray area represents the 5% and 95% quantiles of the weights in a month; and the light-gray area represents the monthly minimum and maximum. The sample period is January 1968 to December 2018.

Table 1: Average returns on double-sorted portfolio in a simulated CAPM economy

		Past Beta											
		Low										High	
		1	2	3	4	5	6	7	8	9	10	10-1	
Low	1	0.25	0.29	0.34	0.26	0.15	0.34	0.25	0.18	0.14	0.23	-0.03	
	2	0.40	0.32	0.39	0.36	0.37	0.39	0.29	0.32	0.25	0.47	0.06	
	3	0.37	0.41	0.42	0.41	0.46	0.28	0.47	0.46	0.48	0.46	0.09	
	4	0.46	0.45	0.45	0.36	0.42	0.44	0.39	0.45	0.58	0.47	0.02	
	5	0.48	0.37	0.48	0.45	0.52	0.51	0.47	0.48	0.44	0.47	-0.01	
	6	0.56	0.51	0.61	0.43	0.47	0.56	0.59	0.47	0.56	0.55	-0.01	
	7	0.52	0.54	0.61	0.58	0.60	0.60	0.51	0.56	0.71	0.59	0.07	
High	8	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	0.68	0.59	0.71	0.66	0.64	0.71	0.69	0.67	0.70	0.03	
	10	0.78	0.74	0.68	0.83	0.75	0.74	0.85	0.78	0.80	0.86	0.08	
	10-1	0.52***	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 $i = 1, \dots, 2,000$ and $t = 1, \dots, 2,000$ with the following calibration: $f_{M,t} \sim \mathcal{N}(\mu_M, \sigma_M^2)$, $\beta_i \sim \mathcal{N}(1, \sigma_\beta^2)$, $\varepsilon_{i,t} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$, where $\mu_M = 5\%/12$, $\sigma_M = \sqrt{(20\%)^2/12}$, $\sigma_\beta = 0.4$, $\sigma_\varepsilon = 2\sigma_M$. Reported numbers are averages over $t = 61, \dots, 2,000$.

Table 2: Firm characteristics by category

<u>Past-returns:</u>		<u>Value:</u>			
(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)	Δ SO	Log change in split-adjusted shares outstanding
			(38)	Debt2P	Total debt to Size
			(39)	E2P	Income before extraordinary items to Size
			(40)	Free CF	Free cash flow to BE
			(41)	LDP	Trailing 12-months dividends to price
			(42)	NOP	Net payouts to Size
			(43)	O2P	Operating payouts to market cap
			(44)	Q	Tobin's Q
			(45)	S2P	Sales to price
			(46)	Sales_g	Sales growth
			<u>Trading frictions:</u>		
			(47)	AT	Total assets
			(48)	Beta	Correlation \times ratio of vols
			(49)	Beta daily	CAPM beta using daily returns
			(50)	DTO	De-trended Turnover - market Turnover
			(51)	Idio vol	Idio vol of Fama-French 3 factor model
			(52)	LME	Price times shares outstanding
			(53)	LME_adj	Size - mean size in Fama-French 48 industry
			(54)	Lturnover	Last month's volume to shares outstanding
			(55)	Rel.to_high_price	Price to 52 week high price
			(56)	Ret_max	Maximum daily return
			(57)	Spread	Average daily bid-ask spread
			(58)	Std turnover	Standard deviation of daily turnover
			(59)	Std volume	Standard deviation of daily volume
			(60)	SUV	Standard unexplained volume
			(61)	Total vol	Standard deviation of daily returns
<u>Investment:</u>					
(6)	Investment	% change in AT			
(7)	Δ CEQ	% change in BE			
(8)	Δ PI2A	Change in PP&E and inventory over lagged AT			
(9)	IVC	Change in inventory over average AT			
(10)	NOA	Net-operating assets over lagged AT			
<u>Profitability:</u>					
(11)	ATO	Sales to lagged net operating assets			
(12)	CTO	Sales to lagged total assets			
(13)	$\Delta(\Delta$ GM- Δ Sales)	Δ (% change in gross margin and % change in sales)			
(14)	EPS	Earnings per share			
(15)	IPM	Pre-tax income over sales			
(16)	PCM	Sales minus costs of goods sold to sales			
(17)	PM	OI after depreciation over sales			
(18)	PM_adj	Profit margin - mean PM in Fama-French 48 industry			
(19)	Prof	Gross profitability over BE			
(20)	RNA	OI after depreciation to lagged net operating assets			
(21)	ROA	Income before extraordinary items to lagged AT			
(22)	ROC	Size + longterm debt - total assets to cash			
(23)	ROE	Income before extraordinary items to lagged BE			
(24)	ROIC	Return on invested capital			
(25)	S2C	Sales to cash			
(26)	SAT	Sales to total assets			
(27)	SAT_adj	SAT - mean SAT in Fama-French 48 industry			
<u>Intangibles:</u>					
(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, Neuhierl, and Weber (2020). 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 December 2018.

Table 3: Portfolio performance statistics

# Eigenvectors	Mean (%)	Standard Deviation (%)	Sharpe Ratio	Skewness	Kurtosis	Maximum Drawdown	Worst Month (%)	Best Month (%)
1	20.08	14.70	1.37	1.24	8.30	22.37	-18.77	29.46
2	24.84	17.51	1.42	0.53	6.07	23.16	-22.44	30.06
3	24.02	14.66	1.64	1.28	10.23	20.91	-19.89	34.05
4	27.54	16.56	1.66	1.07	6.71	22.21	-19.61	30.66
5	28.71	17.94	1.60	1.10	7.11	20.08	-20.08	36.16
6	29.48	18.42	1.60	1.29	8.67	20.84	-19.99	41.88
7	30.13	18.31	1.65	1.34	9.04	21.92	-20.21	42.82
8	29.67	19.84	1.50	1.26	10.38	27.92	-26.02	42.74
9	28.75	17.74	1.62	1.32	8.66	27.05	-20.43	36.56
10	24.93	19.02	1.31	0.29	15.72	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 1 through 10 eigenvectors is estimated every month using the steps outlined in Section 3. The sample period is January 1968 to December 2018.

Table 4: Risk-adjusted returns with one eigenvector

	CAPM	FF3	FF3+UMD	FF5	FF5+UMD	HXZ4	HXZ4+UMD
alpha	1.64*** (0.22)	1.57*** (0.21)	1.38*** (0.20)	1.70*** (0.23)	1.53*** (0.23)	1.58*** (0.25)	1.54*** (0.25)
mktrf	0.06 (0.07)	0.01 (0.06)	0.06 (0.05)	0.00 (0.07)	0.03 (0.06)		
smb		0.36** (0.17)	0.36** (0.14)	0.22 (0.14)	0.22* (0.11)		
hml		0.13 (0.13)	0.22 (0.14)	0.05 (0.14)	0.18 (0.12)		
umd			0.23* (0.12)		0.25** (0.11)		0.33*** (0.11)
rmw				-0.46*** (0.17)	-0.51*** (0.14)		
cma				0.17 (0.21)	0.06 (0.19)		
mkt						0.03 (0.07)	0.05 (0.06)
me						0.29* (0.17)	0.21* (0.11)
ia						0.21 (0.22)	0.22 (0.20)
roe						-0.15 (0.16)	-0.44*** (0.14)
Adj. R ²	0.00	0.06	0.11	0.11	0.17	0.06	0.14
Num. obs.	612	612	612	612	612	612	612

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

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, Xue, and Zhang (2015). The arbitrage portfolio with one eigenvector is estimated every month using the steps outlined in Section 3. Newey and West (1987) standard errors appear in parentheses. The sample period is January 1968 to December 2018.

Table 5: Risk-adjusted returns with six eigenvectors

	CAPM	FF3	FF3+UMD	FF5	FF5+UMD	HXZ4	HXZ4+UMD
alpha	2.49*** (0.28)	2.43*** (0.27)	2.00*** (0.23)	2.52*** (0.30)	2.17*** (0.26)	2.26*** (0.32)	2.18*** (0.28)
mktrf	-0.06 (0.07)	-0.10 (0.06)	-0.01 (0.05)	-0.10 (0.07)	-0.03 (0.06)		
smb		0.31 (0.20)	0.32** (0.15)	0.17 (0.15)	0.15 (0.12)		
hml		0.12 (0.13)	0.30* (0.15)	-0.04 (0.17)	0.25* (0.13)		
umd			0.50*** (0.12)		0.52*** (0.10)		0.61*** (0.11)
rmw				-0.47** (0.20)	-0.59*** (0.15)		
cma				0.32 (0.24)	0.10 (0.22)		
mkt						-0.06 (0.07)	-0.01 (0.06)
me						0.31 (0.20)	0.17 (0.13)
ia						0.34 (0.26)	0.36 (0.24)
roe						0.02 (0.18)	-0.53*** (0.15)
Adj. R ²	0.00	0.03	0.18	0.07	0.23	0.04	0.21
Num. obs.	612	612	612	612	612	612	612

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

This table reports alphas (%/month) and factor loadings on the factors by the Fama and French (1993), Carhart (1997), and Fama and French (2015) and the q -factor model (HXZ4) by Hou, Xue, and Zhang (2015). The arbitrage portfolio with six eigenvectors is estimated every month using the steps outlined in Section 3. Newey and West (1987) standard errors appear in parentheses. The sample period is January 1968 to December 2018.

Table 6: Alphas with lagged characteristics (\mathbf{X}) and lagged θ_α

Lag in \mathbf{X}	Lag in θ_α											
	1	2	3	4	5	6	7	8	9	10	11	12
1	2.17	1.45	1.26	1.32	1.35	1.31	1.39	1.03	1.25	1.38	1.30	1.27
2	1.42	1.00	0.73	0.81	0.52	0.60	0.61	0.42	0.49	0.64	0.39	0.49
3	1.08	0.70	0.64	0.66	0.71	0.56	0.55	0.31	0.31	0.38	0.25	0.24
4	1.14	0.80	0.67	0.85	0.67	0.70	0.68	0.42	0.47	0.58	0.50	0.57
5	1.14	0.80	0.68	0.83	0.75	0.73	0.74	0.47	0.49	0.53	0.49	0.50
6	1.14	0.75	0.68	0.72	0.73	0.84	0.80	0.52	0.47	0.50	0.43	0.52
7	1.11	0.76	0.82	0.79	0.80	0.89	0.97	0.62	0.62	0.62	0.45	0.46
8	1.17	0.76	0.75	0.66	0.73	0.76	0.88	0.59	0.56	0.56	0.40	0.47
9	1.00	0.62	0.66	0.54	0.57	0.60	0.73	0.42	0.39	0.42	0.26	0.25
10	1.11	0.72	0.59	0.71	0.61	0.73	0.81	0.54	0.54	0.66	0.45	0.52
11	1.17	0.83	0.73	0.81	0.75	0.80	0.85	0.62	0.66	0.64	0.63	0.63
12	1.23	0.78	0.77	0.81	0.60	0.69	0.77	0.49	0.40	0.53	0.41	0.58

This table reports alpha (%/month) of our arbitrage portfolios against FF6 model for various lags in characteristics and θ_α . We employ up to 12 lags in the characteristics and in θ_α . The 1-1 lag combination corresponds to the use of the most current information in portfolio construction. The sample period is January 1968 to December 2018.

Table 7: Sharpe ratios with lagged characteristics (\mathbf{X}) and lagged θ_α

Lag in \mathbf{X}	Lag in θ_α											
	1	2	3	4	5	6	7	8	9	10	11	12
1	1.60	1.24	1.16	1.21	1.21	1.12	1.17	0.95	1.10	1.13	1.05	1.04
2	0.87	0.72	0.59	0.65	0.44	0.47	0.43	0.36	0.36	0.45	0.29	0.35
3	0.83	0.65	0.54	0.57	0.55	0.49	0.46	0.33	0.29	0.33	0.23	0.24
4	0.87	0.72	0.61	0.68	0.56	0.56	0.52	0.41	0.42	0.45	0.38	0.44
5	0.88	0.74	0.64	0.69	0.59	0.57	0.56	0.43	0.43	0.45	0.41	0.45
6	0.86	0.71	0.62	0.63	0.61	0.63	0.61	0.44	0.42	0.44	0.38	0.45
7	0.86	0.74	0.72	0.69	0.66	0.68	0.67	0.54	0.51	0.52	0.42	0.46
8	0.88	0.69	0.63	0.61	0.62	0.63	0.64	0.49	0.49	0.46	0.36	0.42
9	0.80	0.61	0.58	0.53	0.53	0.55	0.57	0.42	0.35	0.39	0.29	0.30
10	0.87	0.68	0.56	0.60	0.56	0.62	0.62	0.49	0.46	0.52	0.39	0.44
11	0.91	0.75	0.64	0.64	0.60	0.63	0.64	0.51	0.52	0.51	0.47	0.49
12	0.90	0.70	0.63	0.64	0.51	0.57	0.58	0.45	0.40	0.49	0.37	0.44

This table reports annualized Sharpe ratios of our arbitrage portfolios for different lags in characteristics and lags in θ_α . We employ up to 12 lags in the characteristics and in θ_α . The 1-1 lag combination corresponds to the use of the most current information in portfolio construction. The sample period is January 1968 to December 2018.

Arbitrage Portfolios

Internet Appendix

Not for Publication

Soohun Kim¹ Robert A. Korajczyk² Andreas Neuhierl³

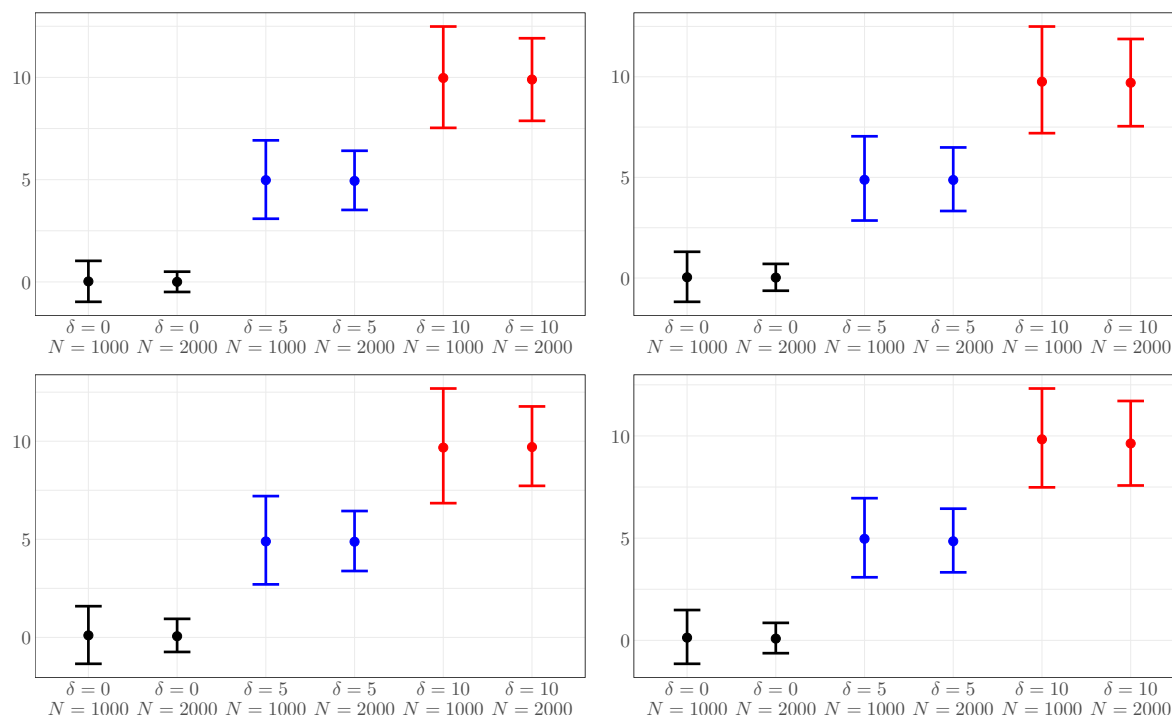
¹College of Business, Korea Advanced Institute of Science and Technology

²Kellogg School of Management, Northwestern University

³Olin Business School, Washington University in St. Louis

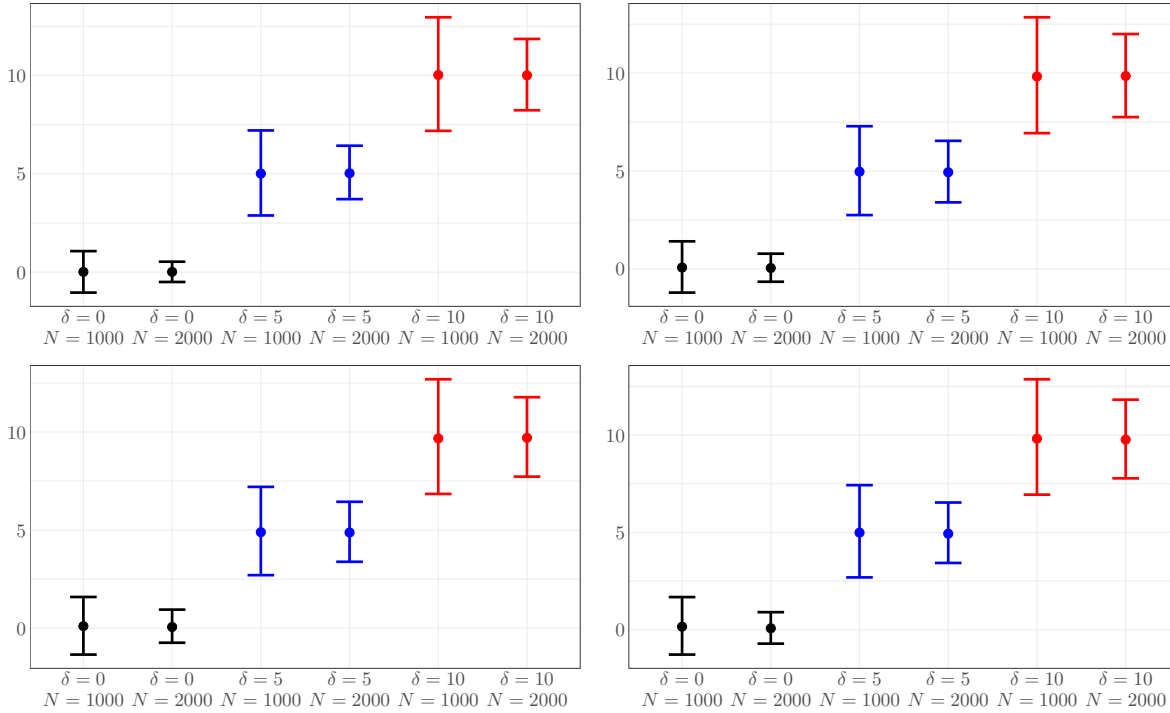
Additional Figures and Tables

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



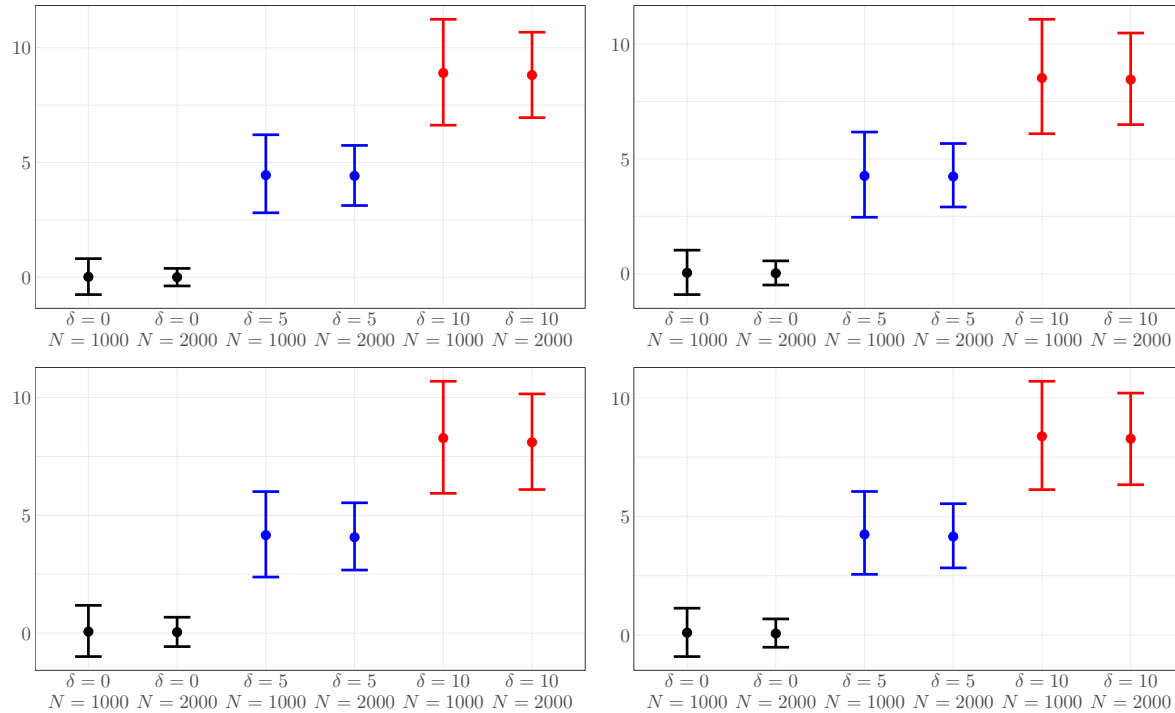
This figure shows the simulation results of the arbitrage portfolio when the return-generating process is calibrated to the CAPM (upper-left panel), the Fama-French three-factor model (upper-right panel), the Fama-French five-factor model (lower-left panel) and the Hou-Xue-Zhang four-factor model (lower-right panel). The arbitrage portfolio \hat{w} is constructed with the returns from $t = 1$ to $t = 12$, and it generates returns for one month out-of-sample. The solid dot represents the mean of the arbitrage portfolio in the out-of-sample period over 10,000 simulations. Error bars represent the 95% confidence interval. In this simulation, we use the correct number of factors in constructing the arbitrage portfolio, that is, $K = 1$ 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 2.2.4.

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



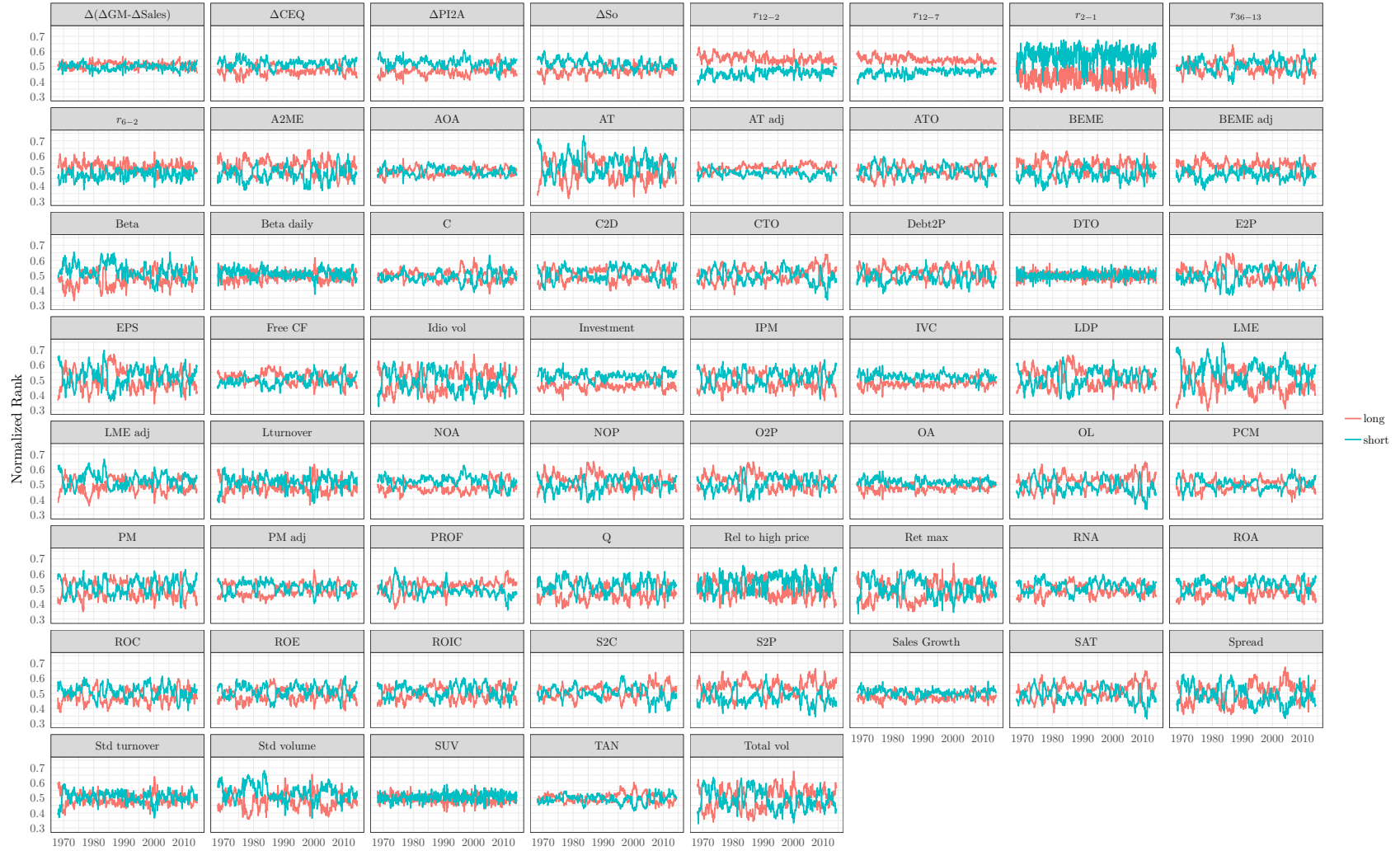
This figure shows the simulation results of the arbitrage portfolio when the return-generating process is calibrated to the CAPM (upper-left panel), the Fama-French three-factor model (upper-right panel), the Fama-French five-factor model (lower-left panel) and the Hou-Xue-Zhang four-factor model (lower-right panel). The arbitrage portfolio $\hat{\mathbf{w}}$ is constructed with the returns from $t = 1$ to $t = 12$, and it generates returns for one month out-of-sample. 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, that is, $K = 1$ 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



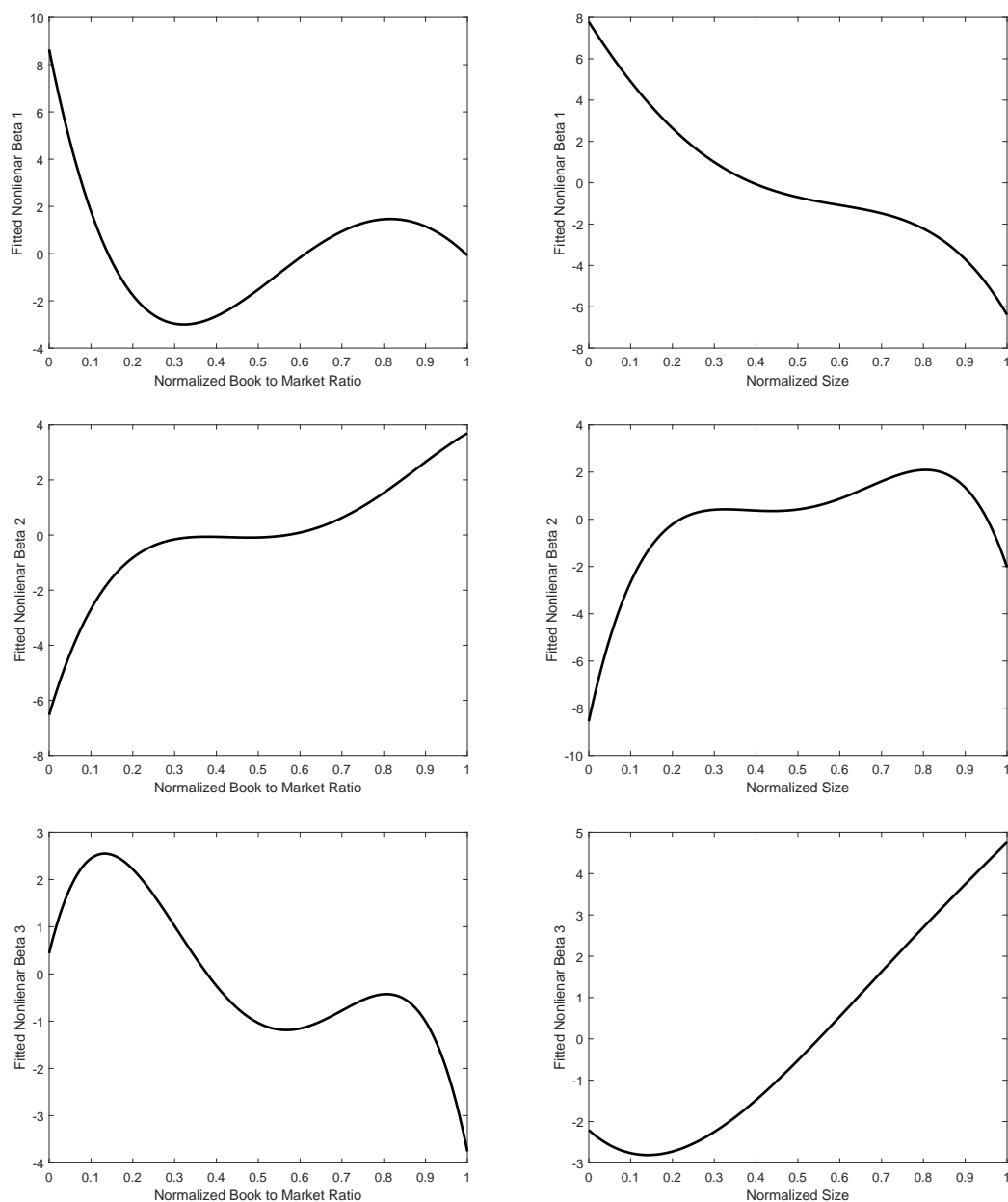
This figure shows the simulation results of the arbitrage portfolio when the return-generating process is calibrated to the CAPM (upper-left panel), the Fama-French three-factor model (upper-right panel), the Fama-French five-factor model (lower-left panel) and the Hou-Xue-Zhang four-factor model (lower-right panel). The arbitrage portfolio $\hat{\mathbf{w}}$ is constructed with the returns from $t = 1$ to $t = 12$, and it generates returns for one month out-of-sample. The solid dot represents the mean of the arbitrage portfolio in the out-of-sample period over 10,000 simulations. Error bars represent the 95% confidence interval. In this simulation, we use the correct number of factors in constructing the arbitrage portfolio, that is, $K = 1$ 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 $\hat{\mathbf{w}}$. The construction is detailed in Section 2.2.4.

Figure A.4: Firm Characteristics of the Arbitrage Portfolio



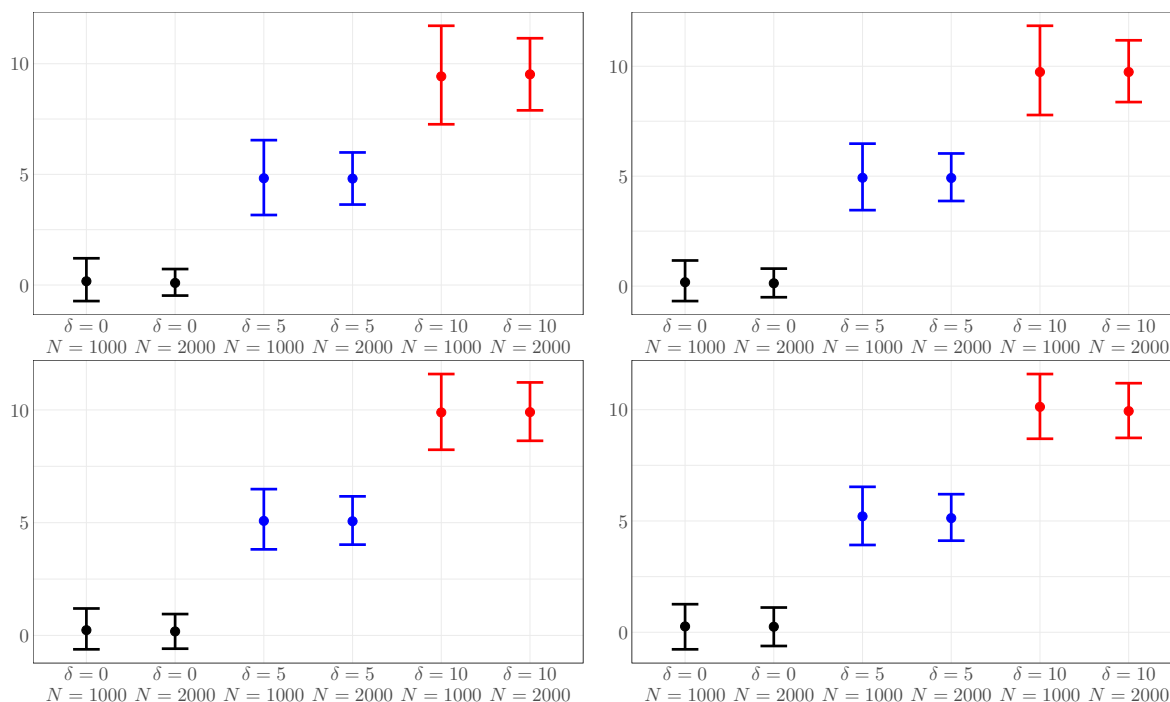
The figure shows the normalized rank of the full set of cross-sectional return characteristics for the long and short legs of the arbitrage portfolio for all the characteristics used in the empirical analysis and described in Table 2. Each month, the characteristics are normalized to be in the unit interval, that is, the normalized characteristics is computed as, $\tilde{c}_{it} = \frac{\text{rank}(c_{it})}{N_t+1}$, 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 cross-sectional comparison over time. The sample period is January 1968 to December 2018.

Figure A.5: Relation between the first nonlinear factor loading function of PC20 and BM&Size characteristics



This (top, middle, bottom) X (left, right) figure represents the relation between first (top), second (middle), and third (bottom) nonlinear beta in PC20 calibration and the normalized book to market ratio (left) or size (right) characteristic.

Figure A.6: Simulated Arbitrage Portfolio Returns in PC10, PC15, PC20, PC25



This figure shows the simulation results of the arbitrage portfolio when the return-generating process is calibrated to the PC10 (upper-left panel), the PC15 (upper-right panel), the PC20 (lower-left panel) and the PC25 (lower-right panel). The arbitrage portfolio $\hat{\mathbf{w}}$ is constructed with the returns from $t = 1$ to $t = 12$, and it generates 1-month out-of-sample results. The solid dot represents the mean of the arbitrage portfolio in the out-of-sample period over 10,000 simulations. Error bars represent the 95% confidence interval. In this simulation, we use $K = 5$ in constructing the arbitrage portfolio. Details of the data-generation are given in Misspecification Analysis.

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

	CAPM	FF3	FF3+UMD	FF5	FF5+UMD	HXZ4	HXZ4+UMD
alpha	2.10*** (0.24)	2.04*** (0.24)	1.59*** (0.22)	2.10*** (0.26)	1.74*** (0.25)	1.85*** (0.28)	1.76*** (0.26)
mktrf	-0.05 (0.08)	-0.09 (0.07)	0.01 (0.06)	-0.08 (0.08)	-0.01 (0.06)		
smb		0.30 (0.20)	0.32** (0.13)	0.18 (0.17)	0.16 (0.11)		
hml		0.11 (0.14)	0.30** (0.14)	-0.08 (0.16)	0.22* (0.12)		
umd			0.52*** (0.13)		0.54*** (0.12)		0.64*** (0.12)
rmw				-0.42** (0.20)	-0.54*** (0.13)		
cma				0.38* (0.23)	0.15 (0.19)		
mkt						-0.04 (0.08)	0.01 (0.06)
me						0.29 (0.21)	0.14 (0.13)
ia						0.38 (0.25)	0.40* (0.22)
roe						0.03 (0.18)	-0.55*** (0.14)
Adj. R ²	0.00	0.03	0.21	0.07	0.26	0.04	0.25
Num. obs.	612	612	612	612	612	612	612

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

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, Xue, and Zhang (2015). The arbitrage portfolio with two eigenvectors is estimated every month using the steps outlined in Section 3. Newey and West (1987) standard errors are given in parentheses. The sample period is January 1968 to December 2018.

Table A.2: Risk-Adjusted Returns with Three Eigenvectors

	CAPM	FF3	FF3+UMD	FF5	FF5+UMD	HXZ4	HXZ4+UMD
alpha	1.99*** (0.23)	1.93*** (0.23)	1.62*** (0.20)	2.05*** (0.25)	1.79*** (0.23)	1.88*** (0.27)	1.81*** (0.24)
mktrf	0.01 (0.06)	-0.03 (0.05)	0.04 (0.05)	-0.04 (0.06)	0.01 (0.05)		
smb		0.33* (0.19)	0.34** (0.14)	0.20 (0.15)	0.19* (0.11)		
hml		0.13 (0.12)	0.26** (0.12)	0.05 (0.14)	0.26** (0.11)		
umd			0.36*** (0.12)		0.39*** (0.11)		0.47*** (0.11)
rmw				-0.46** (0.21)	-0.55*** (0.15)		
cma				0.16 (0.21)	-0.01 (0.18)		
mkt						-0.01 (0.06)	0.03 (0.05)
me						0.31* (0.19)	0.20 (0.13)
ia						0.23 (0.23)	0.25 (0.21)
roe						-0.08 (0.15)	-0.50*** (0.14)
Adj. R ²	-0.00	0.05	0.18	0.10	0.24	0.06	0.22
Num. obs.	612	612	612	612	612	612	612

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

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, Xue, and Zhang (2015). The arbitrage portfolio with three eigenvectors is estimated every month using the steps outlined in Section 3. Newey and West (1987) standard errors are given in parentheses. The sample period is January 1968 to December 2018.

Table A.3: Risk-Adjusted Returns with Four Eigenvectors

	CAPM	FF3	FF3+UMD	FF5	FF5+UMD	HXZ4	HXZ4+UMD
alpha	2.32*** (0.25)	2.25*** (0.25)	1.85*** (0.21)	2.32*** (0.27)	1.99*** (0.22)	2.10*** (0.28)	2.02*** (0.23)
mktrf	-0.04 (0.07)	-0.08 (0.06)	0.01 (0.06)	-0.07 (0.07)	-0.01 (0.06)		
smb		0.31* (0.18)	0.32** (0.13)	0.18 (0.14)	0.17 (0.11)		
hml		0.13 (0.12)	0.30** (0.13)	-0.03 (0.16)	0.24* (0.13)		
umd			0.47*** (0.10)		0.49*** (0.09)		0.58*** (0.10)
rmw				-0.41** (0.19)	-0.53*** (0.14)		
cma				0.34 (0.21)	0.13 (0.18)		
mkt						-0.04 (0.07)	0.00 (0.06)
me						0.31* (0.19)	0.17 (0.11)
ia						0.35 (0.23)	0.37* (0.19)
roe						0.00 (0.15)	-0.52*** (0.13)
Adj. R ²	0.00	0.03	0.20	0.08	0.25	0.05	0.24
Num. obs.	612	612	612	612	612	612	612

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

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, Xue, and Zhang (2015). The arbitrage portfolio with four eigenvectors is estimated every month using the steps outlined in Section 3. Newey and West (1987) standard errors are given in parentheses. The sample period is January 1968 to December 2018.

Table A.4: Risk-Adjusted Returns with Five Eigenvectors

	CAPM	FF3	FF3+UMD	FF5	FF5+UMD	HXZ4	HXZ4+UMD
alpha	2.43*** (0.26)	2.37*** (0.26)	1.93*** (0.22)	2.45*** (0.28)	2.09*** (0.25)	2.20*** (0.31)	2.11*** (0.26)
mktrf	-0.07 (0.07)	-0.12* (0.07)	-0.02 (0.06)	-0.11 (0.07)	-0.04 (0.06)		
smb		0.31 (0.20)	0.32** (0.14)	0.18 (0.16)	0.16 (0.12)		
hml		0.11 (0.13)	0.29* (0.16)	-0.07 (0.17)	0.23* (0.13)		
umd			0.51*** (0.12)		0.53*** (0.11)		0.63*** (0.11)
rmw				-0.44** (0.21)	-0.57*** (0.15)		
cma				0.36 (0.24)	0.13 (0.21)		
mkt						-0.07 (0.08)	-0.02 (0.06)
me						0.32 (0.20)	0.17 (0.12)
ia						0.34 (0.25)	0.36 (0.23)
roe						0.02 (0.18)	-0.55*** (0.14)
Adj. R ²	0.00	0.03	0.20	0.07	0.25	0.04	0.24
Num. obs.	612	612	612	612	612	612	612

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

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, Xue, and Zhang (2015). The arbitrage portfolio with five eigenvectors is estimated every month using the steps outlined in Section 3. Newey and West (1987) standard errors are given in parentheses. The sample period is January 1968 to December 2018.

Table A.5: Risk-Adjusted Returns with Seven Eigenvectors

	CAPM	FF3	FF3+UMD	FF5	FF5+UMD	HXZ4	HXZ4+UMD
alpha	2.55*** (0.27)	2.50*** (0.27)	2.10*** (0.24)	2.59*** (0.30)	2.25*** (0.27)	2.34*** (0.33)	2.25*** (0.29)
mktrf	-0.07 (0.07)	-0.12* (0.06)	-0.03 (0.06)	-0.12 (0.07)	-0.05 (0.06)		
smb		0.34* (0.19)	0.35** (0.14)	0.20 (0.16)	0.18 (0.12)		
hml		0.08 (0.13)	0.25* (0.15)	-0.08 (0.16)	0.19 (0.13)		
umd			0.47*** (0.12)		0.49*** (0.11)		0.59*** (0.11)
rmw				-0.45** (0.20)	-0.57*** (0.14)		
cma				0.33 (0.24)	0.12 (0.22)		
mkt						-0.08 (0.07)	-0.03 (0.06)
me						0.33* (0.20)	0.20 (0.13)
ia						0.31 (0.26)	0.33 (0.24)
roe						0.01 (0.18)	-0.52*** (0.16)
Adj. R ²	0.00	0.03	0.17	0.07	0.21	0.04	0.20
Num. obs.	612	612	612	612	612	612	612

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

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, Xue, and Zhang (2015). The arbitrage portfolio with seven eigenvectors is estimated every month using the steps outlined in Section 3. Newey and West (1987) standard errors are given in parentheses. The sample period is January 1968 to December 2018.

Table A.6: Risk-Adjusted Returns with Eight Eigenvectors

	CAPM	FF3	FF3+UMD	FF5	FF5+UMD	HXZ4	HXZ4+UMD
alpha	2.50*** (0.28)	2.47*** (0.28)	1.98*** (0.23)	2.55*** (0.30)	2.15*** (0.26)	2.26*** (0.33)	2.16*** (0.29)
mktrf	-0.07 (0.08)	-0.12* (0.07)	-0.01 (0.06)	-0.11 (0.07)	-0.03 (0.06)		
smb		0.32* (0.17)	0.33** (0.14)	0.18 (0.16)	0.16 (0.13)		
hml		0.06 (0.13)	0.26* (0.15)	-0.13 (0.16)	0.20 (0.13)		
umd			0.58*** (0.12)		0.60*** (0.10)		0.70*** (0.11)
rmw				-0.47*** (0.18)	-0.61*** (0.14)		
cma				0.38* (0.21)	0.13 (0.21)		
mkt						-0.06 (0.08)	-0.01 (0.06)
me						0.32 (0.20)	0.16 (0.13)
ia						0.30 (0.28)	0.32 (0.24)
roe						0.08 (0.20)	-0.55*** (0.15)
Adj. R ²	0.00	0.02	0.20	0.06	0.24	0.03	0.23
Num. obs.	612	612	612	612	612	612	612

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

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, Xue, and Zhang (2015). The arbitrage portfolio with eight eigenvectors is estimated every month using the steps outlined in Section 3. Newey and West (1987) standard errors are given in parentheses. The sample period is January 1968 to December 2018.

Table A.7: Risk-Adjusted Returns with Nine Eigenvectors

	CAPM	FF3	FF3+UMD	FF5	FF5+UMD	HXZ4	HXZ4+UMD
alpha	2.41*** (0.27)	2.38*** (0.28)	1.96*** (0.24)	2.49*** (0.30)	2.15*** (0.26)	2.26*** (0.34)	2.18*** (0.28)
mktrf	-0.03 (0.06)	-0.08 (0.06)	0.01 (0.05)	-0.09 (0.07)	-0.02 (0.06)		
smb		0.32* (0.17)	0.33** (0.14)	0.18 (0.13)	0.16 (0.10)		
hml		0.03 (0.12)	0.21 (0.14)	-0.10 (0.15)	0.18 (0.11)		
umd			0.49*** (0.12)		0.51*** (0.09)		0.62*** (0.11)
rmw				-0.49*** (0.18)	-0.61*** (0.14)		
cma				0.27 (0.20)	0.05 (0.18)		
mkt						-0.05 (0.07)	0.00 (0.05)
me						0.30 (0.18)	0.15 (0.11)
ia						0.20 (0.24)	0.22 (0.22)
roe						0.00 (0.18)	-0.56*** (0.15)
Adj. R ²	-0.00	0.03	0.18	0.07	0.24	0.03	0.22
Num. obs.	612	612	612	612	612	612	612

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

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, Xue, and Zhang (2015). The arbitrage portfolio with nine eigenvectors is estimated every month using the steps outlined in Section 3. Newey and West (1987) standard errors are given in parentheses. The sample period is January 1968 to December 2018.

Table A.8: Risk-Adjusted Returns with Ten Eigenvectors

	CAPM	FF3	FF3+UMD	FF5	FF5+UMD	HXZ4	HXZ4+UMD
alpha	2.09*** (0.26)	2.09*** (0.28)	1.67*** (0.25)	2.21*** (0.30)	1.86*** (0.28)	1.94*** (0.33)	1.86*** (0.30)
mktrf	-0.02 (0.07)	-0.06 (0.07)	0.04 (0.06)	-0.06 (0.07)	0.01 (0.07)		
smb		0.17 (0.23)	0.18 (0.18)	0.00 (0.19)	-0.01 (0.15)		
hml		-0.01 (0.15)	0.17 (0.14)	-0.17 (0.16)	0.11 (0.12)		
umd			0.49*** (0.13)		0.52*** (0.11)		0.62*** (0.13)
rmw				-0.55*** (0.20)	-0.67*** (0.15)		
cma				0.33 (0.24)	0.11 (0.21)		
mkt						-0.01 (0.08)	0.03 (0.07)
me						0.17 (0.23)	0.02 (0.16)
ia						0.20 (0.26)	0.22 (0.22)
roe						0.03 (0.17)	-0.53*** (0.16)
Adj. R ²	-0.00	0.00	0.14	0.05	0.20	0.00	0.18
Num. obs.	612	612	612	612	612	612	612

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

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, Xue, and Zhang (2015). The arbitrage portfolio with ten eigenvectors is estimated every month using the steps outlined in Section 3. Newey and West (1987) standard errors are given in parentheses. The sample period is January 1968 to December 2018.

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	14.79	16.93	0.87	1.15	7.14	33.32	-24.59	31.17
2	17.44	17.79	0.98	1.02	9.90	33.15	-24.76	41.65
3	21.35	17.11	1.25	0.80	6.83	34.73	-23.04	32.91
4	25.02	18.70	1.34	0.89	6.14	34.88	-24.05	35.59
5	26.22	18.61	1.41	1.00	6.91	32.70	-24.99	39.83
6	27.22	19.87	1.37	1.75	13.05	35.10	-24.06	49.63
7	27.61	19.72	1.40	1.70	12.04	32.10	-23.72	47.76
8	27.53	19.79	1.39	1.37	9.48	39.67	-24.46	44.29
9	28.22	19.80	1.43	1.14	7.00	36.18	-24.89	39.56
10	28.58	19.95	1.43	1.08	6.31	35.26	-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 1 through 10 eigenvectors is estimated every month using the steps outlined in Section 3. The sample period is January 1968 to December 2018. We use 24 months of data to estimate the weights of the arbitrage portfolio and then hold the portfolio for 1 month.

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

# Eigenvectors	CAPM	FF3	FF3+UMD	FF5	FF5+UMD	HXZ4	HXZ4+UMD
1	1.13	1.02	1.11	1.19	1.24	1.33	1.33
2	1.39	1.32	1.16	1.46	1.32	1.42	1.37
3	1.71	1.64	1.48	1.79	1.65	1.71	1.67
4	2.01	1.95	1.74	2.10	1.91	2.01	1.96
5	2.11	2.03	1.81	2.20	2.00	2.11	2.05
6	2.21	2.13	1.88	2.27	2.05	2.19	2.13
7	2.25	2.16	1.91	2.32	2.10	2.23	2.17
8	2.24	2.16	1.92	2.33	2.12	2.24	2.18
9	2.31	2.23	1.99	2.39	2.18	2.30	2.24
10	2.34	2.26	2.02	2.41	2.20	2.32	2.26

This table reports alphas (%/month) against Fama and French (1993), Carhart (1997), Fama and French (2015) and the q -factor model (HXZ4) by Hou, Xue, and Zhang (2015). The arbitrage portfolio is constructed using 1 through 10 eigenvectors. It is estimated every month using the steps outlined in Section 1. The sample period is January 1968 to December 2018. We use 24 months of data to estimate the weights of the arbitrage portfolio and then hold the portfolio for 1 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	11.90	19.73	0.60	1.44	5.73	41.51	-14.03	35.82
2	11.20	17.84	0.63	0.97	4.88	44.63	-19.64	35.33
3	14.91	18.14	0.82	0.94	4.98	36.60	-25.17	31.03
4	16.64	18.78	0.89	0.75	4.41	41.67	-24.92	32.15
5	19.62	18.70	1.05	1.00	4.16	39.66	-16.51	34.54
6	21.03	18.75	1.12	0.88	3.58	36.10	-16.73	33.25
7	21.92	18.73	1.17	0.90	3.94	35.65	-17.05	36.13
8	21.99	19.04	1.15	0.85	3.71	36.53	-17.57	34.89
9	22.49	19.19	1.17	0.83	3.99	39.03	-17.98	36.35
10	23.01	19.17	1.20	0.87	3.90	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 1 through 10 eigenvectors is estimated every month using the steps outlined in Section 3. The sample period is January 1968 to December 2018. We use 36 months of data to estimate the weights of the arbitrage portfolio and then hold the portfolio for 1 month.

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

# Eigenvectors	CAPM	FF3	FF3+UMD	FF5	FF5+UMD	HXZ4	HXZ4+UMD
1	0.82	0.65	1.04	0.91	1.20	1.32	1.36
2	0.80	0.73	0.91	0.95	1.08	1.10	1.11
3	1.08	1.01	1.15	1.27	1.36	1.38	1.39
4	1.23	1.16	1.22	1.41	1.44	1.50	1.50
5	1.47	1.37	1.44	1.63	1.66	1.71	1.71
6	1.59	1.50	1.55	1.76	1.77	1.84	1.83
7	1.67	1.60	1.62	1.85	1.84	1.92	1.91
8	1.69	1.62	1.63	1.85	1.84	1.91	1.90
9	1.74	1.67	1.68	1.90	1.89	1.97	1.95
10	1.78	1.70	1.73	1.94	1.94	2.01	2.00

This table reports alphas (%/month) against Fama and French (1993), Carhart (1997), Fama and French (2015) and the q -factor model (HXZ4) by Hou, Xue, and Zhang (2015). The arbitrage portfolio is constructed using 1 through 10 eigenvectors. It is estimated every month using the steps outlined in Section 1. The sample period is January 1968 to December 2018. We use 36 months of data to estimate the weights of the arbitrage portfolio and then hold the portfolio for 1 month.

Table A.13: Portfolio Performance Statistics for Fourth Order Legendre Polynomials

# Eigenvectors	Mean (%)	Standard Deviation (%)	Sharpe Ratio	Skewness	Kurtosis	Maximum Drawdown	Worst Month (%)	Best Month (%)
1	27.63	17.42	1.59	0.47	6.90	26.78	-24.27	30.39
2	32.60	20.29	1.61	0.17	4.33	29.80	-26.11	29.04
3	30.37	16.33	1.86	0.40	6.11	25.94	-25.94	31.07
4	33.05	18.60	1.78	-0.04	4.06	26.37	-26.37	28.05
5	33.43	19.01	1.76	-0.05	3.76	28.48	-26.77	28.56
6	34.89	19.94	1.75	0.18	5.23	31.73	-31.73	34.20
7	35.83	19.99	1.79	0.14	5.55	32.29	-32.29	31.64
8	34.59	21.90	1.58	-0.81	11.52	47.21	-47.21	33.55
9	32.86	21.12	1.56	-0.78	10.27	41.64	-41.64	33.93
10	29.49	20.34	1.45	-0.12	11.36	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 1 through 10 eigenvectors is estimated every month using the steps outlined in Section 3. The sample period is January 1968 to December 2018.

Table A.14: Alphas for Fourth Order Legendre Polynomials

# Eigenvectors	CAPM	FF3	FF3+UMD	FF5	FF5+UMD	HXZ4	HXZ4+UMD
1	2.31	2.28	1.89	2.36	2.05	2.11	2.04
2	2.76	2.75	2.15	2.78	2.30	2.44	2.32
3	2.54	2.51	2.07	2.57	2.21	2.33	2.24
4	2.78	2.75	2.20	2.76	2.31	2.45	2.34
5	2.82	2.78	2.20	2.80	2.32	2.44	2.33
6	2.94	2.93	2.32	2.96	2.46	2.60	2.49
7	3.01	3.02	2.41	3.01	2.52	2.64	2.53
8	2.91	2.93	2.27	2.92	2.38	2.49	2.37
9	2.76	2.76	2.13	2.74	2.23	2.32	2.21
10	2.49	2.48	1.91	2.44	1.98	2.02	1.92

This table reports alphas (%/month) against Fama and French (1993), Carhart (1997), Fama and French (2015) and the q -factor model (HXZ4) by Hou, Xue, and Zhang (2015). The arbitrage portfolio is constructed using 1 through 10 eigenvectors. It is estimated every month using the steps outlined in Section 1. All estimates are significant at the 1% level with Newey and West (1987) standard errors. The sample period is January 1968 to December 2018.

Table A.15: Portfolio Performance Statistics without Micro-Cap Stocks

# Eigenvectors	Mean (%)	Standard Deviation (%)	Sharpe Ratio	Skewness	Kurtosis	Maximum Drawdown	Worst Month (%)	Best Month (%)
1	15.76	15.19	1.04	0.64	7.18	22.15	-21.62	31.03
2	19.92	18.42	1.08	0.27	5.18	25.33	-25.33	31.05
3	19.25	14.91	1.29	0.82	7.07	22.52	-22.52	31.49
4	22.00	17.45	1.26	0.68	5.49	23.80	-21.63	34.55
5	22.52	18.70	1.20	0.53	5.51	30.49	-21.93	36.71
6	22.01	19.31	1.14	0.38	5.49	27.17	-25.96	36.41
7	23.64	20.15	1.17	0.76	6.47	27.56	-27.56	41.44
8	23.25	20.61	1.13	0.79	6.75	28.05	-28.05	43.37
9	19.26	20.20	0.95	-0.51	12.34	48.41	-37.55	41.01
10	16.69	20.17	0.83	0.33	8.52	58.59	-28.68	42.82

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 1 through 10 eigenvectors is estimated every month using the steps outlined in Section 3. The sample period is January 1968 to December 2018. We exclude micro-cap stocks, smaller than 10% quantile of the market capitalization among NYSE traded stocks.

Table A.16: Alphas without Micro-Cap Stocks

# Eigenvectors	CAPM	FF3	FF3+UMD	FF5	FF5+UMD	HXZ4	HXZ4+UMD
1	1.27	1.21	0.95	1.34	1.12	1.16	1.10
2	1.65	1.62	1.09	1.74	1.30	1.38	1.28
3	1.58	1.54	1.17	1.65	1.34	1.40	1.33
4	1.84	1.82	1.33	1.93	1.52	1.61	1.51
5	1.90	1.89	1.35	1.99	1.54	1.62	1.52
6	1.86	1.85	1.26	1.94	1.45	1.56	1.44
7	2.00	2.01	1.43	2.15	1.66	1.74	1.63
8	1.97	1.99	1.42	2.09	1.62	1.72	1.61
9	1.62	1.66	1.07	1.74	1.25	1.35	1.24
10	1.37	1.39	0.80	1.43	0.95	1.03	0.92

This table reports alphas (%/month) against Fama and French (1993), Carhart (1997), Fama and French (2015) and the q -factor model (HXZ4) by Hou, Xue, and Zhang (2015). The arbitrage portfolio is constructed using 1 through 10 eigenvectors. It is estimated every month using the steps outlined in Section 1. All estimates are significant at the 1% level with Newey and West (1987) standard errors. The sample period is January 1968 to December 2018. We exclude micro-cap stocks, smaller than 10% quantile of the market capitalization among NYSE traded stocks.

Table A.17: Risk-Adjusted Returns with respect to Alternative Factor Models

	SY MP	FF3+UMD+LIQ	FF3+UMD+BAB	FF3+UMD+STREV
alpha	2.08*** (0.28)	2.00*** (0.23)	2.08*** (0.26)	2.05*** (0.24)
mktrf	0.02 (0.08)	-0.01 (0.06)	-0.00 (0.05)	0.01 (0.05)
smb	0.41** (0.20)	0.32** (0.14)	0.32** (0.14)	0.33** (0.15)
mgmt	0.31 (0.21)			
perf	0.25** (0.12)			
hml		0.30** (0.14)	0.37*** (0.12)	0.30* (0.16)
umd		0.50*** (0.12)	0.53*** (0.11)	0.48*** (0.11)
liqf		0.02 (0.06)		
bab			-0.14 (0.11)	
strev				-0.11 (0.10)
Adj. R ²	0.06	0.18	0.18	0.18
Num. obs.	588	612	612	612

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

This table reports alphas (%/month) and factor loadings on the Stambaugh and Yuan (2016), the Fama and French (1993) augmented with the Carhart (1997) momentum factor and Pástor and Stambaugh (2003) liquidity factor, the Fama and French (1993) augmented with the Carhart (1997) momentum and Frazzini and Pedersen (2014) betting-against-beta factor and the Fama and French (1993) augmented with the Carhart (1997) momentum factor and a short-term reversal factor. The arbitrage portfolio with six eigenvectors is estimated every month using the steps outlined in Section 3. Newey and West (1987) standard errors are given in parentheses. The sample period is January 1968 to December 2016 (for the first column of SY MP) and January 1968 to December 2018 (for the other columns).

Table A.18: “Arbitrage Portfolio” without Orthogonality Condition - Regression on Standard Risk Factors

	CAPM	FF3	FF3+UMD	FF5	FF5+UMD	HXZ4	HXZ4+UMD
alpha	2.12*** (0.29)	2.06*** (0.29)	1.30*** (0.25)	2.06*** (0.33)	1.44*** (0.26)	1.78*** (0.36)	1.63*** (0.27)
mktrf	-0.19* (0.10)	-0.23*** (0.09)	-0.06 (0.06)	-0.20** (0.10)	-0.08 (0.07)		
smb		0.28 (0.21)	0.30** (0.13)	0.18 (0.20)	0.15 (0.15)		
hml		0.11 (0.16)	0.44*** (0.15)	-0.13 (0.20)	0.38*** (0.14)		
umd			0.90*** (0.11)		0.92*** (0.11)		1.07*** (0.11)
rmw				-0.31 (0.22)	-0.52*** (0.14)		
cma				0.52* (0.28)	0.12 (0.21)		
mkt						-0.16 (0.10)	-0.08 (0.07)
me						0.30 (0.22)	0.05 (0.14)
ia						0.40 (0.28)	0.43** (0.20)
roe						0.17 (0.20)	-0.80*** (0.14)
Adj. R ²	0.01	0.03	0.32	0.05	0.35	0.03	0.36
Num. obs.	612	612	612	612	612	612	612

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

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, Xue, and Zhang (2015). Newey and West (1987) standard errors are given in parentheses. The sample period is January 1968 to December 2018.

Table A.19: Regression on Statistical Factors

	APC				RP-PC				IPCA			
	1 PC	5 PCs	10 PCs	20 PCs	1 RP-PC	5 RP-PCs	10 RP-PCs	20 RP-PCs	1 KPS-PC	5 KPS-PCs	10 KPS-PCs	20 KPS-PCs
Panel A: Arbitrage Portfolio												
alpha	2.45*** (0.29)	2.08*** (0.25)	1.28*** (0.26)	1.12*** (0.25)	2.45*** (0.29)	1.15*** (0.30)	1.06*** (0.27)	1.01*** (0.26)	2.39*** (0.25)	2.31*** (0.26)	2.15*** (0.42)	1.57*** (0.33)
Adj. R ²	-0.00	0.29	0.34	0.37	-0.00	0.31	0.34	0.36	0.02	0.13	0.14	0.23
Num. obs.	612	612	612	612	612	612	612	612	612	612	612	612
Panel B: Lewellen												
alpha	3.01*** (0.31)	2.25*** (0.20)	0.50* (0.27)	0.58*** (0.22)	3.01*** (0.31)	0.38 (0.28)	0.14 (0.28)	0.49* (0.25)	3.00*** (0.29)	3.06*** (0.28)	2.36*** (0.60)	0.85** (0.41)
Adj. R ²	0.01	0.61	0.72	0.75	0.01	0.68	0.71	0.75	0.03	0.19	0.25	0.46
Num. obs.	612	612	612	612	612	612	612	612	612	612	612	612
Panel C: Stambaugh/Yuan												
alpha	1.44*** (0.20)	0.90*** (0.08)	0.09 (0.07)	-0.05 (0.06)	1.44*** (0.20)	-0.29*** (0.09)	-0.09 (0.08)	-0.12* (0.07)	1.46*** (0.18)	0.95*** (0.18)	0.62** (0.25)	-0.17 (0.16)
Adj. R ²	0.12	0.92	0.96	0.97	0.12	0.94	0.96	0.97	0.31	0.47	0.61	0.85
Num. obs.	612	612	612	612	612	612	612	612	612	612	612	612

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

This table reports alphas relative to 1, 5, 10, and 20 principal components (Connor and Korajczyk (1986), labeled “j PCs”); risk premium principal components (Lettau and Pelger (2020b,a), labeled “j RP-PCs”) of sorted portfolios; and the IPCA method (Kelly, Pruitt, and Su (2019), labeled “j KPS-PCs”). For panels B and C, in each month we sort stocks into 10 portfolios based on their predicted returns using Lewellen (2015)’s rolling Fama-MacBeth approach (panel B) and the Stambaugh and Yuan (2016)’s rank approach (panel C) for our 61 characteristics. We then construct a “high-low” portfolio and report the results for this portfolio. The sample period is January 1968 to December 2018.

Table A.20: Lewellen Portfolio - Regression on Standard Risk Factors

	CAPM	FF3	FF3+UMD	FF5	FF5+UMD	HXZ4	HXZ4+UMD
alpha	3.10*** (0.33)	3.07*** (0.33)	2.25*** (0.25)	3.08*** (0.35)	2.41*** (0.25)	2.79*** (0.46)	2.63*** (0.29)
mktrf	-0.01 (0.11)	-0.15 (0.10)	0.03 (0.07)	-0.11 (0.09)	0.02 (0.07)		
smb		0.73** (0.31)	0.75*** (0.21)	0.59*** (0.18)	0.56*** (0.14)		
hml		-0.00 (0.25)	0.35* (0.19)	-0.32 (0.27)	0.23 (0.17)		
umd			0.97*** (0.09)		0.99*** (0.08)		1.17*** (0.11)
rmw				-0.43 (0.35)	-0.66*** (0.23)		
cma				0.66* (0.35)	0.24 (0.21)		
mkt						-0.09 (0.11)	-0.00 (0.07)
me						0.75*** (0.27)	0.47*** (0.13)
ia						0.18 (0.38)	0.22 (0.19)
roe						0.17 (0.30)	-0.89*** (0.21)
Adj. R ²	-0.00	0.10	0.45	0.13	0.49	0.09	0.50
Num. obs.	612	612	612	612	612	612	612

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

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, Xue, and Zhang (2015). Each month we sort stocks into 10 portfolios based on their predicted returns using Lewellen (2015)'s rolling Fama-MacBeth approach. We then construct a "high-low" portfolio and report the results for this portfolio. Newey and West (1987) standard errors are given in parentheses. The sample period is January 1968 to December 2018.

Table A.21: Lewellen Portfolio - Regression on Alternative Risk Factors

	SY MP	FF3+UMD+LIQ	FF3+UMD+BAB	FF3+UMD+STREV
alpha	2.23*** (0.31)	2.23*** (0.25)	2.33*** (0.27)	2.21*** (0.27)
mktrf	0.13 (0.08)	0.03 (0.07)	0.04 (0.07)	0.02 (0.06)
smb	0.93*** (0.28)	0.75*** (0.21)	0.75*** (0.19)	0.74*** (0.21)
mgmt	0.51** (0.24)			
perf	0.62*** (0.13)			
hml		0.34* (0.18)	0.42*** (0.13)	0.34* (0.17)
umd		0.97*** (0.09)	1.00*** (0.11)	0.99*** (0.09)
liqf		0.06 (0.08)		
bab			-0.15 (0.15)	
strev				0.09 (0.13)
Adj. R ²	0.23	0.45	0.45	0.45
Num. obs.	588	612	612	612

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

This table reports alphas (%/month) and factor loadings on the factors by the several alternative factor models. Each month we sort stocks into 10 portfolios based on their predicted returns using Lewellen (2015)'s rolling Fama-MacBeth approach. We then construct a "high-low" portfolio and report the results for this portfolio. Newey and West (1987) standard errors are given in parentheses. The sample period is January 1968 to December 2016 (for the first column of SY MP) and January 1968 to December 2018 (for the other columns).

Table A.22: Stambaugh/Yuan Portfolio - Regression on Standard Risk Factors

	CAPM	FF3	FF3+UMD	FF5	FF5+UMD	HXZ4	HXZ4+UMD
alpha	1.73*** (0.24)	1.42*** (0.17)	1.41*** (0.20)	1.48*** (0.19)	1.47*** (0.21)	1.66*** (0.25)	1.64*** (0.25)
mktrf	-0.04 (0.07)	-0.14*** (0.05)	-0.14** (0.05)	-0.12** (0.05)	-0.12** (0.06)		
smb		1.05*** (0.09)	1.05*** (0.09)	0.92*** (0.10)	0.92*** (0.10)		
hml		0.63*** (0.13)	0.63*** (0.15)	0.42*** (0.14)	0.43*** (0.14)		
umd			0.01 (0.10)		0.01 (0.09)		0.16** (0.06)
rmw				-0.43*** (0.14)	-0.43*** (0.12)		
cma				0.43* (0.23)	0.42** (0.21)		
mkt						-0.15*** (0.05)	-0.14** (0.06)
me						0.83*** (0.10)	0.79*** (0.10)
ia						0.71*** (0.22)	0.72*** (0.23)
roe						-0.69*** (0.17)	-0.84*** (0.13)
Adj. R ²	-0.00	0.38	0.38	0.42	0.42	0.43	0.45
Num. obs.	612	612	612	612	612	612	612

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

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, Xue, and Zhang (2015). Each month we sort stocks into 10 portfolios based on the Stambaugh and Yuan (2016)'s aggregated rank approach for our 61 characteristics. We then construct a "high-low" portfolio and report the results for this portfolio. Newey and West (1987) standard errors are given in parentheses. The sample period is January 1968 to December 2018.

Table A.23: Stambaugh/Yuan Portfolio - Regression on Alternative Risk Factors

	SY MP	FF3+UMD+LIQ	FF3+UMD+LIQ+BAB	FF3+UMD+STREV
alpha	1.37*** (0.24)	1.40*** (0.20)	1.38*** (0.23)	1.33*** (0.21)
mktrf	-0.09 (0.08)	-0.14*** (0.05)	-0.14*** (0.05)	-0.17*** (0.06)
smb	1.13*** (0.10)	1.05*** (0.10)	1.05*** (0.10)	1.03*** (0.10)
mgmt	0.63*** (0.17)			
perf	-0.18** (0.08)			
hml		0.63*** (0.16)	0.61*** (0.15)	0.62*** (0.16)
umd		0.01 (0.10)	0.00 (0.10)	0.04 (0.09)
liqf		0.04 (0.08)		
bab			0.04 (0.12)	
strev				0.18 (0.11)
Adj. R ²	0.38	0.38	0.38	0.39
Num. obs.	588	612	612	612

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

This table reports alphas (%/month) and factor loadings on the factors by the several alternative factor models. Each month we sort stocks into 10 portfolios based on the Stambaugh and Yuan (2016)'s aggregated rank approach for our 61 characteristics. We then construct a "high-low" portfolio and report the results for this portfolio. Newey and West (1987) standard errors are given in parentheses. The sample period is January 1968 to December 2016 (for the first column of SY MP) and January 1968 to December 2018 (for the other columns).

Misspecification Analysis

Overview Section 2 analyzes the properties of our estimator in a simulated economy with a relatively small number of systematic factors. This section further examine our estimator in a simulated economy with a substantially larger number of factors of 10, 15, 20, and 25, many more than that our estimator handles over a short horizon data. For example, we set the individual stock returns to be generated by 20 factors and we investigate the behavior of our estimator by controlling only five systematic factors. Then, as a result, 15 factors remain uncontrolled. Furthermore, we model a highly nonlinear relationship between factor loadings and characteristics. The overarching goal of the analysis is to understand whether or not, and possibly in what magnitude, the uncontrolled component of nonlinear beta risks leaks to alpha.

Calibration We use the same data of 2,458 individual stock returns over the 36-month period from January 2011 to December 2013 as in Section 2. For the factors, we extract 10 (PC10), 15 (PC15), 20 (PC20), and 25 (PC25) statistical factors from the realized individual stocks' excess returns to address the case that individual stock returns are generated by a large number of factors. Given the factors, we calibrate α_i , β_i , and the variance of residual returns, $\sigma_{i,\varepsilon}^2 = \mathbb{E}[\varepsilon_{i,t}^2]$, of individual stocks for each of PC10, PC15, PC20, and PC25 using the identical calibration procedure in Section 2. Furthermore, we ensure the nonlinear relation between characteristics and factor loading function by fitting the k th column of $\mathbf{G}_\beta(\mathbf{X})$ as $\sum_{l=1}^L \left(\sum_{p=1}^P b_{l,p} \phi_p(\mathbf{x}_l) \right)$, where ϕ_p is the p th order Chebyshev polynomial, $b_{l,p}$ are the solutions of $(b_{l,p}) = \arg \min_{b_{l,p}} \left\| \mathbf{B} \boldsymbol{\iota}_k - \sum_{l=1}^L \left(\sum_{p=1}^P b_{l,p} \phi_p(\mathbf{x}_l) \right) \right\|^2$, and $\boldsymbol{\iota}_k$ is a $(K \times 1)$ vector of zeros except the k th element of one. Figure A.5 plots the relations between the first nonlinear beta of PC20 and book-to-market (left) and size (right) characteristics. We observe that the some of the fitted relations are highly nonlinear.

Return-Generating Process In each repetition, we simulate returns from

$$\mathbf{R} = \boldsymbol{\alpha} \mathbf{1}'_T \sqrt{\delta} + \mathbf{G}_\beta(\mathbf{X}) \mathbf{F}' + \mathbf{E}, \quad (\text{A.1})$$

where α and $\mathbf{G}_\beta(\mathbf{X})$ are calibrated as in the above paragraph,⁴ \mathbf{F} are resampled from the realized statistical factors, and \mathbf{E} are drawn from a normal distribution with the calibrated $\sigma_{i,\varepsilon}^2$ parameters as in the above paragraph. Because our main question is whether or not the uncontrolled component of nonlinear beta risks leaks to alpha, our focus is on the case of $\delta = 0$. But, for the ease of comparison, we consider $\delta = 0, 5, \text{ and } 10$ as in Section 2.

Estimated Arbitrage Portfolio Returns As in our empirical exercise, we estimate the arbitrage portfolio using the simulated returns from (A.1) over 12 periods and hold it in the following 1 period. In the procedure of arbitrage portfolio construction, we set $K = 5$. Given that we simulate the individual stock returns using a model with 10, 15, 20, or 25 factors, our estimator is operating under severe handicaps. Figure A.6 shows the results for using the PC10 (upper-left panel), the PC15 (upper-right panel), the PC20 (lower-left panel) and the PC25 (lower-right panel). The results are not substantially different from those shown in Figures 1–4, where we use a smaller number of systematic factors. Regarding the concern on the leakage of beta to alpha, we find that the average monthly return of arbitrage portfolio is only 0.0009%, 0.0013%, 0.0018%, and 0.0025% for PC10, PC15, PC20, and PC25 when $\delta = 0\%$ and $N = 2,000$. Given that our arbitrage portfolio returns achieve well above 1%/month empirically, the results in Figure A.6 imply that our arbitrage portfolio return is highly unlikely to be driven by the leakage from uncontrolled beta.

⁴We also rescale α as in Section 2 such that the arbitrage portfolio return is normalized to δ .