JOB MARKET TOPIC

A SEMIPARAMETRIC CHARACTERISTICS-BASED FACTOR MODEL

February 17, 2021

Shaoran Li

University of Cambridge

February 17, 2021

- *Part1*

  Dynamic Peer Groups of Arbitrage Characteristics (with Shuyi Ge and Oliver Linton)

  Revised and Resubmitted, Journal of Business & Economic Statistics

  KEYWORDS: Semiparametric; Characteristics-based; Asset pricing; Power-enhanced test;

  - We propose an asset pricing factor model constructed with semi-parametric characteristics-based mispricing and factor loading functions. We approximate the unknown functions by B-splines sieve where the number of B-splines coefficients is diverging. We estimate this model and test the existence of the mispricing function by a power enhanced hypothesis test. The enhanced test solves the low power problem caused by diverging B-spline coefficients, with the strengthened power approaches to one asymptotically. We also investigate the structure of mispricing components through Hierarchical K-means Clusterings. We apply our methodology to CRSP (Center for Research in Security Prices) and FRED (Federal Reserve Economic Data) data for the US stock market with one-year rolling windows during 1967-2017. This empirical study shows the presence of mispricing functions in certain time blocks. We also find that distinct clusters of the same characteristics lead to similar arbitrage returns, forming a "peer group" of arbitrage characteristics.

- *Part2*

  A Dynamic Semiparametric Characteristics-based Model for Optimal Portfolio Selection

  (with Gregory Connor and Oliver Linton)

  Under review

  KEYWORDS: Portfolio Management; Single index; GMM;

  – This paper develops a two-step semiparametric methodology for portfolio weight selection for characteristics-based factor-tilt and factor-timing investment strategies. We build upon the expected utility maximization framework of Brandt (1999) and Aït-sahalia and Brandt (2001). We assume that assets returns obey a characteristics-based factor model with time-varying factor risk premia as in Li and Linton (2020). We prove under our return-generating assumptions that in a market with a large number of assets, an approximately optimal portfolio can be established using a two-step procedure. The first step finds optimal factor-mimicking sub-portfolios using a quadratic objective function over linear combinations of characteristics-based factor loadings. The second step dynamically combines these factor-mimicking sub-portfolios based on a time-varying signal, using the investor's expected utility as the objective function. We develop and implement a two-stage semiparametric estimator. We apply it to CRSP (Center for Research in Security Prices) and FRED (Federal Reserve Economic Data) data and find excellent in-sample and out-sample performance consistent with investors' risk aversion levels.

Dynamic Peer Groups of Arbitrage Characteristics*

Shuyi Ge †1, Shaoran Li ‡2, and Oliver Linton §3

[1 2 3] *Faculty of Economics, University of Cambridge*

February 16, 2021

Abstract

We propose an asset pricing factor model constructed with semi-parametric characteristics-based mispricing and factor loading functions. We approximate the unknown functions by B-splines sieve where the number of B-splines coefficients is diverging. We estimate this model and test the existence of the mispricing function by a power enhanced hypothesis test. The enhanced test solves the low power problem caused by diverging B-spline coefficients, with the strengthened power approaches to one asymptotically. We also investigate the structure of mispricing components through Hierarchical K-means Clusterings. We apply our methodology to CRSP (Center for Research in Security Prices) and FRED (Federal Reserve Economic Data) data for the US stock market with one-year rolling windows during 1967-2017. This empirical study shows the presence of mispricing functions in certain time blocks. We also find that distinct clusters of the same characteristics lead to similar arbitrage returns, forming a "peer group" of arbitrage characteristics.

Keywords: Semiparametric; Characteristics-based; Peer Groups; Power-enhanced test
JEL Classification: C14; G11; G12

*The authors would like to thank Chaohua Dong, Ondrej Tobek, Yuan Liao, Anders Kock, Joachim Freyberger, Alexei Onatski and Weiguang Liu for their useful comments. The editor and two referees deserve considerable credit for their help in improving this manuscript.

†Electronic address: sg751@cam.ac.uk.
‡Corresponding author. Electronic address: sl736@cam.ac.uk.
§Electronic address: obl20@cam.ac.uk

1 Introduction

Stock returns have both common and firm-specific components. Ross (1976) proposed Arbitrage Pricing Theory (APT) to summarize that expected returns on financial assets can be modeled as a linear combination of various factors. In such a model, each asset has a sensitivity beta to the risk factor, The APT model explains the excess returns from both cross-sectional and time-series directions. Fama and French (1993) and Fama and French (2015) approximated those factors by the returns on portfolios sorted by different characteristics, and they developed three-factor and five-factor models. After extracting the common movement parts, they treated the intercept as the mispricing *alpha,* which is asset-specific and cannot be explained by those risk factors. Many papers use a similar method to present other factor models, such as the four-factor model of Carhart (1997), the q-factor model of Hou et al. (2015), and the factor zoo by Feng et al. (2017) among others. All of above papers studied observed factors and did not assigned characteristics-based information to either alpha or beta.

Security-specific characteristics, such as capitalization and book to market ratio, are usually documented to explain asset-specific excess returns. Freyberger et al. (2017) analyzed the non-linear effects of 62 characteristics through pooling regressions. This study concluded that 13 of these characterisitcs have explanatory power on stock excess returns after selecting by adaptive group Lasso. Characterisics-based information are exploited to develop arbitrage portfolios by directly parameterizing the portfolio weights as a linear function of characteristics, as in Hjalmarsson and Manchev (2012) and Kim et al. (2019). Empirically, they showed that their portfolio outperformed other baseline competitors.

This paper's contributions are fourfold. Firstly, we build up a more flexible semi-parametric characteristics-based asset pricing factor model with a focus on mispricing component. Secondly, we extend previous estimation and testing methods, which can fit the current framework better. Especially, we extend the power enhanced test of Fan et al. (2015) in a group manner to strengthen the conventional Wald test for mispricing functions. This test can also select the characteristics that contribute to arbitrage portfolios simultaneously. Thirdly, we construct a two-layer clusterings structure of mispricing components. Finally, our methods are applied to fifty years of monthly US stock data. We detect distinct clusters of the same characteristics resulting in similar arbitrage returns, forming a "peer group" of arbitrage characteristics. This finding supplements existing portfolio management techniques by implying that the development of arbitrage portfolio through the asset weights determined by the linear mispricing function is improvable.

This class of models has a basic regression specification in Equation 1. Consider the panel regression model

$$y_{it} = \alpha_i + \sum_{j=1}^{J} \beta_{ji} f_{jt} + \epsilon_{it}, \tag{1}$$

where y_{it} is the excess return of security i at time t; f_{jt} is the j^{th} risk factor's return at time t; β_{ji} denotes the j^{th} factor loading of asset i; α_i represents the intercept (mispricing) of asset i; and ϵ_{it} is the mean zero idiosyncratic shock. In terms of factor loadings β_{ji}, Connor and Linton (2007) and Connor et al. (2012) studied a characteristic-beta model, which bridges the beta-coefficients and firm-specific characteristics by specifying each beta as an unknown function of one characteristic. In their model, beta functions and unobservable factors are estimated by the backfitting iteration. They concluded that those characteristic-beta functions are significant and non-linear. Their model can be summarized by

$$y_{it} = \sum_{j=1}^{J} g_j(X_{ji}) f_{jt} + \epsilon_{it}, \tag{2}$$

where X_{ji} is the j^{th} observable characteristic of firm i.

They restricted their beta function to be univariate and did not consider the part of factor loading function that cannot be explained by characteristics. To overcome this limitation, Fan et al. (2016) allowed β_{ji} in Equation 1 to have a component explained by observable characteristics as well as an unexplained or stochastic part, written as $\beta_{ji} = g_j(X_i) + u_{ji}$, where u_{ji} is mean independent of X_{ji}. They proposed the Projected Principal Component Analysis (PPCA), which projects stocks excess returns onto the space spanned by firm-specific characteristics and then applies Principal Component Analysis (PCA) to the projected returns to find the unobservable factors. This method has attractive properties even under large n and small T setting. However, they did not study the mispricing part (alpha), which is crucial to both asset pricing theories and portfolio management.

In this paper, we work on a semi-parametric characteristics-based alpha and beta model, which utilizes a set of security-specific characteristics that are similar to Freyberger et al. (2017). We use unknown multivariate characteristic functions to approximate both α_i and β_{ji} in Equation 1. Specifically, we assume α_i and β_{ji} are functions of a large set of asset-specific characteristics as $\alpha_i = h(\mathbf{X}_i) + \gamma_i$ and $\beta_{ji} = g_j(\mathbf{X}_i) + \lambda_{ij}$[1]. We then estimate $h(\mathbf{X}_i)$, $g_j(\mathbf{X}_i)$ and unobservable risk factors f_{jt}. In addition, we design a power enhanced test and Hierarchical K-mean Clusterings for the mispricing function $h(\mathbf{X}_i)$ to study the non-linear behavior of arbitrage characteristics.

Some recent papers such as Kim et al. (2019) and Kelly et al. (2019) analyzed a similar model as ours, which assume that both $h(\mathbf{X}_i)$ and $g_j(\mathbf{X}_i)$ are *linear functions* . They both included around 40 characteristics in $\mathbf{X}_i$. However, they drew different conclusions on the existence of $h(\mathbf{X}_i)$. Kim et al. (2019) determined assets weights of arbitrage portfolios using one-year rolling window estimated $\frac{1}{n}\hat{h}(\mathbf{X}_i)$. They showed that their arbitrage portfolios returns are statistically and economically significant. However, Kelly et al. (2019) applied instrumented principal component analysis (IPCA) to the entire time span from 1965 to 2014, and concluded no evidence to reject the null hypothesis $H_0 : h(\mathbf{X}_i) = \mathbf{X}_i^{\intercal}\mathbf{B} = \mathbf{0}$ through bootstrap. This dispute spurs the deveplopment of a more flexible model and reliable hypothesis tests to investigate the existence and structure of $h(\mathbf{X}_i)$. The introduction of IPCA, which require both large n and T to work, was introduced at Kelly et al. (2017). This method does not fit our setting since we apply rolling window analysis with small T. Furthermore, Kelly et al. (2019) restricted the

[1] $\mathbf{X_i}$ is a vector of a large set of asset-specific characteristics of stock i.

function form of $h(\mathbf{X}_i)$ and $g_j(\mathbf{X}_i)$ to be time-invariant, which is not consistent with our empirical results under a semi-parametric setting. To clarfy the differences with aforementioned research, this paper proposes a semi-parametric model, which allows for both non-linearity and time-variation of $h(\mathbf{X}_i)$ and $g_j(\mathbf{X}_i)$. Furthermore, we consider a different economic question, namely, the existence and structure of mispricing functions. Our empirical study sheds light on why Kelly et al. (2019) and Kim et al. (2019) drew different conclusions: weak, time varying and nonlinear characteristics-based mispricing functions only appear in certain rolling windows, which is hard to be detected without rolling window analysis. However, given the presence of some persistent arbitrage characteristics, portfolios developed through mispricing functions can provide abitrage returns.

The unrestrictive model in this paper brings both opportunities and challenges. According to Huang et al. (2010), the number of B-spline knots must increase in the number of observations to achieve accurate approximation and good asymptotic performance. Therefore, the dimension of B-spline bases coefficients also need to grow with the sample size. Besides, mispricing functions are treated as anomalies. Under a correctly specified factor model, coefficients of these B-splines bases are very likely to be sparse. All of these circumstances make the conventional Wald tests have very low power as discussed in Fan et al. (2015). Therefore, a power enhanced test should be developed to strengthen the power of Wald tests and to detect the most relevant characteristics among a characteristic zoo included in $h(\mathbf{X}_i)$. Kock and Preinerstorfer (2019) illustrated that if the number of coefficients diverges as the number of observations approaches infinity, the standard Wald test is power enhanceable. Fan et al. (2015) proposed a power enhanced test by introducing a screening process on all estimated coefficients one by one. They added significant components as a supplement to the standard Wald test. In this paper, we extend Fan et al. (2015) to a group manner to enhance the hypothesis test on a high dimensional additive semi-parametric function, $H_0 : h(\mathbf{X}_i) = 0$. This method allows all the significant components of $h(\mathbf{X}_i)$ to be selected and contribute to the test statistics, with the testing power approaching to one.

The careful analysis of $h(\mathbf{X}_i)$ is theoretically and practically meaningful. Firstly, the presence of $h(\mathbf{X}_i)$ is an important component of Arbitrage Pricing Theory (APT) and can contribute to asset pricing theories, namely, linking the mispricing functions with security-related characteristics. Secondly, as in Hjalmarsson and Manchev (2012) and Kim et al. (2019) , $h(\mathbf{X}_i)$ can be utilized to construct arbitrage portfolios through observed characteristics. However, both research was built upon the condition that $h(\mathbf{X}_i)$ is linear over characteristics. If the mispricing function $h(\mathbf{X}_i)$ is not monotonic, simply setting portfolio weights to the estimated values of $h(\mathbf{X}_i)$ can be problematic. In this paper, we show that some characteristics with substantially different values give rise to similar arbitrage returns. The distance of arbitrage returns between two assets i and j is $d_{ij} = |h(\mathbf{X}_i) - h(\mathbf{X}_j)|$ and the similarity of characteristics is $\|\mathbf{X}_i - \mathbf{X}_j\|_2$, where $\|\cdot\|_2$ represents L_2 distance. Inspired by Hoberg and Phillips (2016) and Vogt and Linton (2017), we employ a hierarchical K-means clustering to classify these characteristics within each mispricing return group. We present the dynamic of distinct clusters of the same characteristics leading to similar arbitrage returns, forming a "peer group" of arbitrage characteristics. Therefore, under the semiparametric setting, the asset weighting function should rely on these time-varying and non-linear peer groups.

The rest of this paper is organized as follows. Section 2 sets out the semi-parametric model. Section 3 introduces the assumptions and estimation methods. Section 4 constructs a power enhanced test for high dimensional additive semi-parametric functions. Section 5 employs Hierarchical K-Means Clustering to investigate peer groups of arbitrage characteristics. Section 6 describes the asymptotic properties of our estimates and test statistics. Section 6 simulates data to verify the performance of our methodology. Section 7 presents an empirical study. Finally, Section 8 concludes this paper. Characteristics description tables, proofs, mispricing curves and plots of peer groups are arranged to the Appendix.

2 Model setup

We assume that there are n securities observed over T time periods. We also assume that during a short time window, each security has P time-invariant observed characteristics, such as market capitalization, momentum, and book-to-market ratios. Meanwhile, we may omit heteroskedasticity by assuming that each characteristic shares a certain form of variation within each period for all securities. We suppose that

$$y_{it} = (h(\mathbf{X}_i) + \gamma_i) + \sum_{j=1}^{J} (g_j(\mathbf{X}_i) + \lambda_{ij}) f_{jt} + \epsilon_{it}, \tag{3}$$

where y_{it} is the monthly excess return of the i^{th} stock at the month t; $\mathbf{X}_i$ is a $1 \times P$ vector of P characteristics of stock i during time periods $t = 1, \dots T$. T is a small and fixed time block. In practice, most characteristics are updated annually. Thus, we assume $\mathbf{X}_i$ is time-invariant in one-year time window. $h(\mathbf{X}_i)$ is an unknown mispricing function explained by a large set of characteristics whereas γ_i is the random intercept of the mispricing part that cannot be explained by characteristics. Similarly, we have characteristics-beta function $g_j(\cdot)$ to explain the j^{th} factor loadings and the unexplained stochastic part of the loading is λ_{ij} with $E(\lambda_{ij}) = 0$. λ_{ij} is orthogonal to the $g_j(\cdot)$ function. f_{jt} is the realization of the j^{th} risk factor at time t. Finally, ϵ_{it} is homoskedastic zero-mean idiosyncratic residual of the i^{th} stock at time t. Random variables γ_i and λ_{ij} are used to generalize our settings and not to be estimated. They will be treated as noise in the identification assumptions.

To avoid the curse of dimensionality , we impose additive forms on both $h(\cdot)$ and $g_j(\cdot)$ functions: $h(\mathbf{X}_i) = \sum_{p=1}^{P} \mu_p(X_{ip})$ and $g_j(\mathbf{X}_i) = \sum_{p=1}^{P} \theta_{jp}(X_{ip})$, where $\mu_p(X_{ip})$ and $\theta_{jp}(X_{ip})$ are univariate unknown functions of the p^{th} characteristic X_p. We rewrite the model:

$$y_{it} = (\sum_{p=1}^{P} \mu_p(X_{ip}) + \gamma_i) + \sum_{j=1}^{J} (\sum_{p=1}^{P} \theta_{jp}(X_{ip}) + \lambda_{ij}) f_{jt} + \epsilon_{it}, \tag{4}$$

Assumption 1. *We suppose that:*

$$E(\epsilon_{it} | \mathbf{X}, f_{jt}) = 0,$$

$$E(h(\mathbf{X}_i)) = E(g_j(\mathbf{X}_i)) = 0,$$

$$E(\gamma_i|\mathbf{X}) = E(\lambda_{ij}|\mathbf{X}) = 0,$$

$$E(h(\mathbf{X}_i)g_j(\mathbf{X}_i)) = \mathbf{0},$$

Similar to Connor et al. (2012) and Fan et al. (2016), Assumption 1 above is to standardize the model settings, including the zero mean assumption for factor loadings and mispricing functions for identification purposes. We also impose orthogonality between mispricing and factor loading parts for the identification reason. This is because the variation of risk factors can be absorbed into the mispricing part if it is not orthogonal to the factor loadings. More discussion can be found in Connor et al. (2012).

3 Estimation

In this section we discuss the approximation of unknown univariate functions and our estimation methods for model Equation 3. In the semi-parametric setting, we apply the Projected-PCA following Fan et al. (2016) to work on the common factors and characteristics-beta directly. Next, we project the residuals onto the characteristics-based alpha space that is orthogonal to the systematic part. The second step is similar to equality constrained OLS.

3.1 B-Splines Approximation

We use B-splines sieve to approximate unknown functions $\theta(\cdot)$ and $\mu(\cdot)$ in Equation 4. Similar to Huang et al. (2010) and Chen and Pouzo (2012), we have the following procedures. Firstly, suppose that the p^{th} covariate X_p is in the interval $[D_0, D]$, where D_0 and D are finite numbers with $D_0 < D$. Let $\mathbf{D} = \{\underbrace{D_0, D_0, \ldots, D_0}_{l+1} < d_1 < d_2 < \cdots < d_{m_n} < \underbrace{D, D, \ldots, D}_{l+1}\}$ be a simple knot sequence on the interval $[D_0, D]$. Here, $m_n = \lfloor n^v \rceil$ ($\lfloor\cdot\rceil$ gives nearest integer) is a positive integer of the number of internal knots, which is a function of security size n in period t with $0 < v < 0.5$. l is the degree of those bases. Therefore, we have $H_n = l + m_n$ bases in total, which will diverge as $n \to \infty$. Following this setting, a set of B-splines can be built for the space $\Omega_n[\mathbf{D}]$.

Secondly, for the p^{th} characteristic X_p, there is a set of H_n orthogonal bases $\{\phi_{1p}(X_p), \ldots, \phi_{H_np}(X_p)\}$. Those univariate unknown functions can be approximated as linear combinations of these bases as $\mu_p(X_p) = \sum_{q=1}^{H_n} \alpha_q \phi_{qp}(X_p) + R_p^{\mu}(X_p)$ and $\theta_p(X_p) = \sum_{q=1}^{H_n} \beta_{jq}\phi_{qp}(X_p) + R_p^{\theta}(X_p)$, where $R_p^{\mu}(X_p)$ and $R_p^{\theta}(X_p)$ are approximation errors. It is not necessary to use the same bases for both unknown functions and the representation here is for notational simplicity only. Therefore, the model Equation 4 can be written as:

$$y_{it} = \sum_{p=1}^{P}(\sum_{q=1}^{H_n} \alpha_{pq}\phi_{pq}(X_{ip}) + R_p^{\mu}(X_p)) + \gamma_i + \sum_{j=1}^{J}(\sum_{p=1}^{P}(\sum_{q=1}^{H_n} \beta_{jpq}\phi_{pq}(X_{ip}) + R_p^{\theta}(X_p)) + \lambda_{ij})f_{jt} + \epsilon_{it}$$

For each $i = 1, 2, \ldots, n$, $p = 1, 2, \ldots, P$ and $t = 1, 2, \ldots, T$, we have:

$$\mathbf{1_T} = (1, \ldots, 1)^\intercal \in \mathbb{R}^T,$$

$$\beta_j = (\beta_{1,j1}, \ldots, \beta_{H_n,j1}, \ldots, \beta_{1,jP}, \ldots, \beta_{H_n,jP})^\intercal \in \mathbb{R}^{H_nP},$$

$$\mathbf{B} = (\beta_1, \ldots, \beta_J),$$

$$\mathbf{A} = (\alpha_{11}, \ldots, \alpha_{1H_n}, \ldots, \alpha_{P1}, \ldots, \alpha_{PH_n})^\intercal \in \mathbb{R}^{H_nP},$$

$$\boldsymbol{\Phi}(\mathbf{X}) = \begin{bmatrix} \phi_{1,11}(X_{11}) & \cdots & \phi_{1,1H_n}(X_{11}) & \cdots & \phi_{1,P1}(X_{1P}) & \ldots & \phi_{1,PH_n}(X_{1P}) \\ \phi_{2,11}(X_{21}) & \cdots & \phi_{2,1H_n}(X_{21}) & \cdots & \phi_{2,P1}(X_{2P}) & \ldots & \phi_{2,PH_n}(X_{2P}) \\ \vdots & \vdots & \vdots & \ddots & \vdots & & \\ \phi_{n,11}(X_{n1}) & \cdots & \phi_{n,1H_n}(X_{n1}) & \cdots & \phi_{n,P1}(X_{nP}) & \ldots & \phi_{n,PH_n}(X_{nP}) \end{bmatrix},$$

where $\phi_{i,ph}(X_{ip})$ is the h^{th} basis of the p^{th} characteristic of asset i at time t. Therefore, the original model

$$\mathbf{Y} = (h(\mathbf{X}) + \boldsymbol{\Gamma})\mathbf{1}_\mathbf{T}^\intercal + (\mathbf{G}(\mathbf{X}) + \boldsymbol{\Lambda})\mathbf{F}^\intercal + \mathbf{U},$$

can be represented by B-spline sieve as:

$$\mathbf{Y} = (\boldsymbol{\Phi}(\mathbf{X})\mathbf{A} + \boldsymbol{\Gamma} + \mathbf{R}^\mu(\mathbf{X}))\mathbf{1}_\mathbf{T}^\intercal + (\boldsymbol{\Phi}(\mathbf{X})\mathbf{B} + \boldsymbol{\Lambda} + \mathbf{R}^\theta(\mathbf{X}))\mathbf{F}^\intercal + \mathbf{U}, \tag{5}$$

$\mathbf{Y}$ is $n \times T$ matrix of y_{it}; $\boldsymbol{\Phi}(\mathbf{X})$ is the $n \times PH_n$ matrix of B-Spline bases; $\mathbf{A}$ is a $PH_n \times 1$ matrix of mispricing coefficients; $\mathbf{R}^\mu(\mathbf{X})$ is a $n \times 1$ matrix of approximation errors; $\mathbf{B}$ is a $PH_n \times J$ matrix factor loadings' coefficients; $\mathbf{R}^\theta(\mathbf{X})$ is a $n \times J$ matrix of approximation errors. We have $R_p^\mu(X_p) \to^p 0$ and $R_p^\theta(X_p) \to^p 0,$ as $n \to \infty$ as in Huang et al. (2010). Therefore, we omit the approximation errors for simplicity below. $\mathbf{F}$ is the $T \times J$ matrix of f_{tj} and $\mathbf{U}$ is a $n \times T$ matrix of ϵ_{it}. $h(\mathbf{X})$ is a $n \times 1$ vector of characteristics-based mispricing component; $\mathbf{G}(\mathbf{X})$ is a $n \times J$ vector of characteristics-based factor loadings; $\mathbf{1_T}$ is a $T \times 1$ vector of 1. The rest are defined the same as Equation 4.

We define a projection matrix as:

$$\mathbf{P} = \Phi(\mathbf{X})(\Phi(\mathbf{X})^\intercal\Phi(\mathbf{X}))^{-1}\Phi(\mathbf{X})^\intercal.$$

The remaining goals of this paper are to estimate both $h(\boldsymbol{X})$ and $\mathbf{G}(\mathbf{X})$ consistently and conduct a power enhanced test of the hypothesis $H_0 : h(\boldsymbol{X}) = \mathbf{0}$, i.e., to check the existence of mispricing functions under semi-parametric settings. Finally, we cluster peer groups of arbitrage characteristics.

3.2 Two Steps Projected-PCA

In this section, we combine and extend Projected-PCA by Fan et al. (2016) and equality constrained least squares similar to Kim et al. (2019) to estimate the model. To facilitate the estimation, we define a $T \times T$ time series demeaning matrix $\mathbf{D_T} = \mathbf{I_T} - \frac{1}{T}\mathbf{1_T}\mathbf{1_T^\intercal}$.[2] Next, we demean the equation above on both sides. Therefore we have

$$\mathbf{Y}\mathbf{D_T} = \tilde{\mathbf{Y}} = (\mathbf{\Phi}(\mathbf{X})\mathbf{B} + \mathbf{\Lambda})\mathbf{F}^\intercal\mathbf{D_T} + \mathbf{U}\mathbf{D_T}.$$

Mispricing terms disappear since they are time-invariant by $(\mathbf{\Phi}(\mathbf{X})\mathbf{A} + \mathbf{\Gamma})\mathbf{1_T^\intercal}\mathbf{D_T} = \mathbf{0}$. This helps us to work on the systematic part directly. Henceforth, we use $\mathbf{F}$ to represent the time-demeaned factor matrix.

Our procedures are designed to estimate factor loadings $\mathbf{G}(\mathbf{X})$, time-demeaned unobserved factors $\mathbf{F}$ and mispricing coefficients $\mathbf{A}$ in sequence.

Under Assumption 1, we have the following estimation procedures:

1 Projecting $\tilde{\mathbf{Y}}$ onto the spline space spanned by $\{\mathbf{X_{ip}}\}_{i\leqslant n, p\leqslant P}$ through a $n \times n$ projection matrix $\mathbf{P}$ with $\mathbf{P} = \Phi(\mathbf{X})(\Phi(\mathbf{X})^\intercal\Phi(\mathbf{X}))^{-1}\Phi(\mathbf{X})^\intercal$. We then collect the projected data $\hat{\mathbf{Y}} = \Phi(\mathbf{X})(\Phi(\mathbf{X})^\intercal\Phi(\mathbf{X}))^{-1}\Phi(\mathbf{X})^\intercal\tilde{\mathbf{Y}}$.

2 Applying the Principle Component Analysis to the projected data $\hat{\mathbf{Y}}^\intercal\hat{\mathbf{Y}}$. This allows us to work directly on the sample covariance of $\mathbf{G}(\mathbf{X})\mathbf{F}^\intercal$, under the condition $E(g_j(\mathbf{X}_i)\epsilon_{it}) = E(g_j(\mathbf{X}_i)\lambda_{ij}) = 0$.

3 Estimating $\hat{\mathbf{F}}$ as the eigenvectors corresponding to the first J (assumed given) eigenvalues of the $T \times T$ matrix $\frac{1}{n}\hat{\mathbf{Y}}^\intercal\hat{\mathbf{Y}}$ (covariance of projected $\hat{\mathbf{Y}}$).

The method above substantially improves estimation accuracy and facilitates theoretical analysis even under the large n and small T . Small T is preferable in our model setting as we use one-year rolling windows analysis in both simulation and empirical studies, and large n is required for asymptotic analysis.

Factor loadings $\hat{\mathbf{G}}(\mathbf{X})$ are estimated as:

$$\hat{\mathbf{G}}(\mathbf{X}) = \hat{\mathbf{Y}}\hat{\mathbf{F}}(\hat{\mathbf{F}}^\intercal\hat{\mathbf{F}})^{-1}$$

In the next step, we estimate the coefficients of the mispricing bases.

4 The estimator of $\mathbf{A}$ is

$$\hat{\mathbf{A}} = \arg\min_{\mathbf{A}} \mathrm{vec}(\mathbf{Y} - \mathbf{\Phi}(\mathbf{X})\mathbf{A}\mathbf{1_T^\intercal} - \hat{\mathbf{G}}(\mathbf{X})\hat{\mathbf{F}}^\intercal)^\intercal \mathrm{vec}(\mathbf{Y} - \mathbf{\Phi}(\mathbf{X})\mathbf{A}\mathbf{1_T^\intercal} - \hat{\mathbf{G}}(\mathbf{X})\hat{\mathbf{F}}^\intercal),$$

subject to $\hat{\mathbf{G}}(\mathbf{X})^\intercal\mathbf{\Phi}(\mathbf{X})\mathbf{A} = \mathbf{0_J}$.

Let a $PH_n \times 1$ vector $\hat{\mathbf{A}}$ be a closed-form solution:

$$\hat{\mathbf{A}} = \mathbf{M}\tilde{\mathbf{A}},$$

[2] $\mathbf{I_T}$ is a $T \times T$ identity matrix, and $\mathbf{1_T}$ is a $T \times 1$ matrix of 1.

where

$$\mathbf{M} = \mathbf{I} - (\Phi(\mathbf{X})^\intercal\Phi(\mathbf{X}))^{-1}\mathbf{\Phi}(\mathbf{X})^\intercal\hat{\mathbf{G}}(\mathbf{X})(\hat{\mathbf{G}}(\mathbf{X})^\intercal\hat{\mathbf{G}}(\mathbf{X}))^{-1}\hat{\mathbf{G}}(\mathbf{X})^\intercal\mathbf{\Phi}(\mathbf{X}),$$

$$\tilde{\mathbf{A}} = \frac{\mathbf{1}}{\mathbf{T}}(\Phi(\mathbf{X})^\intercal\Phi(\mathbf{X}))^{-1}\Phi(\mathbf{X})^\intercal(\mathbf{Y} - \hat{\mathbf{G}}(\mathbf{X})\hat{\mathbf{F}}^\intercal)\mathbf{1}_\mathbf{T},$$

given $\mathbf{P}\hat{\mathbf{G}}(\mathbf{X}) = \hat{\mathbf{G}}(\mathbf{X})$.

As in Assumption 1, the $h(\mathbf{X})$ is orthogonal to the characteristics-based loadings $\mathbf{G}(\mathbf{X})$.

5 We also estimate the covariance matrix of $\hat{\mathbf{A}}$, i.e., $\mathbf{\Sigma}$, by extending the methods of Liew (1976). This can facilitate theoretical analysis in the next section. According to Liew (1976), $\hat{\mathbf{A}}$ is the equality constrained least-square estimator, which has the covariance matrix as (under $n \leqslant T$ and covariance shrinkage as in Ledoit et al. (2012) and Fan et al. (2013) among others.):

$$\hat{\mathbf{\Sigma}} = \mathbf{M}\hat{\mathbf{\Sigma}}_{\tilde{\mathbf{A}}}\mathbf{M}^\intercal,$$

where:

$$\hat{\mathbf{\Sigma}}_{\tilde{\mathbf{A}}} = (\Phi(\mathbf{X})^\intercal\Phi(\mathbf{X}))^{-1}\Phi(\mathbf{X})^\intercal \begin{bmatrix} \hat{\sigma}_1^2 & & \\ & \ddots & \\ & & \hat{\sigma}_n^2 \end{bmatrix} \Phi(\mathbf{X})(\Phi(\mathbf{X})^\intercal\Phi(\mathbf{X}))^{-1},$$

$$\hat{\sigma}_i^2 = \frac{\sum_1^T \hat{e}_{it}^2}{T-1},$$

where $\sum_1^T \hat{e}_{it}^2 = \sum_1^T (y_{it} - \sum_{p=1}^P \sum_{q=1}^{H_n} \hat{\alpha}_{pq}\phi_{pq}(x_{ip}) - \sum_{j=1}^J (\sum_{p=1}^P \sum_{q=1}^H \hat{\beta}_{jpq}\phi_{pq}(x_{ip}))\hat{f}_{jt})^2$. Heteroskedasticity is caused by γ_i.

4 Power Enhanced Tests

There are considerable discussions about the mispricing phenomenon under factor models while the existence of mispricing functions remains controversial. Namely, whether there are relevant covariates explaining remaining excess returns after subtracting co-movements components captured by risk factors. Recently, Kim et al. (2019) found the characteristics arbitrage opportunities by estimating a linear characteristic mispricing function, without providing theoretical results. However, Kelly et al. (2019) conducted a conventional Wald hypothesis test on the similar mispricing function using bootstrap, concluding that there is no evidence to reject the null hypothesis $H_0 : h(\mathbf{X}) = \mathbf{0}$. Additionally, they applied the bootstrap method to estimate the covariance matrix $\mathbf{\Sigma}$, which caused potential problems for theoretical analysis. Moreover, according to Fan et al. (2015), their test results may have relatively low power when the true coefficient vector of linear mispricing function $\mathbf{A}$ has a sparse structure.

Both studies adopt a parametric framework, which relies on the strong assumption of linearity. However, this assumption is not consistent with Connor et al. (2012), which showed that both characteristic-beta and mispricing functions are very likely to be non-linear. Therefore, we propose a semiparametric model to accommodate the non-linearity to a great extent.

But semi-parametric framework leads to additional challenges for inference. On the one hand, as mentioned above, the number of coefficients of mispricing B-splines diverge as $n \to \infty$, which implies that the power of standard Wald test can be quite low, (see Fan et al. (2015)). On the other hand, according to other research like Fama and French (1993) and Fama and French (2015), mispricing terms can be regarded as anomalies. This means that in our model setting, the true mispricing coefficient vector $\mathbf{A}$ can be high-dimensional but sparse, reducing the power of conventional Wald test further.

According to Kock and Preinerstorfer (2019), conventional hypothesis tests under these circumstances are power enhanceable. The power enhanced Wald test in this paper is an extension of Fan et al. (2015) to a group manner, namely, the hypothesis test under high-dimensional additive semi-parametric settings. The proposed test are power strengthened when the coefficients of the additive regression $\mathbf{A}$ is diverging as $n \to \infty$ without size distortion. Meanwhile, this test is robust to sparse alternatives. On top of that, the proposed test can select the most important components from sparse additive functions. Finally, the proposed method can also be applied when the number of characteristics is diverging, i.e., $P \to \infty$.

We construct a new test:

$$H_0 : h(\mathbf{X}) = \mathbf{0}, \qquad H_1 : h(\boldsymbol{X}) \neq \mathbf{0},$$

equivalently,

$$H_0 : \mathbf{A} = \mathbf{0}, \qquad H_1 : \mathbf{A} \in \mathcal{A},$$

where $\mathcal{A} \subset \mathbb{R}^{PH_n} \backslash \mathbf{0}$.

Here, we have:

$$S_1 = \frac{\hat{\mathbf{A}}\hat{\mathbf{\Sigma}}^{-1}\hat{\mathbf{A}}^{\intercal} - PH_n}{\sqrt{2PH_n}}$$

where S_1 is the "original" Wald test statistics; P is the number of characteristics; PH_n is the total number of B-spline bases, and $\mathbf{A} \in \mathbb{R}^{PH_n}$. The value of H_n is a function of asset number n, therefore, $H_n \to \infty$ as $n \to \infty$. Under H_0, S_1 has nondegenerate limiting distribution F as $n \to \infty$. Given the significance level q, $q \in (0, 1)$ as well as the critical value F_q:

$$S_1 | H_0 \to^d F$$

$$\lim_{N \to \infty} \Pr(S_1 > F_q | H_0) = q.$$

Pesaran and Yamagata (2012) showed that:

$$S_1 | H_0 \to^d \mathcal{N}(0, 1),$$

under regularity conditions.

Potentially, sparse and diverging PH_n means that it is plausible to add a power enhanced component to S_1, which can improve the power of the hypothesis test without any size distortions.

Therefore, we can construct an extra screening component S_0 as:

$$S_0 = H_n \sum_{p=1}^{P} \mathbf{I}(\sum_{h=1}^{H_n} |\hat{\alpha}_{ph}|/\hat{\sigma}_{ph} \geqslant \eta_n),$$

where $\hat{\sigma}_{ph}$ is the ph^{th} entry of the diagonal elements of $\hat{\boldsymbol{\Sigma}}$. $\mathbf{I}(\cdot)$ is an indicator for the screening process while η_n is a data-driven threshold value to avoid potential size-distortion.

Here we discuss the choice of η_n. By construction and assumption of independent characteristics, we assume all B-Spline bases are orthogonal. Our goal is to bound the maximum of those standardized coefficients.

Define $Z = \max_{\{1 \leqslant p \leqslant P, 1 \leqslant h \leqslant H_n\}} \{|\hat{\alpha}_{ph}|/\hat{\sigma}_{ph}\}$. We have

$$\hat{\alpha}_{ph}/\hat{\sigma}_{ph}|\mathbf{H_0} \to^d N(0,1),$$

$$\mathbf{E}(\mathbf{Z}) = \sqrt{2 \log PH_n}.$$

After grouping coefficients of bases used to approximate the unknown function of each characteristic, let $Q = \max(\sum_{h=1}^{H_n} |\hat{\alpha}_{1h}|/\hat{\sigma}_{1h}, \ldots, \sum_{h=1}^{H_n} |\hat{\alpha}_{ph}|/\hat{\sigma}_{ph} \ldots, \sum_{h=1}^{H_n} |\hat{\alpha}_{Ph}|/\hat{\sigma}_{Ph})$. Following this, we may set the threshold as $\eta_n = H_n\sqrt{2\log(PH_n)}$, where $H_n = l + n^v$. As H_n is a slowly diverging sequence, it can control the influence of the group size properly. Meanwhile, η_n also diverges slowly so that η_n is a conservative threshold value used to avoid potential size distortion.

Apart from strengthening the power of conventional hypothesis test, $\mathbf{I}(\cdot)$ is a screening term which can select the most relevant characteristics at the same time.

We then define the arbitrage characteristics set, which includes the characteristics that have the strong explanation power for mispricing functions:

$$\mathcal{M} = \{\mathbf{X_m} \in \mathcal{M} : \sum_{h=1}^{H_n} |\alpha_{ph}|/\sigma_{ph} \geqslant \eta_n, \quad m = 1, 2, \ldots, M\}$$

$$\hat{\mathcal{M}} = \{\mathbf{X_m} \in \hat{\mathcal{M}} : \sum_{h=1}^{H_n} |\hat{\alpha}_{ph}|/\hat{\sigma}_{ph} \geqslant \eta_n, \quad m = 1, 2, \ldots, M\}$$

Therefore, we have $\mathcal{M} \cup \mathbf{0} = \mathcal{A}$ and $\mathcal{M} \cap \mathbf{0} = \emptyset$. When the set $\mathcal{M}$ is relatively small, conventional tests are likely to suffer the lower power problem. The added S_0 strengthens the power of the test and drives the power to one since H_n is slowly diverging.

Therefore, our new test statistics is $S = S_0 + S_1$, and asymptotic properties of S will be discussed later.

To conclude, the advantages of our new statistics $S = S_0 + S_1$ are:

1 The power of the hypothesis test on $H_0 : h(\mathbf{X}) = \mathbf{0}$ is mainly enhanced without size distortions.

2 We can find specific characteristics which cause the mispricing by this screening mechanism.

As designed, S_0 satisfies all three properties of Fan et al. (2015), as $n \to \infty$:

1 S_0 is non-negative, $\Pr(S_0 \geqslant 0) = 1$

2 S_0 does not cause size distortion: under H_0, $\Pr(S_0 = 0 \mid H_0) \to 1$

3 S_0 enhances test power. Under alternative H_1, S_0 diverge quickly in probability given the well chosen $\eta_{n,T}$.

Based on properties of S_0, we have three properties of S listed:

1 No size distortion $\limsup\limits_{n\to\infty} \Pr(S > F_q | H_0) = q$

2 $\Pr(S > F_q|H_1) \geqslant \Pr(S_1 > F_q|H_1)$. Hence, the power of S is at least as large as that of S_1.

3 $\Pr(S > F_q|H_1) \to 1$ when S_0 diverges. This happens, especially, when the true form of $\hat{\mathbf{A}}$ has a sparse structure.

5 Hierarchical K-Means Clustering

This section introduces a Herarchical K-means Clustering method to find peer groups of arbitrage characteristics based on their arbitrage returns. We ask whether distinct groups of the same characteristics may result in similar characteristic-based arbitrage returns in each rolling block, which is an implication for non-monotonic mispricing function, and forms a "peer group" of arbitrage characteristics. Because arbitrage portfolios rely on the linearity of characteristics-bases mispricing components to work, our clustering results can provide new evidence for the effectiveness of these arbitrage porfolios. Introduction of K-means clustering can be found in Cox (1957) and Fisher (1958).

After the screening process in section 4, we obtain the relevant components of mispricing function $h(\boldsymbol{X})$, which is estimated as

$$\hat{\mathcal{M}} = \{\mathbf{X_m} \in \hat{\mathcal{M}} : \sum_{h=1}^{H_n} |\hat{\alpha}_{ph}|/\hat{\sigma}_{ph} \geqslant \eta_n, \quad m = 1, 2, \ldots, M\}.$$

We define a $n \times M$ matrix $\boldsymbol{M}$ of arbitrage characteristics at time window t as :

$$\boldsymbol{M} = \{\boldsymbol{X_1}, \boldsymbol{X_2}, \ldots, \boldsymbol{X_M}\}, \text{ where } \boldsymbol{X_m} \in \hat{\mathcal{M}}.$$

Note that these characteristics are time-invariant within each rolling window. We also set characteristics-based arbitrage returns of asset i in month t as:

$$\ddot{y}_{it} = \boldsymbol{\phi}(\boldsymbol{M_i})\hat{\mathbf{A}}_{\mathrm{M}},$$

where $\boldsymbol{\phi}(\boldsymbol{M_i})$ and $\hat{\mathbf{A}}_{\mathrm{M}}$ are the corresonding parts of matrix $\boldsymbol{\Phi}(\mathbf{X_i})$ and vector $\hat{\mathbf{A}}$. For each rolling window, we classify all n assets through a 2-layer K-means clustering. At the first layer, we cluster these assets into K groups according to the similarity of their characteristics-based arbitrage returns $\ddot{y}_{it}$. At the second layer, we divide R_j subgroups within the j^{th} group from the first layer by the similarity of their arbitrage characteristics, where $j = 1, 2, \ldots, K$. Finally, the peer groups of arbitrage characteristics can be attained. We repreat this method for all rolling blocks to investigate dynamic patterns of these peer groups. These clusterings will provide illustrative evidence of linear/nonlinear and time-invariant/time-varying structure of mispricing function $h(\boldsymbol{X})$.

We give the classification procedures of both layers. We define Δ_{ij} as the difference between characteristics-based arbitrage returns of $\ddot{y}_{it}$ and $\ddot{y}_{jt}$, as well as Υ_{ij} as the difference between arbitrage characteristics:

$$\Delta_{ij} = \ddot{y}_{it} - \ddot{y}_{jt}, \text{ where } i \neq j, j = 1, 2, \ldots, n.$$

$$\Upsilon_{ij} = \|\boldsymbol{M}_i - \boldsymbol{M}_j\|_2, \text{ where } i \neq j, i, j = 1, 2, \ldots, n,$$

$\boldsymbol{M}_i$ represents the i^{th} row of $\boldsymbol{M}$. We set two tolerance thresholds ψ_y and ψ_x, which are used to control the biggest difference within each group of both layers separately. To accelerate the convergence of the K-means Clustering, we first apply a first difference process, which is introduced below, to obtain centroids as in Vogt and Linton (2017).

For the first layer, we have first difference process:

1. **First difference**: We randomly pick i^{th} asset and then we calculate Δ_{ij} with other assets $j = 1, 2, \ldots, n$. Thus we obtain $\Delta_{i(1)} \ldots \Delta_{i(n)}$, with n being the total individuals for classification. Without loss of generality, we assume $\Delta_{i(1)} = \min\{\Delta_{i(1)} \ldots \Delta_{i(n)}\}$, and $\Delta_{i(n)} = \max\{\Delta_{i(1)} \ldots \Delta_{i(n)}\}$.

2. **Ordering**: We rank the values obtained in Step 1 as follows:

$$\begin{aligned}\Delta_{i(1)} \leqslant \ldots \leqslant \Delta_{i(j_1-1)} &< \Delta_{i(j_1)} \leqslant \ldots \leqslant \Delta_{i(j_2-1)} \\ &< \Delta_{i(j_2)} \leqslant \ldots \leqslant \Delta_{i(j_3-1)} \\ &\vdots \\ &< \Delta_{i(j_{K-1})} \leqslant \ldots \leqslant \Delta_{i(n)}.\end{aligned}$$

We use the strict inequality mark to show large jumps of "first difference", all of which are larger than ψ_y , while the weak inequality means that the distance calculated is smaller than ψ_y. We identify $K - 1$ jumps that are larger than ψ_y above. Thus, the initial classification is achieved, and we have a total of K groups with $j_1 - 1$ members in the first group $\mathcal{C}_1$, $j_2 - j_1$ members in the second group $\mathcal{C}_2$, . . . , and $n - j_K + 1$ members in the final group $\mathcal{C}_K$.

In terms of the second layer, for the assets in the k^{th} group $\mathcal{C}_k$, we use the same method to further divide them into r subgroups as $\mathcal{R}_{1k}, \mathcal{R}_{2k}, \dots, \mathcal{R}_{rk}$. Within each subgroup, we have:

$$\Upsilon_{ab} = \|\boldsymbol{M}_a - \boldsymbol{M}_b\|_2 \leqslant \psi_x, \text{ where } a, b \in \mathcal{R}_{ik}, i = 1, 2, \dots, r, \text{ and } k = 1, 2, \dots, K.$$

The K-means algorithm is:

1. Step 1: Determine the starting mean values for each group $\hat{\bar{c}}_1^{[0]}, \dots, \hat{\bar{c}}_K^{[0]}$ and calculate the distances $\hat{D}_k(i) = \Delta(\ddot{y}_{it}, \hat{\bar{c}}_k^{[0]}) = |\ddot{y}_{it} - \hat{\bar{c}}_k^{[0]}|$ for each i and k. Define the partition $\{\mathcal{C}_1^{[0]}, \dots, \mathcal{C}_K^{[0]}\}$ by assigning the i^{th} individual to the k-th group $\mathcal{C}_k^{[0]}$ if $\hat{D}_k(i) = \min_{1 \leqslant k' \leqslant K} \hat{D}_{k'}(i)$.

2. Step l: Let $\{\mathcal{C}_1^{[l-1]}, \dots, \mathcal{C}_K^{[l-1]}\}$ be the partition of $\{1, \dots, n\}$ from the latest iteration step. Calculate mean functions
$$\hat{\bar{c}}_k^{[l]} = \frac{1}{|\mathcal{C}_k^{[l-1]}|} \sum_{i \in \mathcal{C}_k^{[l-1]}} \ddot{y}_{it} \quad \text{for } 1 \leqslant k \leqslant K$$
And then we calculate $\Delta(\ddot{y}_{it}, \hat{\bar{c}}_k^{[l]}) = |\ddot{y}_{it} - \hat{\bar{c}}_k^{[l]}|$ for each i and k. Define the partition $\{\mathcal{C}_1^{[l]}, \dots, \mathcal{C}_K^{[l]}\}$ by assigning the i^{th} individual to the k-th group $\mathcal{C}_k^{[l]}$ if $\hat{D}_k(i) = \min_{1 \leqslant k' \leqslant K_0} \hat{D}_{k'}(i)$.

3. Iterate the above steps until the partition $\{\mathcal{C}_1^{[w]}, \dots, \mathcal{C}_K^{[w]}\}$ does not change anymore.

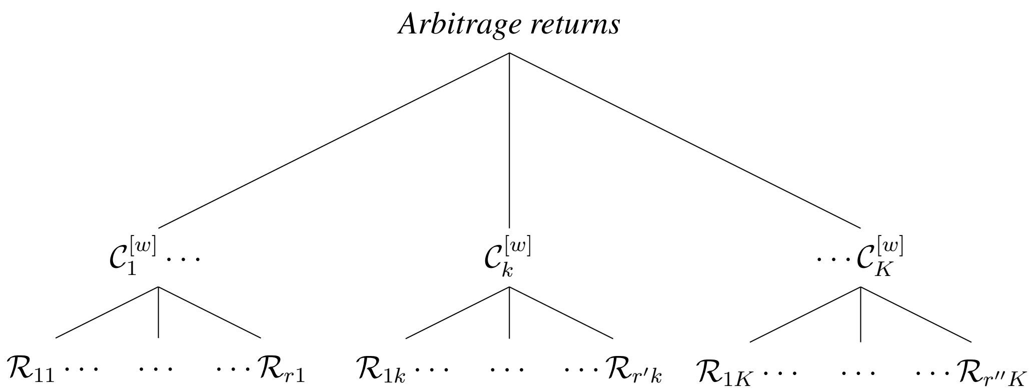
To accelerate the convergence of K-means algorithm, at the step 1, results of first difference are used. As we have already obtained our initial grouping $\{\mathcal{C}_1, \dots, \mathcal{C}_K\}$, therefore starting values for the Step 1 is:

$$\hat{\bar{c}}_k^{[0]} = \frac{1}{|\mathcal{C}_k|} \sum_{i \in \mathcal{C}_k} \ddot{y}_{it} \quad \text{for } 1 \leqslant k \leqslant K,$$

where $|\mathcal{C}_k|$ is the cardinality of the group $\mathcal{C}_k$.

The consistency and other theoretical results of the above procedures can be found in Pollard (1981)Pollard et al. (1982), Sun et al. (2012) and Vogt and Linton (2017).

For the second layer, we repeat the procedures within each group $\mathcal{C}_k^{[w]}$ respect to Υ_{ab}, and the structure of characteristics-based arbitrage returns is:



The first layer is the structure of characteristics-based arbitrage returns, while the second layer gives peer groups of characteristics that can provide similar characteristics-based arbitrage returns.

The number of clusterings are determined by threshold values ψ_y and ψ_x directly. ψ_y and ψ_x are chosen by the tradeoff between the number of clusterings and total within sum of squares.

6 Asymptotic properties

This section discusses assumptions and properties of estimates and power enhanced statistics S.

6.1 Consistency Assumptions

Assumption 2. *As $n \to \infty$, we have:*

$$\frac{1}{n}\mathbf{Y}^\intercal\mathbf{Y} \to_P \mathbf{M_Y},$$

$$\mathbf{F}^\intercal\mathbf{F} = \mathbf{I_J},$$

where $\mathbf{M_Y}$ is a positive definite matrix, and $\mathbf{I_J}$ is a $J \times J$ identity matrix.

We define $\lambda_{min}(M)$ and $\lambda_{max}(M)$ as the largest and the smallest eigenvalue of matrix M, respectively. Additionally, we define C_{min} and C_{max} to be positive constants such that:

$$C_{min} \leqslant \lambda_{min}(\frac{1}{n}\mathbf{\Phi}^\intercal(\mathbf{X})\mathbf{\Phi}(\mathbf{X})) < \lambda_{max}(\frac{1}{n}\mathbf{\Phi}^\intercal(\mathbf{X})\mathbf{\Phi}(\mathbf{X})) \leqslant C_{max}$$

as $n \to \infty$.

We impose these restrictions to avoid non-invertibility of stock returns, characteristics, and rotation indeterminacy.

Assumption 3.

$$\frac{1}{n}\mathbf{G}(\mathbf{X})^\intercal\mathbf{P}\mathbf{G}(\mathbf{X}) \to_P \begin{bmatrix} d_1 & & \\ & \ddots & \\ & & d_{PH_n} \end{bmatrix},$$

as $n \to \infty$, where d_{PH_n} are distinct entries.

Both Assumption 2 and 3 are similar to those in Fan et al. (2016), which are used to separately identify risk factors and factor loadings. Given the orthogonal bases of B-splines and uncorrelated or weakly correlated characteristics, Assumption 3 is mild.

Assumption 4. *K_{min} and K_{max} are positive constants such that:*

$$K_{min} \leqslant \lambda_{min}(\frac{1}{n}\mathbf{G}(\mathbf{X})^\intercal\mathbf{P}\mathbf{G}(\mathbf{X})) < \lambda_{min}(\frac{1}{n}\mathbf{G}(\mathbf{X})^\intercal\mathbf{P}\mathbf{G}(\mathbf{X})) \leqslant K_{max}$$

as $n \to \infty$.

This assumption requires nonvanishing explanatory power of the B-spline bases $\mathbf{\Phi}(\mathbf{X})$ on the factor loading matrix $\mathbf{G}(\mathbf{X})$.

Assumption 5. *ϵ_{it} is realized i.i.d. idiosyncratic shocks with $E(\epsilon_{it}) = 0$ and $\mathrm{var}(\epsilon_{it}) = \sigma^2$.*

Heteroskedasticity is caused by γ_i, namely, $\mathrm{var}(\gamma_i + \epsilon_{it}) = \sigma_i^2$.

6.2 Main Results

Theorem 6.1. *Let $\hat{\mathbf{F}}$ be the $J \times T$ matrix estimate of latent risk factors. Under Assumption 1-4, $\hat{\mathbf{F}} \to^P \mathbf{F}$, as $n \to \infty$.*

Theorem 6.2. *Define the $n \times J$ matrix $\hat{\mathbf{G}}(\mathbf{X})$ as the estimate of factor loadings $\mathbf{G}(\mathbf{X})$. Under Assumption 1-4 and Theorem 6.1 , as $n \to \infty$, then $\hat{\mathbf{G}}(\mathbf{X}) \to^P \mathbf{G}(\mathbf{X})$.*

Theorem 6.3. *Let the $PH_n \times 1$ vector $\hat{\mathbf{A}}$ be the solution of constrained OLS, then*

$$\hat{\mathbf{A}} = \mathbf{M}\tilde{\mathbf{A}},$$

where

$$\mathbf{M} = \mathbf{I} - (\Phi(\mathbf{X})^\intercal\Phi(\mathbf{X}))^{-1}\mathbf{\Phi}(\mathbf{X})^\intercal\hat{\mathbf{G}}(\mathbf{X})(\hat{\mathbf{G}}(\mathbf{X})^\intercal\mathbf{\Phi}(\mathbf{X})(\Phi(\mathbf{X})^\intercal\Phi(\mathbf{X}))^{-1}\mathbf{\Phi}(\mathbf{X})^\intercal\hat{\mathbf{G}}(\mathbf{X}))^{-1}\hat{\mathbf{G}}(\mathbf{X})^\intercal\mathbf{\Phi}(\mathbf{X}),$$
$$\tilde{\mathbf{A}} = \frac{\mathbf{1}}{\mathbf{T}}(\Phi(\mathbf{X})^\intercal\Phi(\mathbf{X}))^{-1}\Phi(\mathbf{X})^\intercal(\mathbf{Y} - \hat{\mathbf{G}}(\mathbf{X})\hat{\mathbf{F}}^\intercal)\mathbf{1}_{\mathbf{T}}^\intercal.$$

Under Assumption 1-4, $\mathbf{\Phi}(\mathbf{X})\hat{\mathbf{A}} \to^P h(\boldsymbol{X})$, as $n \to \infty$.

Theorem 6.4. *Under Assumption 3 and Assumption 5, $\mathbf{E}(\mathbf{Z}) = \sqrt{2\log PH_n}$.*

Theorem 6.5. *Define η_n as the threshold value to control the maximum noise, then:*

$$\inf_{\mathbf{A}=\mathbf{0}} \Pr(\max_{p\leqslant P, h\leqslant H} |\hat{\alpha}_{ph} - \alpha_{ph}|/\hat{\sigma}_{ph} \leqslant \eta_n|\mathbf{A}) \to 1.$$

Under $n \to \infty$ and H_0, given the properties of S_0 and S_1, then:

$$S \to^d N(0,1),$$

The power of S is approaching to 1 as:

$$\inf_{\mathbf{A}\in\mathcal{A}} \Pr(\textit{reject } H_0|\mathbf{A}) \to 1.$$

7 Numerical Study

In this section, we use Compustat and Fama-French three and five factors data to simulate stocks returns and then evaluate the performance of our estimation and hypothesis test procedures.

7.1 Data Generation

Firstly, we use Fama-French three factors monthly returns and all the characteristics that will be included in the empirical study to mimic the stocks excess returns. Most of the characteristics are updated annually so we treat those variables as time-invariant during each one-year rolling block. For the characteristics that are updated every month, we substitute the mean values as their fixed values for each fiscal year. We use Fama-French monthly returns from July of year t to June of year $t+1$ and characteristics of fiscal year $t-1$ to generate the stock returns from July of year t to June of year $t+1$. The periods we generate are the same as the empirical study, namely, 50 years from July 1967 to June 2017. For each rolling block with 12 months we have:

$$y_{it} = h(X_i) + \sum_{j=1}^{3} g_j(X_j) f_{jt} + \epsilon_{it}, \tag{6}$$

y_{it} is the generated stock's return; $h(X_i)$ is the mispricing function consists of a non-linear characteristic function of x_i, which is to mimic the sparse structure of the mispricing function; $g_j(\boldsymbol{X}_j)$ is the j^{th} characteristics-based factor loading, which has an additive semi-parametric structure; $\boldsymbol{X}_j$ is the j^{th} subset consisting of 4 characteristics; f_{jt} is the j^{th} Fama-French factor returns at time t; ϵ_{it} is the idiosyncratic shock for stock i at time t, generated from $N(0, \sigma^2)$.

We generate characteristic functions:

$$h(X_i) = \sin X_i,$$

$$g_j(\boldsymbol{X}_j) = X_{j1}^2 + (3X_{j2}^3 - 2X_{j2}^2) + (3X_{j3}^3 - 2X_{j3}) + X_{j4}^2,$$

X_{ji} is a randomly picked characteristic without replacement from the data in empirical study and $j = 1, 2, 3$, $i = 1, \dots, 4$. Description of these characteristics can be found in the Appendix. Additionally, all $h(X_i)$, $g_j(\boldsymbol{X}_j)$ are rescaled to have zero mean and unit variance. As we use real data to conduct the simulation, the assumption of independent X_i may not be satisfied. Although some characterisitcs are correlated, the semi-paramtric model overcomes this problem properly when compared with the parametric model that has serious size distortion.

We do not specify $h(X_i)$ and $g_j(\boldsymbol{X}_j)$ to be orthogonal explicitly, but we draw characteristics without replacement and employ sine-waves and polynomials to approximate the orthogonality as much as possible. Orthogonality is a strong assumption in reality, which is required for Theorem 6.5. In this simulatin, our method can only estimate the component of $h(X_i)$ that is orthogonal to $g_j(\boldsymbol{X}_j)$. However, results reveal that one can still select the arbitrage characteristics even if we cannot estimate arbitrary $h(X_i)$ unrestrictively.

7.2 Model Misspecification

In this subsection, we show the necessity to consider semi-parametric analysis when the forms of factor loadings and mispricing functions are nonlinear.

Under the data generation process, we consider both semi-parametric and linear analysis to compare Mean Squared Error (MSE) and hypothesis test results under both specifications. We apply our estimation methodology in section 3 to estimate Equation 7.1. For semi-parametric specification, we choose the number of B-Spline bases to be $\lfloor n^{0.3} \rceil$. n is the number of assets in each balanced rolling window, and $\lfloor \cdot \rceil$ means the nearest integer. We orthogonalize these bases, and then use the Projected-PCA and restricted OLS to estimate model Equation 7.1. As for the hypothesis test part, we choose threshold value to be $\eta_n = H_n \sqrt{2 \log(PH_n)} = \lfloor n^{0.3} \rceil \sqrt{2 \log(P \lfloor n^{0.3} \rceil)}$, where P is the number of characteristics, and n is the number of stocks in each rolling block. For the linear specification, each characteristic only has one basis, which is itself. In terms of hypothesis test, we use the same logic as in the semi-parametric settings, and we set $\eta_n = \sqrt{3 \log(P)}$.

In all the estimation above, we assume we know the real number of factors, which is three. We will discuss the situation when the number of factors is unknown in the next subsection. Mean Squared Error (MSE) is also reported to measure the fitness of the model Equation 7.1.

From Table 1, under different noise levels, namely $\sigma^2 = 1$ and $\sigma^2 = 4$, the semi-parametric model outperforms the linear model in the following aspects:

1 The fitness of the semi-parametric model is much better than the linear model, which can be illustrated from MSE.

2 The semi-parametric model can enhance the power of S_1 by non-zero S_0, which can not only select the correct mispricing characteristics but also avoid size distortions. As for the linear model, it is influenced by the correlated characteristics. Therefore, during certain periods we even obtain the non-invertible characteristic matrix. The linear model can also select the relevant covariates with decent probability, but it suffers from serious size distortions. In contrast, our semi-parametric model with orthogonal bases can mitigate this problem to a great extent.

3 Because S_1 can be very small and even negative, especially when the noise σ_i is strong, the additional component S_0 is necessary to strengthen the power of S_1 and select the relevant characteristics that can explain the mispricing function.

Table 1: Simulation Results 1 Part1

		$\sigma^2 = 1$												$\sigma^2 = 4$											
		Linear Model						Semi-parametric Model						Linear Model						Semi-parametric Model					
Window	n	S	S_0	S_1	MSE	Selected %	Distortion%	S	S_0	S_1	MSE	Selected %	Distortion%	S	S_0	S_1	MSE	Selected %	Distortion%	S	S_0	S_1	MSE	Selected %	Distortion%
1	468	24.9	11.5	13.4	6.4	100%	100%	-0.5	6.2	-5.7	6	81.2%	0%	14.2	10.8	3.4	8.6	100%	87.4%	-8.2	0	-8.2	8.1	0%	0%
2	894	32.8	11.6	21.2	2	100%	100%	3.4	8	-4.6	1.6	99.9%	0%	11.4	5.8	5.6	4.3	100%	2.1%	-8.5	0	-8.5	3.7	0%	0%
3	1108	34.4	5.7	28.7	11.9	100%	0%	8.6	9	-0.4	11.5	100%	0%	17.1	5.7	11.4	14.1	100%	0%	-7	0.7	-7.7	13.7	7.3%	0%
4	1199	-0.57	0	-0.57	10.2	0%	0%	9.2	9.1	0.1	9.5	96.8%	4.3%	-1.4	0	-1.4	12.5	0%	0%	-6.1		0.06	-6.2	7%	0%
5	1333	92	19.6	72.4	2.31	100%	100%	10.6	9	1.6	2	100%	0%	28.2	6.1	22	4.5	100%	6.5%	0.2	7.4	-7.2	4.1	82.8%	0%
6	1409	90	28.5	61.5	16	100%	100%	28.6	12.6	15.9	15.8	100%	28%	45.3	16.1	29.2	18.4	100%	73.4%	16.3	10.9	5.4	17.5	68.4%	35.9%
7	1466	78.4	10.6	67.8	6.4	100%	74.2%	19.5	9	10.5	6.2	100%	0%	34.8	5.7	29.1	8.6	100%	0.02%	4.3	9	-4.7	8.4	99.9%	0%
8	1560	133	16.8	116.2	3.3	100%	100%	20.3	10	10.3	3.2	100%	0%	45.2	6.1	39.1	5.5	100%	6.9%	4.2	10	-5.8	5.4	100%	0%
9	1494	117.7	13.6	104.1	3.6	100%	100%	23.1	9	14.1	3.5	100%	0%	44.1	7.6	36.5	5.8	100%	32.4%	6	9	-3	5.6	100%	0.01%
10	1292	90.7	11.5	79.2	3.7	100%	100%	16.2	9	7.2	3.6	100%	0%	39.5	9.3	30.2	5.9	100%	61.1%	3.6	8.9	-5.3	5.7	99.7%	0%
11	1393	84.7	10.6	74.1	6.1	100%	85.1%	20.7	9.1	11.6	5.8	100%	1.1%	37.1	6.5	30.6	8.3	100%	12.9%	8.9	8.9	0	7.8	98.1%	1.3%
12	1340	83.5	28	55.5	2.38	100%	100%	10.6	9	1.6	2	100%	0%	26	6.2	19.8	4.6	100%	7.1%	-1.8	5.7	-7.5	4.1	63.7	0%
13	1285	113.8	16	97.8	1.73	100%	100%	10.6	9	1.6	1.6	100%	0%	34.5	6.6	27.9	4	100%	15.3%	-2.4	5.1	-7.5	3.7	57.1%	0%
14	1181	88.5	12.8	75.7	4.7	100%	100%	15.8	9	6.8	4.5	100%	0%	31.2	5.9	25.3	6.9	100%	2.3%	3.7	9	-5.3	6.6	100%	0%
15	1110	45.7	7.5	38.1	8.9	100%	30.4%	11.5	9	2.5	8.7	100%	0%	23.9	5.8	18.1	11.1	100%	0.6%	-2	4.8	-6.8	10.8	0.54%	0%
16	1044	20.5	5.7	14.8	18.4	100%	0%	9.9	9	0.9	17.9	100%	0%	14.6	5.7	8.9	20.6	100%	0%	1.2	6.1	-4.9	20	68.1%	0.2%
17	1125	59.4	11.5	47.9	9.2	100%	100%	13.2	9	4.2	9	100%	0%	27.2	6.2	21	11.5	100%	8.4%	2.6	8.8	-6.2	11	97.9%	0%
18	2192	NA	NA	NA	NA	NA	NA	23.2	11	12.2	4.3	100%	0%	NA	NA	NA	NA	NA	NA	6.7	11	-4.3	6.4	100%	0%
19	2236	56.1	11.5	44.6	5.8	100%	100%	17.8	11	6.8	5.2	100%	0%	28.3	6.3	22	8	100%	20.3%	4.3	11	-6.7	7.4	100%	0%
20	2273	43.3	5.7	37.6	3.8	100%	0%	22.4	11	11.4	3.2	100%	0%	22.4	5.7	16.7	6.1	100%	0%	5	10.2	-5.2	5.4	92.6%	0%
21	2235	59.8	11.8	48	2.7	100%	100%	20.2	11	9.2	2	100%	0%	25	7.3	17.7	4.9	100%	28.2%	4.6	11	-6.4	4.2	100%	0%
22	2270	40.2	11.5	28.7	2.78	100%	99.5%	17.2	11.6	5.6	2.1	100%	0%	17.1	5.9	11.2	5	100%	3.5%	-6	0.1	-6.1	4.2	1.1%	0%
23	2405	41.4	8.9	32.5	4.1	100%	54.2%	16.3	11	5.3	3.3	100%	0%	18.7	5.8	12.9	6.3	100%	7.1%	-3.3	3	-6.3	5.5	27.3%	0%
24	2376	19	9.7	9.3	1.8	100%	69.9%	23.1	11	12.1	1	100%	0%	7.5	5.7	1.8	4	100%	0%	5.6	11	-5.4	3.2	100%	0%
25	2323	15.9	9.5	6.4	3.5	66.7%	98.6%	20.6	11	9.6	2.7	100%	0%	1.1	0	1.1	5.8	0%	0%	5.3	11	-5.7	4.9	100%	0%
26	2344	NA	NA	NA	NA	NA	NA	24.9	12.9	12	3.3	100%	17.1%	NA	NA	NA	NA	NA	NA	6.5	11	-4.5	5.4	100%	0%
27	2434	NA	NA	NA	NA	NA	NA	27.3	11	16.3	1.2	100%	0%	NA	NA	NA	NA	NA	NA	6.9	11	-4.1	3.4	100%	0%
28	2548	0.9	0	0.9	4.2	0%	0%	26.2	11	15.2	3.3	100%	0%	-1.3	0	-1.3	6.5	0%	0%	6.9	11	-4.1	5.5	100%	0%
29	2741	10.3	5.7	4.5	4.2	100%	0%	58.2	11.1	47.1	3.4	100%	1.3%	6.6	5.7	0.9	6.4	100%	0%	17.6	11	6.6	5.5	100%	0%
30	2928	5.6	4.6	1	7.1	80.4%	0%	59.2	11.8	47.4	6.3	100%	7.8%	-0.4	0.1	-0.5	9.3	2.5%	0%	18.8	11	7.8	8.5	100%	0.3%
31	2894	13.4	5.7	7.7	6.4	100%	0%	61	13.4	47.6	5.7	100%	21.6%	8.1	5.7	2.3	8.6	100%	0%	17.7	11	6.7	7.8	100%	0.2%
32	2905	23.1	11.5	11.6	5.9	100%	100%	33.2	11.3	21.9	5.2	100%	3%	12.9	8.5	4.4	8.1	100%	48.2%	9.8	11	-1.2	7.4	100%	0%
33	2804	9.8	5.7	4.1	9.6	100%	0%	42.7	18.5	24.2	8.9	100%	68.5%	7.3	5.7	1.6	11.9	100%	0%	9.7	11	-1.3	11.2	100%	0%
34	2570	6.9	5.7	1.2	22	99.7%	0%	37.3	12.2	25.1	21.2	100%	10.4%	2	1.9	0.1	24	34.4%	0%	12.7	11	1.7	23.3	100%	0.2%
35	2516	8.3	5.7	2.6	7.9	100%	0%	41.3	11	30.3	7.2	100%	0.4%	5.1	5.02	0.08	10.1	87.3%	0%	12.9	11	1.9	9.4	100%	0%
36	2491	10.7	5.7	4.9	2.1	100%	0%	41.3	11	30.3	1.4	100%	0.4%	0.5	0.25	0.25	4.4	4.5%	0%	12.4	11	1.4	3.6	100%	0%
37	2402	14.1	5.7	8.4	5.6	100%	0%	26.5	11.2	15.3	4.9	100%	2.2%	8.8	5.7	3.1	7.9	100%	0%	7.9	11	-3.1	7.1	100%	0%
38	2326	19.7	9.6	10.1	3	100%	66.8%	28.9	11.3	17.6	2.3	100%	2.1%	8.1	5.8	2.3	5.3	100%	0.3%	8.7	11	-2.3	4.4	99.9%	0.1%
39	2241	17	5.7	16.1	2.9	100%	0.2%	11	11	0	1.7	100%	0%	9.1	5.8	2.3	5.3	100%	0.3%	-7.5	0.1	-7.6	4	1.1%	0%
40	2178	21.8	5.7	16.1	2.9	100%	0%	9.5	11	-1.5	2.2	100%	0.3%	12.2	5.7	6.5	5.2	100%	0%	-8.1	0	-8.1	4.4	0%	0%

Table 2: Simulation Results 1 Part2

		$\sigma^2 = 1$												$\sigma^2 = 4$											
		Linear Model						Semi-parametric Model						Linear Model						Semi-parametric Model					
Window	n	S	S_0	S_1	MSE	Selected %	Distortion%	S	S_0	S_1	MSE	Selected %	Distortion%	S	S_0	S_1	MSE	Selected %	Distortion%	S	S_0	S_1	MSE	Selected %	Distortion%
41	2113	24.1	6.1	18	4.7	100%	7.5%	7.8	10	-2.2	3.9	100%	0%	13.9	5.7	8.2	6.9	100%	0%	-8.1	0	-8.1	6.1	0%	0%
42	2023	18.4	5.7	12.7	6.8	100%	0%	11.3	10	1.3	6	100%	0%	10.8	5.8	5.1	9	100%	0%	-7.1	0.3	7.4	8.2	2.7%	0%
43	2007	18.8	5.7	13.1	4.9	100%	0%	9.1	10	-0.9	4.1	100%	0%	10.5	5.7	4.8	7.1	100%	0%	-8.3	0	-8.3	6.3	0%	0%
44	1924	16.6	5.8	10.8	8.18	100%	0.2%	13.6	10.8	2.8	7.5	100%	8%	11.2	5.8	5.4	10.4	100%	0.3%	-3.5	2.7	-6.2	9.7	26.3%	0.2%
45	1990	27.5	5.7	21.8	2.1	100%	0%	8.1	10	-1.9	1.4	100%	0%	13.3	5.7	7.5	4.4	100%	0%	-8	0	-8	3.6	0%	0%
46	1937	20.3	5.8	14.5	5.4	100%	0.9%	19.7	11.8	7.9	4.7	100%	18%	12.6	5.9	6.7	7.6	100%	3%	8	11.2	-3.2	6.8	100%	12.3%
47	1909	13.2	5.7	7.5	5.2	100%	0%	14.2	10.4	3.8	4.5	100%	3.5%	8.8	5.7	3.1	7.4	100%	0%	2.7	8.4	-5.7	6.7	84.9%	0%
48	1872	21.8	5.7	16.1	2.7	100%	0%	11.4	10	1.4	2	100%	0%	11.1	5.8	5.3	4.9	100%	0%	-6.8	0.6	-7.4	4.2	5.7%	0%
49	1841	16.3	5.7	10.5	2.1	100%	0%	8.7	10	-1.3	1.4	100%	0.1%	8.1	5.7	2.4	4.4	100%	0%	-8.4	0	-8.4	3.6	0%	0%
50	1826	11	5.7	5.3	4.3	100%	0%	12.6	10.6	2	3.5	100%	3.5%	6.5	5.7	0.8	6.6	99.7%	0.3%	-6.9	0	-6.9	5.7	0%	0%

This table documents results under the characteristics-based beta and alpha of Fama-Frech 3 factors model. To mimic the empirical study, we simulated 50 12-month rolling windows, and each window is repeated for 1000 times. Each column summarises the mean value of 1000 estimations and test results. S_1 is the conventional Wald test while S_0 is the power-strengthened component. This table also compares the performance of both semi-parametric and linear models under different noise levels, $\sigma^2 = 1$ and $\sigma^2 = 4$. NA results are causes by non-invertible characteristic matrices. "Selected" means the percentage of selecting the relevant characteristic in the mispricing function in 1000 experiments. Similarly, "distortion" represents the percentage of wrongly selecting irrelevant characteristics in 1000 repetitions.

7.3 Robustness Under Stronger Noise

In Table 1, we set two different noise levels of random shocks, namely $\sigma^2 = 1$ and $\sigma^2 = 4$. Although $\sigma^2 = 1$ is closer to the empirical data, we conduct this comparison to show the robustness of our methods. When the noise level becomes three times bigger, the accuracy of power enhanced tests gets much lower for certain windows. However, there are no size distortions under comparatively high noise level recalling that all the components of our simulation model are rescaled to have unit variance. Another fact is that the stronger noise does deteriorate the power of conventional Wald tests, leading to an even smaller value of S_1, which can be mitigated through adding S_0.

Therefore, we conclude that our methods are robust to a higher noise level regarding no size distortions. However, the accuracy of selecting relevant components and the role of enhancing the power of hypothesis tests will be influenced negatively.

7.4 Number of Factors

In the empirical study, the number of factors is unknown. Therefore, in this subsection we will study whether our methodology is robust to a various numbers of factors considered.

We simulate according to another data generation process:

$$y_{it} = h(X_i) + \sum_{j=1}^{5} g_j(X_j) f_{jt} + \epsilon_{it}, \tag{7}$$

similarly, y_{it} is the generated stock return; $h(X_i)$ is the mispricing function consist of a non-linear characteristic function of X_i, to mimic the sparse structure of the mispricing function; $g_j(\boldsymbol{X}_j)$ is the j^{th} characteristics-based factor loading, which has an additive semi-parametric structure; X_j is a subset consisting of four characteristics; f_{jt} is the j Fama-French 5-factor returns at time t; ϵ_{it} is the idiosyncratic shock, generated from $N(0, \sigma^2)$. Moreover, we generate characteristic functions as:

$$h(X_i) = \sin X_i,$$

$$g_j(\boldsymbol{X}_j) = X_{j1}^2 + (3X_{j2}^3 - 2X_{j2}^2) + (3X_{j3}^3 - 2X_{j3}) + X_{j4}^2,$$

where X_{ji} is a randomly picked characteristic without replacement from the data in empirical study with $j = 1, \ldots, 5$, $i = 1, \ldots, 4$. Furthermore, all $h(X_i)$ and $g_j(\boldsymbol{X}_j)$ are rescaled to have zero mean and unit variance.

Given the above data generation process, together with the data generation process in Section 6.1, we test the influence of over and under-estimated number of factors. We choose the number of factors to be either three or five, and compare the results in Table 3.

The first category column is the scenario of over estimating the number of factors. We simulate the data generation process using the Fama-French three factors model but estimate the number of factors to be five.

However, this does not cause any serious problems. For some rolling blocks, the probability of mistakenly selected irrelevant characteristics is slightly higher under over estimating the number of factors. Moreover, over estimating the number of factors can increase the model fitting marginally. Therefore, we conclude that over estimating the number of factors does not cause severe size distortion using our methods.

On the other hand, under estimating the number of factors can lead to misleading test results. We can conclude this from the last column where we estimate the number of factors to be three in a five-factor model. Compared with the correct specified model, under estimating causes not only higher MSE, but also higher distortions, which means it is more likely to select irrelevant characteristics. Therefore, in the empirical study we prefer the five-factor model rather than the three-factor model.

Table 3: Simulation Results 2 Part1

		Number of factors $J = 3$												Number of factors $J = 5$											
		Number of estimated factors $\hat{J} = 5$						Number of estimated factors $\hat{J} = 3$						Number of estimated factors $\hat{J} = 5$						Number of estimated factors $\hat{J} = 3$					
Window	n	S	S_0	S_1	MSE	Selected %	Distortion%	S	S_0	S_1	MSE	Selected %	Distortion%	S	S_0	S_1	MSE	Selected %	Distortion%	S	S_0	S_1	MSE	Selected %	Distortion%
1	468	2.6	7	-4.4	5.9	99.6%	0%	-0.5	6.2	-5.7	6	81.2%	0%	4.5	6.9	-2.4	6	97.4%	1.3%	-8.5	0	-8.5	6.9	0%	0%
2	894	6	8	-2	1.5	99.9%	0%	3.4	8	-4.6	1.6	99.9%	0%	5.5	8	-2.5	2.3	100%	0%	-6.5	1	-7.5	3	12.9%	0%
3	1108	12.8	9	3.8	11.4	100%	0.1%	8.6	9	-0.4	11.5	100%	0%	14.3	9	5.3	13.6	100%	0.5%	5.3	9	-3.7	14.1	100%	0.1%
4	1199	13.5	10.3	3.2	9.4	99.8%	0%	9.2	9.1	0.1	9.5	96.8%	4.3%	16.6	11.7	4.9	9.8	99.5%	25.6%	2.3	3	-0.7	10.1	23.5%	10%
5	1333	15.4	9	6.4	1.8	100%	0.1%	10.6	9	1.6	2	100%	0%	16.7	9	7.6	2.7	100%	0%	4.5	9	-4.5	3.4	100 %	0%
6	1409	41.6	17.5	24.1	15.6	100%	51.3%	28.6	12.6	15.9	15.8	100%	28%	58.7	28.3	30.4	13.5	100%	90%	106.1	29.9	76.2	13.2	100%	100%
7	1466	26.8	9	17.8	6.1	100%	0.01%	19.5	9	10.5	6.2	100%	0%	26.3	9	17.3	9.2	100%	0.3%	3.5	9	-5.5	11.7	100%	0%
8	1560	27.6	10	17.6	3	100%	0%	20.3	10	10.3	3.2	100%	0%	30.4	10	20.4	5	100%	0.5%	26.7	24	2.7	6.7	100%	100%
9	1494	31.7	9.1	22.6	3.3	100%	0.7%	23.1	9	14.1	3.5	100%	0%	32.1	9.2	22.9	4.4	100%	1.4%	29.1	18	11.1	4.6	100%	100%
10	1292	22.5	9	13.5	3.4	100%	0.1%	16.2	9	7.2	3.6	100%	0%	26.3	10	16.3	4.3	100%	11.3%	46.7	18	28.7	4.4	100%	100%
11	1393	27.8	9.4	18.4	5.7	100%	4%	20.7	9.1	11.6	5.8	100%	1.1%	30	10.7	19.3	5.6	100%	17.2%	49	29.1	19.9	5.8	100%	100%
12	1340	15.2	9	6.2	1.8	100%	0%	10.6	9	1.6	2	100%	0%	15.2	9	6.2	1.8	100%	0%	4	9	-5	2.7	100%	0%
13	1285	15.4	9	6.4	1.4	100%	0.2%	10.6	9	1.6	1.6	100%	0%	15.1	9	6.1	1.4	100%	0.1%	3.5	9	-4.5	2.5	100%	0%
14	1181	21.9	9	12.9	4.4	100%	0.2%	15.8	9	6.8	4.5	100%	0%	21.4	9	12.4	4.7	100%	0.2%	4.5	9	-4.5	6	100%	0%
15	1110	16.4	9	7.4	8.5	100%	0%	11.5	9	2.5	8.7	100%	0%	17.1	9	8.1	9.8	100%	0.1%	5.3	9	-3.7	10.2	100%	0%
16	1044	13.3	9.1	4.3	17.8	100%	0.8 %	9.9	9	0.9	17.9	100%	0%	14.9	9.2	5.7	17.8	100%	2.1%	40	22.4	17.6	16.8	100%	100%
17	1125	18.7	9	9.7	8.8	100%	0.1%	13.2	9	4.2	9	100%	0%	24.1	9.7	14.4	10.7	100%	7.1%	101.4	27	74.4	10.3	100%	100%
18	2192	31.8	11	20.8	4.1	100%	0.2%	23.2	11	12.2	4.3	100%	0%	69.8	28.6	41.2	5.4	100%	77.6%	563.8	33	530.8	3.6	100%	100%
19	2236	24.4	11	13.4	5.1	100%	0%	17.8	11	6.8	5.2	100%	0%	25.1	11	14.1	5.4	100%	0%	10.2	11	-0.8	6.1	100%	0%
20	2273	29.4	11	18.4	3	100%	0.4%	22.4	11	11.4	3.2	100%	0%	30.3	11.1	19.2	4.2	100%	0.7%	61.1	33	28.1	5.3	100%	100%
21	2235	27.5	11	16.5	1.8	100%	0%	20.2	11	9.2	2	100%	0%	29	11	18	2.2	100%	0.3%	5.9	11	-5.1	3.3	100%	0%
22	2270	24.9	13.7	11.2	1.9	100%	23.9%	17.2	11.6	5.6	2.1	100%	0%	43.2	20.4	22.8	2.3	100%	56.7%	41.6	22.1	19.5	2.1	100%	100%
23	2405	22.5	11	11.5	3.2	100%	0.1%	16.3	11	5.3	3.3	100%	0%	21.7	11	10.7	3.3	100%	0%	10.9	11.9	-1	4.3	100%	7.8%
24	2376	30.6	11	19.6	0.8	100%	0.1%	23.1	11	12.1	1	100%	0%	30.3	11	19.3	1.2	100%	0%	20.4	21.4	-1	2.7	100%	94.8%
25	2323	27.2	11.1	16.1	2.5	100%	0.4%	20.6	11	9.6	2.7	100%	0%	26.8	11	15.8	2.8	100%	0%	8.5	11	-2.5	3.9	100%	0%
26	2344	36.4	16.7	19.7	3.1	100%	51.3%	24.9	12.9	12	3.3	100%	17.1%	36.1	17	19.1	3.2	100%	54%	47.5	23.3	24.2	4.3	100%	100%
27	2434	36.3	11.1	25.2	1	100%	0.9%	27.3	11	16.3	1.2	100%	0%	38.3	11.3	27	1.3	100%	2.6%	89.5	33	56.5	1.7	100%	100%
28	2548	34.5	11	23.5	3.2	100%	0.1%	26.2	11	15.2	3.3	100%	0%	34.8	11	23.8	3.3	0%	0.2%	50.3	22	28.3	4	100%	100%
29	2741	73	12.3	60.7	3.2	100%	10.9%	58.2	11.1	47.1	3.4	100%	1.3%	79.4	15.4	64	3.5	100%	36.8%	439.7	62.7	377	3.6	100%	100%
30	2928	73.9	13.8	60.1	6.1	24.1%	0%	59.2	11.8	47.4	6.3	100%	7.8%	84.6	18.7	65.9	7.4	100%	52.2%	94	32.6	61.4	7.2	100%	100%
31	2894	77.3	16.3	61	5.5	100%	45.4%	61	13.4	47.6	5.7	100%	21.6%	77.2	16.3	60.9	5.5	100%	45.9%	28.6	11	17.6	6.5	100%	0%
32	2905	42.4	12.9	29.5	5	100%	16%	33.2	11.3	21.9	5.2	100%	3%	41.7	12.8	28.9	6.1	100%	15.7%	8.7	11	-2.3	9.4	100%	0%
33	2804	53.8	20.5	33.3	8.8	100%	86.8%	42.7	18.5	24.2	8.9	100%	68.5%	54.1	20.4	33.6	10.1	100%	85.6%	35.5	22	13.5	12.3	100%	100%
34	2570	47.6	14.2	33.4	21.1	27.2%	0%	37.3	12.2	25.1	21.2	100%	10.4%	49.4	14.5	34.9	41.2	100%	28.9%	53.8	22	31.8	38.8	100%	100%
35	2516	50.9	11.3	39.6	7	100%	2.9%	41.3	11	30.3	7.2	100%	0.4%	38.4	11	27.4	18.4	100%	0.4%	51.2	33	18.2	20.8	100%	100%
36	2491	51.3	11.8	39.5	1.3	100%	6.8%	41.3	11	30.3	1.4	100%	0.4%	50.5	11.3	39.2	1.6	100%	3%	15.9	11	4.9	3.3	100%	0%
37	2402	34.4	12.2	22.2	4.7	100%	10.5%	26.5	11.2	15.3	4.9	100%	2.2%	37.4	14.2	23.2	5.1	100%	29.2%	68.8	22	46.8	6.1	100%	0%
38	2326	37.4	12.3	25.1	2.1	100%	10.3%	28.9	11.3	17.6	2.3	100%	2.1%	37.2	12.2	25	2.8	100%	9.4%	44.6	22	22.6	3.4	100%	0%
39	2241	14.8	11	3.8	1.6	100%	0.1%	11	11	0	1.7	100%	0%	14.9	11	3.9	1.7	100%	0%	23.4	22	1.4	2.4	100%	100%
40	2178	13.1	11.1	2	2	100%	1.1%	9.5	11	-1.5	2.2	100%	0.3%	12.9	11.2	1.8	2.2	100%	1.3%	20.4	13.1	7.3	3.4	13.3%	100%

Table 4: Simulation Results 2 Part2

		Number of factors $J = 3$												Number of factors $J = 5$											
		Number of estimated factors $\hat{J} = 5$						Number of estimated factors $\hat{J} = 3$						Number of estimated factors $\hat{J} = 5$						Number of estimated factors $\hat{J} = 3$					
Window	n	S	S_0	S_1	MSE	Selected %	Distortion%	S	S_0	S_1	MSE	Selected %	Distortion%	S	S_0	S_1	MSE	Selected %	Distortion%	S	S_0	S_1	MSE	Selected %	Distortion%
41	2113	11	10	1	3.8	100%	0.2%	7.8	10	-2.2	3.9	100%	0%	11.5	10	1.5	4.5	100%	0.2%	41.1	32.4	8.7	5.4	99.9%	100%
42	2023	15.2	10	5.2	5.9	100%	0%	11.3	10	1.3	6	100%	0%	15.7	10	5.7	6.4	100%	0%	-8	0	-8	9.2	0%	0%
43	2007	12.6	10	2.6	4	100%	0.5%	9.1	10	-0.9	4.1	100%	0%	13.4	10.2	3.2	4.7	100%	1.7%	-0.1	6.4	-6.5	5.6	64.4%	0%
44	1924	19.9	13.1	6.8	7.3	100%	30.7%	13.6	10.8	2.8	7.5	100%	8%	19.5	12.9	6.6	7.5	100%	28.9%	20	20	0	8.3	100%	100%
45	1990	11.4	10	1.4	1.2	100%	0.1%	8.1	10	-1.9	1.4	100%	0%	20.7	14.6	6	1.8	100%	45.2%	116	20	96	1.7	100%	100%
46	1937	27.1	14	13.1	4.5	100%	37.7%	19.7	11.8	7.9	4.7	100%	18%	28.3	14.8	13.5	5.4	100%	45.8%	24.6	20	4.6	6.2	100%	100%
47	1909	19.5	11.7	7.8	4.4	100%	16.1%	14.2	10.4	3.8	4.5	100%	3.5%	24	14	10	4.4	100%	38.1%	51.7	35.2	16.5	5.4	100%	100%
48	1872	15.2	10	5.1	1.8	100%	0.2%	11.4	10	1.4	2	100%	0%	15	10	5	2.1	100%	0.1%	5	10	-5	2.9	100%	0%
49	1841	12.3	10.1	2.2	1.2	100%	1.1%	8.7	10	-1.3	1.4	100%	0.1%	11.8	10.1	1.7	4.4	100%	0.8%	-10	0	-10	4.4	0%	0%
50	1826	18.5	12.4	6.1	3.3	100%	15%	12.6	10.6	2	3.5	100%	3.5%	20.2	13.3	6.9	3.7	100%	19.3%	-3.9	0.4	-4.3	4.4	3.9%	0%

This table presents results under the characteristics-based beta and alpha of both Fama-French 3 and 5 factors model. To mimic the empirical study, we simulated 50 12-month rolling windows, and each window is repeated for 1000 times. Each column summarises the mean value of 1000 estimation and test results. We compare the results of both over and under estimating the number of factors, namely,$\hat{J} = 3$ and $\hat{J} = 5$. S_1 is the conventional Wald test while S_0 is the power-strengthened component. NA results are caused by non-invertible characteristic matrices. "Selected" means the percentage of selecting the relevant characteristic in mispricing functions in 1000 repetitions. Similarly, "distortion" represents the percentage of wrongly selecting irrelevant characteristics in 1000 experiments.

8 Empirical Study

8.1 Data

We use monthly stock returns from CRSP and firms' characteristics from Compustat, ranged from 1965 to 2017. We construct 33 characteristics following the methods of Freyberger et al. (2017). Details of these characteristics can be found in the appendix. We use characteristics from fiscal year $t-1$ to explain stock returns between July of year t to June of year $t+1$. After adjusting the dates from the balance sheet data, we merge two data sets from CRSP and Compustat. We require all of the firms included in our analysis to have at least three years of characteristics data in Compustat.

Data is modified with regards to the following aspects:

1 Delisting is quite common for CRSP data. We use the way of Hou et al. (2015) to correct the returns of delisting stocks for all the delisted assets before 2018. Detailed methods can be found in the appendix.

2 Projected-PCA works well, even under small T circumstances. Thus, we choose the width of our window to be 12 months. Another reason for the short window width is that we assume mispricing functions are time-invariant in each window. One of the limitations of Projected-PCA is that it can only be used for a balanced panel, which means the number of stocks will vary when we applied one-year rolling windows to obtain a short time balanced panel. Meanwhile, we take monthly updated characteristics' mean values of 12 months as fixed characteristic values in each window. We also use rolling window method to detect peer groups of arbitrage characteristics.

3 B-splines are based on each time-invariant characteristic among n firms which are not delisted in each window.

4 Rolling windows are moving at a 12-month step from Jul. 1967 to Jun. 2017. The first 24 months returns are not included as they do not have corresponding characteristics.

5 Excess returns are obtained by the difference between monthly stock returns and Fama-French risk-free monthly returns.

8.2 Estimation

We construct B-spline bases based on evenly distributed knots, and the degree of each basis is three. We choose $v = 0.3$, which means the number of bases for each characteristic in each window is $\lfloor n^{0.3} \rceil$, and n is the number of stocks. To get a relatively large balanced panel in each window, some characteristics with too many missing values are eliminated. Therefore, only 33 characteristics are left. Firms kept in balanced panels in our dataset range from 468 to 2928, which means that both n and $\hat{\mathbf{A}} \in \mathbb{R}^{PH_n}$ are diverging. Large n can satisfy asymptotic requirements.

These facts emphasize the necessity of introducing a power enhanced component into the hypothesis test. Before the next step, we use time-demeaning matrix $\mathbf{D_T}$ to demean excess return matrix in each window.

Next, we project the time-demeaned monthly excess return matrix $\tilde{\mathbf{Y}}$ to the B-spline space spanned by characteristics $\mathbf{\Phi}(\mathbf{X})$, and then we collect the fitted value $\hat{\mathbf{Y}}$. We apply Principle Component Analysis on $\frac{1}{n}\hat{\mathbf{Y}}^\intercal\hat{\mathbf{Y}}$, and attain the first five eigenvectors corresponding to the first five biggest eigenvalues as the estimates of unobservable factors $\mathbf{F}$. We choose the number of factors to be five according to simulation results.

Then, we estimate factor loading matrix by:

$$\hat{\mathbf{G}}(\mathbf{X}) = \tilde{\mathbf{Y}}\hat{\mathbf{F}}(\hat{\mathbf{F}}^\intercal\hat{\mathbf{F}})^{-1}.$$

Moveover, we use equality-constrained OLS to estimate the mispricing function. We project excess monthly return matrix on the characteristic space $\mathbf{\Phi}(\mathbf{X})$ that is orthogonal to factor loading matrix $\hat{\mathbf{G}}(\mathbf{X})$.

Another goal of this paper is to conduct a power enhanced test on the mispricing function. Therefore, our final step is to estimate covariance matrix $\mathbf{\Sigma}$ of $\hat{\mathbf{A}}$.

8.3 Power Enhanced Hypothesis Tests

In this section, we conduct a power enhanced test in each rolling block. Firstly, we set threshold value for each window, $\eta_n = H_n\sqrt{2\log(PH_n)}$, where H_n is the number of bases for each characteristic whereas P is the number of total characteristics in each window, with $P = 33$. η_n is data-driven critical value and it diverges as the number of firms increases. We use indicator function $\mathbf{I}(\sum_{h=1}^{H_n} |\hat{\alpha}_{ph}|/\hat{\sigma}_{ph} \geqslant \eta_n)$ with critical value $\eta_n = H_n\sqrt{2\log(PH_n)}$ to achieve three goals.

1 This indicator function select the most relevant characteristics that can explain the variation of the mispricing function. Results of last column in Table 5 are characteristics selected in $\hat{\mathcal{M}} = \{\mathbf{X_p} \in \hat{\mathcal{M}} : \sum_{h=1}^{H} |\hat{\alpha}_{ph}|/\hat{\sigma}_{ph} \geqslant \eta_n, \quad h = 1, 2, \ldots, H_n, \quad p = 1, 2, \ldots, P\}$.

2 It contributes to the test statistics S by adding a diverging power enhanced component S_0. As $T = 12$ is small in this empirical study, we assume the homoskedasticity of ϵ_{it}. We also specify a overshrunk covariance matrix by setting off-diagonal elements to be zeros.

3 It avoids size-distortion by the conservative critical value η_n.

The diagonal elements of $\hat{\mathbf{\Sigma}}$ are estimated variances of mispricing coefficients. These elements can be substituted into the indicator function $\mathbf{I}(|\hat{\alpha}_{ph}|/\hat{\sigma}_{ph} \geqslant \eta_n)$, where $\hat{\sigma}_{ph}$ is the ph^{th} diagonal element of $\hat{\mathbf{\Sigma}}$.

Finally, the new statistics S can be calculated as:

$$S = S_0 + S_1,$$

$$S_0 = H_n \sum_{p=1}^{P} \mathbf{I}(\sum_{h=1}^{H_n} |\hat{\alpha}_{ph}|/\hat{\sigma}_{ph} \geqslant \eta_n), \quad S_1 = \frac{\hat{\mathbf{A}}\hat{\mathbf{\Sigma}}^{-1}\hat{\mathbf{A}}^\intercal - PH_n}{\sqrt{2PH_n}}.$$

8.4 Test Results

This section presents the empirical results. Details can be found in Table 5, which list the results of 50 rolling windows from Jul.1967 to Jun.2017. Generally, the number of firms included in the 12-month rolling block is increasing period by period. The number of our characteristic B-spline bases is a function of the number of firms n in each block, which is $\lfloor n^{0.3} \rceil$. Therefore, the dimension of mispricing coefficient vector $\hat{\mathbf{A}} \in \mathcal{R}^{PH_n}$ is also diverging. This verifies the necessity of using power enhanced component S_0.

Recalling that $\mathbf{S}|\mathbf{H_0} \rightarrow^{\mathbf{d}} \mathbf{N}(\mathbf{0}, \mathbf{1})$, some of the test statistics S are big enough to reject the null hypothesis. However, for some testing windows, there are no strong signals showing the existence of characteristics-based mispricing functions after subtracting systematic effects. Moreover, most S_1 values are small and even negative, which may be caused by the sparsity structure of the mispricing function or/and the low power problems due to diverging dimension of mispricing coefficients.

The power enhanced component S_0 works well in the empirical study. It selects the most important explaining characteristics and strengthens the power of S_1, mitigating the low power problem.

Apart from contributing to the power of tests, the indicator function in the power enhanced component can also screen out the most relevant explanatory characteristics, which are concluded as "Characteristics Selected" in Table 5.

Some empirical findings are worth discussing. Although short-term cumulative returns like r_{2_1} are always selected, we cannot take this as evidence of arbitrage opportunities as we construct r_{2_1} as time-invariant monthly average of the last month returns. Higher average one month lagged returns imply higher monthly returns. However, this is not the case for long-term and mid-term cumulative returns like r_{12_2}, r_{12_7} and r_{6_2}, because these average returns of these variables contain a lot of information from another rolling window.

Apart from the cumulative returns, some other characteristics contribute to the arbitrage opportunities as well. PCM (Price to Cost Margin) appears twice. From Figure 2, we find that the PCM mispricing curve is nonlinear and generally decreasing as the value of PCM increases. ROA (Return-on-asset) also plays a role during 1988-1989. It behaves like a parabola with fluctuations near zero in Figure 3. As for Lev (ratio of long-term debt and debt in the current liabilities), it is decreasing for Lev<0 and increasing afterwards as in Figure 7. In Figure 8, IPM (pre-tax profit margin) function behaves like a "V" shape with the turning point zero during 2004-2005. DelGmSale (Difference in the percentage in gross margin and the percentage change in sales) experiences a bump at the zero during 2015-2016 in Figure 9. C2D curve behaves like "V" around the zero in 2016-2017, (see Figure 10). All characteristics in above figures are standardized as uniform distributed characteristics in the interval $[-100, 100]$. This is for presentation purpose only since most characteristics are unevenly distributed.

Table 5: Empirical Study Results

Time period	n	S	S_0	S_1	MSE	Characteristics Selected
Jul.1967-Jun.1968	468	-9.6	0	-9.6	0.005	NONE
Jul.1968-Jun.1969	951	-0.45	8	-8.45	0.004	r_{2_1}
Jul.1969-Jun.1970	1108	1.7	9	-7.3	0.005	r_{2_1}
Jul.1970-Jun.1971	1199	-8.7	0	-8.7	0.006	NONE
Jul.1971-Jun.1972	1333	-10	0	-10	0.004	NONE
Jul.1972-Jun.1973	1409	12.7	18	-5.3	0.005	r_{12_2},r_{6_2}
Jul.1973-Jun.1974	1466	2.1	9	-6.9	0.005	r_{2_1}
Jul.1974-Jun.1975	1560	-10.7	0	-10.7	0.01	NONE
Jul.1975-Jun.1976	1494	0.1	9	8.9	0.05	r_{2_1}
Jul.1976-Jun.1977	1292	0.1	9	-9	0.004	r_{2_1}
Jul.1977-Jun.1978	1393	-9.4	0	-9.4	0.005	NONE
Jul.1978-Jun.1979	1340	8.6	18	-9.4	0.005	r_{2_1},r_{12_7}
Jul.1979-Jun.1980	1285	1	9	-8	0.005	r_{2_1}
Jul.1980-Jun.1981	1181	9.7	18	-8.2	0.006	r_{12_7},r_{12_2}
Jul.1981-Jun.1982	1110	1.2	9	-7.8	0.01	r_{2_1}
Jul.1982-Jun.1983	1044	33.1	36	-3	0.01	r_{12_2},r_{12_7},r_{6_2},r_{2_1}
Jul.1983-Jun.1984	1125	-0.9	9	-9.9	0.006	r_{2_1}
Jul.1984-Jun.1985	2192	-0.2	11	-11.2	0.01	r_{2_1}
Jul.1985-Jun.1986	2236	13.1	22	-8.94	0.01	r_{12_7},r_{12_2}
Jul.1986-Jun.1987	2273	1.7	11	-9.3	0.01	PCM
Jul.1987-Jun.1988	2235	0.9	11	-10.1	0.01	r_{2_1}
Jul.1988-Jun.1989	2270	1.2	11	-9.8	0.01	ROA
Jul.1989-Jun.1990	2405	-0.1	11	-11.1	0.01	r_{2_1}
Jul.1990-Jun.1991	2376	1.1	11	-9.9	0.02	r_{2_1}
Jul.1991-Jun.1992	2323	2.1	11	-8.9	0.02	r_{2_1}
Jul.1992-Jun.1993	2344	12.2	22	-9.8	0.02	r_{12_7},r_{12_2}
Jul.1993-Jun.1994	2434	0.4	11	-10.6	0.01	r_{2_1}
Jul.1994-Jun.1995	2548	2.4	11	-8.6	0.01	r_{2_1}
Jul.1995-Jun.1996	2741	14.1	22	-7.9	0.02	BEME,r_{2_1}
Jul.1996-Jun.1997	2928	18.1	22	-3.9	0.01	BEME,r_{2_1}
Jul.1997-Jun.1998	2894	26.5	33	-6.5	0.02	r_{2_1},r_{12_7},r_{12_2}
Jul.1998-Jun.1999	2905	24.6	33	-8.4	0.02	AT,LME,r_{2_1}
Jul.1999-Jun.2000	2804	13.8	22	-8.2	0.03	r_{2_1},r_{12_7}
Jul.2000-Jun.2001	2570	37.7	44	-6.3	0.02	AT,LME, r_{2_1}, r_{6_2}
Jul.2001-Jun.2002	2516	1.3	11	-9.7	0.02	r_{2_1}
Jul.2002-Jun.2003	2491	15	22	-7	0.02	Lev, r_{2_1}
Jul.2003-Jun.2004	2402	3.9	11	-7.1	0.01	r_{2_1}
Jul.2004-Jun.2005	2326	1.8	11	-9.2	0.01	IPM
Jul.2005-Jun.2006	2241	2.5	11	-8.5	0.01	r_{2_1}
Jul.2006-Jun.2007	2178	1.5	11	-9.5	0.01	r_{2_1}
Jul.2007-Jun.2008	2113	12.6	20	-7.4	0.01	r_{12_2},r_{2_1}
Jul.2008-Jun.2009	2023	1.7	10	-8.3	0.02	r_{2_1}
Jul.2009-Jun.2010	2007	1	10	-9	0.01	r_{2_1}
Jul.2010-Jun.2011	1924	13.6	20	-6.4	0.01	r_{2_1}
Jul.2011-Jun.2012	1990	2.5	10	-7.5	0.01	r_{2_1}
Jul.2012-Jun.2013	1937	23.7	30	-6.3	0.01	r_{2_1},r_{12_7},r_{12_2}
Jul.2013-Jun.2014	1909	2.3	10	-7.7	0.01	r_{2_1}
Jul.2014-Jun.2015	1872	5.5	10	-4.5	0.01	r_{2_1}
Jul.2015-Jun.2016	1841	12.4	20	-7.6	0.01	DelGmSale,r_{2_1}
Jul.2016-Jun.2017	1826	26.1	30	-3.9	0.01	C2D,PCM,r_{12_7}

This table summaries the empirical results, where n represents the number of stocks in this rolling window.

Table 6: First layer 1986-1987 (clusterings of $\ddot{y}_{it}$)

Group number	Group centeroid	Group size
1	0.0059	435
2	0.1205	26
3	-0.0082	428
4	0.0399	189
5	0.0697	71
6	-0.1018	29
7	-0.0617	110
8	-0.0390	250
9	-0.0225	349
10	0.0208	386

Another finding is the persistence of some arbitrage characteristics. Arbitrage characteristics can be persistent for two years once appear, such as BEME (Ratio of the book value of equity and market value of equity) in Figure 4. Some persistent arbitrage characteristics even have similar shapes of mispricing functions in different rolling windows, such as AT (Total asset) in Figure 6 and LME (Total market capitalization of the previous month) in Figure 5.

8.5 Dynamic Peer Groups of Arbitrage Characteristics

In this section, we illustrate that there are distinguishable peer groups of the same arbitrage characteristics resulting in similar unsystematic returns. We apply the methods in section 5 and take two rolling windows, namely, Jul.1986- Jun.1987 and Jul.2004-Jun.2005 as demonstrative examples.

In the rolling window Jul.1986-Jun.1987, PCM is selected as only arbitrage characteristic that explains arbitrage returns. We reveal that similar characteristic-based arbitrage returns are determined by distinguishable groups of the characteristic PCM. We first divide arbitrage returns $\ddot{y}_{it}$ into different return groups. And then, we detect whether there are some clustering structures within groups of the highest and the lowest of characteristic-based arbitrage returns, respectively. As we have 2326 assets, for the visualization purpose, we set the threshold value of the K-means method to be relatively small to have as many as ten groups.

In Table 6, group 2 has the largest positive average return while group 6 has the worst. Next, we detect the clusterings of characteristic "PCM" within each group individually, which is the second layer in section 5.

In Table 7, there are two clusterings of PCM that provide the highest positive characteristic-based arbitrage returns. Group 2.2, which has an extreme negative PCM value but a high characteristic-based arbitrage return, is

Table 7: Second layer 1986-1987 (clusterings of characteristic PCM)

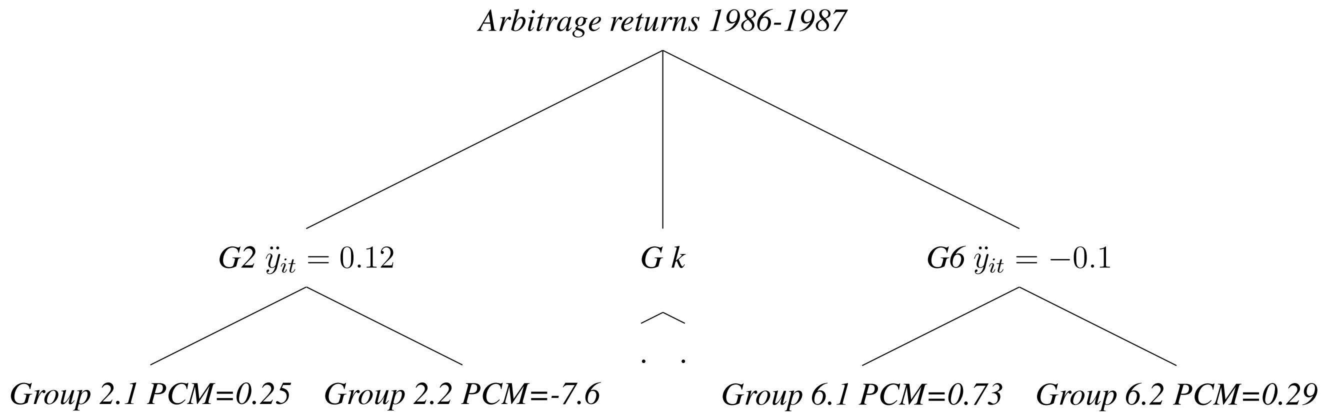
Group number	Centeroids of Arbitrage returns	Centeroids of PCM	Group size
2.1	0.1211	0.2452	25
2.2	0.1039	-7.630	1

Table 8: Second layer 1986-1987 (clusterings of characteristic PCM)

Group number	Centeroids of Arbitrage returns	Centeroids of PCM	Group size
6.1	-0.1085	0.728	9
6.2	-0.0989	0.288	20

an outlier. Members in group 2.1 with excellent arbitrage performance have positive and small PCM values.

Table 8 gives groups of PCM in group 6. Members of this group are divided into two clusterings. Group 6.1 has a relatively large PCM value, while group 6.2 has a smaller PCM, which is close to that in group 2.2 with the highest arbitrage return. This is an evident illustration of the nonlinear strucure of $h(\mathbf{X})$ in this window. The structure of characteristic-based arbitrage returns during Jul.1986- Jun.1987 is:



The classification can be found at Figure 11, where assets are labeled by their "PERMNO," and both axes are rescaled.

Another example is the characteristic-based arbitrage return $\ddot{y}_{it}$ during the year 2004-2005. Power enhanced test selects characteristic "IPM" as the only explanatory variable.

We apply the Hierarchical K-means method. The results of the first layer classification can be found in Table 9. There are ten groups in total according to the similarity of charateristic-based arbitrage returns. Next, we pick two groups with the highest and the lowest returns, respectively, to check clusters of "IPM" in these two groups.

Similarly, we show classification results in Table 10 and Table 11. Positive IPM values give higher characteristic-based arbitrage returns. On the contrary, when IPM is close to zero or negative, the characteristic-based arbitrage returns fall into the lowest group (group 8).

Table 9: First layer (clusterings of $\ddot{y}_{it}$)

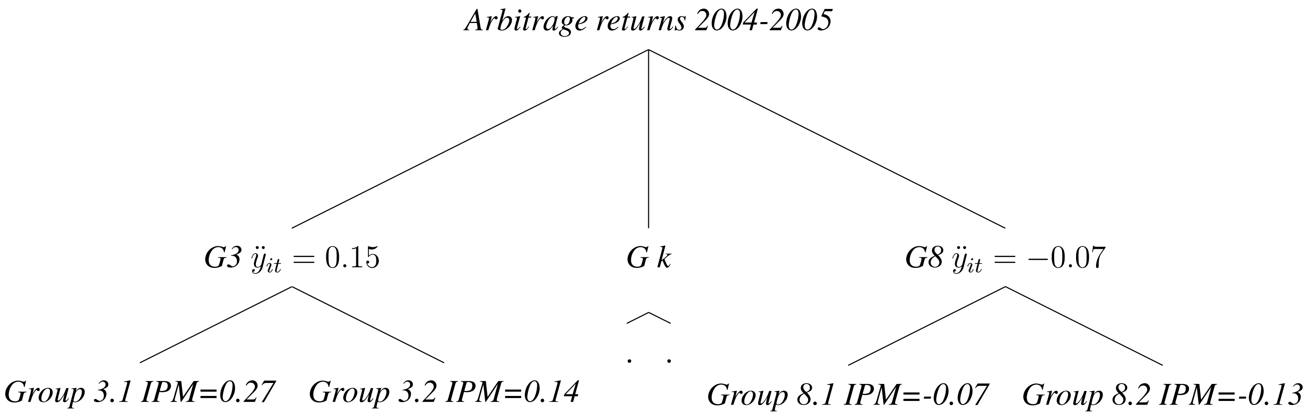
Group number	Group centeroid	Group size
1	0.0421	276
2	0.0059	459
3	0.1537	26
4	-0.024	367
5	0.0659	166
6	0.023	387
7	0.0999	120
8	-0.0758	67
9	-0.0437	244
10	-0.0082	436

Table 10: Second layer (clusterings of characteristic IPM)

Group number	Centeroids of Arbitrage returns	Centeroids of PCM	Group size
3.1	0.1681	0.266	5
3.2	0.1502	0.143	21

Table 11: Second layer (clusterings of characteristic IPM)

Group number	Centeroids of Arbitrage returns	Centeroids of PCM	Group size
8.1	-0.0713	-0.07	10
8.2	-0.1016	-0.134	57



The plots of and IPM can be found at Figure 12, where the axes are rescaled and assets are labeled by their "PERMNO" code with five digits.

Finally, it is obvious that peer groups of arbitrage characteristics are dynamic in two aspects. Firstly, the selected arbitrage characteristics are time-varying. Although some of the arbitrage characteristics can show up for more than one block once appear, no arbitrage characteristic can be substantially persistent. Secondly, as in Figure 4, the same arbitrage characteristic can have different function forms in various rolling windows. However, the patterns of some characteristics show strong persistence in different time periods, such as AT in Figure 6 and LME in Figure 5. In a word, under the flexible semiparametric setting, methods for contructing arbitrage portfolio in Kim et al. (2019) may be improvable, although the characteristic-based mispricing function is significant for certain time periods. The arbitrage portfolios can perform better by considering peer groups of arbitrage characteristics.

9 Conclusion

We proposed a semi-parametric characteristics-based factor model, with a focus on the existence and structure of the mispricing function. Both unknown characteristics-based factor loadings and the mispricing component are approximated by B-spline sieve. We also develops a power enhanced test to investigate whether there are mispricing components, orthogonal to the main systematic factors. This is necessary because when the B-spline coefficients of the mispricing functions are diverging, the conventional Wald test has very low power. Our proposed methods work well for both simulations and the US stock market. Empirically, we found distinct clusters of the same characteristics resulting in similar arbitrage returns. The mispricing function and selected arbitrage characteristics are time-varying. We conclude that the traditional way of developing arbitrage portfolios can be improved by considering peer groups of arbitrage characteristics

10 Appendix

10.1 Characteristic Description

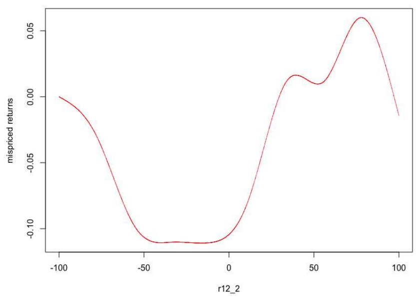
Table 12: Characteristic Details

Name	Description	Reference
A2ME	We define assets-market cap as total assets (AT) over market capitalization as of December t-1. Market capitalization is the product of shares outstanding (SHROUT) and price(PRC).	Bhandari (1988)
AT	Total assets (AT)	Gandhi and Lusting (2015)
ATO	Net sales over lagged net operating assets. Net operating assets are the difference between operating assets and operating liabilities. Operating assets are total assets (AT) minus cash and short-term investments (CHE), minus investment and other advances (IVAO). Operating liabilities are total assets (AT), minus debt in current liabilities(DLC),minus long-term debt (DLTT),minus minority interest (MIB), minus preferred stock (PSTK), minus common equity (CEQ).	Soliman(2008)
BEME	Ratio of book value of equity to market value of equity. Book equity is shareholder equity (SH) plus deferred taxes and investment tax credit (TXDITC), minus preferred stock (PS). SH is shareholder‘s equity (SEQ). If missing, SH is the sum of common equity (CEQ) and preferred stock (PS). If missing, SH is the difference between total assets (AT) and total liabilities (LT). Depending on availability, we use the redemption (item PSTKRV), liquidating (item PSTKL), or par value (item PSTK) for PS. The market value of equity is as of December t-1. The market value of equity is the product of shares outstanding (SHROUT) and price (PRC).	Rosenberg, Reid and Lanstein (1985) Davis, Fama, and French (2000)

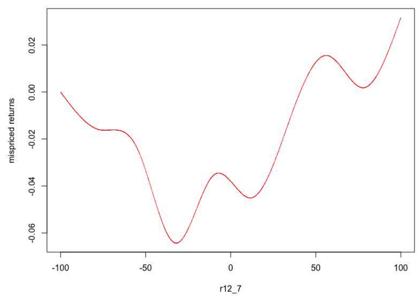
C	Ration of cash and short-term investments (CHE) to total assets (AT)	Palazzo
C2D	Cash flow to price is the ratio of income and extraoridinary items (IB) and depreciation and amortization (dp) to total liabilities (LT).	
CTO	We define caoital turnover as ratio of net sales (SALE) to lagged total assets (AT).	Haugen and Baker (1996)
Debt2P	Debt to price is the radio of long-term debt (DLTT) and debt in current liabilities (DLC) to the market capitalization as of December t-1 . Market capitalization is the product of shares outstanding (SHROUT) and price (PRC).	Litzenberger and Ramaswamy (1979)
Δceq	The percentage change in the book value of equity (CEQ).	Richardson et al. (2005)
$\Delta(\Delta Gm - Sales)$	The difference in the percentage change in gross margin and the percentage change in sales (SALE). We define gross margin as the difference in sales (SALE) and costs of goods sold (COGS).	Abarbanell and Bushee (1997)
$\Delta Shrout$	The definition of the percentage change in shares outstanding (SHROUT).	Pontiff and Woodgate (2008)
$\Delta PI2A$	We define the change in property, plants ,and equipment as changes in property,plants,and equipment (PPEGT) and inventory (INVT) over lagged total assets (TA).	Lyandres , Sun, and Zhang (2008)
DTO	We define turnover as ratio of daily volume (VOL) to shares outstanding (SHROUT) minus the daily market turnover and de-trend it by its 180 trading day median. We scale down the volume of NASDAQ securities by 38% after 1997 and by 50% before that to address the issue of double-counting of volume for NASDAQ securities.	Garfinkel (2009); Anderson and Dyl (2005)

E2P	We define earnings to price as the ratio of income before extraordinary items (IB) to the market capitalization as December t-1 Market capitalization is the product of share outstanding (SHROUT) and price (PRC).	Basu (1983)
EPS	We define earnings per share as the ratio of income before extraordinary items (IB) to share outstanding (SHROUT) as of December t-1	Basu (1997)
Investment	We define investment as the percentage year-on-year growth rate in total assets (AT).	Cooper, Gulen and Schill(2008)
IPM	We define pre-tax profit margin as ratio of pre-tax income (PI) to sales (SALE).	
Lev	leverage is the ratio of long-term debt (DLTT) and debt in the current liabilities (DLC) to the sum of long-term debt, debt in current liabilities, and stockholders' equity (SEQ)	Lewenllen (2015)
LME	Size is the total market capitalization of the previous month defined as price (PRC) times shares outstanding (SHROUT)	Fama and French (1992)
Turnover	Turnover is last month's volume (VOL) over shares outstanding (SHROUT).	Datar, Naik and Radcliffe (1998)
OL	Operating leverage is the sum of cost of goods sold (COGS) and selling, general, and administrative expenses (XSGA) over total assets.	Novy-Marx (2011)
PCM	The price-to-cost margin is the difference between net sales (SALE) and costs of goods sold (COGS) divided by net sales (SALE).	Gorodnichenko and Weber (2016) and D'Acunto, Liu, Pflucger and Wcber (2017)
PM	The profit margin is operating income after depreciation (OIADP) over sales (SALE)	Soliman (2008)
Q	Tobin's Q is total assets (AT), the market value of equity (SHROUT times PRC) minus cash and short-term investments (CEQ) minus deferred taxes (TXDB) scaled by total assets (AT).	
ROA	Return-on-assets is income before extraordinary items (IB) to lagged total assets (AT).	Balakrishnan, Bartov and Faurel (2010)

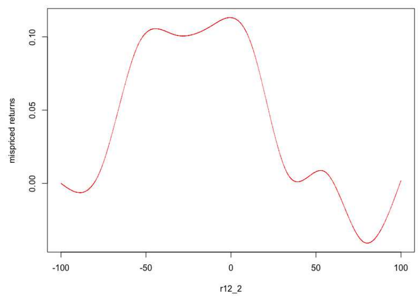
ROC	ROC is the ratio of market value of equity (ME) plus long-term debt (DLTT)minus total assets to Cash and Short-Term Investments (CHE).	Chandrashekar and Rao (2009)
ROE	Return-on-equity is income before extraordinary items (IB) to lagged book-value of equity.	in Haugen and Baker (1996)
r_{12-2}	We define momentum as cumulative return from 12 months before the return prediction to two months before.	Fama and French (1996)
r_{12-7}	We define intermediate momentum as cumulative return from 12 months before the return prediction to seven months before.	Novy-Marx (2012)
r_{6-2}	We definer_{6-2} as cumulative return from 6 months before the return prediction to two months before.	Jegadeesh and Titman (1993)
r_{2-1}	We define short-term reversal as lagged one-month return.	Jegadeesh(1990)
S2C	Sales-to-cash is the ratio of net sales (SALE) to Cash and Short-Term Investments (CHE).	following Ou and Penman (1989)
Sales-G	Sales growth is the percentage growth rate in annual sales (SALE).	Lakonishok, Shleifer , and Vishmy (1994)
SAT	We define asset turnover as the ratio of sales (SALE) to total assets (AT).	Soliman (2008)
SGA2S	SGA to sales is the ratio of selling ,general and administrative expenses (XSGA) to net sales (SALE).	



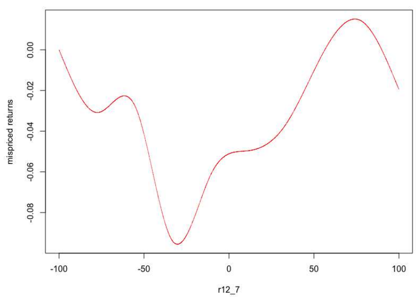
(a) r_{12-2} Curve 1972-1973



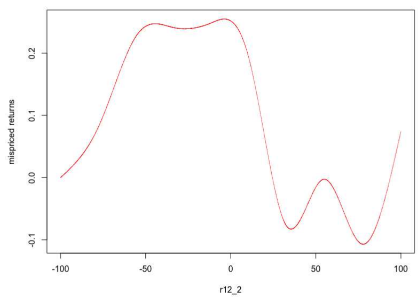
(b) r_{12-7} Curve 1978-1979



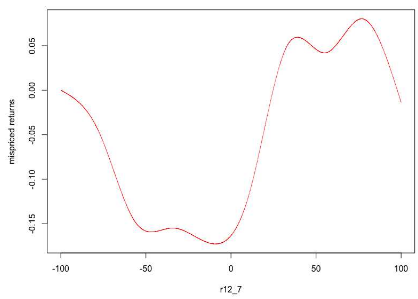
(c) r_{12-2} Curve 1980-1981



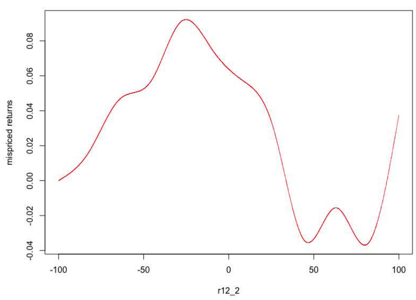
(d) r_{12-7} Curve 1985-1986



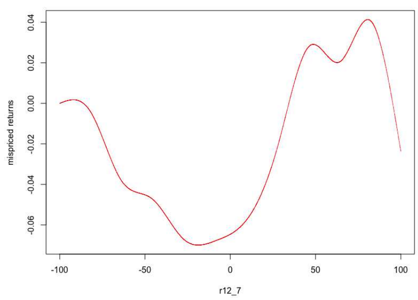
(e) r_{12-2} Curve 1982-1983



(f) r_{12-7} Curve 1982-1983

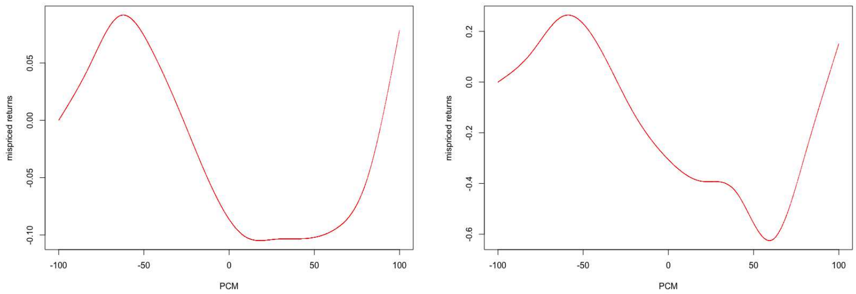


(g) r_{12-2} Curve 1985-1986



(h) r_{12-2} Curve 1985-1986

Figure 1: Mispricing Characteristic Curve of standardized r_{12-2} and r_{12-7}



(a) PCM Curve 1984-1985 (b) PCM Curve 2016-2017

Figure 2: Mispricing Characteristic Curve of standardized PCM

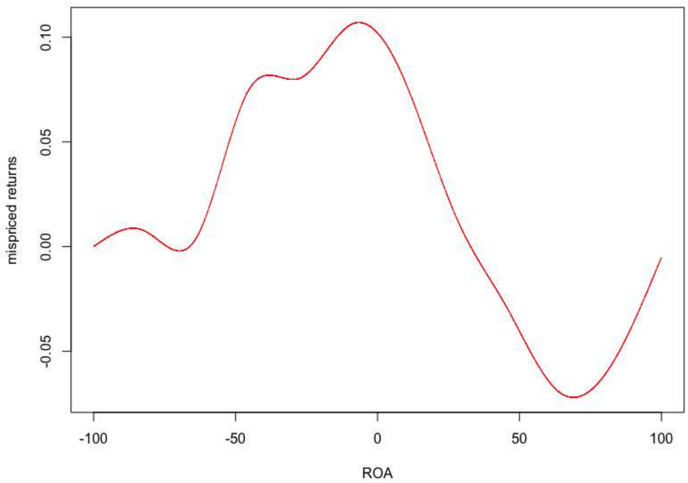
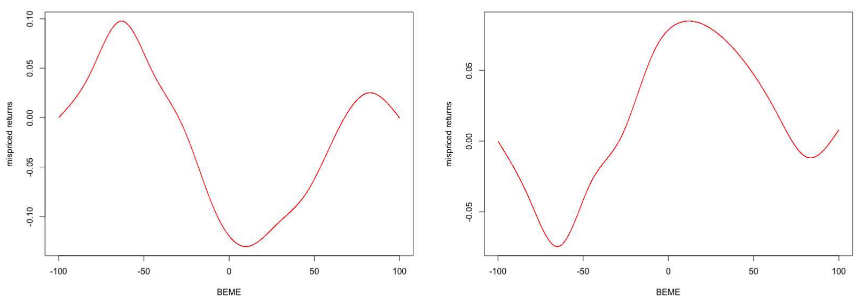
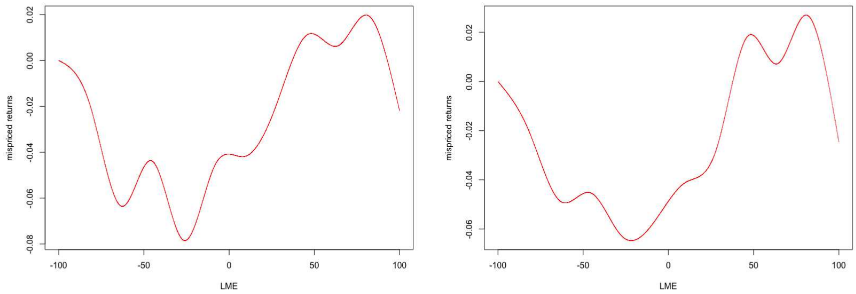


Figure 3: Mispricing Characteristic Curve of standardized ROA in 1988-1989



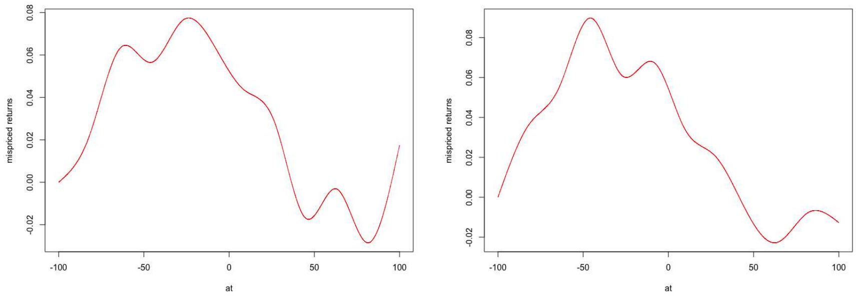
(a) BEME Curve 1995-1996 (b) BEME Curve 1996-1997

Figure 4: Mispricing Characteristic Curve of standardized BEME



(a) LME Curve 1998-1999 (b) LME Curve 2000-2001

Figure 5: Mispricing Characteristic Curve of standardized LME



(a) AT Curve 1998-1999 (b) AT Curve 2000-2001

Figure 6: Mispricing Characteristic Curve of standardized AT

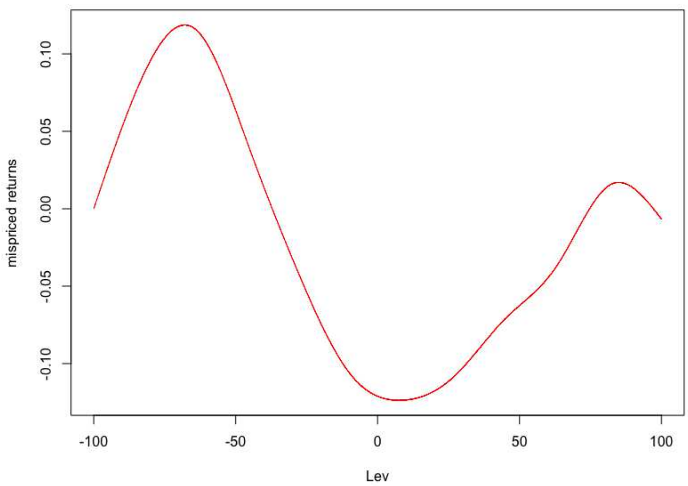


Figure 7: Mispricing Characteristic Curve of standardized LEV in 2002-2003

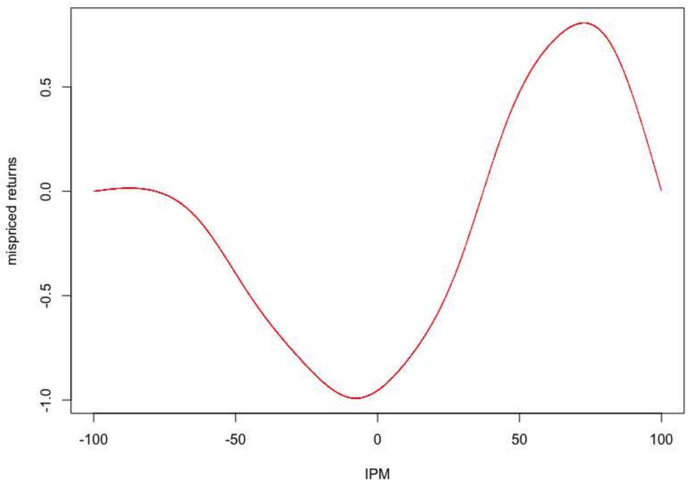


Figure 8: Mispricing Characteristic Curve of standardized IPM in 2004-2005

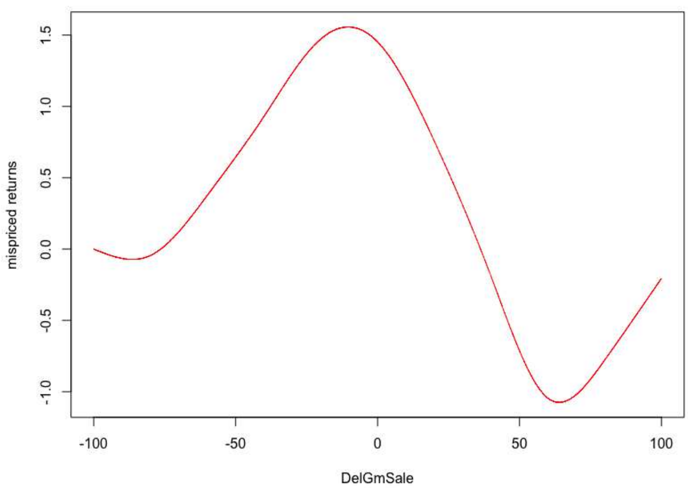


Figure 9: Mispricing Characteristic Curve of standardized DelGmSale in 2015-2016

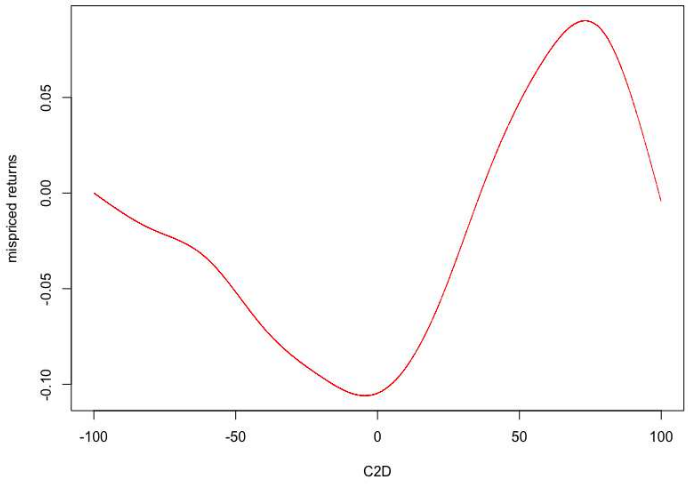
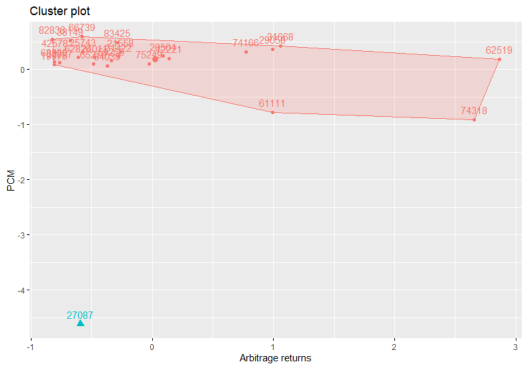
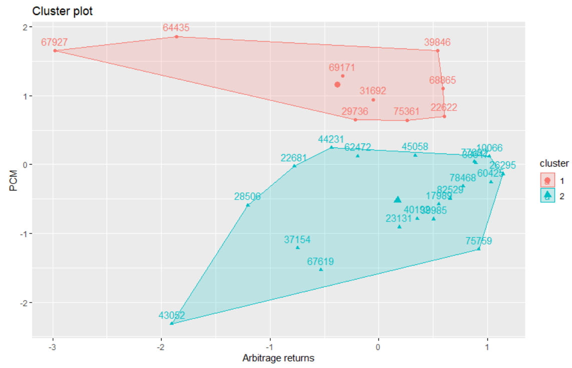


Figure 10: Mispricing Characteristic Curve of standardized C2D in 2016-2017

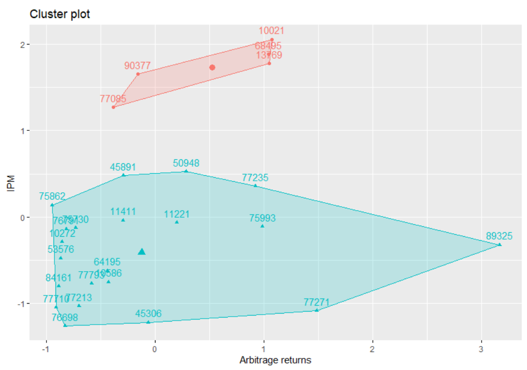


(a) Clustering of PCM with highest returns

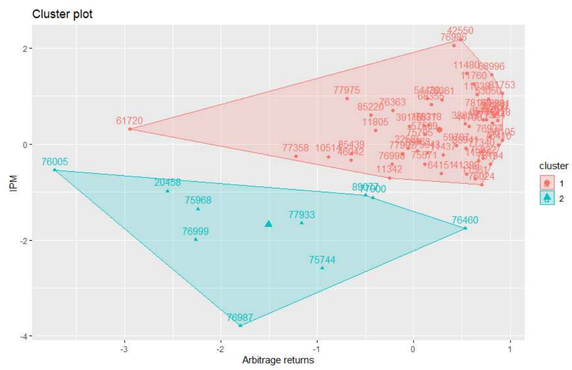


(b) Clustering of PCM with lowest returns

Figure 11: Clustering of PCM 1986-1987



(a) Clustering of IPM with highest returns



(b) Clustering of IPM with lowest returns

Figure 12: Clustering of IPM 2004-2005

10.2 Proofs

Through out the proofs, we have the number of observations $n \to \infty$, and time T is fixed.

Proof of Theorem 6.1 : In equation 5, we have

$$\mathbf{Y} = (\mathbf{\Phi}(\mathbf{X})\mathbf{A} + \mathbf{\Gamma} + \mathbf{R}^{\mu}(\mathbf{X}))\mathbf{1}_{\mathbf{T}}^{\intercal} + (\mathbf{\Phi}(\mathbf{X})\mathbf{B} + \mathbf{\Lambda} + \mathbf{R}^{\theta}(\mathbf{X}))\mathbf{F}^{\intercal} + \mathbf{U},$$

Multiply time-demeaned matrix $\mathbf{D_T}$ on both sides, where $\mathbf{D_T} = \mathbf{I_T} - \frac{1}{T}\mathbf{1}_{\mathbf{T}}^{\intercal}\mathbf{1}_{\mathbf{T}}$. Given time-invariant mispricing components, we obtain:

$$\mathbf{Y}\mathbf{D_T} = (\mathbf{\Phi}(\mathbf{X})\mathbf{B} + \mathbf{\Lambda} + \mathbf{R}^{\theta}(\mathbf{X}))\mathbf{F}^{\intercal}\mathbf{D_T} + \mathbf{U}\mathbf{D_T},$$

Onwards, we define $\mathbf{Y}\mathbf{D_T} = \tilde{\mathbf{Y}}$ and $\mathbf{F}^{\intercal} = \mathbf{F}^{\intercal}\mathbf{D_T}$. Time-demeaned factors do not change their properties.

Next, multiple both sides by $\mathbf{P} = \Phi(\mathbf{X})(\Phi(\mathbf{X})^{\intercal}\Phi(\mathbf{X}))^{-1}\Phi(\mathbf{X})^{\intercal}$,

$$\hat{\mathbf{Y}} = (\mathbf{\Phi}(\mathbf{X})\mathbf{B} + \mathbf{P}\mathbf{\Lambda} + \mathbf{P}\mathbf{R}^{\theta}(\mathbf{X}))\mathbf{F}^{\intercal} + \mathbf{P}\mathbf{U}\mathbf{D_T}.$$

We decompose:

$$\mathbf{P}\tilde{\mathbf{Y}} = \hat{\mathbf{Y}} = \mathbf{\Phi}(\mathbf{X})\mathbf{B}\mathbf{F}^{\intercal} + \mathbf{P}\mathbf{\Lambda}\mathbf{F}^{\intercal} + \mathbf{P}\mathbf{U}\mathbf{D_T} + \mathbf{P}\mathbf{R}^{\theta}(\mathbf{X})\mathbf{F}^{\intercal} = \mathbf{e_1} + \mathbf{e_2} + \mathbf{e_3} + \boldsymbol{e_4},$$

as $n \to \infty$ and $n^{v} \to \infty$, approximation error $\mathbf{R}^{\theta}(\mathbf{X}) \to_P \mathbf{0}$ as in Huang et al. (2010). Thus, $\mathbf{e}_4^{\intercal} \to^P \mathbf{0}$.

Under Assumption 1, we have following results:

for $\frac{1}{n}\sum_{j=1}^{3} \boldsymbol{e}_2^{\intercal}\boldsymbol{e}_j$,

$$\frac{1}{n}\mathbf{P}\mathbf{\Lambda} \to^P \mathbf{0},$$

therefore,

$$\frac{1}{n}\sum_{j=1}^{3}\mathbf{e}_2^{\intercal}\mathbf{e_j} + \frac{1}{n}\sum_{j=1}^{3}\mathbf{e}_j^{\intercal}\mathbf{e_2} \to^P \mathbf{0}.$$

For $\frac{1}{n}\sum_{j=1}^{3} \boldsymbol{e}_3^{\intercal}\boldsymbol{e}_j$,

$$\frac{1}{n}\mathbf{P}\mathbf{U} \to^P \mathbf{0},$$

therefore,

$$\frac{1}{n}\sum_{j=1}^{3}\mathbf{e}_2^{\intercal}\mathbf{e_j} + \frac{1}{n}\sum_{j=1}^{3}\mathbf{e}_j^{\intercal}\mathbf{e_2} \to^P \mathbf{0}.$$

And only $\frac{1}{n}\mathbf{e}_1^{\intercal}\mathbf{e_1}$ left, namely,

$$\frac{1}{n}\mathbf{e}_1^{\intercal}\mathbf{e_1} = \mathbf{F}\frac{\mathbf{B}^{\intercal}\mathbf{\Phi}^{\intercal}(\mathbf{X})\mathbf{\Phi}(\mathbf{X})\mathbf{B}}{\mathbf{n}}\mathbf{F}^{\intercal}.$$

Under Assumption 2-4 and fixed T. A much smaller $T\times T$ matrix $\frac{1}{n}\hat{\mathbf{Y}}^{\intercal}\hat{\mathbf{Y}}$ can be sovled by asymptotic principal component by Connor and Korajczyk (1986). $\hat{\mathbf{F}} = \frac{1}{\sqrt{T}}\{\psi_1, \psi_2, \dots, \psi_J\}$, where$\{\psi_1, \psi_2, \dots, \psi_J\}$ are eigenvectors corresponding to the first J eigenvalues of $\frac{1}{n}\hat{\mathbf{Y}}^{\intercal}\hat{\mathbf{Y}}$.

Thus, $\hat{\mathbf{F}} \to_P \mathbf{F}$ follows. $\square$

Proof of Theorem 6.2 : Given $\hat{\mathbf{F}}$, we have:

$$\hat{\mathbf{G}}(\mathbf{X}) = \hat{\mathbf{Y}}\hat{\mathbf{F}}(\hat{\mathbf{F}}'\hat{\mathbf{F}})^{-1},$$

as $\hat{\mathbf{F}}^{\intercal}\hat{\mathbf{F}} = \mathbf{I_J}$, therefore,

$$\hat{\mathbf{G}}(\mathbf{X}) = \tilde{\mathbf{Y}}\hat{\mathbf{F}}.$$

Then we need to show:

$$E((\hat{\mathbf{G}}(\mathbf{X_i}) - \mathbf{G}(\mathbf{X_i}))^{\mathbf{2}}) = 0.$$

Take the sample analogue,

$$\frac{1}{n}((\hat{\mathbf{G}}(\mathbf{X}) - \mathbf{G}(\mathbf{X})))^{\intercal}((\hat{\mathbf{G}}(\mathbf{X}) - \mathbf{G}(\mathbf{X}))).$$

Given:

$$\mathbf{G}(\mathbf{X}) = \mathbf{\Phi}(\mathbf{X})\mathbf{B} + \mathbf{R}^{\theta}(\mathbf{X}).$$

$$\hat{\mathbf{G}}(\mathbf{X}) = (\mathbf{\Phi}(\mathbf{X})\mathbf{B} + \mathbf{P\Lambda} + \mathbf{PR}^{\theta}(\mathbf{X}))\mathbf{F}^{\intercal}\hat{\mathbf{F}} + \mathbf{PUD_T}\hat{\mathbf{F}}$$

Furthermore,

$$\mathbf{G}(\mathbf{X}) - \hat{\mathbf{G}}(\mathbf{X}) = (\mathbf{\Phi}(\mathbf{X})\mathbf{B} + \mathbf{P\Lambda} + \mathbf{PR}^{\theta}(\mathbf{X}))\mathbf{F}^{\intercal}\hat{\mathbf{F}} + \mathbf{PUD_T}\hat{\mathbf{F}} - \mathbf{\Phi}(\mathbf{X})\mathbf{B} - \mathbf{R}^{\theta}(\mathbf{X}) = \mathbf{q_1} + \mathbf{q_2} + \mathbf{q_3} + \mathbf{q_4}.$$

Similar to the Proof of Theorem 6.1,

$$\frac{1}{n}((\hat{\mathbf{G}}(\mathbf{X}) - \mathbf{G}(\mathbf{X})))^{\intercal}((\hat{\mathbf{G}}(\mathbf{X}) - \mathbf{G}(\mathbf{X}))) \to^P \frac{1}{n}\mathbf{q_1^{\intercal}q_1} + \frac{\mathbf{1}}{\mathbf{n}}\mathbf{q_3^{\intercal}q_3} + \frac{\mathbf{1}}{\mathbf{n}}\mathbf{q_1^{\intercal}q_3} + \frac{\mathbf{1}}{\mathbf{n}}\mathbf{q_3^{\intercal}q_1}.$$

For the first term,

$$\frac{1}{n}\mathbf{q_1^{\intercal}q_1} = \hat{\mathbf{F}}^{\intercal}\mathbf{F}(\mathbf{\Phi}(\mathbf{X})\mathbf{B} + \mathbf{P\Lambda} + \mathbf{PR}^{\theta}(\mathbf{X}))^{\intercal}(\mathbf{\Phi}(\mathbf{X})\mathbf{B} + \mathbf{P\Lambda} + \mathbf{PR}^{\theta}(\mathbf{X}))\mathbf{F}^{\intercal}\hat{\mathbf{F}},$$

due to

$$\frac{1}{n}\sum_{j=1}^{3}\mathbf{e_2^{\intercal}e_j} + \frac{1}{n}\sum_{j=1}^{3}\mathbf{e_j^{\intercal}e_2} \to^P \mathbf{0},$$

and

$$\frac{1}{n}\mathbf{e_1^T e_1} \to^P \mathbf{F}\frac{\mathbf{B}^{\intercal}\mathbf{\Phi}^{\intercal}(\mathbf{X})\mathbf{\Phi}(\mathbf{X})\mathbf{B}}{\mathbf{n}}\mathbf{F}^{\intercal}$$

then,

$$\frac{1}{n}\mathbf{q_1^T q_1} \to^P \hat{\mathbf{F}}^{\intercal}\mathbf{F}\frac{\mathbf{B}^{\intercal}\mathbf{\Phi}^{\intercal}(\mathbf{X})\mathbf{\Phi}(\mathbf{X})\mathbf{B}}{\mathbf{n}}\mathbf{F}^{\intercal}\hat{\mathbf{F}}.$$

Theorem 6.1 and Assumption 2 give $\hat{\mathbf{F}} \to \mathbf{F}$ and $\mathbf{F}^{\mathbf{T}}\mathbf{F} = \mathbf{I_J}$, therefore:

$$\frac{1}{n}\mathbf{q_1^T q_1} \to^P \frac{\mathbf{B}^\intercal\mathbf{\Phi}^\intercal(\mathbf{X})\mathbf{\Phi}(\mathbf{X})\mathbf{B}}{\mathbf{n}},$$

Similarly,

$$\frac{1}{n}\mathbf{q_3^T q_3} \to^P \frac{\mathbf{B}^\intercal\mathbf{\Phi}^\intercal(\mathbf{X})\mathbf{\Phi}(\mathbf{X})\mathbf{B}}{\mathbf{n}},$$
$$\frac{1}{n}\mathbf{q_1^T q_3} \to^P -\frac{\mathbf{B}^\intercal\mathbf{\Phi}^\intercal(\mathbf{X})\mathbf{\Phi}(\mathbf{X})\mathbf{B}}{\mathbf{n}},$$
$$\frac{1}{n}\mathbf{q_3^T q_1} \to^P -\frac{\mathbf{B}^\intercal\mathbf{\Phi}^\intercal(\mathbf{X})\mathbf{\Phi}(\mathbf{X})\mathbf{B}}{\mathbf{n}}.$$

Therefore,

$$\frac{1}{n}\mathbf{q}_1^\intercal\mathbf{q}_1 + \frac{\mathbf{1}}{\mathbf{n}}\mathbf{q}_3^\intercal\mathbf{q}_3 + \frac{\mathbf{1}}{\mathbf{n}}\mathbf{q}_1^\intercal\mathbf{q}_3 + \frac{\mathbf{1}}{\mathbf{n}}\mathbf{q}_3^\intercal\mathbf{q}_1 \to 0.$$

Then,

$$\frac{1}{n}((\hat{\mathbf{G}}(\mathbf{X}) - \mathbf{G}(\mathbf{X})))^\intercal((\hat{\mathbf{G}}(\mathbf{X}) - \mathbf{G}(\mathbf{X}))) \to^P 0$$

thus,

$$\hat{\mathbf{G}}(\mathbf{X}) \to^{\mathbf{P}} \mathbf{G}(\mathbf{X}).$$

Then Theorem 6.2 follows.

□

Proof of Theorem 6.3 : Let $\dot{\mathbf{Y}} = \frac{1}{\mathbf{T}}(\mathbf{Y} - \hat{\mathbf{G}}(\mathbf{X})\hat{\mathbf{F}})\mathbf{1_T}$. By substituting the restriction, we have the Lagrangian equation:

$$\min_{\mathbf{A}}(\dot{\mathbf{Y}} - \mathbf{\Phi}(\mathbf{X})\mathbf{A})^\intercal(\dot{\mathbf{Y}} - \mathbf{\Phi}(\mathbf{X})\mathbf{A}) + \lambda\hat{\mathbf{G}}^\intercal(\mathbf{X})\mathbf{\Phi}(\mathbf{X})\mathbf{A} \tag{8}$$

Then we take the first order condition with respect to $\mathbf{A}$ and λ separately, and we obtain:

$$\begin{pmatrix} \mathbf{2\Phi}(\mathbf{X})^\intercal\mathbf{\Phi}(\mathbf{X}) & \mathbf{\Phi}(\mathbf{X})^\intercal\hat{\mathbf{G}}(\mathbf{X}) \\ \hat{\mathbf{G}}(\mathbf{X})^\intercal\mathbf{\Phi}(\mathbf{X})^\intercal & \mathbf{0} \end{pmatrix} \begin{pmatrix} \hat{\mathbf{A}} \\ \lambda \end{pmatrix} = \begin{pmatrix} \mathbf{2\Phi}(\mathbf{X})^\intercal\dot{\mathbf{Y}} \\ \mathbf{0} \end{pmatrix}. \tag{9}$$

Under Assumption 2, the above matrice are invertible, which can be written as:

$$\begin{pmatrix} \hat{\mathbf{A}} \\ \boldsymbol{\lambda} \end{pmatrix} = \begin{pmatrix} \mathbf{2\Phi}'(\mathbf{X})\mathbf{\Phi}(\mathbf{X}) & \mathbf{\Phi}'(\mathbf{X})\hat{\mathbf{G}}(\mathbf{X}) \\ \hat{\mathbf{G}}(\mathbf{X})^\intercal\mathbf{\Phi}(\mathbf{X})^\intercal & \mathbf{0} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{2\Phi}'(\mathbf{X})\dot{\mathbf{Y}} \\ \mathbf{0} \end{pmatrix}. \tag{10}$$

Therefore, we obtain:

$$\hat{\mathbf{A}} = \mathbf{M}\tilde{\mathbf{A}},$$

where

$$\mathbf{M} = \mathbf{I} - (\Phi(\mathbf{X})^\intercal\Phi(\mathbf{X}))^{-1}\mathbf{\Phi}(\mathbf{X})^\intercal\hat{\mathbf{G}}(\mathbf{X})(\hat{\mathbf{G}}(\mathbf{X})^\intercal\hat{\mathbf{G}}(\mathbf{X}))^{-1}\hat{\mathbf{G}}(\mathbf{X})^\intercal\mathbf{\Phi}(\mathbf{X}),$$
$$\tilde{\mathbf{A}} = \frac{\mathbf{1}}{\mathbf{T}}(\Phi(\mathbf{X})^\intercal\Phi(\mathbf{X}))^{-1}\Phi(\mathbf{X})^\intercal\dot{\mathbf{Y}}\mathbf{1_T}.$$

Furthermore, let $\Xi = \mathbf{\Phi}(\mathbf{X})\hat{\mathbf{A}} - \mathbf{h}(\mathbf{X}) = \mathbf{\Phi}(\mathbf{X})\mathbf{M}\tilde{\mathbf{A}} - \mathbf{\Phi}(\mathbf{X})\mathbf{A} - \mathbf{R}^{\mu}(\mathbf{X})$.

Under the restriction $\hat{\mathbf{G}}'(\mathbf{X})\mathbf{\Phi}(\mathbf{X})\mathbf{A} = \mathbf{0}$, we can obtain:

$$\mathbf{\Xi} = \mathbf{\Phi}(\mathbf{X})\mathbf{M}(\mathbf{\Phi}(\mathbf{X})^{\intercal}\mathbf{\Phi}(\mathbf{X}))^{-1}\mathbf{\Phi}(\mathbf{X})^{\intercal}\frac{\mathbf{1}}{\mathbf{T}}(\mathbf{\Phi}(\mathbf{X})\mathbf{A} + \mathbf{R}^{\mu}(\mathbf{X}) + \mathbf{\Gamma} + (\mathbf{\Lambda} + \mathbf{R}^{\theta}(\mathbf{X}))\mathbf{F}')\mathbf{1}_{\mathbf{T}} - \mathbf{\Phi}(\mathbf{X})\mathbf{A} - \mathbf{R}^{\mu}(\mathbf{X}). \tag{11}$$

Furthermore, we have:

$$\mathbf{\Phi}(\mathbf{X})\mathbf{M}(\mathbf{\Phi}(\mathbf{X})^{\intercal}\mathbf{\Phi}(\mathbf{X}))^{-1}\mathbf{\Phi}(\mathbf{X})^{\intercal} = (\mathbf{I} - (\Phi(\mathbf{X})^{\intercal}\Phi(\mathbf{X}))^{-1}\mathbf{\Phi}(\mathbf{X})^{\intercal}\hat{\mathbf{G}}(\mathbf{X})(\hat{\mathbf{G}}(\mathbf{X})^{\intercal}\hat{\mathbf{G}}(\mathbf{X}))^{-1}\hat{\mathbf{G}}(\mathbf{X})^{\intercal})\mathbf{P}. \tag{12}$$

And then, substituteEquation 12 into Equation 11 and under Assumption 1 and Theorem 6.2:

$$\mathbf{\Xi} = \mathbf{\Phi}(\mathbf{X})\mathbf{A} - \mathbf{\Phi}(\mathbf{X})\mathbf{A} - \mathbf{R}^{\mu}(\mathbf{X}).$$

$$\mathbf{R}^{\mu}(\mathbf{X}) \to \mathbf{0} \text{ as } n \to \infty,$$

therefore,

$$\frac{1}{n}\mathbf{\Xi}^{\intercal}\mathbf{\Xi} \to \mathbf{0}.$$

And the Theorem 6.3 follows. □

Proof of Theorem 6.4 : Define $Z = \max_{\{1\leqslant p\leqslant P, 1\leqslant h\leqslant H_n\}}\{|\hat{\alpha}_{ph}|/\hat{\sigma}_{ph}\}$. Under Assumption 3.2.3, we have

$$\hat{\alpha}_{ph}/\hat{\sigma}_{ph}|\mathbf{H_0} \to^{d} N(0,1).$$

Therefore, under the $\mathbf{H_0}$, we have:

$$\begin{aligned}
e^{tE(Z)} &\leqslant E[e^{tZ}] \\
&= E[max\{t|\hat{\alpha}_{ph}|/\hat{\sigma}_{ph}\}] \\
&\leqslant \sum_{p=1,h=1}^{p=P,h=H_n} E[e^{t|\hat{\alpha}_{ph}|/\hat{\sigma}_{ph}}] \\
&= ne^{t^2/2}.
\end{aligned}$$

Then take the logarithm of both sides we can obtain:

$$E[Z] \leqslant \frac{\log n}{t} + \frac{t}{2}.$$

If we set $t = \sqrt{2\log n}$ to minimise $\frac{\log n}{t} + \frac{t}{2}$, then we have:

$$E[Z] \leqslant \sqrt{2\log n}.$$

Therefore, we can bound the $|\hat{\alpha}_{ph}|/\hat{\sigma}_{ph}$ by $\sqrt{2\log n}$. □

Proof of Theorem 6.5 : To proof

$$\inf_{\mathbf{A}\in\mathcal{A}} P(\text{reject } H_0|\mathbf{A}) \to 1,$$

equivalently, we need to prove

$$\inf_{\mathbf{A}\in\mathcal{A}} P(S_0 + S_1 > F_q|\mathbf{A}) \to 1.$$

$S_0 = H_n \sum_{p=1}^{P} \mathbf{I}(\sum_{h=1}^{H_n} |\hat{\alpha}_{ph}|/\hat{\sigma}_{ph} \geqslant \eta_n)$, as $H_n = n_v \to \infty$ as $n \to \infty$.

Under Theorem 6.4 and $n \to \infty$, we have:

$$E(S_0|\mathbf{A}) \to \infty.$$

Meanwhile $F_q = O(1)$, according to Fan et al. (2015) and Kock and Preinerstorfer (2019), we can show that:

$$\inf_{\mathbf{A}\in\mathcal{A}} P(S_0 + S_1 > F_q|\mathbf{A}) \to 1.$$

Then the Theorem 6.5 follows. □

References

Carhart, M. M. (1997), 'On persistence in mutual fund performance', *The Journal of finance* **52**(1), 57–82.

Chen, X. and Pouzo, D. (2012), 'Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals', *Econometrica* **80**(1), 277–321.

Connor, G., Hagmann, M. and Linton, O. (2012), 'Efficient semiparametric estimation of the fama–french model and extensions', *Econometrica* **80**(2), 713–754.

Connor, G. and Korajczyk, R. A. (1986), 'Performance measurement with the arbitrage pricing theory: A new framework for analysis', *Journal of financial economics* **15**(3), 373–394.

Connor, G. and Linton, O. (2007), 'Semiparametric estimation of a characteristic-based factor model of common stock returns', *Journal of Empirical Finance* **14**(5), 694–717.

Cox, D. R. (1957), 'Note on grouping', *Journal of the American Statistical Association* **52**(280), 543–547.

Fama, E. F. and French, K. R. (1993), 'Common risk factors in the returns on stocks and bonds', *Journal of financial economics* **33**(1), 3–56.

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

Fan, J., Liao, Y. and Mincheva, M. (2013), 'Large covariance estimation by thresholding principal orthogonal complements', *Journal of the Royal Statistical Society. Series B, Statistical methodology* **75**(4).

Fan, J., Liao, Y. and Wang, W. (2016), 'Projected principal component analysis in factor models', *Annals of statistics* **44**(1), 219.

Fan, J., Liao, Y. and Yao, J. (2015), 'Power enhancement in high-dimensional cross-sectional tests', *Econometrica* **83**(4), 1497–1541.

Feng, G., Giglio, S. and Xiu, D. (2017), 'Taming the factor zoo', *Fama-Miller Working Paper* **24070**.

Fisher, W. D. (1958), 'On grouping for maximum homogeneity', *Journal of the American statistical Association* **53**(284), 789–798.

Freyberger, J., Neuhierl, A. and Weber, M. (2017), Dissecting characteristics nonparametrically, Technical report, National Bureau of Economic Research.

Hjalmarsson, E. and Manchev, P. (2012), 'Characteristic-based mean-variance portfolio choice', *Journal of Banking & Finance* **36**(5), 1392–1401.

Hoberg, G. and Phillips, G. (2016), 'Text-based network industries and endogenous product differentiation', *Journal of Political Economy* **124**(5), 1423–1465.

Hou, K., Xue, C. and Zhang, L. (2015), 'Digesting anomalies: An investment approach', *The Review of Financial Studies* **28**(3), 650–705.

Huang, J., Horowitz, J. L. and Wei, F. (2010), 'Variable selection in nonparametric additive models', *Annals of statistics* **38**(4), 2282.

Kelly, B. T., Pruitt, S. and Su, Y. (2017), 'Instrumented principal component analysis', *Available at SSRN 2983919* .

Kelly, B. T., Pruitt, S. and Su, Y. (2019), 'Characteristics are covariances: A unified model of risk and return', *Journal of Financial Economics* .

Kim, S., Korajczyk, R. A. and Neuhierl, A. (2019), 'Arbitrage portfolios', *The Review of Financial Studies* .

Kock, A. B. and Preinerstorfer, D. (2019), 'Power in high-dimensional testing problems', *Econometrica* **87**(3), 1055–1069.

Ledoit, O., Wolf, M. et al. (2012), 'Nonlinear shrinkage estimation of large-dimensional covariance matrices', *The Annals of Statistics* **40**(2), 1024–1060.

Liew, C. K. (1976), 'Inequality constrained least-squares estimation', *Journal of the American Statistical Association* **71**(355), 746–751.

Pesaran, M. H. and Yamagata, T. (2012), Testing capm with a large number of assets, *in* 'AFA 2013 San Diego Meetings Paper'.

Pollard, D. (1981), 'Strong consistency of k-means clustering', *The Annals of Statistics* pp. 135–140.

Pollard, D. et al. (1982), 'A central limit theorem for k-means clustering', *The Annals of Probability* **10**(4), 919–926.

Ross, S. (1976), 'The arbitrage theory of capital asset pricing', *Journal of Economic Theory* **13**(3), 341–360.

Sun, W., Wang, J., Fang, Y. et al. (2012), 'Regularized k-means clustering of high-dimensional data and its asymptotic consistency', *Electronic Journal of Statistics* **6**, 148–167.

Vogt, M. and Linton, O. (2017), 'Classification of non-parametric regression functions in longitudinal data models', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **79**(1), 5–27.

A Dynamic Semiparametric Characteristics-based Model for Optimal Portfolio Selection

Gregory Connor*[1], Shaoran Li†[2], and Oliver Linton‡[3]

[1]*Maynooth University, Ireland*
[2]*University of Cambridge, UK*
[3]*University of Cambridge, UK*

February 10, 2021

Abstract

This paper develops a two-step semiparametric methodology for portfolio weight selection for characteristics-based factor-tilt and factor-timing investment strategies. We build upon the expected utility maximization framework of Brandt (1999) and Aït-sahalia and Brandt (2001). We assume that asset returns obey a characteristics-based factor model with time-varying factor risk premia as in Li and Linton (2020). We prove under our return-generating assumptions that in a market with a large number of assets, an approximately optimal portfolio can be established using a two-step procedure. The first step finds optimal factor-mimicking sub-portfolios using a quadratic objective function over linear combinations of characteristics-based factor loadings. The second step dynamically combines these factor-mimicking sub-portfolios based on a time-varying signal, using the investor's expected utility as the objective function. We develop and implement a two-stage semiparametric estimator. We apply it to CRSP (Center for Research in Security Prices) and FRED (Federal Reserve Economic Data) data and find excellent in-sample and out-sample performance consistent with investors' risk aversion levels.

KEYWORDS: Portfolio management; Single index; GMM;

JEL CLASSIFICATION: C14; G11.

*Electronic address: gregory.connor@mu.ie
†Electronic address: sl736@cam.ac.uk
‡Electronic address: obl20@cam.ac.uk

1 Introduction

The traditional portfolio choice model proceeds by estimating the parameters of an asset return distribution and then finding the portfolio that maximizes expected payoffs for a given risk level, such as the optimal mean-variance portfolio choice model proposed by Markowitz et al. (1952). This approach can produce biases in portfolio weights since the portfolio selection process ignores the estimation error in the empirically-derived return distribution parameters. Furthermore, as the number of assets increases, the estimation of the high-dimensional covariance matrix becomes intractable. Some notable methods have been proposed to solve this issue, such as linear (Ledoit and Wolf (2004)) and nonlinear shrinkage (Ledoit and Wolf (2017)) of the target covariance matrix or selecting main elements by threshold (Fan et al. (2013)). However, these approaches may cause information loss and lead to unsatisfactory results, as illustrate by Ao et al. (2019). At the same time, Ao et al. (2019) studied a method called MAXSER, which is a sparse regression which sets the optimal sharpe ratio as the regressand. Their method also requires a sparsity assumption and can be problematic when the number of assets n is large. Meanwhile, all aforementioned papers ignore the importance of predictive variables, which have been documented by many researchers, such as Fama and French (1989), who analyzed the forecasting ability of dividend yield, default spread and term spread on asset returns, as well as Keim and Stambaugh (1986), Campbell and Shiller (1988), and Hodrick (1992) among others. The goal of this paper is to construct a two-step optimal portfolio that takes advantage of both a large number of assets and dynamic predictors.

Brandt (1999) used nonparametric tools to directly estimate the portfolio weights that maximize expected utility of the observed data, without first estimating the return distribution. He estimated the dynamic portfolio weights of the assets in a two-asset model as a nonparametric function of the univariate time-series predictor of the future excess returns of the risky assets. Aït-sahalia and Brandt (2001) replaced the univariate time series predictor with an index-based set of predictors: the time-varying portfolio weights in a three-asset model were assumed to be a nonlinear function of a linear fix combination of a vector of predictive variables. However, the number of assets included in their portfolio was quite limited.

Brandt et al. (2009) developed a characteristic-based model for portfolio selection with a large cross-section of assets. They assumed that optimal portfolio weights were linearly related to a small set of observable characteristics, such as book-to-market ratio, momentum and market equity. They found the linear coefficients that maximized expected utility under this assumption.

In this paper, we develop a new semiparametric model of portfolio selection, which combines the advantages of a large cross-section of assets and dynamic predictive variables. This is achieved by a characteristics-based asset pricing model. We generalize the methodologies in the above-mentioned papers since we do not impose the assumption that optimal weights are linear in the characteristics. Furthermore, the firm-specific characteristics included in our model can be significantly broadened. There are 33 characteristics in our empirical study, which provides more potential abnormal return opportunities. Also, as in Aït-sahalia and Brandt (2001), we also allow information-based dynamically-varying portfolio allocation based on a single-index function of predictors. We replace weighting across asset classes in Aït-sahalia and Brandt (2001) with weighting across our optimally-constructed characteristics-based sub-portfolios.

We estimate the model using a new, two-stage semiparametric procedure. The first step involves the estimation of the factor-mimicking sub-portfolios. This is a high-dimensional estimation problem since the number of assets is diverging, but the objective function is quadratic, allowing us to solve it using semiparametric techniques. That step compacts those assets into several sub-portfolios rather than discarding some of them, and also helps to reduce the dimensionality, which simplifies the next step. The second step maximizes the dynamic expected utility of a risk-free asset and those sub-portfolios conditional on a set of predictors, similar to Aït-sahalia and Brandt (2001). Our two-step statistical methodology accounts fully for the estimation error in both semiparametric steps, and we show that it approximates the intractable single-stage, asset-by-asset portfolio weight estimation problem in a well-defined sense.

Our model is not entirely general: we do not allow individual asset selection in response to asset-specific valuation information. We essentially allow for factor-tilt strategies, which means weighting securities according to their factor exposure in response to the associated factor risk premia, and factor timing, which means dynamically varying factor-tilt strategies, accounting for predictability in factor risk premia, but not individual asset selection. This method keeps most of the information contained in individual assets, while benefitting greatly from dimensionality reduction.

We base our model on a dynamic, characteristics-based factor model of returns. This kind of model was first studied by Connor and Linton (2007) and Connor et al. (2012), where they specified their model as:

$$y_{it} = \alpha_i + \sum_{j=1}^{J} g_j(X_{ji}) f_{jt} + \epsilon_{it}, \tag{1}$$

where y_{it} is the excess return on security i at time t; f_{jt} is the j^{th} risk factor's return at time t; X_{ji} is the j^{th} observable characteristic of firm i; α_i represents the intercept (mispricing) part of i^{th} asset return; and ϵ_{it} are the mean zero idiosyncratic shocks. They restricted characteristic-based loading $g_j(\cdot)$ to be a univariate nonparametric function. To extend the dimension of the factor loading function $g_j(\cdot)$, Kelly et al. (2019) and Kim et al. (2019) specify both mispricing and factor loading parts as a parametric linear function of a large set of firm-specific characteristics as:

$$y_{it} = h(\boldsymbol{X_i}) + \sum_{j=1}^{J} g_j(\boldsymbol{X_i}) f_{jt} + \epsilon_{it}. \tag{2}$$

They illustrated the validility of characteristics-based factor models and provided relevant empirical results. Li and Linton (2020) generalized the parametric part of Equation 2 as semiparametric functions to be consistent with earlier research. They also proposed power enhanced tests to verify their model, concluding that the semiparametric mispricing component $h(\boldsymbol{X_i})$ was only significant during certain rolling windows.

Section two describes the econometric framework for our model. We assume the returns are generated by the asset pricing model in Li and Linton (2020) and that the factor risk premia are predictable based on a single-index function involving a set of both stationary and nonstationary predictors.

Section three presents the general portfolio management problem and our restricted class of portfolio selection rules in which the problem is divided into two steps. In the first step, the investors choose a set of characteristics-based sub-portfolios that are well-diversified and mimic the returns of the underlying unobservable factors. In the second step, the investors choose a dynamic combination of these sub-portfolios and a risk-free asset dependent upon their time-varying information set and utility function. The information is specified as a single-index function, which is well-approximated by orthogonal series, allowing both stationary and nonstationary covariates. We show that under reasonable conditions on risk preference, the two-step selection rule has asymptotically zero impact on an investor's expected utility as the number of assets grows to infinity, relative to the unattainable true optimal choice.

Section four derives estimators for both steps. In the factor-tilt step, the factor-mimicking portfolios are constructed by the linear combination of estimated characteristics-based factor loadings. To diversify the idiosyncratic shocks further, the weight for each factor loading function is estimated through a constrained quadratic objective function. In the second step, called factor timing, the optimization of the expected utility function is solved by the methodology of the continuously-updating GMM, as in Hansen et al. (1996). The weights allocated to the risk-free asset and sub-portfolios are determined by the single-index function approximated through Hermite polynomials, which allows for both stationary and nonstationary predictors, as in Dong et al. (2016b). The coefficients of those orthogonal bases are estimated by solving the sample counterpart under the continuously-updating GMM framework. Section five documents the hypothesis tests on the significance of these predictors included in the single-index function.

Section six presents the empirical findings. We apply our approaches to monthly CRSP and FRED data and reveal some popular predictive variables' nonstationarity and significance. Furthermore, we find our portfolios have different but outstanding performance under various levels of risk aversion. Finally, the results of the in-sample and out-sample are similar and reflect the risk preference of the investor.

Section seven concludes and discusses the paper, while proofs of theorems and supplementary tables are arranged in the Appendix.

2 Econometric Framework

We assume there is a large panel of monthly stocks' excess returns generated by the characteristics-based model:

$$y_{it} = \sum_{j=1}^{J} g_j(\boldsymbol{X_i})(f_{jt} + \phi_{jt}) + \epsilon_{it}, \tag{3}$$

where y_{it} is i^{th} stock's excess return at time t while $\boldsymbol{X_i}$ is a large set of assets' P-vector of characteristics, which is regarded as time-invariant within a short time window; $g_j(\boldsymbol{X_i})$ is the j^{th} characteristics-based factor loading, which is specified as a multivariate additive semiparametric function. The factor returns $\boldsymbol{F_t} = (f_{1t}, \ldots, f_{Jt})^\intercal$ are the common sources of risk in asset returns at time t with associated means $\boldsymbol{\phi_t} = \{\phi_{1t}, \ldots, \phi_{Jt}\}$. The asset-specific return ϵ_{it} is conditional zero mean, i.e., $E(\epsilon_{it}|\boldsymbol{X_i}, \boldsymbol{F_t}) = 0$.

This framework is an extension of Connor and Linton (2007) and Connor et al. (2012), who assumed the factor beta function $g(\cdot)$ to be univariate. This model is a special case of Li and Linton (2020) by replacing the mispricing component with the mean value ϕ_{jt} of the j^{th} risk factor.

We allow for time variation in the characteristics of the assets across rolling windows. We treat the $n \times J$ matrix of characteristics in the t^{th} rolling window $\boldsymbol{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)^\intercal$, as a random draw from a multivariate population distribution. Furthermore, the investor can observe $\boldsymbol{X}$ before rolling block t, and then choose his time t portfolio.

We define the $n \times J$ matrix $G(\boldsymbol{X}) = (g_1(\boldsymbol{X}), \ldots, g_J(\boldsymbol{X}))$ and $g_j(\boldsymbol{X}) = (g_j(\boldsymbol{X_1}), \ldots, g_j(\boldsymbol{X_n}))^\intercal$, and the matrix form of the **demeaned** assets' returns at time t is :

$$\boldsymbol{Y_t} = G(\boldsymbol{X})\boldsymbol{F_t} + \boldsymbol{\epsilon_t}, \qquad t = 1, 2, \ldots, T, \tag{4}$$

where $\boldsymbol{Y_t}$ is a $n \times 1$ matrix of the demeaned assets' excess returns at time t, $G(\boldsymbol{X}) = (g_1(\boldsymbol{X}), \ldots, g_J(\boldsymbol{X}))$ is a $n \times J$ factor loading matrix, and $\boldsymbol{\epsilon_t}$ is a $n \times 1$ vector of asset-specific returns.

Furthermore, we assume that there exists a nonsingular $J \times J$ matrix $\boldsymbol{M^G}$ such that:

$$E(G(\boldsymbol{X_i})^\intercal G(\boldsymbol{X_i})) = \boldsymbol{M^G}, \tag{5}$$

where the expectation is over the multivariate probability density of characteristics and the off-diagonal elements are all zeros. The Equation 5 simply imposes that for any large random population of the assets, the factor-loading matrix is nonsingular with the probability of one.

For a finite value of n and realized characteristics matrix $\boldsymbol{X}$, under weak assumption, the finite sample second moment matrix approaches the population value as:

$$p \lim_{n\to\infty} \frac{1}{n} \sum_{1=1}^{n} G(\boldsymbol{X_i})^\intercal G(\boldsymbol{X_i}) = \boldsymbol{M^G}, \tag{6}$$

There is a scale indeterminacy in Equation 3 since one could multiply the function $g_j(\boldsymbol{X}_i)$ by any non-zero constant c and $(f_{jt} + \phi_{jt})$ by $1/c$, and the model of returns would be identical. We resolve the this indeterminacy by setting $\boldsymbol{M}^G_{jj} = 1,$ where $j = 1, \ldots, J$.

The estimation of Equation 5 under Equation 6 is discussed in Li and Linton (2020); it is achieved mainly through "Projected Principal Component Analysis" (PPCA) by Fan et al. (2016). The idea is to project the $n \times T$

asset excess returns onto the B-spline space spanned by $\boldsymbol{X}$, where $\mathbf{X} = (\mathbf{X_1}, \mathbf{X_2}, \ldots, \mathbf{X_n})$, and then, we collect the projected returns $\hat{\mathbf{Y}}$. Furthermore, we perform PCA on $\frac{1}{\mathbf{T}}\hat{\mathbf{Y}}\hat{\mathbf{Y}}^{\intercal}$. Therefore, the $\hat{G}(\boldsymbol{X})$ is estimated as the largest J eigenvectors of $\frac{1}{\mathbf{T}}\hat{\mathbf{Y}}\hat{\mathbf{Y}}^{\intercal}$. Due to the property of the PCA, the assumption Equation 6 is satisfied.

Let $\boldsymbol{M^f} = E(\boldsymbol{F_t}\boldsymbol{F_t}^{\intercal})$ denote the nonsingular covariance matrix of the factors and $\boldsymbol{M^\epsilon} = E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^{\intercal})$ denote the covariance matrix of the idiosyncratic returns; note these two sources of returns are statistically independent. We assume that the asset-specific risks are diversifiable, in particular:

$$\lambda_{max}(\boldsymbol{M^\epsilon}) < c_\lambda < \infty,$$

where the λ_{max} denotes the largest eigenvalue whereas c_λ is a constant. This follows readily from standard assumptions on weak correlations of the asset-specific risks.

We allow dynamic variation in the mean value of factor return premia. At the beginning of each period, a $K \times 1$ vector of random signal $\boldsymbol{z_t} = (z_{1t}, \ldots, z_{Kt})^{\intercal}$ is observed by the investor before he chooses his portfolio. The expected return on the j^{th} factor in Equation 3 is a nonlinear function of a fixed linear combination of these dynamic signals by coefficients $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)^{\intercal}$ as:

$$\phi_{jt+1} = \pi_j(\boldsymbol{\theta}^{\intercal}\boldsymbol{z_t}). \tag{7}$$

The three vectors $\boldsymbol{F_t}$,$\boldsymbol{z_t}$ and $\boldsymbol{\epsilon_t}$ are assumed to be statistically independent. At each time t, the investor observes the characteristics of those assets $\boldsymbol{X}$, which is treated as time-invariant during this time block, and the dynamic signal $\boldsymbol{z_t}$. Then, the investor chooses his time t portfolio based on this information. Finally, at the $(t+1)^{th}$ period, his portfolio return depends upon the realized assets' returns, which in turn depends on the realized factor returns and asset-specific returns $\boldsymbol{F_{t+1}}$ and $\boldsymbol{\epsilon_{t+1}}$ respectively, according to Equation 3.

3 A Two-Step Version of the Portfolio Choice Problem

This section first defines the utility function of a rational decision-maker and then describes how the optimal portfolio weights are chosen through a two-step procedure. In step one, the investor chooses characteristics-based factor-mimicking sub-portfolios based on a linear combination of the beta function$\sum_{j=1}^{J} g_j(\boldsymbol{X})$ in Equation 3. Step two combines these sub-portfolios optimally using expected utility as the investor's objective function, based on a dynamic index.

3.1 Utility Function of the Investor

The investor in our model is myopic and he chooses his portfolio for time t to maximize one-period expected utility of return. We assume his return at time t is W_t and his risk-averse von Neumann Morgenstern preference is defined over W_t with a lower bound on the second derivative:

$$\frac{d}{dW}u(W) > 0, -c < \frac{d^2}{W^2}u(W) < 0 \tag{8}$$

Additionally, we define the optimal portfolios weights $n \times 1$ vector $\mathbf{w}^*$ such that:

$$\mathbf{w}^* = \arg\max_{\mathbf{w}} E[u(r_{ft} + \mathbf{w}^{*\intercal}\mathbf{r_t})|\boldsymbol{X_t}, \boldsymbol{z_t}], \tag{9}$$

where r_{ft} is the risk-free return at time t and $\mathbf{r_t}$ is a $n \times 1$ vector of stock returns at time t. In practice , the optimal $\mathbf{w}^*$ is hard to determine and unstable when n is large or the trading frequency is high, as discussed in the Introduction. Therefore, we consider optimal portfolio choice under a restriction on portfolio weights. Rather than choosing asset weights directly, the investor chooses a set of J characteristics-based portfolios to approximately mimic the factors. Then, in the second step, the investor combines these factor-mimicking subportfolios optimally using his expected utility function conditional on a group of predictors.

3.2 Step 1: Factor-mimicking Sub-portfolios

In this subsection, we propose a method to construct factor-mimicking sub-portfolios based on Equation 3 and discuss the properties of these sub-portfolios.

We propose a semiparametric weighting function to mimic the risk factors $\boldsymbol{F_t}$, which is in the form of a linear combination of characteristics-based factor loadings as in Equation 3:

$$b_j(\boldsymbol{X}_i) = \gamma_{j1}g_1(\boldsymbol{X}_i) + \cdots + \gamma_{jJ}g_J(\boldsymbol{X}_i), \tag{10}$$

therefore, the portfolio weight of i^{th} asset to construct the j^{th} subportfolio is $\frac{1}{n}b_j(\boldsymbol{X}_i)$.

The weighting matrix of assets to mimic all J factors is as follows:

$$B(\boldsymbol{X}_i) = \frac{1}{n}\boldsymbol{\Gamma} G(\boldsymbol{X}_i)^\intercal, \tag{11}$$

where

$$B(\boldsymbol{X}_i) = \frac{1}{n}\begin{Bmatrix} b_1(\boldsymbol{X}_i) \\ b_2(\boldsymbol{X}_i) \\ \vdots \\ b_J(\boldsymbol{X}_i) \end{Bmatrix},$$

$$\boldsymbol{\Gamma} = \begin{Bmatrix} \gamma_{11} & \dots & \gamma_{1J} \\ \gamma_{21} & \dots & \gamma_{2J} \\ \dots & \dots & \dots \\ \gamma_{J1} & \dots & \gamma_{JJ} \end{Bmatrix}, \quad G(\boldsymbol{X}_i)^\intercal = \begin{Bmatrix} G_1(\boldsymbol{X}_i) \\ G_2(\boldsymbol{X}_i) \\ \vdots \\ G_J(\boldsymbol{X}_i) \end{Bmatrix}.$$

Thus, the $J \times 1$ factor-mimicking portfolio return vector at time t is calculated:

$$\boldsymbol{Q_t}(\boldsymbol{X}) = \begin{Bmatrix} q_{1t}(\boldsymbol{X}) \\ q_{2t}(\boldsymbol{X}) \\ \vdots \\ q_{Jt}(\boldsymbol{X}) \end{Bmatrix} = \sum_{i=1}^{n} B(\boldsymbol{X}_i)y_{it}. \tag{12}$$

The factor-mimicking portfolio vector has at least two attractive properties, which are listed as theorems.

Theorem 1. *Each subportfolio in the $J \times 1$ factor-mimicking vector defined by Equation 12 is a linear combination of risk factors f_{jt} directly.*

Theorem1 implies that we can control the similarity between subportfolios and risk factors by adjusting coefficients matrix Γ, which provides us with considerable flexibility.

Theorem 2. *The returns of portfolio defined by Equation 12 have asymptotically zero idiosyncratic variance.*

Theorem 2 illustrates that portfolio returns of factor-mimick sub-portfolios can diversify the asset-specific returns completely as the number of assets goes to infinity.

An investor who uses a semiparametric characteristics-based weight function to choose sub-portfolios rather than individual assets i sacrifices the flexibility to weight assets differently based on the properties of their asset-specific returns ϵ_{it}, since the sub-portfolio weight function Equation 12 only differentiates assets by their characteristic vectors. However, for both hedge fund managers and researchers, there are no satisfactory rules for choosing thousands of assets robustly. Furthermore, some weighting strategies have to be rebalanced once per trading day, and even more frequently for some strategies. This high-speed decision-making problem is intractable without some simplifying applicable rules like Equation 12.

3.3 Step 2: Factor-timing Portfolio Based on Dynamic Signals

This subsection describes how to approximate the dynamic signal function $\pi_j(\boldsymbol{\theta}^\intercal \boldsymbol{z_t})$ in Equation 7, and how to use this function as dynamic weights assigned to those factor-mimicking sub-portfolios in subsection 3.2, to reflect information about their overperformance/underperformance on a risk-adjusted basis. This subsection captures the particular "factor-timing" strategy used by the investor.

Here, we define the objective function as:

$$\arg\max_{\boldsymbol{\theta}} E[u(\alpha r_{ft} + \boldsymbol{\Pi}(\boldsymbol{\theta}^\intercal \boldsymbol{z_t})^\intercal \boldsymbol{Q_{t+1}}(\boldsymbol{X}))], \tag{13}$$

subject to

$$\|\boldsymbol{\theta}\|_2 = 1 \text{ and } \theta_1 > 0$$

and

$$\alpha + \sum_{j=1}^{J} \boldsymbol{\pi_j}(\boldsymbol{\theta}^\intercal \boldsymbol{z_t}) = 1$$

where r_{ft} is the risk-free return at time t and α is its portfolio weight, and

$$\boldsymbol{\Pi}(\boldsymbol{\theta}^\intercal \boldsymbol{z_t}) = (\pi_1(\boldsymbol{\theta}^\intercal \boldsymbol{z_t}), \ldots, \pi_J(\boldsymbol{\theta}^\intercal \boldsymbol{z_t}))^\intercal.$$

The first restriction is for identification purposes while the second is for unit investment. We do not restrict short selling and leverage. Equation 13 is a transformation of the objective function Equation 9. The $n \times 1$ vector of assets' returns $\boldsymbol{r_t}$ is replaced by the vector of sub-portfolios' returns $\boldsymbol{Q_{t+1}}$ conditional on $\boldsymbol{X}$, which compacts the information of $\mathbf{r_t}$ through observed characteristics $\boldsymbol{X}$. Similarly, the dynamic weights of each asset $\mathbf{w}^*$ is substituted by the dynamic information function $\boldsymbol{\Pi}(\boldsymbol{\theta}^\intercal \boldsymbol{z_t})$, which is the mean function for the risky factors $\boldsymbol{\phi_t}$ as in Equation 3. In other words, the objective function Equation 9 is a transformation of the utility function Equation 13 by incorporating conditional variables $\boldsymbol{z_t}, \boldsymbol{X}$.

Our purpose is to maximize the conditional expectation of the investor's utility function. The investment allocation to the j^{th} factor-tilt sub-portfolios is determined by the j^{th} information indicator $\pi_j(\boldsymbol{\theta}^\intercal \boldsymbol{z_t})$, which is a single-index function, to avoid the problem of "curse of dimensionality" caused by fully nonparametric methods. We specify fixed linear combinations as information input in an unknown function $\pi_j(\cdot)$, as stated by Aït-sahalia and Brandt (2001), for at least two reasons. Statistically, this can achieve a better convergence rate for estimates, and economically, a univariate index value provides meaningful and convenient descriptions of current investment opportunities. Meanwhile, these index functions' effects on each sub-portfolio can be highly nonlinear, as documented by Aït-sahalia and Brandt (2001). Therefore, we do not specify the functional form of $\pi_j(\cdot)$, allowing a parametric index function to influence each sub-portfolio's weight nonparametrically.

To facilitate our estimation procedures, we approximate those unknown functions $\pi_j(\cdot)$ by orthonormal bases similar to Dong et al. (2016b). Their methods can allow the elements of information vector $\boldsymbol{z_t}$ to be nonstationary. As pointed by Gao et al. (2013) and Gao and Phillips (2013), conventional kernel estimation as in the Brandt (1999) and Aït-sahalia and Brandt (2001) method may not be workable due to the breakdown of limit theory, when $\boldsymbol{z_t}$ is a multivariate $I(1)$ process. In practice, some time series predictors are likely to be nonstationary, like the unemployment rate, inflation and exchange rates, among other economic indicators. Therefore, we apply a similar method as in the Dong et al. (2016b) to validate a more comprehensive application of our model.

Suppose all the link functions π_j belong to $L^2(\mathbb{R}) = \{f(x) : \int f^2(x)dx < \infty\}$. The Hermite function sequence $\{\mathcal{H}_i\}$ is an orthonormal basis in $L^2(\mathbb{R})$:

$$\mathcal{H}_i(x) = (\sqrt{\pi}2^i i!)^{-1/2} H_i(x) \exp(-\frac{x^2}{2}), \quad i \geqslant 0, \tag{14}$$

where $H_i(x)$ are Hermite polynomials orthogonal with density $\exp(-x^2)$. The orthogonality reads $\int H_i(x)H_j(x)dx = \delta_{ij}$, the Kronecker delta.

Therefore, any countinuous function $\pi_j(\cdot) \in L^2(\mathbb{R})$ can be expanded into a linear combination of orthogonal series:

$$\pi_j(\boldsymbol{\theta}^\intercal \boldsymbol{z_t}) = \sum_{l=0}^{\infty} \beta_{jl} \mathcal{H}_l(\boldsymbol{\theta}^\intercal \boldsymbol{z_t}). \tag{15}$$

We keep the first $L-1$ terms and leave the rest as approximate residues:

$$\pi_j(\boldsymbol{\theta}^\intercal \boldsymbol{z_t}) = \sum_{l=0}^{L-1} \beta_{jl}\mathcal{H}_l(\boldsymbol{\theta}^\intercal \boldsymbol{z_t}) + \psi(\boldsymbol{\theta}^\intercal \boldsymbol{z_t}), \tag{16}$$

where $\psi(\boldsymbol{\theta}^\intercal \boldsymbol{z_t})$ is the approximation residues.

Furthermore, all J dynamic indicator functions can be approximated (we assume the same truncation parameters for all functions for purposed of notation simplicity only):

$$\boldsymbol{\Pi}(\boldsymbol{\theta}^\intercal \boldsymbol{z_t}) = \boldsymbol{B}\boldsymbol{\mathcal{H}_L}(\boldsymbol{\theta}^\intercal \boldsymbol{z_t}) + \Psi(\boldsymbol{\theta}^\intercal \boldsymbol{z_t}), \tag{17}$$

where

$$\boldsymbol{B} = \begin{Bmatrix} \beta_{10} & \dots & \beta_{1(L-1)} \\ \dots & \dots & \dots \\ \beta_{J0} & \dots & \beta_{J(L-1)} \end{Bmatrix}, \quad \boldsymbol{\mathcal{H}_L}(\boldsymbol{\theta}^\intercal \boldsymbol{z_t}) = \begin{Bmatrix} \mathcal{H}_0(\boldsymbol{\theta}^\intercal \boldsymbol{z_t}) \\ \dots \\ \mathcal{H}_{L-1}(\boldsymbol{\theta}^\intercal \boldsymbol{z_t}) \end{Bmatrix},$$

and $\Psi(\boldsymbol{\theta}^\intercal \boldsymbol{z_t})$ is the approximation error.

Therefore, the objective function Equation 13 is transformed through replacing $\boldsymbol{\Pi}(\boldsymbol{\theta}^\intercal \boldsymbol{z_t})$ by $\boldsymbol{B}\boldsymbol{\mathcal{H}_L}(\boldsymbol{\theta}^\intercal \boldsymbol{z_{t-1}})$ as:

$$\arg\max_{\alpha,\boldsymbol{B},\boldsymbol{\Theta}} E[u(\alpha r_{ft} + (\boldsymbol{B}\boldsymbol{\mathcal{H}_L}(\boldsymbol{\theta}^\intercal \boldsymbol{z_{t-1}}))^\intercal \boldsymbol{Q_t}(\boldsymbol{X}))], \tag{18}$$

subject to

$$\|\boldsymbol{\theta}\|_2 = 1 \text{ and } \theta_1 > 0$$

$$\alpha + \sum_{j=1}^{J} \beta_{\mathrm{j}}^\intercal \mathcal{H}_{\mathbf{L}}(\boldsymbol{\theta}^\intercal \boldsymbol{z_{t-1}}) = 1$$

Theorem 3. *The restricted optimal portfolio weight function chosen by Equation 18 gives an approximately optimal portfolio.*

Theorem 3 demonstrates that, as the number of assets $n \to \infty$, our two-step procedure is approximately equivalent to Equation 9, which is the completely unrestricted asset-by-asset portfolio optimization because these two methods give the same expected utility asymptotically.

4 Methodology

This section illustrates procedures for estimating $\boldsymbol{\Gamma}$ in Equation 11, and $\boldsymbol{B}, \boldsymbol{\theta}$ in Equation 18.

We assume that the investor chooses Γ based on the following objective:

$$\hat{\Gamma} = \arg\min_{\Gamma} \sum_{j=1}^{J} E(b_j(\boldsymbol{X_i})^2) \tag{19}$$

subject to

$$E[Q_t(\boldsymbol{X})Q_t(\boldsymbol{X})^\intercal] = \boldsymbol{I}_J,$$

where $\boldsymbol{I}_J$ is a $J \times J$ identity matrix.

In words, we choose the linear combination coefficients $J \times J$ matrix to maximize the spread of the portfolio weights, specifically by minimizing the expected sum of squared portfolio weights, in the class of semiparametric functions of the characteristics, subject to an orthogonality constraint on the vector of sub-portfolios' returns. These portfolios are an econometrically-derived variant of the widely popular Small-Minus-Big (SMB) and High-minus-Low (HML) portfolios designed by Fama and French (1993) to capture the size-related and value-related return factors. Fama and French (1993) did not minimize the sum of squared portfolio weights as was done in Equation 19, but they instead set the portfolio weights using capitalization weight, which, in the highly diversified US equity market, have a very low sum-of-squared relative to the number of assets. Fama and French (1993) did not explicitly impose the orthogonality condition applied in Equation 19, but, as they noted, they chose their size and value breakpoints so that the portfolio returns would have very low correlation. The reason that we set the orthogonal constraint here is to diversify idiosyncratic risks further.

Next we show that Equation 19 can have a Langrangian solution. After expanding the constraint and under the independence between $\boldsymbol{F_t}$ and $\boldsymbol{\epsilon_t}$, we have:

$$\begin{aligned} E[Q_t(\boldsymbol{X})Q_t(\boldsymbol{X})^\intercal] &= E[(\tfrac{1}{n}\boldsymbol{\Gamma G(X)^\intercal G(X) F_t} + \tfrac{1}{n}\boldsymbol{\Gamma G(X)^\intercal \epsilon_t})(\tfrac{1}{n}\boldsymbol{F_t^\intercal G(X)^\intercal G(X)\Gamma^\intercal} + \tfrac{1}{n}\boldsymbol{\epsilon_t^\intercal G(X)\Gamma^\intercal})] \\ &= E(\boldsymbol{\Gamma}\tfrac{\boldsymbol{G(X)^\intercal G(X)}}{\boldsymbol{n}}\boldsymbol{F_t F_t^\intercal}\tfrac{\boldsymbol{G(X)^\intercal G(X)}}{\boldsymbol{n}}\boldsymbol{\Gamma^\intercal}) + \tfrac{1}{n^2}E(\boldsymbol{\Gamma G(X)^\intercal \epsilon_t \epsilon_t^\intercal G(X)\Gamma^\intercal}) \\ &\to \boldsymbol{\Gamma M^G E(F_t F_t^\intercal) M^G \Gamma^\intercal} + \tfrac{1}{n^2}\boldsymbol{\Gamma G(X)^\intercal E(\epsilon_t\epsilon_t^\intercal) G(X)\Gamma^\intercal} \end{aligned},$$

which is a quadratic form in $\boldsymbol{\Gamma}$.

As for the objective function, we have:

$$\sum_{j=1}^{J} E(b_j(\boldsymbol{X_i})^2) = \sum_{j=1}^{J} E(\boldsymbol{\Gamma^2 G(X_i)^2}),$$

which is linear in $\boldsymbol{\Gamma^2}$.

Therefore, we write this constrained optimization problem of sample analogues in Lagrangian form:

$$L(\boldsymbol{\Gamma}) = \frac{1}{n}\sum_{1}^{J}\sum_{i=1}^{n}\boldsymbol{\Gamma} G(\boldsymbol{X_i})^\intercal G(\boldsymbol{X_i})\boldsymbol{\Gamma}^\intercal - \boldsymbol{\Lambda}^\intercal vec((\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{Q_t}(\boldsymbol{X})\boldsymbol{Q_t}(\boldsymbol{X})^\intercal) - \boldsymbol{I}_J), \tag{20}$$

where Λ is the $\frac{1}{2}J(J+1)$ vector of Lagrangian multipliers, and vec is the vecterization of a matrix.

The optimal $\boldsymbol{\Gamma}$ and associated Lagrangian multipliers will solve the first order conditions:

$$\frac{\partial L}{\partial \boldsymbol{\Gamma}} = \boldsymbol{0}^{J\times J},$$

$$\frac{\partial L}{\partial \boldsymbol{\Lambda}} = \boldsymbol{0}^{\frac{1}{2}J(J+1)}.$$

Meanwhile, we collect the estimate $\hat{\mathbf{\Gamma}}$ to obtain the factor-mimicking sub-portfolios' returns as:

$$\hat{\boldsymbol{Q}}_{\boldsymbol{t}}(\boldsymbol{X}) = \begin{Bmatrix} \hat{q}_{1t}(\boldsymbol{X}) \\ \hat{q}_{2t}(\boldsymbol{X}) \\ \vdots \\ \hat{q}_{Jt}(\boldsymbol{X}) \end{Bmatrix} = \sum_{i=1}^{n} \hat{B}(\boldsymbol{X}_i) y_{it} = \frac{1}{n} \sum_{i=1}^{n} \hat{\mathbf{\Gamma}} \hat{G}(\boldsymbol{X}_{\boldsymbol{i}})^{\intercal} y_{it}, \tag{21}$$

where $\hat{G}(\boldsymbol{X}_{\boldsymbol{i}})$ is the consistent estimate of Equation 3 as in Li and Linton (2020), where they specify $\boldsymbol{G}(\boldsymbol{X}_{\boldsymbol{i}})$ as an additive semiparametric function of asset-specified characteristics.

The next step derives an estimator for the dynamic portfolio allocation weighting functions $\mathbf{\Pi}(\boldsymbol{\theta}^{\intercal}\boldsymbol{z}_{\boldsymbol{t-1}})$. The portfolio weight estimation problem is to find:

$$(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{B}}, \hat{\boldsymbol{\theta}}) = \arg\max_{\alpha, \boldsymbol{B}, \boldsymbol{\theta}} E[u(\alpha r_{ft} + (\boldsymbol{B}\boldsymbol{\mathcal{H}}_{\boldsymbol{L}}(\boldsymbol{\theta}^{\intercal}\boldsymbol{z}_{\boldsymbol{t-1}}))^{\intercal}\hat{\boldsymbol{Q}}_{\boldsymbol{t}}(\boldsymbol{X}))], \tag{22}$$

subject to

$$\|\boldsymbol{\theta}\|_2 = 1 \text{ and } \theta_1 > 0$$

$$\alpha + \sum_{j=1}^{J} \beta_{\mathrm{j}}^{\intercal} \mathcal{H}_{\mathbf{L}}(\boldsymbol{\theta}^{\intercal}\boldsymbol{z}_{\boldsymbol{t-1}}) = 1$$

where $\hat{\boldsymbol{Q}}_{\boldsymbol{t}}(\boldsymbol{X})$ is the estimate of sub-portfolios from Equation 21. This is essentially the same semiparametric estimation problem analyzed by Aït-sahalia and Brandt (2001). The procedure relies on the profile estimation of the single-index function. We iterate the first order condition to convergence after choosing initial values arbitrarily.

With respect to identification issues, we need to solve another constrained optimization problem as:

$$\arg\max_{\alpha, \boldsymbol{B}, \boldsymbol{\theta}} E[u(\alpha r_{ft} + (\boldsymbol{B}\boldsymbol{\mathcal{H}}_{\boldsymbol{L}}(\boldsymbol{\theta}^{\intercal}\boldsymbol{z}_{\boldsymbol{t-1}}))^{\intercal}\hat{\boldsymbol{Q}}_{\boldsymbol{t}}(\boldsymbol{X}))], \tag{23}$$

subject to

$$\|\boldsymbol{\theta}\|_2 = 1 \text{ and } \theta_1 > 0$$

$$\alpha + \sum_{j=1}^{J} \beta_{\mathrm{j}}^{\intercal} \mathcal{H}_{\mathbf{L}}(\boldsymbol{\theta}^{\intercal}\boldsymbol{z}_{\boldsymbol{t-1}}) = 1$$

Therefore, the first order condition of the maximization with respect to $\boldsymbol{B}, \boldsymbol{\theta}$ is:

$$E[\boldsymbol{M}_t] = \begin{Bmatrix} u'(\cdot)\hat{\boldsymbol{Q}}_{\boldsymbol{t}}(\boldsymbol{X}) \otimes \boldsymbol{\mathcal{H}}_{\boldsymbol{L}}(\boldsymbol{\theta}^{\intercal}\boldsymbol{z}_{\boldsymbol{t-1}})] = \mathbf{0}_{\boldsymbol{JL\times 1}} \\ u'(\cdot)\hat{\boldsymbol{Q}}_{\boldsymbol{t}}(\boldsymbol{X})\boldsymbol{I}_{\boldsymbol{L\times L}}(\boldsymbol{B}\boldsymbol{\mathcal{H}'}_{\boldsymbol{L}}(\boldsymbol{\theta}^{\intercal}\boldsymbol{z}_{\boldsymbol{t-1}}) \otimes \boldsymbol{z}_{\boldsymbol{t-1}})] = \mathbf{0}_{\boldsymbol{JL\times 1}} \\ u'(\cdot) r_{ft} = 0 \end{Bmatrix},$$

where $\boldsymbol{I}_{\boldsymbol{L\times L}}$ is a $L \times L$ identity matrix while $\boldsymbol{\mathcal{H}'}_{\boldsymbol{L}}$ and $u'(\cdot)$ are the first derivatives of the truncated orthonormal series and the investor's utility function respectively.

As we can see, there are $2 \times \boldsymbol{JL} + 1$ moment conditions to maximize the objective function.

These moment conditions can be used to construct standard GMM problem as was done in Hansen (1982):

$$(\alpha, \hat{\boldsymbol{B}}, \hat{\boldsymbol{\theta}}) = \min_{\alpha, \boldsymbol{B}, \boldsymbol{\theta}} E[\boldsymbol{M_t}]^\intercal \boldsymbol{S} E[\boldsymbol{M_t}]$$

subject to

$$\|\boldsymbol{\theta}\|_2 = 1 \text{ and } \theta_1 > 0$$

$$\alpha + \sum_{j=1}^{J} \beta_{\mathrm{j}}^\intercal \mathcal{H}_{\mathbf{L}}(\boldsymbol{\theta}^\intercal \boldsymbol{z_{t-1}}) = 1$$

where $\boldsymbol{S}$ is the optimal weighting positive definite matrix as $\boldsymbol{S} = cov(\boldsymbol{M_t})^{-1}$.

Then, we substitute these moment conditions $E[\boldsymbol{M_t}]$ with corresponding sample counterparts as:

$$\boldsymbol{m_t} = \left\{ \begin{array}{c} \frac{1}{T} \sum_{t=1}^{T} u'(\cdot) \hat{\boldsymbol{Q}}_{\boldsymbol{t}}(\boldsymbol{X}) \otimes \boldsymbol{\mathcal{H}_{L-1}}(\boldsymbol{\theta}^\intercal \boldsymbol{z_{t-1}})] = \boldsymbol{0_{JL \times 1}} \\ \frac{1}{T} \sum_{t=1}^{T} u'(\cdot) \hat{\boldsymbol{Q}}_{\boldsymbol{t}}(\boldsymbol{X}) \boldsymbol{I_{L \times L}}(\boldsymbol{B}\boldsymbol{\mathcal{H}'_{L-1}}(\boldsymbol{\theta}^\intercal \boldsymbol{z_{t-1}}) \otimes \boldsymbol{z_{t-1}})] = \boldsymbol{0_{JL \times 1}} \\ u'(\cdot) r_{ft} = 0 \end{array} \right\}.$$

Similarly, we obtain the estimate of weighting as:

$$\hat{\boldsymbol{S}} = \frac{1}{T} \sum_{t=1}^{T} \left\{ \begin{array}{c} u'(\cdot) \hat{\boldsymbol{Q}}_{\boldsymbol{t}}(\boldsymbol{X}) \otimes \boldsymbol{\mathcal{H}_{L-1}}(\boldsymbol{\theta}^\intercal \boldsymbol{z_{t-1}})] = \boldsymbol{0_{JL \times 1}} \\ u'(\cdot) \hat{\boldsymbol{Q}}_{\boldsymbol{t}}(\boldsymbol{X}) \boldsymbol{I_{L \times L}}(\boldsymbol{B}\boldsymbol{\mathcal{H}'_{L-1}}(\boldsymbol{\theta}^\intercal \boldsymbol{z_{t-1}}) \otimes \boldsymbol{z_{t-1}})] = \boldsymbol{0_{JL \times 1}} \\ u'(\cdot) r_{ft} = 0 \end{array} \right\} \left\{ \begin{array}{c} u'(\cdot) \hat{\boldsymbol{Q}}_{\boldsymbol{t}}(\boldsymbol{X}) \otimes \boldsymbol{\mathcal{H}_{L-1}}(\boldsymbol{\theta}^\intercal \boldsymbol{z_{t-1}})] = \boldsymbol{0_{JL \times 1}} \\ u'(\cdot) \hat{\boldsymbol{Q}}_{\boldsymbol{t}}(\boldsymbol{X}) \boldsymbol{I_{L \times L}}(\boldsymbol{B}\boldsymbol{\mathcal{H}'_{L-1}}(\boldsymbol{\theta}^\intercal \boldsymbol{z_{t-1}}) \otimes \boldsymbol{z_{t-1}})] = \boldsymbol{0_{JL \times 1}} \\ u'(\cdot) r_{ft} = 0 \end{array} \right\}^\intercal.$$

Finally, we substitute the sample analogues and $\hat{\boldsymbol{S}}$ into the objective function, and estimate $\hat{\boldsymbol{B}}, \hat{\boldsymbol{\theta}}$:

$$(\hat{\alpha}, \hat{\boldsymbol{B}}, \hat{\boldsymbol{\theta}}) = \min_{\alpha, \boldsymbol{B}, \boldsymbol{\theta}} \boldsymbol{m}_{\boldsymbol{t}}^\intercal \hat{\boldsymbol{S}} \boldsymbol{m_t}, \tag{24}$$

subject to

$$\|\boldsymbol{\theta}\|_2 = 1 \text{ and } \theta_1 > 0$$

$$\alpha + \sum_{j=1}^{J} \beta_{\mathrm{j}}^\intercal \mathcal{H}_{\mathbf{L}}(\boldsymbol{\theta}^\intercal \boldsymbol{z_{t-1}}) = 1$$

Furthermore, we substitute Equation 24 into the optimization iteration, which is called the continuously-updating estimator; details can be found in Dong et al. (2018).

5 Hypothesis Tests

This section introduces the hypothesis tests that help us to understand which index variables are important to guide the construction of factor-timing portfolios. We apply a Wald test to infer the significance of θ_j. We have the null and alternative hypotheses as follows:

$$\mathcal{H}_0 : \boldsymbol{C\theta} = \mathbf{0}_{\boldsymbol{D}\times\mathbf{1}}, \quad \text{against} \quad \mathcal{H}_1 : \boldsymbol{C\theta} \neq \mathbf{0}_{\boldsymbol{D}\times\mathbf{1}},$$

where $\boldsymbol{C}$ is a $D \times K$ fix matrix indicating the number of constraints $\boldsymbol{D}$.

We denote the value of the objective function Equation 24 under $\hat{\boldsymbol{B}}, \hat{\boldsymbol{\theta}}$ as $\boldsymbol{\mathcal{V}}(\hat{\boldsymbol{B}}, \hat{\boldsymbol{\theta}})$ while under the null hypothesis $\mathcal{H}_0$ as $\boldsymbol{\mathcal{V}}(\hat{\boldsymbol{B}}, \hat{\boldsymbol{\theta}}^*)$.

Therefore, if the null hypothesis is correct, we have:

$$T(\boldsymbol{\mathcal{V}}(\hat{\boldsymbol{B}}, \hat{\boldsymbol{\theta}}^*) - \boldsymbol{\mathcal{V}}(\hat{\boldsymbol{B}}, \hat{\boldsymbol{\theta}})) \sim \boldsymbol{\chi}^2(\boldsymbol{D}), \tag{25}$$

where $\boldsymbol{\chi}^2(\boldsymbol{D})$ is the chi-square distribution with degree of freedom D. This method is a minimum-χ^2 test, the purpose of which is to check the minimized values of objective function Equation 24 after imposing some restrictions.

We reject the null hypothesis if the test statistic exceeds the critical value.

6 Empirical Study

6.1 Data Description

6.1.1 Index Variables

We use the same index variable set as Aït-sahalia and Brandt (2001). These variables are all at a monthly frequency:

- The **Default Spread** is the yield difference between Moody's Baa and Aaa rated bonds, observed from 1967-07-01 to 2017-06-01 (600 months in total) denoted as DS.
- The **Term Spread** is the yield difference between 10 and 1 year government bonds, observed from 1967-07-01 to 2017-06-01 (600 months in total) denoted as TS.
- The **Trend** is the difference between the log of the current S&P 500 index level and the log of the average index level over the previous 12 months, observed from 1967-07-01 to 2017-06-01 (600 months in total).
- The **Dividend Yield**, also called Dividend-to-Price, is the sum of dividends paid on the S&P 500 index over the past 12 months divided by the current level of the index observed from 1967-07-01 to 2017-06-01 (600 months in total). We use the percentage natural logarithm form of Dividend Yield, denoted as Ln(DY%).
- The **Risk Free** rate is obtained from the Fama-French factor model's risk-free rate, observed from 1967-07-01 to 2017-06-01, denoted in the percentage form as RF%.

In Table 1, both "Trend" and "RF%" have small variation while "RF%" has some strong correlation with "TS" and " Ln(DY%)". Apart from these, we also find that all of the index variables are not symmetrically distributed, which is shown by the non-zero skewness. As for the kurtosis, the table indicates that outliers are quite common among these variables.

In Table 2, we conclude the results of the unit root test and autocorrelation. After the Dickey-Fuller tests, we fail to reject the null hypotheses that there are no unit roots among all index variables, especially for the " Ln(DY%)". That can also be found from Figure 4. In terms of autocorrelation, almost all of the index series present persistent autocorrelation even for lag nine, " Ln(DY%)" showing a strong signal of autocorrelation coefficient of 0.94, as shown in Figure 4, Figure 5. However, "Trend" is an exception, where the autocorrelation decays to zero and is negative after lag ten as shown in Figure 3. These test results verify the necessity of applying orthogonal series to approximate the single index function with nonstationary covariates, as in subsection 3.3.

The data above was collected from the websites of FRED and Multpl.

6.2 Monthly Stock Data

We collected monthly stock returns from CRSP and firms' characteristics from Compustat, from 1965 to 2017. We constructed 33 characteristics following the methods of Freyberger et al. (2020). Details of these characteristics can be found in the appendix of Li and Linton (2020). We construct characteristics from fiscal year $t-1$ to explain stock returns between July of year t to June of year $t+1$. Following Hou et al. (2015), we adjust returns of delisted stocks. The method that we apply to estimate the Equation 4 is similar to Li and Linton (2020). We only include firms with at least three years of data in Compustat. The values of firm-specific characteristics are updated annually as most characteristic data are reported every year. We use rolling windows to accommodate these characteristics-based loadings and the risk factors are estimated correspondingly.

The time span of our in-sample analysis is 50 years, from July 1967 to June 2017 (600 months).

6.3 In-sample Factor-mimicking Portfolios

This section presents portfolios that mimic the annually updated risk factors estimated through Equation 5. In this study, we choose the number of unobservable factors in Equation 3 to be three. In Li and Linton (2020), they compared the effects of the number of factors through a simulation study, concluding that underestimating the number of factors can be problematic. However, their discussions mainly focused on the estimation of mispricing functions. We only have four dynamic index variables, and therefore, we follow the renowned research of Fama and French (1993) to set the number of factors to be three. According to the literature, three factors can capture the most important common variation in asset excess returns.

The methods are introduced in subsection 3.2, and we utilize all 600 months of data and construct three such portfolios every year, assuming the number of risk factors in Equation 3 to be three. Then, we conclude the descriptive statistics of these three sub-portfolios in Table 3. The zero mean and unit variance are determined by the constraints in estimation. As for the correlation between risk factors and those sub-portfolios using observations of all 600 months, our first step works very well because the diagonal elements of correlation are quite high while the off-diagonal elements are negligible. That demonstrates that each sub-portfolio imitates only the target risk factor's variation accurately and leaves the rest uncorrelated. The weights put on each asset for these sub-portfolios are calculated through a constrained optimization, which restricts the similarity between sub-portfolios and risk factors. Furthermore, during certain years, the sub-portfolios behaved in the opposite direction of the imitated factor, which can also influence the average correlation over 600 months. The annual correlation can be found Table 7, where some negatively correlated periods are presented.

6.4 Utility Function

We utilitze the classic Constant Relative Risk Aversion (CRRA) utility function to model function $u(W)$ in Equation 8:

$$u(W) = \begin{cases} \frac{W^{1-\xi}}{1-\xi} & \text{if } \xi > 1; \\ \ln(W) & \text{if } x = 1, \end{cases}$$

where ξ is an integer and $\xi = W\frac{\partial^2 u(W)/\partial W^2}{\partial u(W)/\partial W}$, measuring the level of risk aversion. Therefore, under this setting, the investor is risk-averse and tries to maximize his expected utility function through factor-mimicking and factor-timing portfolio strategy. The CRRA utility function is twice differentiable, which can further facilitate our optimization algorithm.

6.5 Selection of Truncation Number

The value of L in Equation 16, which is the truncation number in polynomials, needs to be determined here. Unfortunately, to the best of our knowledge, there is no rule of thumb for the best choice of L. We refer to Dong et al. (2015) and Dong et al. (2016a), where the authors determined L according to the number of observations n. However, the n in this study ranged from 468 to 2928. After trading off the computation burden and approximation accuracy, we choose L to be four throughout the empirical study.

Table 1: Index Variable Summary

Index name	T	Descriptive Statistics Mean	Variance	Median	Max	Min	Skewness	Kurtosis	Correlation Matrix DS	TS	Trend	Ln(DP%)	RF%
DS	600	1.08	0.2	0.94	3.38	0.55	1.82	7.34	1.00				
TS	600	1.12	1.39	1.23	3.40	-3.07	-0.32	2.72	0.09	1.00			
Trend	600	0.03	0.01	0.05	0.22	-0.4	-1.24	5.63	-0.27	0.07	1.00		
Ln(DY%)	600	1.00	0.17	1.05	1.83	0.10	-0.09	2.07	0.46	-0.26	-0.12	1.00	
RF%	600	0.39	0.08	0.41	1.35	0.00	0.51	3.44	0.23	-0.68	-0.01	0.65	1.00

This table documents the descriptive statistics of the index variables that are used in this empirical study as well as the correlations among them. To be consistent with most of the literature, we use the percentage values of DP and RF.

Table 2: Tests Summary

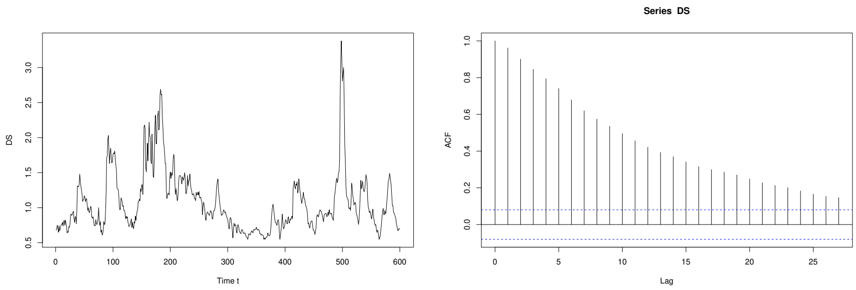
Index name	T	Unit root test time trend	p-value (Trend)	ΔY_t	p-value (ΔY_t)	Autocorrelation ρ_3	ρ_6	ρ_9
DS	600	-2.3×10^{-5}	-0.81	-3.82×10^{-2}	-3.43	0.85	0.68	0.54
		(2.8×10^{-5})		(1.11×10^{-2})				
TS	600	9.27×10^{-5}	1.23	-3.6×10^{-2}	-3.26	0.88	0.77	0.69
		(7.51×10^{-5})		(1.1×10^{-2})				
Trend	600	4.1×10^{-6}	0.47	-7.59×10^{-2}	-4.86	0.70	0.37	0.12
		(8.75×10^{-6})		(1.56×10^{-2})				
Ln(DY%)	600	-1.8×10^{-5}	-1.44	-8.86×10^{-3}	-1.72	0.98	0.96	0.94
		(1.25×10^{-5})		(5.16×10^{-3})				
RF%	600	-6.6×10^{-5}	-3.22	-5.48×10^{-2}	-4.23	0.94	0.90	0.87
		(2.05×10^{-5})		(5.16×10^{-3})				

This table summarizes the results of unit root tests and autocorrelations of those index variables. It reports the estimates, standard errors (in parentheses) and t-statistics of Dickey-Fuller test with trend individually. Autocorrelation column illustrates the correlation between the series and lag 3, lag 6 and lag 9 respectively, dennoted as ρ_3, ρ_6, ρ_9.

Table 3: Factor-mimicking Portfolios Summary

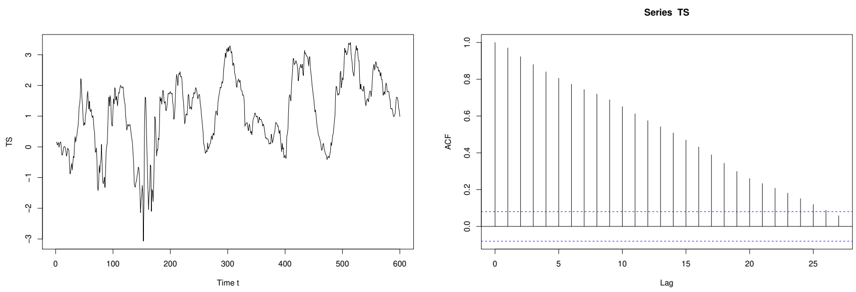
Index name	T	Descriptive Statistics Mean	Variance	Median	Max	Min	Skewness	Kurtosis	Average Correlation $\hat{\boldsymbol{f}}_1$	$\hat{\boldsymbol{f}}_2$	$\hat{\boldsymbol{f}}_3$
$\hat{\boldsymbol{q}}_1$	600	0.00	1.00	0.08	2.81	-3.10	-0.38	3.33	0.48	-0.01	0.13
$\hat{\boldsymbol{q}}_2$	600	0.00	1.00	0.03	2.70	-3.20	-0.15	3.42	-0.11	0.59	0.05
$\hat{\boldsymbol{q}}_3$	600	0.00	1.00	0.05	2.60	-2.79	-0.10	2.82	0.06	-0.03	0.63

This table presents the descriptive statistics of factor-mimicking portfolios and their correlations with estimated risk factors. $\hat{\boldsymbol{q}}_1$, $\hat{\boldsymbol{q}}_2$, and $\hat{\boldsymbol{q}}_3$ are constructed portfolios through all 600 months' data while $\boldsymbol{f}_1$, $\boldsymbol{f}_2$, and $\boldsymbol{f}_3$ are three $T \times 1$ factors estimated by rolling windows.



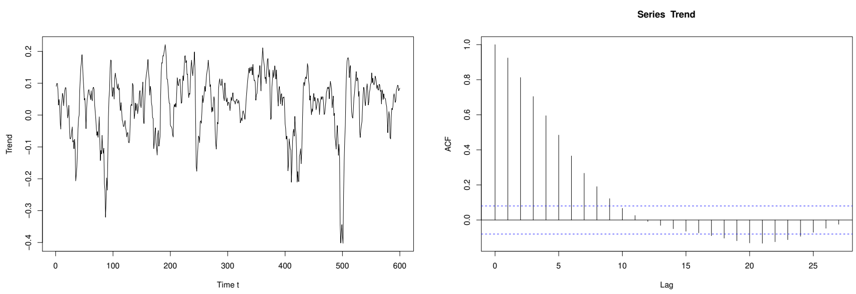
(a) Series Plot of DS (b) Autocorrelation Plot of DS

Figure 1: The Plot of DS



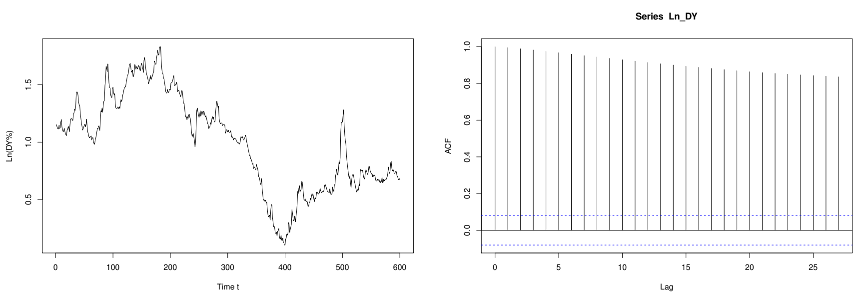
(a) Series Plot of TS (b) Autocorrelation Plot of TS

Figure 2: The Plot of DS



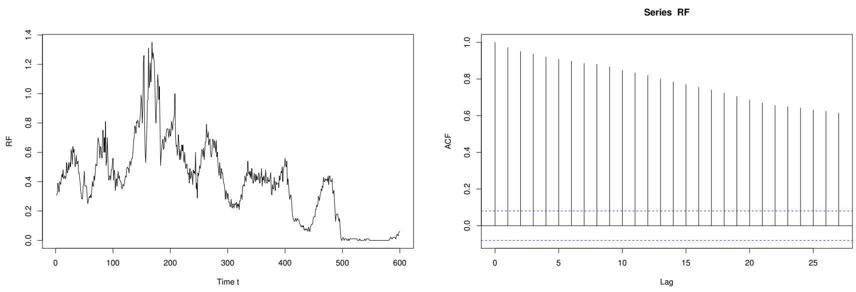
(a) Series Plot of Trend (b) Autocorrelation Plot of Trend

Figure 3: The Plot of Trend



(a) Series Plot of Ln(DY%) (b) Autocorrelation Plot of Ln(DY%)

Figure 4: The Plot of Ln(DY%)



(a) Series Plot of RF (b) Autocorrelation Plot of RF

Figure 5: The Plot of RF

6.6 Estimation of Dynamic Signals

This section presents the estimates of single-index coefficient vector $\boldsymbol{\theta}$ and the results of the corresponding hypothesis tests. We use the CRRA utility function with various risk aversion levels ξ. Meanwhile, we also test the null hypothesis in section 5 to examine whether some of the coefficients of dynamic variables are significantly different from 0. These tests are used to show the importance of this dynamic information during the second step of portfolio management, namely factor-timing.

During our estimation, all the optimization processes converged, and the optimized values are reported. The in-sample results are based on the data for all 600 months and the estimation procedures are repeated under different risk-aversion levels $\xi = 2, \xi = 5$, and $\xi = 10$. We then obtain the values of the objective function Equation 24, denoted as $\boldsymbol{\mathcal{V}}(\hat{\boldsymbol{B}}, \hat{\boldsymbol{\theta}})$. The hypothesis tests are conducted by setting $\theta_i = 0$, where i indicates the i^{th} index variable. We denote the value of the objective function Equation 24 under $\mathcal{H}_0 : \theta_i = 0$ as $\boldsymbol{\mathcal{V}}(\hat{\boldsymbol{B}}, \hat{\theta}_i = 0)$. In addition, χ^2 statistics are calculated as $T\Delta\boldsymbol{V} = T(\boldsymbol{\mathcal{V}}(\hat{\boldsymbol{B}}, \hat{\theta}_i = 0) - \boldsymbol{\mathcal{V}}(\hat{\boldsymbol{B}}, \hat{\boldsymbol{\theta}}))$.

Table 4: Index Variable Summary

		$\xi=2$				$\xi=5$				$\xi=10$			
Index name	T	$\hat{\theta}_i$	$\boldsymbol{\mathcal{V}}(\hat{\boldsymbol{B}}, \hat{\theta}_i=0)$	$\boldsymbol{\mathcal{V}}(\hat{\boldsymbol{B}}, \hat{\boldsymbol{\theta}})$	$T\Delta\boldsymbol{V}$	$\hat{\theta}_i$	$\boldsymbol{\mathcal{V}}(\hat{\boldsymbol{B}}, \hat{\theta}_i=0)$	$\boldsymbol{\mathcal{V}}(\hat{\boldsymbol{B}}, \hat{\boldsymbol{\theta}})$	$T\Delta\boldsymbol{V}$	$\hat{\theta}_i$	$\boldsymbol{\mathcal{V}}(\hat{\boldsymbol{B}}, \hat{\theta}_i=0)$	$\boldsymbol{\mathcal{V}}(\hat{\boldsymbol{B}}, \hat{\boldsymbol{\theta}})$	$T\Delta\boldsymbol{V}$
DS	600	0.15	0.23	0.006	134.4	0.06	0.0001	2.5×10^{-8}	0.066	0.08	0.0026	4.9×10^{-10}	1.56
TS	600	0.19	0.064	0.006	34.8	-0.34	0.0008	2.5×10^{-8}	0.481	-0.06	5.2×10^{-9}	4.9×10^{-10}	0
Trend	600	0.03	0.01	0.006	2.4	0.06	1×10^{-6}	2.5×10^{-8}	0.006	0.03	8.1×10^{-10}	4.9×10^{-10}	0
Ln(DY%)	600	-0.97	0.06	0.006	32.4	-0.93	3.2×10^{-6}	2.5×10^{-8}	0.002	-0.995	2×10^{-8}	4.9×10^{-10}	0

This table reports the estimates and hypothesis test of dynamic index variables. $\hat{\theta}_i$ is the estimate of the coefficient of the i^{th} index variable while $\boldsymbol{V}$ represent the value of the objective function. $\Delta\boldsymbol{V} = \boldsymbol{\mathcal{V}}(\hat{\boldsymbol{B}}, \hat{\theta}_i = 0) - \boldsymbol{\mathcal{V}}(\hat{\boldsymbol{B}}, \hat{\boldsymbol{\theta}})$.

The findings in Table 4 differ across the risk-aversion levels. When the magnitude of risk aversion is low, the influence of the dynamic index variables is significant. With $\xi = 2$, nearly all of the values of $T\Delta\boldsymbol{V}$ exceed the 95% critical value of $\chi^2(1)$, which is 3.84, except for DS. As the risk-aversion becomes larger, the importance of these dynamic variables declines. This can be confirmed when $\xi = 5$ and $\xi = 10$, where all the four variables are insignificant. We compare the values of the objective function Equation 24, and most of them are quite similar, both close to zero. That means the moment conditions in Equation 24 can be satisfied even if we restrict the coefficient of the i^{th} index variable to zero. Nevertheless, we cannot reject their joint significance.

6.7 In-sample Performance of Factor-timing Portfolios

This section presents portfolio performance estimated using in-sample data. As mentioned previously, there are two steps in the construction of our dynamic portfolio, namely, factor-tilt and factor-timing steps. In subsection 6.3, we describe how to build the sub-portfolios that mimic the behavior of risk factors. This section solves the second step, factor-timing: choosing the time-varying weights for the risk-free asset and risky sub-portfolios. The dynamic weights are determined by a single-index function with a set of index variables. These variables capture investment opportunities. We standardize the amount of investment to be 1 unit and take the monthly returns as the wealth gleaned by the investor. We do not restrict leverage or short-selling in order to check the influence of the risk-aversion level.

As we have 600 months in total, we record the average returns every year and annual standard deviations in Table 5 to save the space, and we calculate the Sharpe-ratio directly through $mean(Return_t)/SD_{annual}$. Table 5 shows the in-sample results from July 1967 to June 2017 under all three risk-aversion levels defined in subsection 6.4, and these results are compared with monthly S&P 500 returns.

Some findings here are significant and worth discussing. Firstly, for investors who have relatively lower risk-aversion, the average portfolio returns are more rewarding, with some extremely high returns appearing as well. For example, when the risk-aversion level $\xi = 2$, the twelve month average monthly returns can be 10.61and 8. 65. As for $\xi = 10$, the average monthly returns are more normal. Most monthly returns are around 5$\%$ except for some outliers. Secondly, a higher risk-aversion level corresponds to more volatile returns, such as losing -2.33 monthly during the whole year when $\xi = 2$, provided the standard deviation of the monthly return is 6.44. But the circumstances can be much more favourable when ξ increases to 5 and 10, with the standard deviation of the monthly return being 3.98 and 2.81, respectively. Especially under $\xi = 10$, the returns are quite acceptable and stable. Thirdly, all of the portfolios under various ξ have a relatively low Sharpe-ratio, compared with S&P 500 returns, which may be due to the high volatility.

In this empirical study, we optimize over three risky sub-portfolios and one risk-free asset without restricting leveraging or short-selling, and the weights for each asset are plotted in Figure 6. As we can see in (a), when the relative risk aversion level is low, $\xi = 2$, the weights for each asset are variable, while the scale of the vertical axis here is wider than (b) and (c). As we increase ξ, the weights become more stable. Specifically, when the ξ increases to 10, the only substantial volatility in the weights appeared around the stock crash in March 2000.

6.8 Out-sample Performance of Factor-timing Portfolios

This section examines the out-sample performance of our two-step portfolio selection procedure. We test the last six months of the last ten years in our data set for various risk aversion levels. The coefficients of the dynamic information function are estimated using all of the past information while the sub-portfolios are estimated using the first six months each year. The "Return" in Table 6 is calculated by substituting the predictors observed at the beginning of time $t+1$. The sub-portfolios are constructed at time t, based on all the available data at the target year before time $t+1$. Table 6 also lists the assigned weights to each sub-portfolios and the risk-free asset using 1 unit of investment, represented by c_1,c_2,c_3 and c_0. To summarize each column, we also provide the mean and standard deviation values at the end of the table, indicated by "ColMeans" and "ColStd".

As we can see from Table 6, most of the out-sample performance is quite similar to the in-sample performance Table 5. When the risk-aversion level is low such as $\xi = 2$, the variation of assets' weights is the largest and with an extensive range. Correspondingly, the realized monthly returns are also variable and high on average. The mean return of all 60 months is 0.36, which is very similar to that of the in-sample result which is 0.37. Not surprisingly, the out-sample standard deviation 8.88 is bigger than that of the in-sample result (6.44).

When the risk-aversion level increases to $\xi = 5$, the weights' volatility decreases, and the mean return also falls from 0.36 to 0.27, which is similar to the in-sample result (0.29). Compared with the $\xi = 2$ situation, the standard deviation of all the assigned weights and the monthly returns decline.

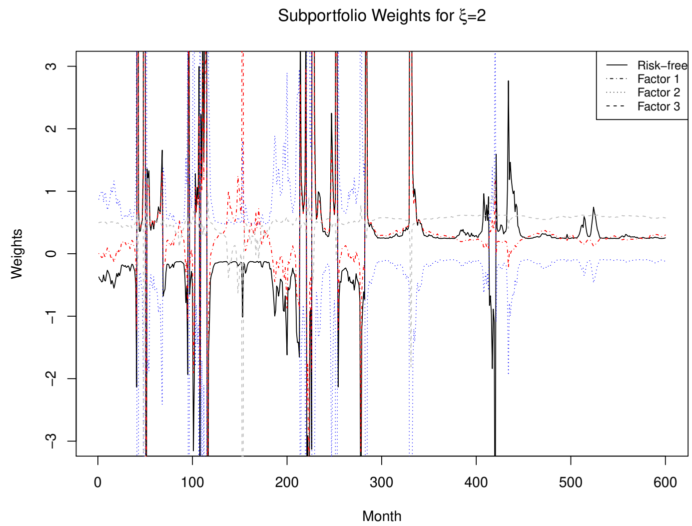
In the case of $\xi = 10$, all of the weights and monthly returns become more stable and less volatile. However, the average monthly return here is much lower than the in-sample result (0.17), with a smaller standard deviation of 1.51.

From the above analysis, we can conclude that our out-sample results are robust and vary according to the risk aversion levels. When the risk-aversion level is low, the investor reassigns his weights broadly and frequently, with an high average monthly return but high volatility. As the risk aversion level increases, the investor adjusts his weights more moderately, and the monthly average return and its standard deviation are reduced.

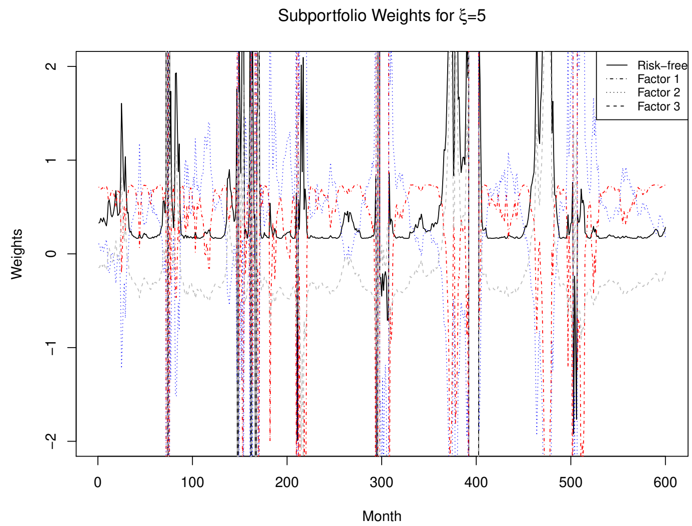
Table 5: Average Annual In-sample Results

	$\xi = 2$			$\xi = 5$			$\xi = 10$			Average Monthly S&P 500		
n	Return	SD	Sharpe-ratio	Return	SD	Sharpe-ratio	Return	SD	Sharpe-ratio	return	SD	Sharpe-ratio
468	-0.13	1.07	-0.12	0.15	0.70	0.22	0.06	0.61	0.10	0.01	0.03	0.01
894	-0.15	1.07	-0.14	0.20	0.47	0.41	0.08	0.63	0.13	-0.00	0.03	-0.00
1108	-0.13	1.11	-0.11	0.45	0.66	0.69	0.06	0.72	0.09	-0.02	0.04	-0.02
1199	-2.33	9.36	-0.25	0.10	0.88	0.12	0.04	0.59	0.08	0.02	0.03	0.02
1333	10.61	37.73	0.28	0.09	0.74	0.12	0.04	0.80	0.04	0.01	0.03	0.01
1409	-0.23	1.38	-0.16	0.10	0.77	0.13	0.06	0.82	0.08	-0.00	0.03	-0.00
1466	-0.16	1.03	-0.16	-1.02	3.50	-0.29	0.08	0.70	0.12	-0.01	0.04	-0.01
1560	-0.10	1.12	-0.09	0.35	1.19	0.30	0.05	0.70	0.07	0.00	0.07	0.00
1494	0.19	2.89	0.06	0.32	1.25	0.26	0.06	0.59	0.10	0.01	0.04	0.01
1292	0.52	7.48	0.07	0.14	1.29	0.11	0.03	0.60	0.05	-0.00	0.02	-0.00
1393	-0.12	0.73	-0.16	0.08	0.92	0.09	0.03	0.58	0.06	-0.00	0.03	-0.00
1340	-0.10	0.84	-0.12	0.38	0.94	0.41	0.05	0.79	0.06	0.00	0.03	0.00
1285	0.39	2.92	0.13	-0.13	5.27	-0.02	0.09	0.61	0.16	0.01	0.04	0.01
1181	-0.12	0.76	-0.16	-0.68	8.15	-0.08	0.03	0.57	0.06	0.01	0.03	0.01
1110	-0.15	0.83	-0.18	0.82	2.08	0.39	0.05	0.75	0.07	-0.01	0.04	-0.01
1044	-0.09	1.34	-0.07	-0.05	1.64	-0.03	-0.07	0.59	-0.11	0.04	0.03	0.04
1125	-0.73	1.72	-0.42	0.12	1.15	0.10	0.05	0.75	0.07	-0.01	0.02	-0.01
2192	-0.88	1.83	-0.48	-1.83	13.38	-0.14	0.05	0.73	0.06	0.02	0.03	0.02
2236	8.65	16.35	0.53	-0.87	2.35	-0.37	0.06	0.77	0.07	0.02	0.03	0.02
2273	0.07	0.53	0.13	0.08	0.87	0.09	0.05	0.95	0.05	0.02	0.03	0.02
2235	0.60	2.48	0.24	0.02	0.63	0.04	0.04	0.73	0.05	-0.01	0.06	-0.01
2270	-0.02	1.15	-0.02	0.28	0.95	0.30	0.09	0.58	0.16	0.02	0.02	0.02
2405	-0.33	1.19	-0.28	0.24	0.82	0.29	0.09	0.61	0.15	0.01	0.02	0.01
2376	2.03	3.87	0.53	0.02	1.00	0.02	0.03	0.63	0.05	0.01	0.05	0.01
2323	0.09	0.62	0.14	3.80	9.31	0.41	0.05	0.66	0.08	0.01	0.02	0.01
2344	0.05	0.68	0.08	2.45	3.21	0.76	0.03	0.61	0.05	0.01	0.01	0.01
2434	0.06	0.69	0.09	0.03	1.07	0.03	0.05	0.63	0.08	0.00	0.01	0.00
2548	-0.87	11.71	-0.07	-0.04	0.95	-0.04	0.10	0.56	0.17	0.01	0.02	0.01
2741	0.26	1.06	0.25	0.15	0.59	0.25	0.15	0.69	0.21	0.02	0.02	0.02
2928	0.10	0.55	0.19	0.14	0.74	0.18	0.25	0.53	0.47	0.02	0.04	0.02
2894	0.10	0.68	0.14	-0.03	1.90	-0.01	0.38	0.67	0.57	0.02	0.03	0.02
2905	0.11	0.73	0.16	-0.14	2.53	-0.05	0.25	0.64	0.39	0.02	0.05	0.02
2804	0.13	0.89	0.15	2.60	7.64	0.34	7.55	18.53	0.41	0.01	0.03	0.01
2570	0.10	0.69	0.14	-1.02	4.43	-0.23	0.95	1.64	0.58	-0.01	0.04	-0.01
2516	0.26	1.33	0.20	0.11	0.94	0.12	0.07	0.61	0.12	-0.02	0.05	-0.02
2491	0.11	0.89	0.13	0.07	1.05	0.07	0.03	0.60	0.05	-0.00	0.05	-0.00
2402	0.20	0.97	0.20	0.07	0.97	0.07	-0.04	0.58	-0.08	0.01	0.02	0.01
2326	0.06	0.90	0.07	0.01	0.50	0.03	0.11	0.58	0.18	0.01	0.02	0.01
2241	0.06	0.69	0.09	0.46	0.96	0.48	0.13	0.59	0.23	0.00	0.02	0.00
2178	0.12	0.73	0.17	7.04	15.93	0.44	0.15	0.60	0.25	0.02	0.02	0.02
2113	0.05	0.68	0.07	0.14	0.66	0.20	0.11	0.57	0.18	-0.01	0.04	-0.01
2023	-0.00	0.57	-0.00	-0.86	2.32	-0.37	0.04	0.85	0.04	-0.03	0.08	-0.03
2007	0.03	0.53	0.05	0.23	1.87	0.13	0.02	0.69	0.03	0.01	0.04	0.01
1924	-0.02	0.87	-0.02	0.02	1.14	0.02	-0.04	0.75	-0.05	0.01	0.02	0.01
1990	0.01	0.88	0.01	0.04	0.66	0.07	0.03	0.80	0.03	0.00	0.04	0.00
1937	0.00	0.89	0.00	-0.02	0.59	-0.03	0.01	0.79	0.01	0.02	0.02	0.02
1909	0.01	0.89	0.01	-0.02	0.94	-0.02	-0.01	0.70	-0.01	0.02	0.01	0.02
1872	0.00	0.69	0.01	-0.03	0.93	-0.04	-0.00	0.63	-0.00	0.01	0.02	0.01
1841	0.00	0.70	0.01	0.05	0.93	0.06	0.00	0.63	0.01	0.00	0.04	0.00
1826	0.01	0.70	0.02	-0.02	0.90	-0.02	-0.01	0.61	-0.01	0.01	0.01	0.01
Total mean	0.37	6.44	0.06	0.29	3.98	0.07	0.23	2.81	0.08	0.01	0.04	0.17

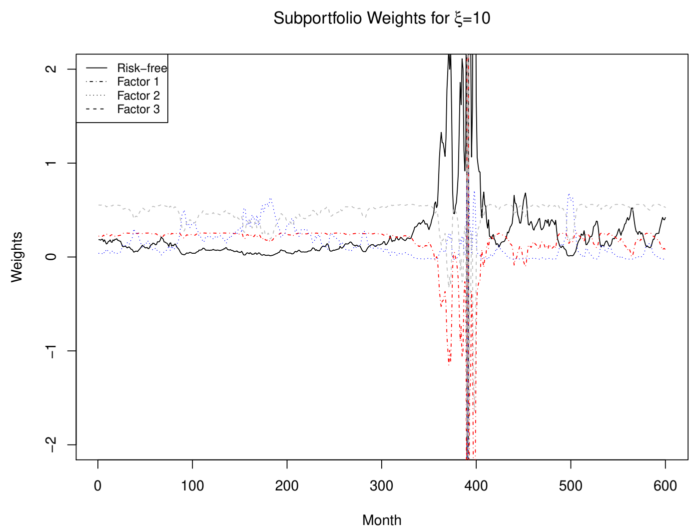
This table illustrates the in-sample results under various risk-aversion levels annually from July 1967- June 2017. n represents the number of stocks included in the portfolio. Both returns and standard deviations are calculated based on each year's results. The results of monthly S&P 500 returns are reported for comparison. No restrictions on leverage or short-selling.



(a) Monthly Weights Change Under $\xi = 2$



(b) Monthly Weights Change Under $\xi = 5$



(c) Monthly Weights Change Under $\xi = 10$

Figure 6: The Plot of Subportfolio Weights

Table 6: Monthly Out-sample Results Comparison

	$\xi=2$					$\xi=5$					$\xi=10$				
Year	c_0	c_1	c_2	c_3	Return	c_0	c_1	c_2	c_3	Return	c_0	c_1	c_2	c_3	Return
2008	1.12	0.04	0.05	-0.21	-0.04	-0.14	0.59	0.22	0.32	0.12	0.07	0.73	0.01	0.20	0.65
	2.03	-0.13	-0.41	-0.49	0.21	-0.11	0.63	0.18	0.30	0.60	0.11	2.12	-0.37	-0.86	2.67
	2.41	-0.21	-0.64	-0.55	0.67	-0.23	0.89	-0.09	0.43	0.22	-0.14	-2.37	1.00	2.51	0.84
	2.31	-0.19	-0.58	-0.54	-0.82	-0.20	0.82	-0.02	0.40	0.69	-0.14	-2.31	0.99	2.46	4.20
	2.24	-0.17	-0.54	-0.53	-1.28	-0.14	0.69	0.12	0.33	1.35	0.62	11.15	-3.11	-7.65	-6.09
	1.85	-0.09	-0.31	-0.45	0.77	-0.11	0.63	0.18	0.30	0.49	0.11	2.02	-0.34	-0.79	3.93
2009	-0.00	1.47	0.79	-1.26	-1.60	0.15	-1.56	2.56	-0.15	6.52	-0.29	0.20	0.24	0.85	2.69
	-0.00	1.45	0.83	-1.28	0.73	0.10	-1.08	2.03	-0.06	-0.30	-0.26	0.31	0.03	0.93	0.31
	-0.00	1.46	0.81	-1.28	-1.44	0.11	-1.15	2.11	-0.07	0.92	-0.26	0.30	0.04	0.92	-0.44
	-0.00	1.49	0.77	-1.25	-0.86	0.06	-0.66	1.60	-0.00	-1.90	-0.25	0.39	-0.13	1.00	-1.87
	-0.00	1.45	0.83	-1.28	-0.05	0.03	-0.25	1.19	0.03	1.21	-0.26	0.67	-0.66	1.25	0.82
	-0.00	1.39	0.94	-1.33	-0.48	0.02	0.03	0.93	0.03	0.33	-0.39	1.86	-2.80	2.33	-0.63
2010	-0.00	0.22	0.02	0.75	-0.21	-0.01	2.06	-0.73	-0.31	-1.17	-0.00	2.17	-0.84	-0.33	-1.23
	-0.00	0.23	0.02	0.75	0.50	-0.01	2.04	-0.72	-0.31	-0.00	-0.00	2.18	-0.84	-0.34	-0.05
	-0.00	0.23	0.02	0.75	-0.72	-0.01	2.07	-0.74	-0.31	-2.10	-0.00	2.17	-0.84	-0.33	-2.17
	-0.00	0.22	0.03	0.75	0.31	-0.01	1.97	-0.67	-0.30	-2.02	-0.00	2.18	-0.84	-0.34	-2.31
	-0.00	0.27	-0.01	0.74	-0.86	-0.01	2.11	-0.78	-0.32	-1.49	-0.00	2.18	-0.83	-0.35	-1.51
	-0.00	0.30	-0.03	0.74	-0.71	-0.01	2.39	-1.03	-0.36	-1.00	-0.00	2.17	-0.84	-0.32	-0.97
2011	1.49	-0.90	0.61	-0.20	1.01	0.00	1.07	0.28	-0.35	-0.11	-0.02	0.37	0.66	0.00	0.36
	1.02	-0.91	1.10	-0.20	1.42	0.00	1.09	0.26	-0.35	-0.35	-0.02	0.37	0.64	0.01	0.12
	1.18	-0.92	0.96	-0.22	-1.19	0.00	1.06	0.28	-0.35	0.45	-0.03	0.39	0.62	0.02	0.09
	1.05	-0.92	1.07	-0.21	-1.56	0.00	1.08	0.27	-0.35	0.54	-0.03	0.39	0.62	0.02	0.05
	1.31	-0.92	0.83	-0.21	-2.11	0.00	1.04	0.31	-0.34	1.24	-0.03	0.40	0.61	0.02	0.73
	1.52	-0.89	0.56	-0.19	-0.78	0.00	1.01	0.33	-0.35	0.78	-0.03	0.43	0.56	0.04	0.70
2012	-33.41	5.01	8.21	21.20	53.27	0.04	0.75	0.17	0.04	1.64	0.19	0.12	0.30	0.39	0.96
	31.14	-4.95	-6.39	-18.80	-37.15	0.05	0.76	0.17	0.02	1.23	0.19	0.12	0.30	0.39	0.77
	22.59	-3.62	-4.46	-13.50	-6.39	0.05	0.76	0.17	0.03	0.30	0.17	0.12	0.33	0.39	0.16
	13.19	-2.17	-2.36	-7.67	13.97	0.05	0.77	0.17	0.01	-1.08	0.18	0.12	0.31	0.39	-0.93
	11.36	-1.88	-1.95	-6.53	9.04	0.07	0.79	0.17	-0.03	-0.80	0.23	0.12	0.26	0.38	-0.72
	14.33	-2.34	-2.61	-8.37	-4.71	0.07	0.79	0.17	-0.04	0.33	0.27	0.12	0.23	0.38	0.26
2013	1.06	-0.71	2.04	-1.40	0.44	0.05	1.18	-0.87	0.64	-0.30	0.09	0.19	0.36	0.36	-0.04
	1.40	-1.02	2.53	-1.91	-0.90	0.05	1.17	-0.85	0.63	0.43	0.09	0.19	0.36	0.36	0.03
	1.12	-0.76	2.12	-1.48	0.89	0.05	1.14	-0.77	0.58	-0.64	0.09	0.18	0.37	0.35	-0.17
	0.43	-0.12	1.10	-0.41	-0.26	0.05	1.12	-0.72	0.55	0.92	0.10	0.18	0.37	0.35	0.32
	1.09	-0.73	2.08	-1.43	-5.69	0.05	1.07	-0.60	0.48	2.90	0.10	0.17	0.40	0.33	0.15
	-0.42	0.65	-0.08	0.85	1.06	0.04	1.05	-0.53	0.44	0.94	0.10	0.17	0.40	0.33	0.46
2014	0.01	0.24	0.63	0.13	-1.20	-0.04	2.39	1.34	-2.68	-4.24	0.08	-0.76	1.23	0.45	-3.96
	0.01	0.25	0.63	0.12	0.02	-0.05	2.39	1.34	-2.68	0.29	0.09	-0.36	0.89	0.38	0.14
	0.01	0.25	0.63	0.12	0.35	-0.05	2.39	1.34	-2.68	0.43	0.09	-0.28	0.82	0.37	0.34
	0.01	0.25	0.63	0.12	0.75	-0.05	2.39	1.34	-2.68	1.40	0.09	-0.37	0.90	0.38	0.97
	0.01	0.26	0.62	0.11	1.29	-0.06	2.40	1.34	-2.69	1.29	0.11	0.00	0.59	0.29	1.15
	0.01	0.25	0.62	0.12	-0.01	-0.05	2.39	1.34	-2.68	0.33	0.10	-0.19	0.75	0.34	0.08
2015	0.37	0.36	1.09	-0.82	-1.66	-0.35	0.80	0.34	0.21	-0.13	0.40	0.20	0.29	0.12	-0.06
	0.50	0.54	1.42	-1.46	2.83	-1.35	4.17	-1.20	-0.62	4.29	0.37	0.20	0.24	0.20	-0.21
	0.51	0.55	1.43	-1.48	-0.58	-1.32	4.08	-1.16	-0.60	-0.60	0.36	0.20	0.22	0.22	-0.05
	0.42	0.43	1.21	-1.05	2.27	-0.56	1.51	0.01	0.04	3.29	0.39	0.19	0.27	0.15	1.19
	2.79	3.58	7.06	-12.44	-3.46	0.33	-1.46	1.36	0.77	-0.37	0.31	0.23	0.15	0.31	-0.09
	-1.06	-1.50	-2.36	5.92	2.20	0.17	-0.91	1.11	0.64	-0.17	0.28	0.27	0.10	0.35	-0.04
2016	1.26	-0.15	-0.86	0.75	-1.11	-0.01	1.45	-1.28	0.84	-0.05	0.00	0.09	-0.11	1.02	-0.86
	1.37	-0.29	-0.40	0.32	0.19	-0.01	1.12	-0.89	0.78	0.50	0.00	0.11	-0.10	0.99	0.37
	1.37	-0.30	-0.35	0.28	1.24	-0.01	1.09	-0.85	0.78	-1.09	0.00	0.11	-0.10	0.99	-0.01
	1.36	-0.28	-0.43	0.34	0.80	-0.01	1.03	-0.78	0.76	0.84	0.00	0.13	-0.10	0.97	1.02
	1.37	-0.29	-0.40	0.32	-0.18	-0.01	0.94	-0.67	0.74	-2.72	0.00	0.17	-0.09	0.92	-1.29
	1.37	-0.33	-0.19	0.15	-0.06	-0.01	0.85	-0.56	0.72	2.38	0.00	0.17	-0.09	0.92	0.18
2017	-0.00	-1.51	2.14	0.37	-0.92	0.60	-1.03	-2.37	3.80	-0.65	0.08	0.25	0.46	0.21	0.04
	-0.00	-1.43	2.06	0.37	0.23	0.59	-1.00	-2.32	3.73	0.01	0.10	0.21	0.51	0.18	0.23
	-0.00	-1.09	1.72	0.37	1.21	0.34	-0.24	-1.03	1.93	-0.14	0.14	0.10	0.65	0.11	0.53
	-0.00	-0.71	1.35	0.37	1.56	0.25	0.10	-0.40	1.06	0.52	0.21	-0.08	0.87	0.00	0.85
	-0.00	-0.60	1.23	0.37	1.18	0.24	0.16	-0.27	0.87	-0.08	0.26	-0.20	1.01	-0.07	1.18
	-0.00	-0.30	0.93	0.37	0.28	0.24	0.25	-0.02	0.54	0.07	0.60	-1.08	2.03	-0.55	0.45
ColMean	1.58	-0.18	0.47	-0.88	0.36	-0.02	0.95	0.03	0.04	0.27	0.08	0.53	0.15	0.25	0.07
Colstd	7.15	1.42	2.03	4.75	8.88	0.30	1.18	1.02	1.21	1.58	0.19	1.67	0.81	1.22	1.51

This table demonstrates the out-sample results under various risk-aversion levels of the last six months from 2008- 2017. c_0 represents the weights of the risk-free asset while c_1 c_2 and c_3 show the weights of three sub-portfolios individually. "Return" represents the monthly return. "ColMean" and "ColStd" show the column means and standard deviations, respectively. No restrictions on leverage or short-selling.

7 Conclusion

This paper develops and tests a two-step portfolio selection procedure which relies on a large universe of investable assets and a set of dynamic predictors of factor-related returns. The first step in the procedure creates a collection of well-diversified mimicking portfolios to approximate the returns of pervasive risk factors. The second step uses a set of predictors including default spread, term spread, price trend, and dividend yield. These predictors are combined into a single-index function, which in turn determines a dynamic allocation of portfolio weights across the factor-mimicking portfolios in order to maximize investor's expected utility. Due to the nonstationarity of some predictive variables, we apply orthogonal series to approximate the single-index function in estimation. We apply the technique to fifty years of monthly U.S. data and find very good performance both in-sample and out-of-sample. We show empirically that the factor mimicking portfolios have high correlation with the targeted factors and low correlation with each other. Our dynamic portfolios perform well, both for high risk-aversion and low risk aversion investors, providing high average returns and also high return volatility for the less risk-averse and correspondingly lower average returns and lower volatility for the more risk-averse investor.

References

Y. Aït-sahalia and M. W. Brandt. Variable selection for portfolio choice. *The Journal of Finance*, 56(4):1297–1351, 2001.

M. Ao, L. Yingying, and X. Zheng. Approaching mean-variance efficiency for large portfolios. *The Review of Financial Studies*, 32(7):2890–2919, 2019.

M. W. Brandt. Estimating portfolio and consumption choice: A conditional euler equations approach. *The Journal of Finance*, 54(5):1609–1645, 1999.

M. W. Brandt, P. Santa-Clara, and R. Valkanov. Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns. *The Review of Financial Studies*, 22(9):3411–3447, 2009.

J. Y. Campbell and R. J. Shiller. Stock prices, earnings, and expected dividends. *The Journal of Finance*, 43(3): 661–676, 1988.

G. Connor and O. Linton. Semiparametric estimation of a characteristic-based factor model of common stock returns. *Journal of Empirical Finance*, 14(5):694–717, 2007.

G. Connor, M. Hagmann, and O. Linton. Efficient semiparametric estimation of the fama–french model and extensions. *Econometrica*, 80(2):713–754, 2012.

C. Dong, J. Gao, and B. Peng. Semiparametric single-index panel data models with cross-sectional dependence. *Journal of Econometrics*, 188(1):301–312, 2015.

C. Dong, J. Gao, and B. Peng. Another look at single-index models based on series estimation. *Available at SSRN 2858624*, 2016a.

C. Dong, J. Gao, D. Tjøstheim, et al. Estimation for single-index and partially linear single-index integrated models. *The Annals of Statistics*, 44(1):425–453, 2016b.

C. Dong, J. Gao, and O. B. Linton. High dimensional semiparametric moment restriction models. *Available at SSRN 3045063*, 2018.

E. F. Fama and K. R. French. Business conditions and expected returns on stocks and bonds. *Journal of financial economics*, 25(1):23–49, 1989.

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

J. Fan, Y. Liao, and M. Mincheva. Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society. Series B, Statistical methodology*, 75(4), 2013.

J. Fan, Y. Liao, and W. Wang. Projected principal component analysis in factor models. *Annals of statistics*, 44 (1):219, 2016.

J. Freyberger, A. Neuhierl, and M. Weber. Dissecting characteristics nonparametrically. *The Review of Financial Studies*, 33(5):2326–2377, 2020.

J. Gao and P. C. Phillips. Semiparametric estimation in triangular system equations with nonstationarity. *Journal of Econometrics*, 176(1):59–79, 2013.

J. Gao, C. Phillips, et al. Functional coefficient nonstationary regression with non-and semi parametric cointegration. 2013.

L. P. Hansen. Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society*, pages 1029–1054, 1982.

L. P. Hansen, J. Heaton, and A. Yaron. Finite-sample properties of some alternative gmm estimators. *Journal of Business & Economic Statistics*, 14(3):262–280, 1996.

R. J. Hodrick. Dividend yields and expected stock returns: Alternative procedures for inference and measurement. *The Review of Financial Studies*, 5(3):357–386, 1992.

K. Hou, C. Xue, and L. Zhang. Digesting anomalies: An investment approach. *The Review of Financial Studies*, 28(3):650–705, 2015.

D. B. Keim and R. F. Stambaugh. Predicting returns in the stock and bond markets. *Journal of financial Economics*, 17(2):357–390, 1986.

B. T. Kelly, S. Pruitt, and Y. Su. Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics*, 2019.

S. Kim, R. A. Korajczyk, and A. Neuhierl. Arbitrage portfolios. *Georgia Tech Scheller College of Business Research Paper*, (18-43), 2019.

O. Ledoit and M. Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of multivariate analysis*, 88(2):365–411, 2004.

O. Ledoit and M. Wolf. Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets goldilocks. *The Review of Financial Studies*, 30(12):4349–4388, 2017.

S. Li and O. B. Linton. A dynamic network of arbitrage characteristics. *Available at SSRN 3638105*, 2020.

H. M. Markowitz et al. Portfolio selection. *Journal of Finance*, 7(1):77–91, 1952.

8 Appendix

8.1 Proof

Proof of Theorem 1 : Write the subportfolio vector $\boldsymbol{Q_t}(\boldsymbol{X})$ as:

$$\boldsymbol{Q_t}(\boldsymbol{X}) = \sum_{i=1}^{n} \boldsymbol{B}(\boldsymbol{X_i}) y_{it}.$$

Because:

$$y_{it} = G(\boldsymbol{X_i})\boldsymbol{F_t} + \epsilon_{it}.$$

Then, substitute y_{it} into $\boldsymbol{Q_t}(\boldsymbol{X})$:

$$\begin{aligned}
\boldsymbol{Q_t}(\boldsymbol{X}) &= \textstyle\sum_{i=1}^{n} \boldsymbol{B}(\boldsymbol{X_i})(G(\boldsymbol{X_i})\boldsymbol{F_t} + \epsilon_{it}) \\
&= \tfrac{1}{n}\textstyle\sum_{i=1}^{n} \Gamma G(\boldsymbol{X_i})^\intercal (G(\boldsymbol{X_i})\boldsymbol{F_t} + \epsilon_{it}) \\
&= \Gamma(\tfrac{1}{n}\textstyle\sum_{i=1}^{n}(G(\boldsymbol{X_i})^\intercal G(\boldsymbol{X_i}))\boldsymbol{F_t} + \tfrac{1}{n}\textstyle\sum_{i=1}^{n}(G(\boldsymbol{X_i})^\intercal \epsilon_{it}))
\end{aligned}$$

Given:

$$p \lim_{n\to\infty} \frac{1}{n}\sum_{1=1}^{n} G(\boldsymbol{X_i})^\intercal G(\boldsymbol{X_i}) = \boldsymbol{M^G},$$

where $\boldsymbol{M^G}$ is an identity matrix, and

$$E(\epsilon_{it}|\boldsymbol{X_i}, \boldsymbol{F_t}) = 0.$$

Therefore, we have:

$$p \lim_{n\to\infty} \frac{1}{n}\sum_{1=1}^{n} G(\boldsymbol{X_i})^\intercal \epsilon_{it} = 0.$$

Thus,

$$p \lim_{n\to\infty} \boldsymbol{Q_t}(\boldsymbol{X}) = \Gamma \boldsymbol{M^G} \boldsymbol{F_t} = \Gamma \boldsymbol{F_t}.$$

This shows that the factor-mimicking portfolio is a linear combination of risk factors given Γ is a non-zero matrix. □

Proof of Theorem 2 : Let $\tilde{\boldsymbol{F}}$ represent the demeaned risk factor matrix while

$$\tilde{\boldsymbol{y}}_t = G(\boldsymbol{X})\tilde{\boldsymbol{F}}_t + \boldsymbol{\epsilon_t}.$$

Correspondingly, we have:

$$\begin{aligned}
\tilde{\boldsymbol{Q}}_t(\boldsymbol{X}) &= \boldsymbol{B}(\boldsymbol{X}_i)\tilde{\boldsymbol{y}}_t \\
&= \tfrac{1}{n}\Gamma G(\boldsymbol{X})^\intercal \tilde{\boldsymbol{y}}_t.
\end{aligned}$$

And then,

$$\begin{aligned}
E(\tilde{\boldsymbol{Q}}_t(\boldsymbol{X})\tilde{\boldsymbol{Q}}_t(\boldsymbol{X})^\intercal|\boldsymbol{X}) = \;& \Gamma(\tfrac{1}{n}G(\boldsymbol{X})^\intercal G(\boldsymbol{X}))E(\tilde{\boldsymbol{F}}\tilde{\boldsymbol{F}}^\intercal)(\tfrac{1}{n}G(\boldsymbol{X})^\intercal G(\boldsymbol{X}))\Gamma^\intercal + \\
& \Gamma(\tfrac{1}{n}G(\boldsymbol{X})^\intercal \tfrac{1}{n}E(\boldsymbol{\epsilon_t}\boldsymbol{\epsilon_t}^\intercal)(G(\boldsymbol{X}))\Gamma^\intercal.
\end{aligned}$$

Given $E(\epsilon_{it}|\boldsymbol{X_i}, \boldsymbol{F_t}) = 0$, the cross terms are $E(\tilde{\boldsymbol{F}}\tilde{\boldsymbol{\epsilon}}_t) = 0$.

Taking the second term and using the Euclidian matrix norm:

$$||\Gamma(\tfrac{1}{n}G(\boldsymbol{X})^\intercal \tfrac{1}{n}E(\tilde{\boldsymbol{\epsilon}}_t\tilde{\boldsymbol{\epsilon}}_t{}^\intercal)G(\boldsymbol{X}))\Gamma^\intercal|| \leqslant$$
$$\tfrac{1}{n}||\Gamma\tfrac{1}{n}(G(\boldsymbol{X})^\intercal G(\boldsymbol{X}))\Gamma^\intercal|| \times ||E(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\intercal)|| \rightarrow_{n\to\infty}$$
$$\tfrac{1}{n}||\Gamma\Gamma^\intercal|| \times ||E(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\intercal)|| \rightarrow_{n\to\infty} 0$$

The conclusion of the above formula is due to

$$p\lim_{n\to\infty}\frac{1}{n}\sum_{1=1}^{n}G(\boldsymbol{X_i})^\intercal G(\boldsymbol{X_i}) = \boldsymbol{M^G},$$

and $||E(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\intercal)||$ has bounded eigenvalues for all n.

Furthermore, the well-chosen coefficient matrix Γ can give:

$$E(\tilde{\boldsymbol{Q}}_{\boldsymbol{t}}(\boldsymbol{X})\tilde{\boldsymbol{Q}}_{\boldsymbol{t}}(\boldsymbol{X})^\intercal|\boldsymbol{X}) = \boldsymbol{I}_{JJ}$$

□

Proof of Theorem 3 : We decompose the investment returns of optimal asset-by-asset portfolio and risk-free rate as:

$$r_{ft} + \boldsymbol{R_F} + \boldsymbol{\epsilon_t^*},$$

where the R_F is the optimal factor returns since the return generation function states the risk premiums come from risk factors. The $\boldsymbol{\epsilon_t^*}$ is the zero mean idiosyncratic returns.

Since

$$E(\epsilon_{it}|\boldsymbol{X_i}, \boldsymbol{F_t}) = 0,$$

it follows from the second-order stochastic dominance that the expected utility has the following relationship:

$$E(u(r_{ft} + \boldsymbol{R_F})) > E(u(r_{ft} + \boldsymbol{R_F} + \boldsymbol{\epsilon_t^*})).$$

because zero mean $\boldsymbol{\epsilon_t^*}$ only contributes variance rather than returns.

According to Theorem 1 and Equation 17, the restricted portfolio optimally combines the factors' returns. Therefore, our two-stage portfolio's return can be written as:

$$r_{ft} + \boldsymbol{R_F} + \boldsymbol{\epsilon_t^{**}},$$

where the only difference is the idiosyncratic returns. Our goal now is to show that:

$$E(u(r_{ft} + \boldsymbol{R_F} + \boldsymbol{\epsilon_t^{**}})|\boldsymbol{X}, \boldsymbol{z_t}) \rightarrow^{n\rightarrow\infty} E(u(r_{ft} + \boldsymbol{R_F})|\boldsymbol{X}, \boldsymbol{z_t}).$$

Next, we take the Taylor expansion of $u(r_{ft} + \boldsymbol{R_F} + \boldsymbol{\epsilon_t^{**}})$ around $r_{ft} + \boldsymbol{R_F}$:

$$u(r_{ft} + \boldsymbol{R_F} + \boldsymbol{\epsilon_t^{**}}) = u(r_{ft} + \boldsymbol{R_F}) + \frac{d}{d(r_{ft} + \boldsymbol{R_F})}u(r_{ft} + \boldsymbol{R_F})\boldsymbol{\epsilon_t^{**}} + \frac{d^2}{d(r_{ft} + \boldsymbol{R_F})}u(r_{ft} + \boldsymbol{R_F})^2\boldsymbol{\epsilon_t^{**2}}.$$

We take the expection on both sides, given $E(\boldsymbol{\epsilon_t^{**}}) = 0$ and $\frac{d^2u(\cdot)}{dW^2} \geqslant -c$. Therefore, we have:

$$E(u(r_{ft} + \boldsymbol{R_F} + \boldsymbol{\epsilon_t^{**}})) \geqslant E(u(r_{ft} + \boldsymbol{R_F})) - cE(\boldsymbol{\epsilon_t^{**2}}),$$

where $p\lim_{n\rightarrow\infty} E(\boldsymbol{\epsilon_t^{**2}}) = 0$, according to Theorem2. Therefore, we have :

$$p\lim_{n\rightarrow\infty} E(u(r_{ft} + \boldsymbol{R_F} + \boldsymbol{\epsilon_t^{**}})|\boldsymbol{X}, \boldsymbol{z_t}) - E(u(r_{ft} + \boldsymbol{R_F})|\boldsymbol{X}, \boldsymbol{z_t}) = 0,$$

which completes the proof. □

Table 7: Annual Correlation Between Subportfolios and Risk Factors 1-20

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
f_1	0.61	-0.49	0.68	0.69	0.66	0.73	-0.10	-0.16	0.72	0.79	0.73	0.69	0.77	-0.17	0.64	0.87	0.65	0.71	0.51	0.75
f_2	0.89	0.64	0.70	0.75	0.65	0.79	0.60	0.91	0.71	0.79	0.90	0.92	0.97	-0.01	0.79	0.74	0.79	0.86	0.88	0.41
f_3	0.80	0.81	0.63	0.40	0.66	0.39	0.75	0.64	0.40	0.72	0.86	0.74	0.96	0.56	0.72	0.86	0.84	0.66	0.78	0.55

This table shows the annual correlation between factor-mimicking subportfolios and corresponding risk factors from Jul. 1967- Jun.1987.

Table 8: Annual Correlation Between Subportfolios and Risk Factors 21-40

	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
f_1	0.78	0.86	0.54	0.58	0.22	0.76	0.68	0.16	0.60	0.74	0.75	0.67	0.72	0.73	0.69	0.67	0.96	0.66	0.73	0.76
f_2	0.77	0.79	0.73	-0.24	0.93	0.66	0.86	-0.06	0.87	0.81	0.94	0.76	0.22	-0.06	0.97	0.94	0.92	0.36	0.93	0.93
f_3	-0.40	0.72	0.76	0.64	0.71	0.83	0.70	0.57	0.73	0.67	0.72	0.58	0.66	0.60	0.79	0.77	0.92	0.68	0.79	0.72

This table shows the annual correlation between factor-mimicking subportfolios and corresponding risk factors from Jul. 1987- Jun.2007.

Table 9: Annual Correlation Between Subportfolios and Risk Factors 41-50

	41	42	43	44	45	46	47	48	49	50
f_1	-0.22	0.73	-0.15	0.72	-0.55	-0.08	0.33	0.74	0.63	0.77
f_2	0.63	0.76	0.29	0.73	0.65	0.34	-0.23	0.81	0.83	0.81
f_3	0.64	0.58	0.65	0.61	0.82	0.51	0.87	0.86	0.82	0.53

This table shows the annual correlation between factor-mimicking subportfolios and corresponding risk factors from Jul. 2007- Jun.2017.