PHBS WORKING PAPER SERIES

No Sparsity in Asset Pricing: Evidence from a Generic Statistical Test

Junnan He
Sciences Po

Lingxiao Zhao
Peking University

Guofu Zhou
Washington University in St. Louis

December 2024

Working Paper 20241202

Abstract

We present a novel test to determine sparsity in characteristic-based factor models. Applying the test to industry and pseudo-random portfolios, we reject the null hypothesis that fewer than ten factors are sufficient to explain returns, and show that at least forty factors are needed for the various sample periods examined. We find that dense models outperform sparse ones in both pricing and investing. Testing with tree-based portfolios also indicates no sparsity. Our results suggest that most existing factor models, which have fewer than six factors, are questionable, and that future research on such low-dimensional models is unlikely to be fruitful.

Keywords: factor models, characteristic-based factors, sparsity, test
JEL Classification: G12, C12

Peking University HSBC Business School
University Town, Nanshan District
Shenzhen 518055, China

No Sparsity in Asset Pricing:

Evidence from a Generic Statistical Test*

Junnan He †, **Lingxiao Zhao** ‡, **Guofu Zhou** §

September 2024

*We thank Lin Cong, Guanhao Feng, Jingyu He, Xin He, Dashan Huang, Jiantao Huang (discussant), Nan Lin, Werner Ploberger, Lianjie Shu (discussant), Jun Tu, and Dacheng Xiu; seminar participants at CUHK Business School, Peking University, Washington University in St. Louis, and Wuhan University; as well as participants at the 36th Asian Finance Association (AsianFA) Annual Conference, the 2024 Hong Kong Conference for Fintech, AI, and Big Data in Business, and the 8th PKU-NUS Annual International Conference for valuable comments and suggestions.

†Sciences Po, Department of Economics, 28 Rue des Saints-Peres, Paris, France 75007. E-mail: junnan.he@sciencespo.fr.

‡Peking University HSBC Business School, University Town, Nanshan District, Shenzhen, P.R.China 518055. E-mail: lingxiao@phbs.pku.edu.cn.

§Olin Business School, Washington University in St. Louis, 1 Brookings Drive, St. Louis, MO, USA 63130. E-mail: zhou@wustl.edu.

No Sparsity in Asset Pricing: Evidence from a Generic Statistical Test

Abstract

We present a novel test to determine sparsity in characteristic-based factor models. Applying the test to industry and pseudo-random portfolios, we reject the null hypothesis that fewer than ten factors are sufficient to explain returns, and show that at least forty factors are needed for the various sample periods examined. We find that dense models outperform sparse ones in both pricing and investing. Testing with tree-based portfolios also indicates no sparsity. Our results suggest that most existing factor models, which have fewer than six factors, are questionable, and that future research on such low-dimensional models is unlikely to be fruitful.

JEL Classification: G12, C12

Keywords: factor models, characteristic-based factors, sparsity, test

1 Introduction

One crucial question in empirical asset pricing is to identify a set of factors that are capable of explaining the cross-section expected equity returns. After the pioneering capital asset pricing model (CAPM) model of Sharpe (1964) which suggests the market factor is the single ancestor for everything, there are the well-known three- (five-, or six-) factor models proposed by Fama and French (1993, 2015, 2018), the investment-based four-factor model proposed by Hou, Xue, and Zhang (2015), the mispricing-based four-factor model proposed by Stambaugh and Yuan (2017), the behavior-based three-factor model proposed by Daniel, Hirshleifer, and Sun (2020), and so on. The list of models in the literature is large given the growing number of factors proposed. Indeed, there are more than one hundred potential factors or anomalies discussed by Green, Hand, and Zhang (2017) and Hou, Xue, and Zhang (2020) and more are analyzed by Feng, Giglio, and Xiu (2020).

Given the large number of candidate factors, there is also a large literature on how to search for a small subset of relevant factors, such as Barillas and Shanken (2018), Chib, Zeng, and Zhao (2020), Bryzgalova, Huang, and Julliard (2023), He, Huang, Li, and Zhou (2023). Many such studies adapt concepts and methods of sparse estimators from statistics, examples of which include the AIC (Akaike (1974)), BIC (Schwarz (1978)), LASSO (Tibshirani (1996)), SCAD (Fan and Li (2001)), Elastic Net (Zou and Hastie (2005)). Sparse asset pricing models have the benefit of easy interpretation. However, the assumption of model sparsity is challenged. Feng, Giglio, and Xiu (2020) emphasize that "the asset pricing literature has adopted the concept of sparsity without always explicitly acknowledging it." Such an implicit assumption can be problematic, and a sparse model would be inappropriate if the underlying DGP is dense, i.e., at the opposite

end of the spectrum from sparse (see e.g., Tibshirani (1996) and Hsu, Kakade, and Zhang (2014)). Consistent with this point, Chib, Zhao, and Zhou (2024) find that a relatively complex linear model is necessary to price more than a hundred anomalies. Kozak, Nagel, and Santosh (2020) also point out that characteristic-sparse models cannot adequately summarize the cross-section of stock returns. In the general context of quantitative economic research using big data, Giannone, Lenza, and Primiceri (2021) also emphasize that sparsity is an illusion in many modeling contexts.

In this paper, we propose the first formal statistical test for sparsity in the number of factors in asset pricing. The null hypothesis is that an unknown sparse set of factors can adequately explain the cross-sectional variation of asset returns. If the sparsity is rejected, it will imply that that a large number of factors are needed to explain the asset returns. It also rejects the earlier well known models that rely on a few factors. Theoretically, our test is generic and directly addresses the question of whether there exists an unknown small set of candidate factors that can explain the cross-section of expected equity returns. This setting diverges from the typical setting in the literature, which focuses on the number of non-zero coefficients in a regression of a set of given factors (see e.g., Jin and Cai (2007); Jin (2008); Cai and Jin (2010); Carpentier and Verzelen (2019, 2021)). When factors are sparse, it implies there is a small proportion of non-zero coefficients, but not the converse. In addition, the pricing factors can be highly correlated, falling outside the typical assumption of no or limited correlation (see e.g., Cai and Jin (2010); Carpentier and Verzelen (2021)). To test against the sparse null hypothesis, we adopt the alternative hypothesis that the cross-sectional variation is represented by a dense model. This means that while many factors can have small effects, they are collectively non-negligible and important in explaining the cross-section of asset returns.

The intuition of our test is motivated from using a measure to summarize the number of "relevant pricing factors." Under the null hypothesis, only a small number of such factors exist, and, whereas under the alternative hypothesis, there is a large number of factors with small effects. The test statistic then aims to distinguish the two scenarios. We reject sparsity when the measure is large. Simulation evidence shows that the test can accurately differentiate whether the simulated data originates from a sparse or dense data-generating process (DGP), even in extreme cases of highly correlated factors.

To address the sparsity hypothesis on the number of characteristic-based factors, we conduct a further empirical analysis using the proposed methodology. We consider two sets of pricing objectives to price industry and pseudo-random portfolios, respectively. Instead of using the excess returns data directly, we "de-market" each time series using the value-weighted market factor and pass on the residuals to the proposed test following Kozak, Nagel, and Santosh (2020) and use the monthly observations of 153 characteristic-based factors in the U.S. following Jensen, Kelly, and Pedersen (2023). With data from January 1974 to December 2023, both whole sample and subsample analysis reject the sparsity hypothesis that the number of factors in the US equity market is less than 10, and it is likely that at least 40 factors are needed. Our conclusion is important, since almost all existing factor models, such as Fama and French (1993, 2015, 2018), Hou, Xue, and Zhang (2015), Stambaugh and Yuan (2017) and Daniel, Hirshleifer, and Sun (2020), focus on fewer than 10 factors. Our test rejects all of these models and suggests that future research in searching for a single-digit number of factors to price the equities will most likely be doomed.

To provide the economic insight for our test, we evaluate the performance of dense models and sparse models in two ways, their ability to price the cross-section and their investment potential. To

compare the pricing ability, we include two different measures, one that captures the pricing error and another that gauges the goodness-of-fit. The results show that the dense model representation performs better in pricing for both the industry portfolios and the pseudo-random portfolios when compared to sparse models in both the whole sample and subsamples. For the investment potential, we consider the annualized Sharpe ratio of tangency portfolios constructed using factors of dense and sparse models, respectively. We find that the dense model has significantly higher annualized Sharpe ratio, both in- and out-of-sample, supporting again the relative advantage of dense models.

Our empirical result of no-sparsity in the US equity market is robust in a number of aspects. First, we consider alternative test assets. Using a novel set of test assets proposed by Cong, Feng, He, and He (2023) who employ tree-based models (P-Trees) to create P-Tree portfolios as test assets, we find that our test continues to support dense models. Second, we provide further analysis on scenarios where only a limited number of firm characteristics are involved. The associated empirical findings demonstrate the effectiveness of the test in determining the complexity of any set of pricing objectives and in identifying the relevant signals for constructing test assets that represent the cross-section. Third, we also apply our test to contexts in which the "de-market" procedure is generalized to other sparse linear factor models, such as the Fama and French three-factor (FF3) or six-factor (FF6) models. By using the residuals from these models instead of the original data, the test results remain consistent, providing further evidence to reject the sparsity assumption.

Our test is related to the well-known Gibbons, Ross, and Shanken (1989) test on the validity of a given factor model, but differs entirely. The GRS test requires a few factors to be given, and, to prove non-sparsity, it will require to test all possible specifications of a few factor models out of

a given set of candidates. Chib, Zhao, and Zhou (2024) show that it is manageable to do so only for a set of 12 candidates, but impossible when the set has more than 100. Hence, the GRS test and the like are unsuitable for our large dimensional case with 154 potential factors. In contrast to the GRS, our test is specification-free in that it does not require the identification of a small set of factors from the potential candidates. Our test is designed to answer the question whether there exists an unknown set of sparse factors to price the assets out of potentially a large set of candidates.

The rest of the paper is organized as follows. Section 2 provides the methodology and the simulation evidence. Section 3 presents the empirical results. Section 4 concludes.

2 Methodology

2.1 Model Setup

We start from the multi-factor asset pricing regression given by

$$\underbrace{\boldsymbol{Y}}_{T\times N} = \underbrace{\boldsymbol{X}}_{T\times L}\underbrace{\boldsymbol{B}'}_{L\times N} + \underbrace{\boldsymbol{E}}_{T\times N}, \tag{1}$$

where T denotes the number of observations, N denotes the number of test assets, and L denotes the total number of candidate characteristic-based factors. The parameter matrix $\boldsymbol{B}$ represents the factor loadings, where each element β_{ij} denotes the i-th (out of N) test asset's loading on the j-th (out of L) factor. Throughout the paper, we use j to index the columns of $\boldsymbol{B}$. The j-th column vector of $\boldsymbol{B}$ is therefore $\boldsymbol{\beta}_j$, which summarizes those N test assets' loadings on the j-th factor, for

$j = 1, 2, \ldots, L$. Without loss of generality, we can assume each column of $\boldsymbol{Y}$ and $\boldsymbol{X}$ have mean zero, and we will be focusing on the β_{ij} coefficients. The ordinary least square (OLS) estimate of $\boldsymbol{B}'$ is given by

$$\hat{\boldsymbol{B}}' = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}. \tag{2}$$

Loosely speaking, we refer to the above multi-factor model as sparse if the variation in $\boldsymbol{Y}$ is explained by only K columns of $\boldsymbol{X}$, where $K << L$ denotes the number of relevant factors. In other words, only K columns of $\boldsymbol{B}$ are non-zero vectors. In contrast, a dense model is the other extreme, for which all columns of $\boldsymbol{X}$ have non-zero coefficients. Below our null and alternative hypotheses formalize the two extreme paradigms of models mathematically. Our test then asks the question of whether the estimated model falls closer to the sparse paradigm or the dense paradigm.

2.2 The Hypotheses

2.2.1 Null Hypothesis

Our null hypothesis can be thought of as a large family of sparse generating distributions for $\boldsymbol{B}$ that is consistent with our definition for sparsity-in-factors, i.e., only K factors are relevant out of L candidate factors in total. Let $\mathscr{F}$ be the set of all distributions defined on $\mathbb{R}^{N\times L}$. Formally, the null hypothesis is

$$\begin{aligned} H_0(\kappa): \quad & \boldsymbol{\beta}_j = z_j\boldsymbol{\gamma}_j, \text{ for } j = 1, 2, \ldots, L, \\ & z_j \overset{\text{IID}}{\sim} Bernoulli(\kappa), \text{ where } \kappa = K/L, \\ & (\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_L) \sim F \text{ for some distribution } F \in \mathscr{F}. \end{aligned} \tag{3}$$

According to this null hypothesis, each $\boldsymbol{\beta}_j$ is a N-dimensional zero vector with probability $1-\kappa$. That is, when $z_j = \mathbf{0}$, the corresponding j-th characteristic-based factor has no explanation power for any of these test assets. However, when $z_j \neq \mathbf{0}$ with probability $\kappa = K/L$, the associated j-th factor is indeed relevant, and $\boldsymbol{\beta}_j = \boldsymbol{\gamma}_j$ is allowed to be any vectors in $\mathbb{R}^N$.[1] If we fixed F to be a normal distribution with large variance, the generating distribution for each $\boldsymbol{\beta}_j$ can be thought of as a N-dimensional "spike-and-slab" prior in Mitchell and Beauchamp (1988). Our hypothesis is more general, in which F can be arbitrary, including a Dirac-delta measure. In the case where F is a Dirac-delta measure, each realized $\boldsymbol{\beta}_j$ is then either the N-dimensional zero vector or a fixed N-dimensional vector.

The above null hypothesis can also be interpreted as a family of Bayesian priors for the matrix $\boldsymbol{B}$ conditional on the knowledge that "about at least $1-\kappa$ fraction of factors have zero coefficients for test assets". Since this prior prescribes that each $\boldsymbol{\beta}_j = \mathbf{0}$ with probability $1-\kappa$, it is rather informative about the location of each $\boldsymbol{\beta}_j$. Under this paradigm, ex-ante, only a few candidate factors have non-zero explanation power of these test assets which represent the cross-section.

2.2.2 Alternative Hypothesis

The alternative hypothesis is designed to capture the other extreme situation of the "many small effects". Given that there is little ex-ante information about the parameters, we require homogeneity across the coefficients for different vectors in $\boldsymbol{X}$. That is, the generating distributions for $\boldsymbol{\beta}_j$ are i.i.d. across $j = 1, \ldots, L$. In other words, ex-ante, the vector of factor loadings for each factor is drawn independently from the same distribution.

[1] Since F can be arbitrary, the null hypothesis $H_0(\kappa)$ is contained in $H_0(\kappa')$ whenever $\kappa < \kappa'$.

We adopt the normal distribution to model this generating process. The normal distribution is the maximum entropy distribution and is therefore conventionally the distribution of the coefficients when a minimal amount of information is known (see, e.g. Jaynes (1968)). The only requirement now is to pin down its mean and variance. Due to symmetry, we take the mean to be zero. Without any additional assumption, the variance can be determined using the moment identity for every i and t,

$$Var_t[y_{ti}] = Var_t\left[\sum_j x_{tj}\beta_{ij}\right] + \sigma_i^2 \tag{4}$$

where $\sigma_i^2 := Var_t[e_{ti}]$. Since β_j is i.i.d. for different j's, it holds that $Var[\beta_{ij}] = Var[\beta_{ij'}]$ for all $j \neq j'$. Therefore, summing up (4) across t, we have for every i,

$$\mathbb{E}[y_i'y_i] - \mathbb{E}[e_i'e_i] = tr(\boldsymbol{X}'\boldsymbol{X})Var[\beta_{ij}] \tag{5}$$

or equivalently,

$$Var[\beta_{ij}] = \frac{T}{tr(\boldsymbol{X}'\boldsymbol{X})}\left(\mathbb{E}[y_i'y_i]/T - \sigma_i^2\right) \tag{6}$$

for every i, j. Therefore, our alternative hypothesis is summarized as

$$H_a: \quad \beta_j \overset{\text{IID}}{\sim} \mathcal{N}\left(0, \frac{T\Omega}{tr(\boldsymbol{X}'\boldsymbol{X})}\right), \text{ for } j = 1, 2, \ldots, L, \tag{7}$$

where $\Omega = \text{diag}\left(\mathbb{E}[y_i'y_i]/T - \sigma_i^2\right)$ is a $N \times N$ diagonal matrix.

Under this alternative hypothesis, not only all β_{ij}'s are non-zero vectors, but also most of them are of order $\frac{1}{\sqrt{L}}$. Hence this hypothesis can be interpreted as a prior for a large number of small coefficients. In empirical implementations, we take the direct approach and plug-in the standard

unbiased plug-in estimates to obtain Ω.[2]

2.3 The Max-min Likelihood Ratio between the Hypotheses

To separate from the notation of the column vectors, we use subscript i to index the row vectors of $\boldsymbol{B}$ in the paper. The i-th row is therefore denoted by $\boldsymbol{\beta}_i$, $i = 1, ..., N$, and the variance of the OLS estimate $\hat{\boldsymbol{\beta}}_i = \left(\hat{\beta}_{i1}, \hat{\beta}_{i2}, \cdots, \hat{\beta}_{iL}\right)'$ is given by $\sigma_i^2 \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$. The variance of the estimate $\hat{\beta}_{ij}$ is the j-th diagonal of $\sigma_i^2 (X'X)^{-1}$. Therefore, we denote the j-th diagonal element of matrix $T\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$ by s_j^2 for some $s_j \geq 0$. To construct our test statistic, we first standardize the OLS estimates $\hat{\boldsymbol{\beta}}_i$ to $\tilde{\boldsymbol{\beta}}_i = (\tilde{\beta}_{i1}, \tilde{\beta}_{i2}, \cdots, \tilde{\beta}_{iL})'$, according to

$$\tilde{\beta}_{ij} = \frac{\sqrt{T}}{\sigma_i s_j} \hat{\beta}_{ij}, \text{ for } i = 1, \cdots, N \text{ and } j = 1, \ldots, L. \tag{8}$$

Under this scaling, conditional on the true parameters $\boldsymbol{B}$, each $\tilde{\beta}_{ij}$ has the same marginal variance.

When the factor loadings matrix is sparse, in the sense that all but a few, for an illustrative special example: the first K columns of matrix $\boldsymbol{B}$ are zero vectors, we could remove the non-zero columns, i.e., $\boldsymbol{\beta}_j \neq \mathbf{0}$ for $j \leq K$. The remaining standardized estimators $\tilde{\beta}_{ij}$, for $i = 1, \ldots, N$ and $j > K$, all center at zero with unit variance. Figure 1 plots the empirical distribution of $\tilde{\beta}_{ij}$ from an illustrative simulation. In this illustration, the distribution of $\tilde{\beta}_{ij}$ with $1 \leq i \leq N$ and $j > K$ lies under the standard normal density as indicated by the dashed blue curve.[3] And on the other hand, the $\tilde{\beta}_{ij}$'s for $j \leq K$ are shown to fall outside the distribution in Figure 1.

[2] That is, we use sample variance $Var[y_i]$ for $\mathbb{E}[y_i'y_i]/T$, and the residual (degree of freedom adjusted) mean-squared error $Var[\hat{e}_i]$ for σ_i^2.

[3] The limiting normality of the empirical distribution in this illustration can be shown as in Azriel and Schwartzman (2015). Nonetheless, our test holds in a more general context since our derivation does not rely on the empirical distribution of $\tilde{\beta}_{ij}$ across i to be normal.

In general, under our null hypothesis, $\tilde{\beta}_j$ can be interpreted as a standard normal vector except for a small number K of indices j with $\beta_j \neq 0$. This leads to an interpretation of the null as follows: each $\tilde{\beta}_j$ is drawn from some unknown distribution with probability κ, and is drawn from the standard normal distribution otherwise. Formally, this means that under the original null hypothesis (3) simplifies to

$$\begin{aligned} H_0(\kappa): \quad & \tilde{\beta}_j \overset{\text{d}}{=} (1-z_j)u_j + z_j\gamma_j, \text{ for } j = 1,2,\ldots,L, \\ & z_j \overset{\text{IID}}{\sim} Bernoulli(\kappa), \text{ where } \kappa = K/L, \\ & u_j \overset{\text{IID}}{\sim} \mathcal{N}(0, I_N), \\ & (\gamma_1,\ldots,\gamma_L) \sim F \text{ for some distribution } F \in \mathcal{F}. \end{aligned} \tag{9}$$

Therefore, $\tilde{\beta}_j$ follows an N-dimensional standard normal distribution with probability $1-\kappa$, and follows some other distributions with probability κ. In other words, under $H_0(\kappa)$, $\tilde{\beta}_j$ is a random vector in the so-called κ-contaminated neighborhood of $\mathcal{N}(0, I_N)$.

Under the alternative hypothesis (7), a direct calculation simplifies the alternative to

$$H_a: \quad \tilde{\beta}_j \overset{\text{d}}{=} \mathcal{N}\left(0, I_N + \Omega_j\right), \; j = 1,\ldots,L, \tag{10}$$

where

$$\Omega_j = \frac{T}{s_j^2}\text{diag}\left(\frac{1}{\sigma_1},\ldots,\frac{1}{\sigma_N}\right)\frac{T\Omega}{tr(X'X)}\text{diag}\left(\frac{1}{\sigma_1},\ldots,\frac{1}{\sigma_N}\right). \tag{11}$$

A test for the pair of hypotheses (9) and (10) is the robust test for the κ-contaminated neighborhood. Below we follow the derivation as in Huber (2004) to obtain the max-min optimal test statistic for each $\tilde{\beta}_j$.

We start with the likelihood ratio without κ-contamination. Denote by p_{aj} the density of $\tilde{\boldsymbol{\beta}}_j$ under the alternative, i.e., the density of $\mathcal{N}(0, I_N + \Omega_j)$. Also, denote by p_0 the density for $\mathcal{N}(0, I_N)$. Without contamination, the likelihood ratio between the distributions $\mathcal{N}\left(0, I_N + \Omega_j\right)$ and $\mathcal{N}\left(0, I_N\right)$ is proportional to

$$\frac{p_{aj}(\tilde{\boldsymbol{\beta}}_j)}{p_0(\tilde{\boldsymbol{\beta}}_j)} = \det\left(I_N + \Omega_j\right)^{-1/2} \exp\left\{\frac{1}{2}\tilde{\boldsymbol{\beta}}_j' \left[I_N - \left(I_N + \Omega_j\right)^{-1}\right] \tilde{\boldsymbol{\beta}}_j\right\}. \tag{12}$$

After taking the log of the likelihood ratio, other terms become irrelevant and it suffices if we consider the statistic

$$x_j(\tilde{\boldsymbol{\beta}}_j) = \tilde{\boldsymbol{\beta}}_j' \left[I_N - \left(I_N + \Omega_j\right)^{-1}\right] \tilde{\boldsymbol{\beta}}_j. \tag{13}$$

Under the null hypothesis, we have seen previously that the density of $\tilde{\boldsymbol{\beta}}_j$ is $q = (1-\kappa)p_0 + \kappa f$ for some arbitrary density function f.[4] Following section 10.3 of Huber (2004), the max-min optimal choice of q is the one that

$$q_j(.) = \begin{cases} (1-\kappa)p_0 & \text{for all } \tilde{\boldsymbol{\beta}}_j \text{ s.t. } x_j \leq x_j^*; \\ \\ c_j p_{aj} & \text{for all } \tilde{\boldsymbol{\beta}}_j \text{ s.t. } x_j > x_j^*. \end{cases} \tag{14}$$

for some constants x_j^* and c_j such that $\int q_j = 1$ and $c_j p_{aj} = (1-\kappa)p_0$ whenever $x_j(\tilde{\boldsymbol{\beta}}_j) = x_j^*$.

Proposition 1 *For each $j = 1, ..., L$, if $1 \leq (1-\kappa)\det(I_N + \Omega_j)^{1/2}$ there is a unique pair of*

[4] This is without loss of generality since any distribution F can be approximated by a sequence of continuous distributions in total variation.

constants x_j^* and c_j that simultaneously solve for q_j defined as above that

$$\begin{cases} \int q_j = 1; \\ \\ c_j p_{aj}(b) = (1-\kappa) p_0(b) \text{ for all } b \text{ s.t. } x_j(b) = x_j^*. \end{cases} \tag{15}$$

In particular, the solution x_j^* satisfies

$$\frac{1}{1-\kappa} = \int_{x_j(b)\le x_j^*} p_0(b)\mathrm{d}b + \det(I_N+\Omega_j)^{1/2}\exp(-\frac{1}{2}x_j^*)\int_{b'\Omega_j b > x_j^*} p_0(b)\mathrm{d}b. \tag{16}$$

Now it follows from Theorem 3.2 in Huber (2004) that the max-min test statistic for each $\tilde{\beta}_j$ is simply $\ln\frac{p_{aj}}{q_j}(\tilde{\beta}_j)$. Further simplification gives an equivalent max-min test statistic: $\min\{x_j(\tilde{\beta}_j), x_j^*\}$.

For implementation, we need to numerically approximate the integrals in (16). Denote by u_i for $i = 1,...,N$ independent standard normal random variables $u_i \sim \mathcal{N}(0,1)$, then the integrals in (16) can be expressed as

$$\begin{aligned} \int_{x_j(b)\le x_j^*} p_0(b)\mathrm{d}b &= \Pr\left(\begin{bmatrix} u_1 \\ \vdots \\ u_N \end{bmatrix}^T \left[I_N - \left(I_N+\Omega_j\right)^{-1} \right] \begin{bmatrix} u_1 \\ \vdots \\ u_N \end{bmatrix} \le x_j^* \right) \\ &= \Pr\left(\sum_{i=1}^N \frac{\omega_j(i)}{1+\omega_j(i)} \chi_1^2 \le x_j^* \right), \end{aligned} \tag{17}$$

where $\omega_j(i)$ is the ith eigenvalue of Ω_j. And similarly,

$$\int_{b'\Omega_j b > x_j^*} p_0(b)\mathrm{d}b = \Pr\left(\sum_{i=1}^{N} \omega_j(i)\chi_1^2 > x_j^*\right). \tag{18}$$

In other words, they are respectively the cumulative distribution function (CDF) and the complementary CDF of sums of scaled Chi-squared random variable with degree of freedom one. Numerically, the solutions x_j^*'s for (16) in this paper are approximately solved for, using an R-package "momentchi2" approximating the above CDF functions.

2.4 The Test Statistic

We propose the overall test statistic to be a weighted sum of each of the individual test statistic, that is,

$$T = \frac{1}{L}\sum_{j=1}^{L} w_j \min\left\{x_j\left(\tilde{\beta}_j\right), x_j^*\right\}. \tag{19}$$

for some positive weights w_j, $j = 1,\ldots,L$, that does not depend on $\tilde{\beta}_j$. Different weights put different emphasis on each factor j. For instance, if $w_j = 1$ for every j, then a j with a large x_j^* would be critical. If this x_j^* is much larger than every other threshold, then the outcome of the test depends heavily upon $\tilde{\beta}_j$, and for instance, the null would be much more easily rejected if $x_j(\tilde{\beta}_j) \geq x_j^*$. To ensure each term in the sum has equal influence on the rejection of the null, we can choose w_j so that when each $\tilde{\beta}_j$ is in favor of the alternative, its expectation is similar to others. More specifically, we choose w_j so that

$$\mathbb{E}_a[w_j x_j(\tilde{\beta}_j)] = 1, \tag{20}$$

which gives

$$w_j = \frac{1}{tr(\Omega_j)}. \tag{21}$$

The value of the test statistic would be larger under the alternative than under the null. So the test will reject the sparse null hypothesis when T is greater than a critical threshold. The exact distribution of T under the null is difficult to express, and we do not have an analytical α level. However, the following result allows us to simulate the rejection region.

Theorem 2 *Under $H_0(\kappa)$, T is first-order stochastically dominated by*

$$S := \frac{1}{L}\sum_{j=1}^{L} w_j \left[(1-z_j)\min\{x_j(\boldsymbol{e}_j), x_j^*\} + z_j x_j^*\right], \tag{22}$$

where $z_j \overset{IID}{\sim} Bernoulli(\kappa)$ and $\boldsymbol{e}_j$ is the jth column of the $N \times L$ matrix $[e_1, e_2, ..., e_L]$, where each $i \in \{1, \ldots, N\}$ row is independently drawn from

$$e^i \sim \mathcal{N}\left(0, diag\left(\sqrt{\frac{T}{s_1^2}}, \ldots, \sqrt{\frac{T}{s_L^2}}\right)(X'X)^{-1} diag\left(\sqrt{\frac{T}{s_1^2}}, \ldots, \sqrt{\frac{T}{s_L^2}}\right)\right). \tag{23}$$

In particular, this first-order stochastic upper bound is tight.

Since S is an upper bound for the test statistic, we can obtain a proper alpha level of the test, i.e. the region of rejection for $T \geq t_\alpha$, through simulation. Specifically, we can simulate S to obtain t_α such that $\Pr(S \geq t_\alpha) \leq \alpha$. Alternatively, we can also decide on the test by simulating the p-value. For any given T, the p-value can be approximated by simulating $\Pr(S \geq T)$.

2.5 Simulation Evidence

In this section we report the results of simulation experiments. In each subsection we simulate datasets from the following model

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{B}' + \boldsymbol{E}$$

for $L = 150$, $N = 50$, and various $T = 300$ and $T = 600$. Those dimension parameters are selected to match the empirical analysis. For each setting, we simulate $G = 1000$ datasets and conduct the test accordingly. We report the probabilities of rejecting the null hypotheses at the 5% significance level for different null hypothetical levels of κ.

2.5.1 Simulation under sparse and dense DGP: no correlation across factors

In the standard case where there is no correlation across factors, all x's and error terms are simulated from the standard normal distribution $\mathcal{N}(0,1)$. For each data generating process (DGP), we set $\beta_j = \frac{3}{\sqrt{K}}\mathbf{1}$ for $j = 1,2,\ldots,K$, and $\beta_j = \mathbf{0}$ otherwise. The DGP is set therefore only the first K factors are relevant, while the remaining $L-K$ factors are redundant. The scalar $\frac{3}{\sqrt{K}}$ is multiplied in front of the one vector to make sure that the signal to noise ratios are the same across different DGPs. By choosing different values of K, we can simulate from a sparse or a dense data generating process.

Table 1 reports the rejection probabilities under the sparse settings. The non-zero parameter vectors are set to be $K = 3,4,5,6$ in both panels, and are tested under various $\kappa \times L$ specification. It can be seen that the rejection probabilities are properly under the significance level for all $K < \kappa \times L$. Therefore, Table 1 presents simulation evidence supporting the theoretical requirements for

conducting our test.

Table 2 reports the simulations for which a significant proportion of parameters are distinct from zero. In particular, there are $K = 30, 60, 90, 120, 150$ non-zero parameter vectors respectively and the rejection probabilities tested against each κ specification are reported. We mark the cells in the table bold if the correct decision is to reject the null. It can be seen that the rejection probability are close to one when $K > \kappa \times L$, and zero when $K < \kappa \times L$. The table shows our test has good power against various non-sparse settings where the null is violated.

2.5.2 Simulation under sparse DGP: correlated factors

Since the test statistic is obtained through numerical approximation, we would examine below how the test controls the Type I error under correlated design scenarios.

We consider three cases $\rho \in \{0.2, 0.5, 0.8\}$ where ρ parameterizes the correlation coefficient across factors. Specifically in each case, $\boldsymbol{x}_t$ is drawn i.i.d. from $\mathcal{N}(\mathbf{0}, \Sigma)$, where $\Sigma = (\sigma_{ij})_{1 \leq j, j' \leq L}$ and $\sigma_{jj'} = \rho^{|j-j'|}$. As before, error terms are simulated from the standard normal distribution $\mathcal{N}(0,1)$. For each case, we perform four simulations under the sparse DGP where the number of non-zero parameter vectors is $K = 3, 4, 5, 6$, respectively.

To examine our test under the most extreme scenarios, we almost evenly distribute the non-zero factors among the x's. Since $\sigma_{jj'} = \rho^{|j-j'|}$, evenly distributing the non-zero factors will allow these factors to correlate with the zero factors as much as possible. Our simulation is more extreme than an alternative scenario where only the first K factors are non-zero. In that scenario, the set of non-zero factors would be more correlated with each other and less with the remaining zero factors. More precisely, we set $\boldsymbol{\beta}_j = \frac{3}{\sqrt{K}}\mathbf{1}$ for $j = 1, 50, 100$ if $K = 3$; for $K = 4$ we set $\boldsymbol{\beta}_j = \frac{3}{\sqrt{K}}\mathbf{1}$

for $j = 1, 50, 100, 150$; for $K = 5$ we set $\beta_j = \frac{3}{\sqrt{K}}\mathbf{1}$ for $j = 1, 30, 60, 90, 150$; for $K = 6$ we set $\beta_j = \frac{3}{\sqrt{K}}\mathbf{1}$ for $j = 1, 30, 60, 90, 120, 150$; and set all the remaining $\beta_j = \mathbf{0}$. The above DGPs are pushing to extreme our numerical implementation of the test to see how it controls the rejection level when each genuine factor is highly correlated with some irrelevant factors. In particular, in the simulation when $\rho = 0.8$, the number of highly correlated pairs of $\boldsymbol{x}$'s is similar to that number in the data.[5]

Table 3 reports the rejection probabilities under such simulation design. Although in a relatively shorter sample of $T = 300$ with a high correlation of $\rho = 0.8$, we do observe over rejection at $\kappa = 0.05$, the performance improves when we test for a moderate sparsity level of $\kappa = 0.10$. Moreover, for a larger sample size of $T = 600$, the rejection probabilities are all properly under the significance level for $K < \kappa \times L$, indicating practical reliability. Therefore, Table 3 presents additional simulation evidence supporting the theoretical requirements for conducting our test even if there is a high correlation between factors.

3 Empirical Analysis

In empirical asset pricing, it is common to specify a three-factor (or four-, five-, six-factor) model to explain the expected return of equities. For example, Fama and French use three, five, and six factors in Fama and French (1993, 2015, 2018); Hou, Xue, and Zhang (2015) have a four-factor model; Stambaugh and Yuan (2017) have their own four-factor model; and Daniel, Hirshleifer, and Sun (2020) have their own three-factor model. Such sparse asset pricing models only use less than

[5]That is, specifically, if we count the number of pairs $(x_j, x_{j'})$, $j \neq j'$, with pairwise correlation coefficient $|\rho_{x_j, x_{j'}}| \geq 0.8$, this number from the simulation is approximately the number we would obtain from the data.

ten factors to illustrate the cross-section. Different approaches have been developed to compare these sparse models, trying to prove the superiority of one model over the others.

In this section, we present empirical evidence that the asset pricing model defined over these commonly used characteristic-based factors is not as sparse as we thought. Instead of choosing a sparse model from the existing literature, it is preferable to adopt a dense representation.

3.1 Data

For the $\boldsymbol{X}$ variables on the right-hand side of equation (1), we focus on common "factors" (sometimes also called as "anomalies") in the asset pricing literature (Novy-Marx and Velikov (2016), McLean and Pontiff (2016), Tian (2021), Hou, Xue, and Zhang (2020), Kozak, Nagel, and Santosh (2020)). We use the comprehensive 153 characteristic-based factors in the U.S. proposed by Jensen, Kelly, and Pedersen (2023), and use the monthly observations starting from January 1974 to December 2023 to have a relatively stable sequence. Therefore, $L = 153$ and $T = 600$ for the whole sample analysis. Besides, we also conduct sub-sample analysis, i.e., three twenty-five-year rolling time windows (1974:01-1998:12, 1987:01-2011:12, and 1999:01-2023:12) each covering 25 years of observations. For the sub-sample analysis, $L = 153$ and $T = 300$ are kept the same for each sub-sample.

For the pricing objective, the left-hand side variables $\boldsymbol{Y}$ in equation (1), we consider two different sets of test assets. We conduct the sparsity analysis based on the 49 industry portfolios from the Kenneth R. French data library, as well as another set of pseudo-random portfolios constructed using the company codes. For the set of pseudo-random portfolios (abbreviated as random portfolios in tables and figures), to match the number of industry portfolios, we use each

company code to mod 49, resulting in 49 groups of stocks. Within each group, we calculate the value-weighted portfolio returns at each given point in time. Therefore, $N = 49$ is kept the same for those two sets of portfolios.

Given the widespread acceptance of the Capital Asset Pricing Model (CAPM), our analysis will exclude the dominant market factor and focus on assessing the sparsity among the 153 characteristic-based factors. To achieve this, we employ a "de-market" procedure described by Kozak, Nagel, and Santosh (2020). Specifically, for each factor or test portfolio (x or y) mentioned earlier, we "de-market" it using the value-weighted market factor. That is, for each time series, we run a CAPM regression, and keep the residual terms instead of using the original data directly. Through this "de-market" procedure, the effect of the market factor is removed from all the x's and y's simultaneously, and the resulting data are automatically mean-zero and the test can be run using those residual terms described above. We also show that our empirical findings are robust to this procedure in Section 3.5.

3.2 Empirical Results

In the original null hypothesis of our sparsity test, we specify the degree of sparsity κ. That is equivalent by assuming that the true model has K number of factors, where $K = L \times \kappa$. Therefore, we select $\kappa = K/153$ for different null hypotheses, accordingly, for different positive integers $K \leq 153$. We then report the p-values for different hypothetical numbers of factors for both of the whole and the sub-sample analysis in Figure 2 and Figure 3.

In Figure 2, we present the results for two separate sequences of tests for the sets of industry portfolios and pseudo-random portfolios using the whole sample from January 1974 to December

2023. While the red line presents the results for the 49 industry portfolios, the turquoise line presents the results for the 49 pseudo-random portfolios. For each line, the x-axis represents different levels of sparsity, which is measured by the number of factors $K = \kappa/153$ in the null hypothesis, and the y-axis plots the corresponding p-value for each K. It is worth noticing that the horizontal dashed blue line is the cutoff of $p = 0.05$. From a statistical perspective, we reject the null hypothesis if $p < 5\%$ at 5% significance level. For both the industry portfolios and the pseudo-random portfolios, we can reject the sparsity assumption up to $K = 105$. Roughly speaking, it is necessary to include more than 100 characteristic-based factors to explain the cross-section.

Similarly, in Figure 3, we demonstrate the results for these tests described above using three twenty-five-year rolling time windows in three sub-figures. The three time-windows are January 1974 to December 1998, January 1987 to December 2011, and January 1999 to December 2023, representing the starting, middle, and ending periods of the whole sample, respectively. Although the number of factors needed seems not as large as the whole sample, given the shortened length of the observations, the original results still hold. In short, we can reject the sparsity assumption up to $K = 40$ for all these test assets and subsamples. In other words, those sparse models using less than 10 characteristic-based factors are all rejected even using successive 25 years of data. Hence the test suggests a dense characteristic-based factor model over the class of sparse models with less than 10 factors.

3.3 Economic Significance

In this subsection, we illustrate the economic advantages of using a dense model representation compared to sparse ones. We present two pieces of empirical evidence: (1) dense models outper-

form sparse ones in terms of pricing the 49 industry portfolios and 49 pseudo-random portfolios, and (2) tangency portfolios constructed using factors of a dense model exhibit substantially higher annualized Sharpe ratios compared to sparse ones, indicating greater investment potential.

From an economic standpoint, we consistently include the market factor in any asset pricing model and utilize the original characteristic-based factors without employing a prescribed "de-market" procedure.

3.3.1 Pricing Performance

To evaluate the pricing performance of dense and sparse models, with the same two sets of test portfolios, we need proper model representations. To achieve a dense model representation, we first notice that as by-products of the test, in equation (19) which defines the test statistic, we obtain intermediate outputs $x_j(\tilde{\beta}_j)$, x_j^*, and their associated weights w_j for $j = 1, 2, \cdots, L$. By calculating the normalized difference $w_j \left(x_j(\tilde{\beta}_j) - x_j^* \right)$, we can approximately assess the relative importance of these 153 factors given the set of test assets. From the previous subsection, we conclude that for the whole sample, it is necessary to include more than 100 characteristic-based factors to accurately price both the industry and the pseudo-random portfolios. Therefore, for these two sets of portfolios, the industry and the pseudo-random portfolios, we identify the top 100 factors that matter for each set, respectively.[6] Similarly, we conduct the procedure for those three subsamples: January 1974 to December 1998, January 1987 to December 2011, and January 1999 to December 2023. The κ and the number of relevant factors are set according to the earlier test results.

[6]Here, we use the by-products of each test by setting $\kappa = 100/153$.

Following the previous selection step, we take the set of industry portfolios as an example. Excluding the market factor (Mkt), we transform the raw data of these top 100 significant factors into their principal component (PC) counterparts, $\tilde{\boldsymbol{X}} = \boldsymbol{X}Q$, where $\boldsymbol{X}$ represents the original data and Q is the loading matrix derived from principal component analysis (PCA). Subsequently, we retain only the first 7 PCs, denoted as PC1, PC2, $\cdots$, and PC7, which collectively explain over 80% of the variation of the data. Hence, combining the market factor with the first 7 PCs results in a dense model representation, as these principal components assign weights across all 100 factors. Similarly, for the set of pseudo-random portfolios, we denote the first 7 PCs of the top 100 significant factors as PC1*, PC2*, $\cdots$, and PC7*. When it comes to the sub-samples, we only keep the first several PCs that account for over 80% of the variation in the top candidate factors selected in the first step.

For sparse model representation, we stick to the typical number of factors in sparse asset pricing models, such as those with three, four, five, and six factors. Given that the 153 factors are all potential factors for any sparse model, we employ a random sampling approach. Specifically, if our objective is to generate three-factor models, in addition to the market factor, we randomly select two factors out of 153 candidates to construct the three-factor model. For four-, five-, and six-factor models, the logic is exactly the same. For a given number of factors, we repeat this sampling process $G = 1000$ times and compute their economic performance measure.

In Table 4 and Table 5, we report the pricing performance of dense and sparse models using 49 industry portfolios and 49 pseudo-random portfolios for the whole sample and three subsamples. The pricing performance is evaluated using two different measures: the cross-sectional average of the absolute value of estimated intercepts ($|\hat{\alpha}|$), which measures the pricing error, and the cross-

sectional average of the adjusted R-squared (R^2), which measures the goodness of fit of any given model. The dense models outperform the sparse ones in terms of smaller pricing errors and higher goodness of fit for both sets of test assets, indicating the superiority of using a dense model in pricing the cross-section. To conclude, we observe that for both the industry portfolios and the pseudo-random portfolios, using the whole sample as well as the three subsamples, the pricing performance of the dense models is marginally better.

3.3.2 Investment Potential

The Sharpe ratio is a widely recognized measure that helps investors assess the excess return on an investment relative to its risk. Consequently, the annualized Sharpe ratios of tangency portfolios constructed using factors from any given model offer valuable insights for investment evaluation. We check both the in-sample and out-of-sample performance in the following analysis.

We first calculate the in-sample annualized Sharpe ratios for both dense and sparse model representations, following the method used for pricing performance evaluation. For the in-sample analysis, dense models using principal components provide a reasonable representation. Therefore, we report the in-sample annualized Sharpe ratios for the tangency portfolios constructed using factors from dense and sparse model representations for the whole sample in Panel A of Table 6. Compared to the six-factor model, the resulting in-sample annualized Sharpe ratios of dense models with a similar number of PC factors can reach as high as 1.29 and 1.27, representing an improvement of over 33% compared to the sparse models. This in-sample analysis supports the use of dense models as investment strategies.

However, for the out-of-sample analysis, using principal components introduces a forward-

looking bias when comparing the annualized out-of-sample Sharpe ratios of dense and sparse models. To evaluate the dense model, recognizing that principal components put weights on all relevant factors selected by the test, we can omit the principal component transformation step. Consequently, we work with the characteristic-based factors directly when evaluating the out-of-sample performance of the dense model. While a rolling window of forty years of observations is used for weight computation of the tangency portfolio, the last ten years of data are reserved for out-of-sample evaluation.

Putting more preciously, for each $t = 481, \ldots, 600$ in the out-of-sample period, we begin by performing the test using forty years of historical data ending at $t-1$. From the test, we select the fifty most relevant factors[7], and the optimal weight vector for these factors is then computed as follows:

$$\boldsymbol{w}_{t-1} = \frac{\Omega_{t-1}^{-1}\boldsymbol{\mu}_{t-1}}{\mathbf{1}'\Omega_t^{-1}\boldsymbol{\mu}_t}, \tag{24}$$

where $\boldsymbol{\mu}_{t-1}$ and Ω_{t-1} are the sample mean and covariance of these fifty most relevant factors estimated by the test. Therefore, given the realized excess returns of these fifty factors $\boldsymbol{r}_t$, the realized return of the tangency portfolio can be obtained by

$$r_{p,t} = \boldsymbol{w}_{t-1}'\boldsymbol{r}_t. \tag{25}$$

Finally, the annualized out-of-sample Sharpe ratio can be computed

$$SR_{\text{OOS}} = \sqrt{12}\,\frac{\text{mean}(r_{p,t})}{\text{sd}(r_{p,t})}, \text{ for } t = 481, \ldots, 600. \tag{26}$$

[7]Here, we set $\kappa = 50$ and select a moderate number of top-ranked factors.

The same procedure is conducted for any given sparse model, except for the redundant step involving the test and relevant factor selection. For a single randomly sampled sparse model, we calculate the optimal weights using historical data for the sampled factors and then compute the realized excess returns for the next period. Once the ten years of realized excess returns of tangency portfolios are obtained, the annualized out-of-sample Sharpe ratio can be computed using the equation (26).

We report the annualized out-of-sample Sharpe ratios of both dense and sparse models in Panel B of Table 6. The dense model without a star indicates that the relevant factors are selected using the industry portfolios, while the dense model marked with a star indicates that the relevant factors are selected using the pseudo-random portfolios. The results are calculated using the realized one-period-ahead return, where the optimal weight of tangency portfolio is computed using historical data to mimic a practical out-of-sample investment strategy. It can be seen that the dense model can perform more than 20% better compared to sparse models, indicating that the dense models have better investment performance for practitioners.

To conclude, we observe that the test results and the economic performance from both the pricing and investment perspectives roughly align in supporting a dense model.

3.4 Alternative Pricing Objective: P-Tree Portfolios

For the pricing objective discussed earlier, we examine two distinct sets of test assets: the industry portfolios and the pseudo-random portfolios. It is important to highlight that the construction of these two sets of test assets has no relationship with any firm characteristic.

To illustrate the application of our test and to further analyze the "no sparsity" conclusion with the two asset sets previously discussed, we consider a novel series of test assets proposed by Cong, Feng, He, and He (2023). They employ tree-based models (P-Trees) to create P-Tree portfolios as test assets. Instead of using the traditional double sorting method based on two characteristics, their model develops a single tree based on multiple characteristics (typically more than five), resulting in 10 branches, i.e., 10 P-Tree portfolios, for each tree. When a different set of characteristics is used, the 10 resulting P-Tree portfolios also vary, leading us to denote each variant as another tree.

3.4.1 Grouped P-Trees

Following the approach of the original paper, we conduct our analysis using forty years of monthly observations from January 1981 to December 2020, with $T = 480$.[8] The authors create a total of 20 different trees, denoted as P-Tree k, where $k = 1, 2, \cdots, 20$. As stated, each tree comprises 10 P-Tree portfolios. Following their methodology, we group 5 trees as one set of test assets, resulting in $N = 50$. Consequently, we have four distinct groups of test assets in total: P-Trees 1-5, P-Trees 6-10, P-Trees 11-15, and P-Trees 16-20.

Consistent with the previous sections, we maintain the 153 characteristic-based factors as the $\boldsymbol{X}$ variables throughout our analysis. Before applying our sparsity test, both $\boldsymbol{X}$ and $\boldsymbol{Y}$ variables are "de-market".

In Figure 4, we present the test results from four distinct sequences of tests, each utilizing a group of five P-Trees. The shapes and trends of these curves are similar to those obtained using both industry and pseudo-random portfolios in the earlier subsection. It is also essential to

[8] We thank the authors for providing the data.

incorporate over 100 characteristic-based factors to adequately explain the cross-section defined by grouped P-trees. Considering the construction of grouped P-Tree portfolios based on five distinct signal (or, characteristic) sets, it is evident that their complex structure embodies the inherent characteristics of these novel test assets. Therefore, the test results are both convincing and straightforward to interpret. The evidence confidently rejects the sparsity assumption that the number of factors is less than ten in asset pricing.

3.4.2 Further Discussion: What Happens for One Single P-Tree?

To understand how our sparsity test can assist in determining the complexity of the cross-section, we present the test results using P-Tree portfolios derived from a single tree as test assets. Specifically, we examine 10 P-Tree portfolios grown from P-Tree 1. To further illustrate the point, we compare the test results against those using 10 industry portfolios as benchmarks. This benchmark aims to match the dimension of the cross-section defined by one single tree, where $N = 10$. It is evident that categorizing firms into 10 industries is less detailed than the previously utilized categorization into 49 industries.

In Figure 5, we provide the test results for null hypotheses of various K, where $K = 5, 10, \cdots, 150$, using both 10 industry portfolios and 10 P-Tree portfolios of P-Tree 1. The results indicate that the number of characteristic-based factors required to explain the 10 industry portfolios significantly exceeds that needed for the 10 P-Tree portfolios from P-Tree 1. While the industry portfolios are notably dense, the P-Tree portfolios exhibit less density.

To clarify these results, it's important to remember that the construction of industry portfolios is independent of firm characteristics. In contrast, P-Tree 1's construction involves dividing based

on 6 specific characteristics: "SUE," "DOLVOL," "BM_IA," "ROA," "ZEROTRADE," and "ME" (following the notations used in the original paper). As illustrated in the figure, the sparsity test is rejected for the null hypothesis assuming $K = 5$ at 5% significance level. Given the similarity in the construction of many characteristic-based factors, it can be inferred that the test effectively identifies the degree of sparsity in P-Tree 1. This observation is noteworthy, especially when considering P-Trees 1-5, where the other four trees are developed using different characteristic sets. When these five trees are combined as test assets, their collective density increases compared to the scenario with one single tree.

Furthermore, by using the normalized difference $w_j(x_j(\tilde{\beta}_j) - x_j^*)$ described in the previous economic pricing subsection, we can identify the top ten factors that matter for P-Tree portfolios as "*ami*_126*d*", "*at_gr*1", "*at_turnover*", "*dolvol*_126*d*", "*ivol_capm*_21*d*", "*market_equity*", "*ni_me*", "*niq_at_chg*1", "*prc*", and "*zero_trades*_21*d*".[9] Comparing these to the six characteristics mentioned earlier, we observe that the top-ranked factors are closely aligned with the characteristics Cong, Feng, He, and He (2023) utilized to develop P-Tree 1.

This empirical evidence further underscores the practical application of our sparsity test: it assists in determining the complexity of any set of pricing objectives and identifying the relevant signals for constructing portfolios that represent the cross-section.

[9]The factors are listed alphabetically using notations following the original paper. Here, we use the by-product of the test by setting $\kappa = 10/153$.

3.5 Further Empirical Results

In the previous data description subsection, we mentioned that, following Kozak, Nagel, and Santosh (2020), for each factor or test portfolio (x or y), we "de-market" it using the value-weighted market factor. This means that for each time series, we run a CAPM regression and use the residual terms instead of the original data.

This "de-market" procedure can be generalized to any well-known sparse model. For example, consider the Fama and French three-factor (FF3) model or, alternatively, the Fama and French six-factor (FF6) model. For each time series, we can perform a multivariate regression by putting the FF3 model on the right-hand side and then using the FF3 residuals instead of the original data. Through this "de-FF" process, the effect of the FF3 model is removed from all the x and y values simultaneously. The resulting data automatically has a zero mean, allowing the test to be run confidently using these residual terms. Therefore, if we can still reject the sparsity, it provides strong evidence of the dense nature of asset pricing, reinforcing the robustness of our findings.

In this subsection, we present the test results for using FF3 or FF6 residuals for different sets of test assets. In Figure 6, we present the test results using monthly data of the whole sample from January 1974 to December 2023 using 49 industry portfolios and 49 pseudo-random portfolios. In Figure 7, we present the test results using monthly data over 40 years, from January 1981 to December 2020 using four distinct datasets of P-Tree test portfolios. In Figure 8, we present the test results using monthly data over 40 years, from January 1981 to December 2020, 10 industry portfolios and 10 P-Tree portfolios of P-Tree 1. From those figures, we can see that all the conclusions in the earlier subsections remain. Therefore, we can conclude that, through a strong version of the test, we provide further evidence of the non-sparse nature of asset pricing.

4 Conclusion

This paper proposes the first test for sparsity in the number of factors in a factor model. Our test is specification-free in that it does not require the identification of a small set of factors from the multitude of characteristic-based factors proposed in the growing literature. Our test is relevant because there are many studies that search for a few (often less than 10) factors to explain the cross-section of expected equity returns in the US or global markets, perhaps motivated by an attempt to simplify the understanding of the economic driving forces. Our test helps to answer whether such research is even possible or is likely to be successful.

Theoretically, our test distinguishes between two processes: one in which the cross-sectional asset return variation is attributed to a small number of unknown factors, and another in which the variation can be attributed to a large set of factors, each with perhaps small effects. We provide extensive simulation evidence under various settings, including cases where the factors are highly correlated, to demonstrate that the proposed test is highly effective in assessing sparsity.

Empirically, we conduct a comprehensive analysis using the 153 characteristic-based factors created by Jensen, Kelly, and Pedersen (2023) with U.S. monthly data from January 1974 to December 2023. On the test assets that are on the left-hand side of the equation, we consider two distinct pricing objectives, i.e., pricing a set of industry portfolios and a set of pseudo-random portfolios, respectively. Carrying out our test for both the whole sample and three sub-samples, we find that the sparsity assumption, that the true model comprises fewer than 10 factors, is rejected. And at least 40 factors are needed to explain the returns. Furthermore, we illustrate the advantages of utilizing a dense model representation over sparse ones in terms of both model fitting and Sharpe

ratio, and find the same robust result with P-Tree portfolios (Cong, Feng, He, and He (2023)). In addition, we apply the test to "de-market" assets, which takes the residuals from a CAPM regression or from a multivariate regression using FF3 or FF6 factors. Again, we find that there is no sparsity.

The central message of our paper is that we reject the hypothesis that the number of factors in the US equity market is sparse. Our finding is important in two ways. First, it implies existing factor models, such as Fama and French (1993, 2015, 2018), Hou, Xue, and Zhang (2015), Stambaugh and Yuan (2017) and Daniel, Hirshleifer, and Sun (2020) which rely on less than six factors, are rejected. Second, future research aimed at identifying sparse factor models with fewer than ten factors is likely to be unsuccessful. While we focus on the equity market in the paper, our test is readily applied to other asset classes such as corporate bonds, currencies, and commodities where the existing factor models also have only a few factors. Whether there exists sparsity or not for these assets is of importance for pricing and is of interest for future research.

Figure 1: The empirical distribution of standardized estimators $\tilde{\beta}_{ij}$

This figure displays the empirical distribution of standardized estimators $\tilde{\beta}_{ij}$ for $i = 1, \ldots, N$ and $j = 1, \ldots, L$ using simulated data. In this simulation, the first $K = 5$ columns of $\boldsymbol{B}$ are set to be $\frac{3}{\sqrt{5}}\mathbf{1}$, while the remaining $\boldsymbol{\beta}_j$'s are zero vectors. The data set is simulated with $N = 50$, $L = 150$, and $T = 600$, where all x's and error terms are simulated from $\mathcal{N}(0,1)$.

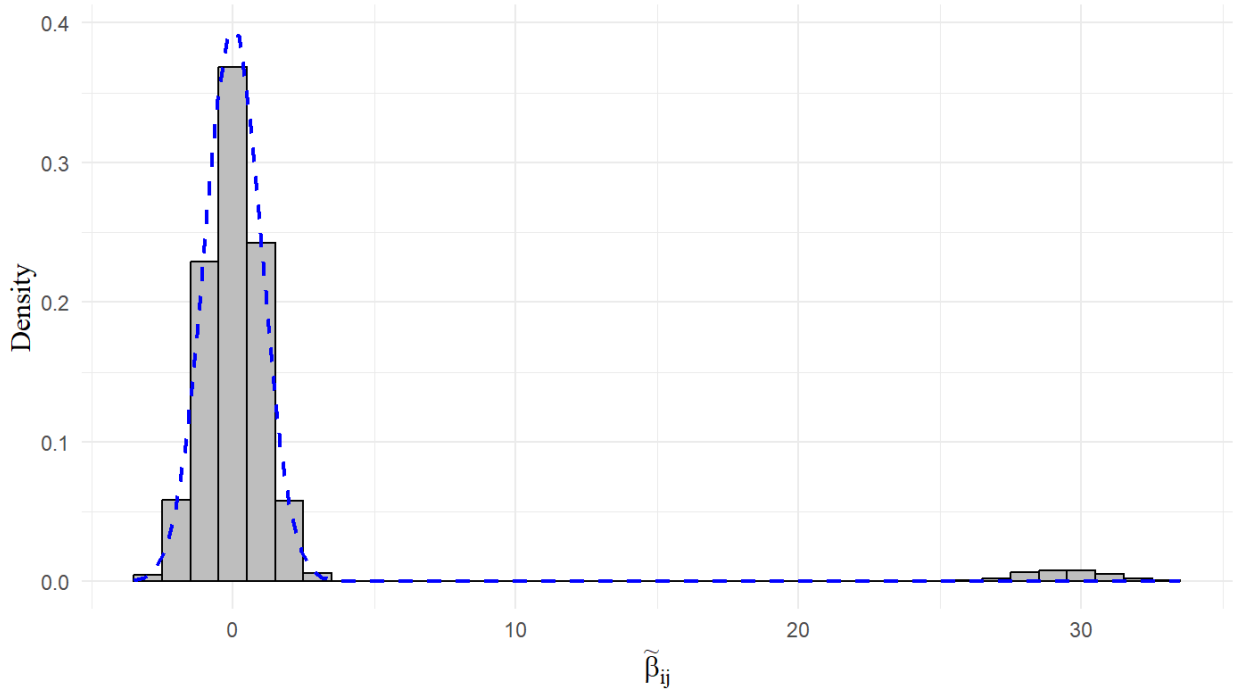


Table 1: Rejection probabilities under sparse settings: no correlation across factors
This table reports the rejection probabilities for different hypothetical sparsity levels of κ at the 5% significance level under sparse settings, where there is no correlation across factors. In each column of Pane A (or, Panel B), given the number of relevant factors K, we generate $G = 1000$ simulated datasets for $L = 150$ and $T = 300$ (or, $T = 600$). We report the rejection frequencies under given specification for a null hypothesis $H_0(\kappa)$ in each cell. We report the rejection frequencies under given specification for a particular null hypothesis $H_0(\kappa)$ in each cell.

Panel A: $T = 300$				
	$\kappa = 0.05$	$\kappa = 0.10$	$\kappa = 0.15$	$\kappa = 0.20$
$K = 3$	0.00%	0.00%	0.00%	0.00%
$K = 4$	0.10%	0.00%	0.00%	0.00%
$K = 5$	0.50%	0.00%	0.00%	0.00%
$K = 6$	1.30%	0.10%	0.00%	0.00%
Panel B: $T = 600$				
	$\kappa = 0.05$	$\kappa = 0.10$	$\kappa = 0.15$	$\kappa = 0.20$
$K = 3$	0.00%	0.00%	0.00%	0.00%
$K = 4$	0.00%	0.00%	0.00%	0.00%
$K = 5$	0.00%	0.00%	0.00%	0.00%
$K = 6$	0.00%	0.00%	0.00%	0.00%

Table 2: Rejection probabilities under dense settings: no correlation across factors

This table reports the rejection probabilities for different hypothetical sparsity levels of κ at the 5% significance level under dense settings, where there is no correlation across factors. In each column of Pane A (or, Panel B), given the number of relevant factors K, we generate $G = 1000$ simulated datasets for $L = 150$ and $T = 300$ (or, $T = 600$). We report the rejection frequencies under given specification for a particular null hypothesis $H_0(\kappa)$ in each cell. Cells are marked in bold if the test rejects the corresponding null hypothesis correctly.

Panel A: $T = 300$									
	$\kappa = 0.10$	$\kappa = 0.20$	$\kappa = 0.30$	$\kappa = 0.40$	$\kappa = 0.50$	$\kappa = 0.60$	$\kappa = 0.70$	$\kappa = 0.80$	$\kappa = 0.90$
$K = 30$	**100.00%**	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
$K = 60$	**100.00%**	**100.00%**	**100.00%**	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
$K = 90$	**100.00%**	**100.00%**	**100.00%**	**100.00%**	**100.00%**	0.00%	0.00%	0.00%	0.00%
$K = 120$	**100.00%**	**100.00%**	**100.00%**	**100.00%**	**100.00%**	**100.00%**	**100.00%**	0.00%	0.00%
$K = 150$	**100.00%**	**100.00%**	**100.00%**	**100.00%**	**100.00%**	**100.00%**	**100.00%**	**100.00%**	**100.00%**
Panel B: $T = 600$									
	$\kappa = 0.10$	$\kappa = 0.20$	$\kappa = 0.30$	$\kappa = 0.40$	$\kappa = 0.50$	$\kappa = 0.60$	$\kappa = 0.70$	$\kappa = 0.80$	$\kappa = 0.90$
$K = 30$	**100.00%**	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
$K = 60$	**100.00%**	**100.00%**	**100.00%**	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
$K = 90$	**100.00%**	**100.00%**	**100.00%**	**100.00%**	**100.00%**	0.00%	0.00%	0.00%	0.00%
$K = 120$	**100.00%**	**100.00%**	**100.00%**	**100.00%**	**100.00%**	**100.00%**	**100.00%**	0.00%	0.00%
$K = 150$	**100.00%**	**100.00%**	**100.00%**	**100.00%**	**100.00%**	**100.00%**	**100.00%**	**100.00%**	**100.00%**

Table 3: Rejection probabilities under sparse settings: correlated factors
This table reports the rejection probabilities for different hypothetical sparsity levels of κ at the 5% significance level under sparse settings with correlated factors. In each column of sub-panel of Panel A (or, Panel B), given the number of relevant factors K and the correlation coefficient ρ across factors, we generate $G = 1000$ simulated datasets for $L = 150$ and $T = 300$ (or, $T = 600$). We report the rejection frequencies under given specification for a particular null hypothesis $H_0(\kappa)$ in each cell.

Panel A-1: $T = 300,\ \rho = 0.20$			**Panel B-1:** $T = 600,\ \rho = 0.20$		
	$\kappa = 0.05$	$\kappa = 0.10$		$\kappa = 0.05$	$\kappa = 0.10$
$K = 3$	0.00%	0.00%	$K = 3$	0.00%	0.00%
$K = 4$	0.20%	0.00%	$K = 4$	0.00%	0.00%
$K = 5$	4.10%	0.00%	$K = 5$	0.00%	0.00%
$K = 6$	4.60%	0.00%	$K = 6$	0.00%	0.00%
Panel A-2: $T = 300,\ \rho = 0.50$			**Panel B-2:** $T = 600,\ \rho = 0.50$		
	$\kappa = 0.05$	$\kappa = 0.10$		$\kappa = 0.05$	$\kappa = 0.10$
$K = 3$	0.80%	0.00%	$K = 3$	0.00%	0.00%
$K = 4$	1.60%	0.00%	$K = 4$	0.00%	0.00%
$K = 5$	7.00%	0.00%	$K = 5$	0.00%	0.00%
$K = 6$	7.90%	0.00%	$K = 6$	0.00%	0.00%
Panel A-3: $T = 300,\ \rho = 0.80$			**Panel B-3:** $T = 600,\ \rho = 0.80$		
	$\kappa = 0.05$	$\kappa = 0.10$		$\kappa = 0.05$	$\kappa = 0.10$
$K = 3$	4.90%	0.10%	$K = 3$	0.00%	0.00%
$K = 4$	6.70%	0.40%	$K = 4$	0.00%	0.00%
$K = 5$	12.90%	0.00%	$K = 5$	0.20%	0.00%
$K = 6$	14.10%	0.60%	$K = 6$	0.30%	0.00%

Figure 2: Industry and pseudo-random portfolios: whole sample analysis
This figure presents the test results using monthly data of the whole sample: January 1974 to December 2023. Given $L = 153$ candidate characteristic-based factors, we plot the p-values for different hypothetical numbers of factors K, each corresponding to the null hypothesis $H_0(\kappa)$, where $\kappa = K/153$. The analysis uses two distinct datasets of test assets: 49 industry portfolios and 49 pseudo-random portfolios. The blue dashed line indicates the 5% significance level.

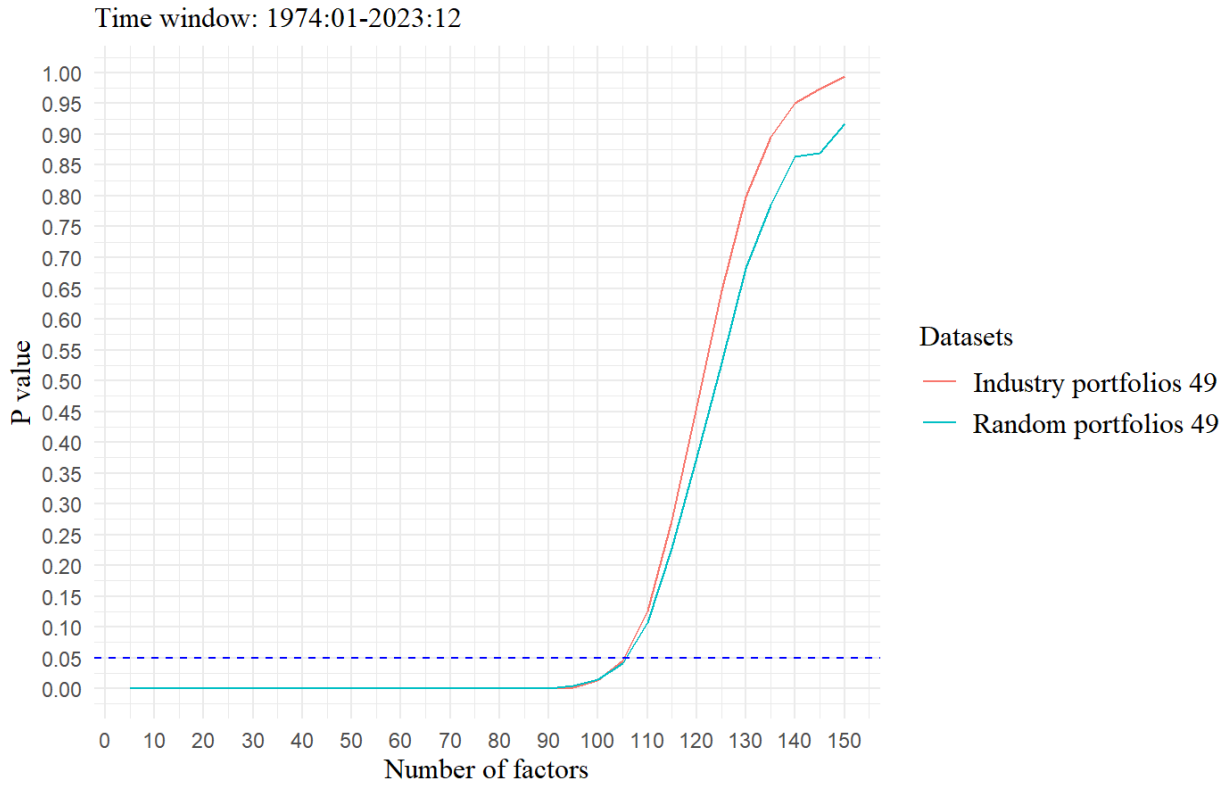


Figure 3: Industry and pseudo-random portfolios: subsample analysis
This figure presents the test results using monthly data of three subsamples: January 1974 to December 1998, January 1987 to December 2011, and January 1999 to December 2023. Given $L = 153$ candidate characteristic-based factors, we plot the p-values for different hypothetical numbers of factors K, each corresponding to the null hypothesis $H_0(\kappa)$, where $\kappa = K/153$. The analysis uses two distinct datasets of test assets: 49 industry portfolios and 49 pseudo-random portfolios. The blue dashed line indicates the 5% significance level.

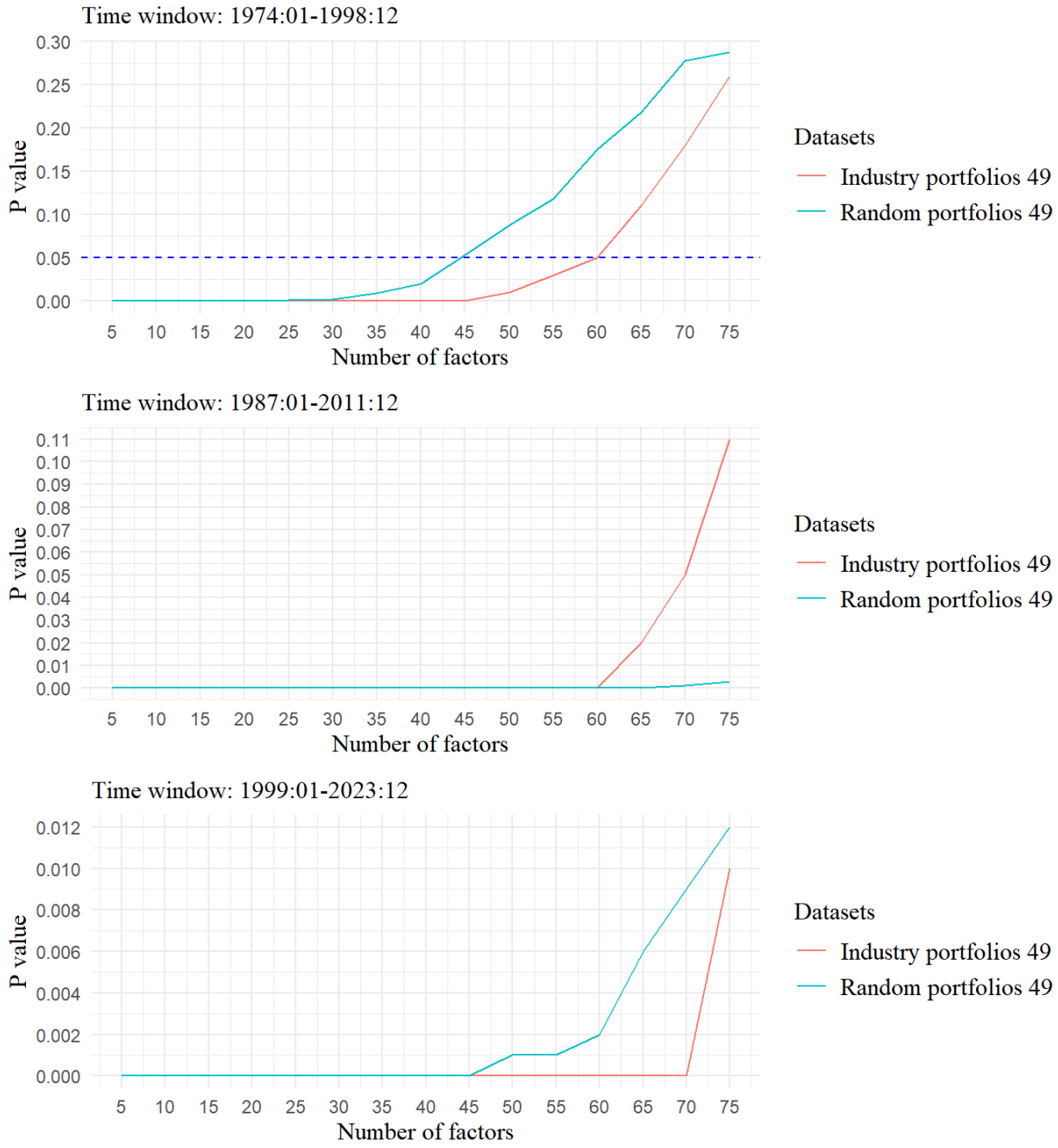


Table 4: Pricing performance of dense and sparse models: whole sample analysis
This table presents the pricing performance of dense and sparse models using the whole sample period from January 1974 to December 2023. The dense models are represented by the first seven principal components of the top one hundred candidate factors, which are selected based on their relative importance to the 49 industry portfolios and 49 pseudo-random portfolios, respectively. The sparse models are obtained by the random sampling approach, and their average performance is reported. For these two sets of test portfolios, we report both the cross-sectional average of absolute pricing errors and the adjusted R-squared.

Pricing measure %	mean($\lvert\hat{\alpha}\rvert$)	mean(adj. R^2)
Panel A: 49 Industry Portfolios		
Mkt+PC1+PC2+$\cdots$+PC7	0.11	0.63
Three-factor model	0.15	0.57
Four-factor model	0.15	0.58
Five-factor model	0.15	0.59
six-factor model	0.15	0.59
Panel B: 49 Random Portfolios		
Mkt+PC1*+PC2*+$\cdots$+PC7*	0.07	0.82
Three-factor model	0.08	0.79
Four-factor model	0.08	0.80
Five-factor model	0.08	0.80
six-factor model	0.08	0.80

Table 5: Pricing performance of dense and sparse models: subsample analysis

This table presents the pricing performance of dense and sparse models using three subsamples: January 1974 to December 1998, January 1987 to December 2011, and January 1999 to December 2023. The dense models are represented by the first several principal components (PCs) that account for over 80% of the variation in the top candidate factors. These top factors are selected based on their relative importance to the 49 industry portfolios and 49 pseudo-random portfolios within the given subsample, respectively. The sparse models are obtained by the random sampling approach, and their average performance is reported. For these two sets of test portfolios, we report both the cross-sectional average of absolute pricing errors and the adjusted R^2.

Pricing measure %	mean($\lvert\hat{\alpha}\rvert$)	mean(adj. R^2)	Pricing measure %	mean($\lvert\hat{\alpha}\rvert$)	mean(adj. R^2)
Panel A: 49 Industry Portfolios			**Panel B: 49 Random Portfolios**		
Panel A-1: 1974:01-1998:12			**Panel B-1: 1974:01-1998:12**		
Mkt+PC1+$\cdots$+PC5	0.21	0.68	Mkt+PC1*+$\cdots$+PC5*	0.10	0.85
Three-factor model	0.22	0.63	Three-factor model	0.11	0.82
Four-factor model	0.22	0.64	Four-factor model	0.12	0.82
Five-factor model	0.22	0.65	Five-factor model	0.12	0.83
Six-factor model	0.22	0.65	Six-factor model	0.12	0.83
Panel A-2: 1987:01-2011:12			**Panel B-2: 1987:01-2011:12**		
Mkt+PC1+$\cdots$+PC4	0.18	0.62	Mkt+PC1*+$\cdots$+PC4*	0.10	0.81
Three-factor model	0.22	0.55	Three-factor model	0.11	0.79
Four-factor model	0.22	0.56	Four-factor model	0.11	0.79
Five-factor model	0.22	0.57	Five-factor model	0.11	0.80
Six-factor model	0.22	0.58	Six-factor model	0.11	0.80
Panel A-3: 1987:01-2023:12			**Panel B-3: 1987:01-2023:12**		
Mkt+PC1+$\cdots$+PC3	0.20	0.60	Mkt+PC1*+$\cdots$+PC3*	0.10	0.80
Three-factor model	0.23	0.54	Three-factor model	0.12	0.78
Four-factor model	0.23	0.55	Four-factor model	0.12	0.78
Five-factor model	0.22	0.56	Five-factor model	0.12	0.79
Six-factor model	0.22	0.57	Six-factor model	0.12	0.79

Table 6: Annualized Sharpe ratio of tangency portfolios constructed using factors of dense and sparse models

This table reports the annualized in-sample and out-of-sample Sharpe ratios of the dense and sparse model representations. For the in-sample analysis, the dense model is represented using the principal components of the top hundred relevant factors for the industry and pseudo-random portfolios, respectively. For the out-of-sample analysis, to avoid forward-looking bias, we omit the principal component transformation step and work directly with characteristic-based factors.

Models	Annualized Sharpe ratio
Panel A: In-sample Analysis	
Mkt + PC1 + PC2 + $\cdots$ + PC7	1.29
Mkt + PC1* + PC2* + $\cdots$ + PC7*	1.27
Three-factor model	0.71
Four-factor model	0.80
Five-factor model	0.88
Six-factor model	0.95
Panel B: Out-of-sample Analysis	
Dense model	0.95
Dense model*	0.86
Three-factor model	0.67
Four-factor model	0.71
Five-factor model	0.75
Six-factor model	0.78

Figure 4: P-Tree portfolios of Grouped P-Trees
This figure presents the test results using monthly data over 40 years, from January 1981 to December 2020. Given $L = 153$ candidate characteristic-based factors, we plot the p-values for different hypothetical numbers of factors K, each corresponding to the null hypothesis $H_0(\kappa)$, where $\kappa = K/153$. The analysis uses four distinct datasets of test assets, each containing 50 P-Tree portfolios. Specifically, we have 20 P-Trees in total, with each P-Tree containing 10 portfolios. We group five trees into one set of test assets, resulting in 50 P-Tree portfolios for each dataset. The blue dashed line indicates the 5% significance level.

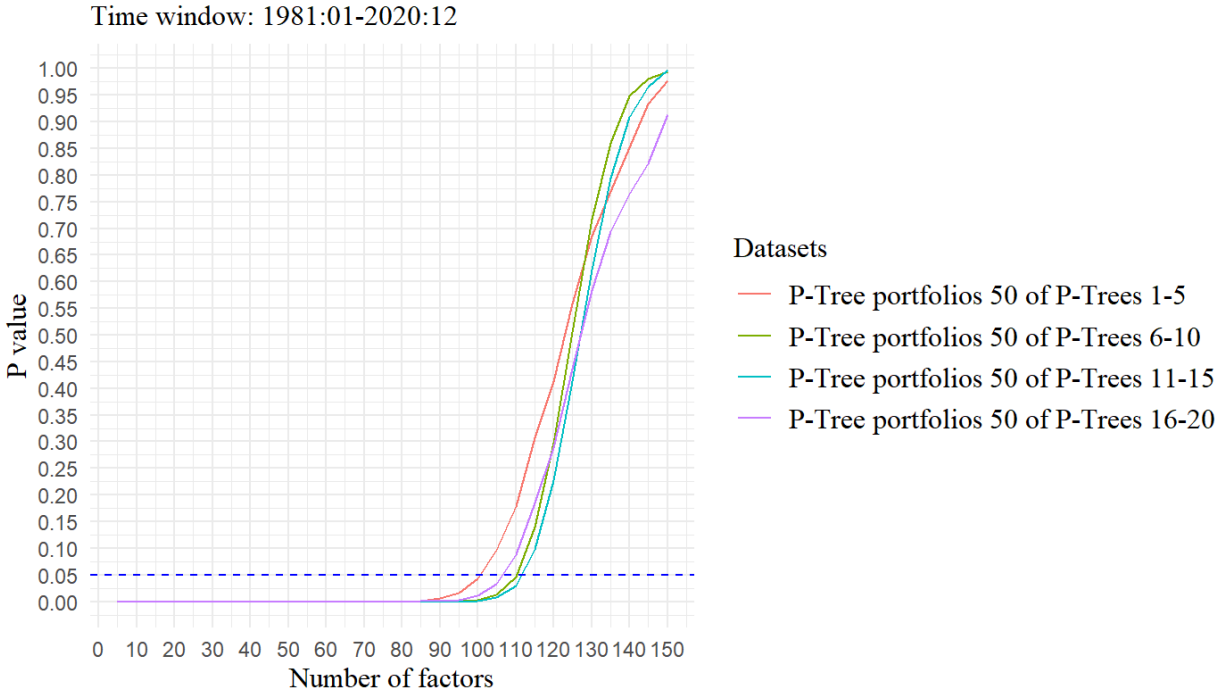


Figure 5: Industry and P-Tree portfolios of P-Tree 1
This figure presents the test results using monthly data over 40 years, from January 1981 to December 2020. Given $L = 153$ candidate characteristic-based factors, we plot the p-values for different hypothetical numbers of factors K, each corresponding to the null hypothesis $H_0(\kappa)$, where $\kappa = K/153$. The analysis uses two distinct datasets of test assets: 10 industry portfolios and 10 P-Tree portfolios of one single tree, i.e., P-Tree 1. The blue dashed line indicates the 5% significance level.

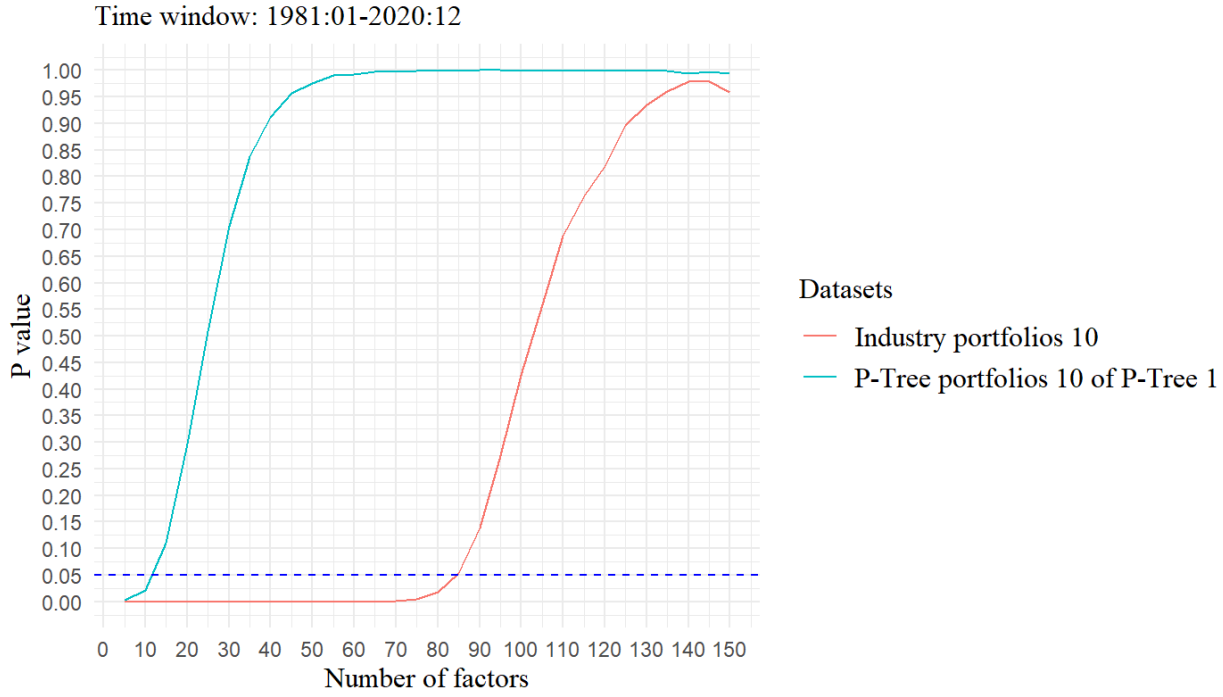
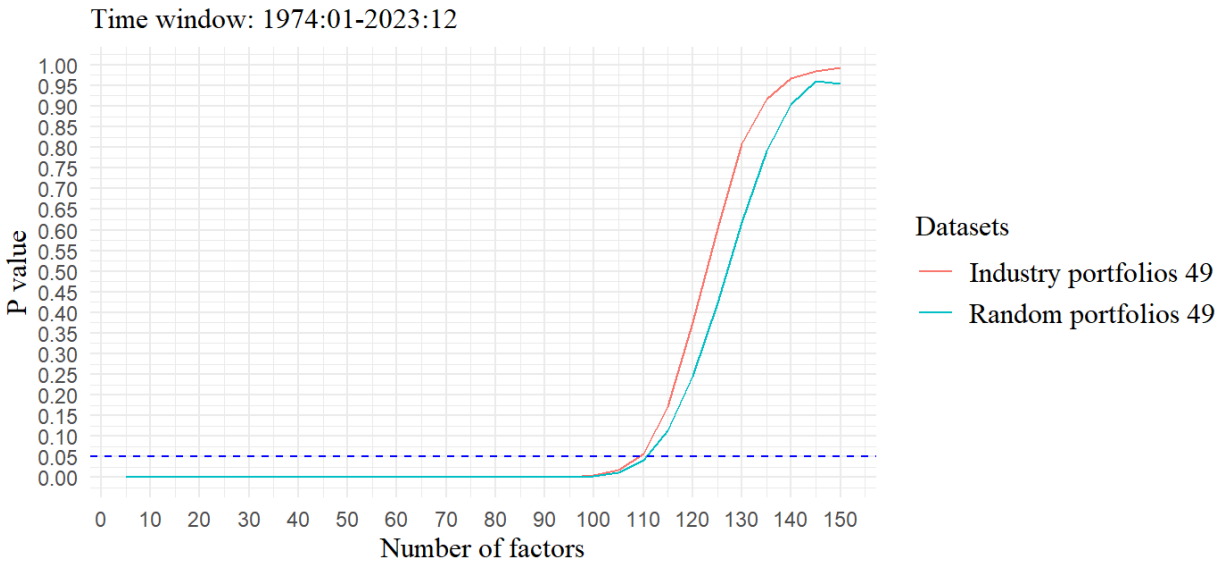
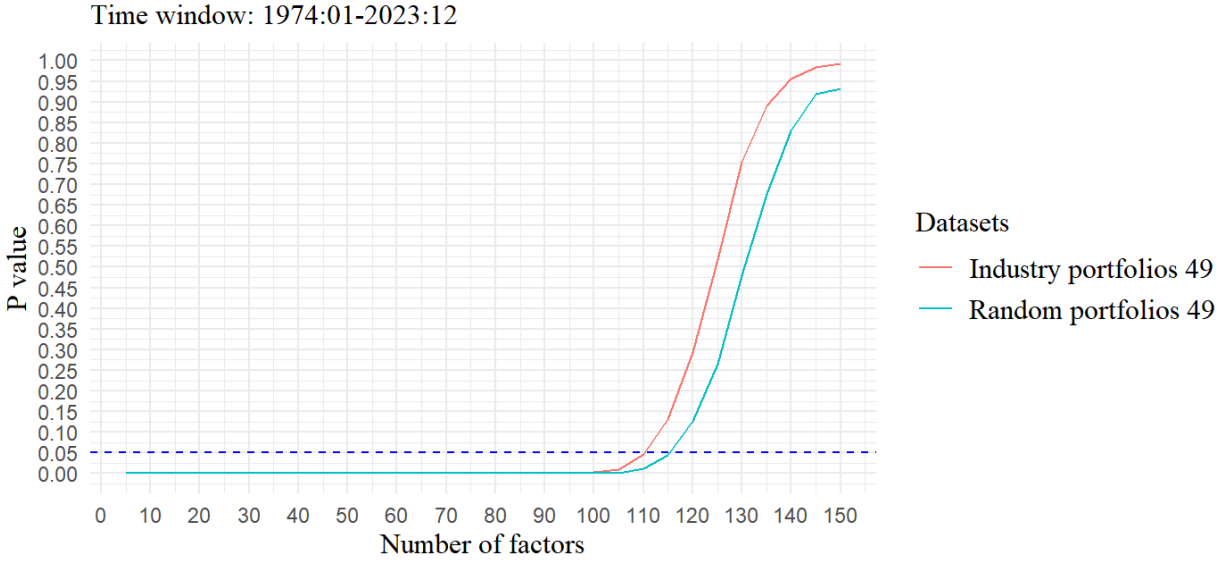


Figure 6: Industry and pseudo-random portfolios: whole sample analysis using FF residuals This figure presents the test results using monthly data of the whole sample: January 1974 to December 2023. Before the test is conducted, all the x and y are residuals after taking a regression using either the FF3 or FF6 models. Given $L = 153$ candidate characteristic-based factors, we plot the p-values for different hypothetical numbers of factors K, each corresponding to the null hypothesis $H_0(\kappa)$, where $\kappa = K/153$. The analysis uses two distinct datasets of test assets: 49 industry portfolios and 49 pseudo-random portfolios. In Panel A, we plot the results for residuals from FF3 model, and in Panel B, we plot the results for residuals from FF6 model. The blue dashed line indicates the 5% significance level.

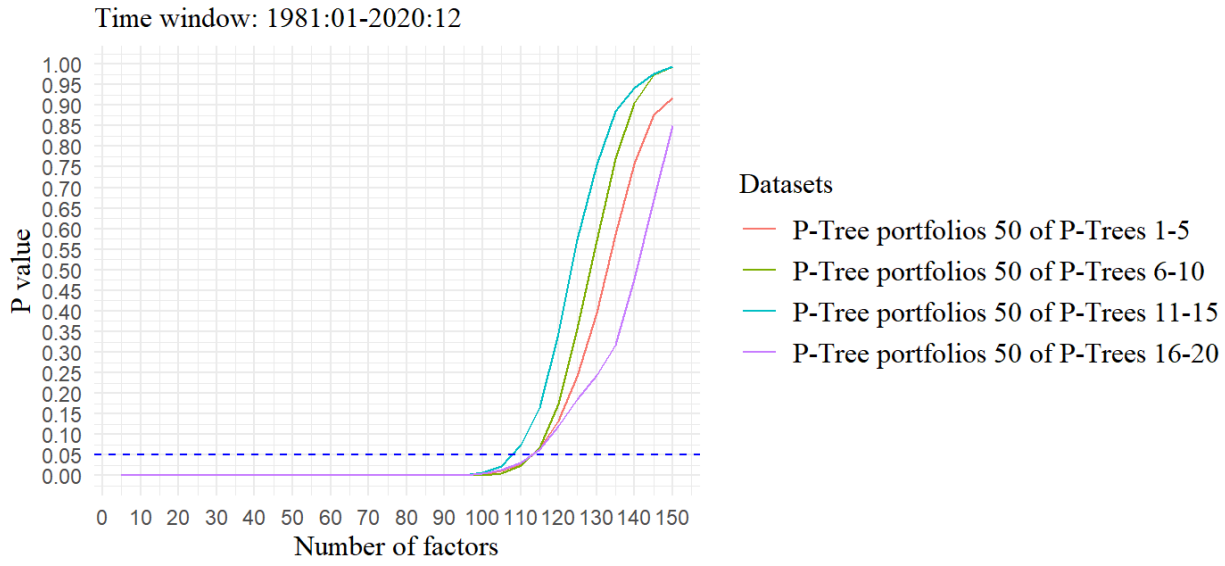


Panel A: FF3 Residuals

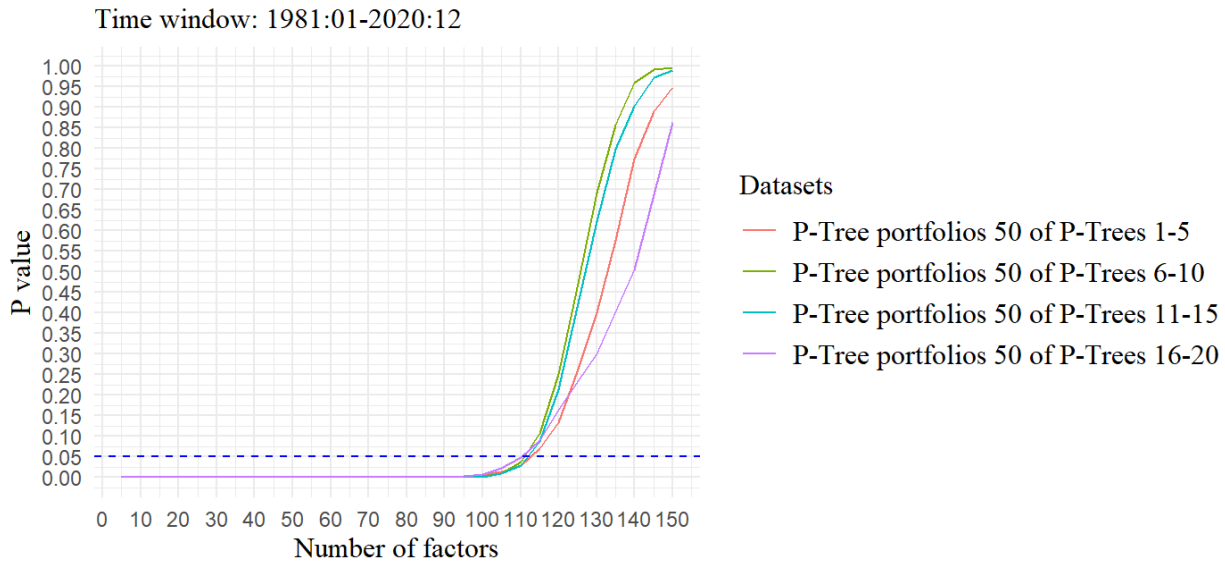


Panel B: FF6 Residuals

Figure 7: P-Tree portfolios of Grouped P-Trees: using FF residuals
This figure presents the test results using monthly data over 40 years, from January 1981 to December 2020. Before the test is conducted, all the x and y are residuals after taking a regression with respect to either the FF3 or FF6 model. Given $L = 153$ candidate characteristic-based factors, we plot the p-values for different hypothetical numbers of factors K, each corresponding to the null hypothesis $H_0(\kappa)$, where $\kappa = K/153$. The analysis uses four distinct datasets of test assets, each containing 50 P-Tree portfolios. Specifically, we have 20 P-Trees in total, with each P-Tree containing 10 portfolios. We group five trees into one set of test assets, resulting in 50 P-Tree portfolios for each dataset. In Panel A, we plot the results for residuals from FF3 model, and in Panel B, we plot the results for residuals from FF6 model. The blue dashed line indicates the 5% significance level.

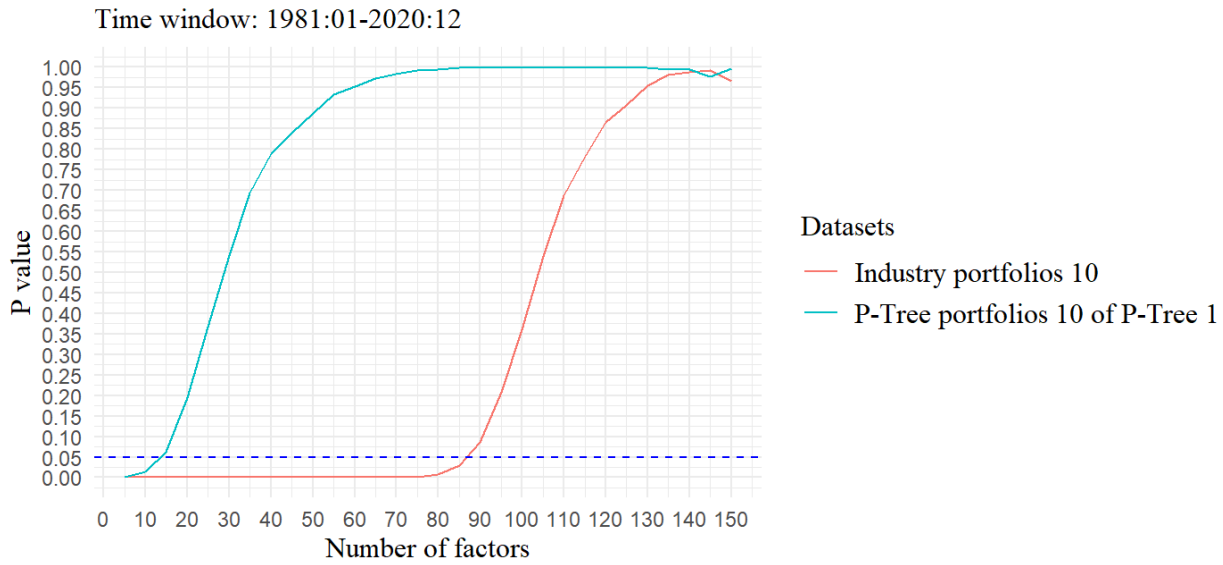


Panel A: FF3 Residuals

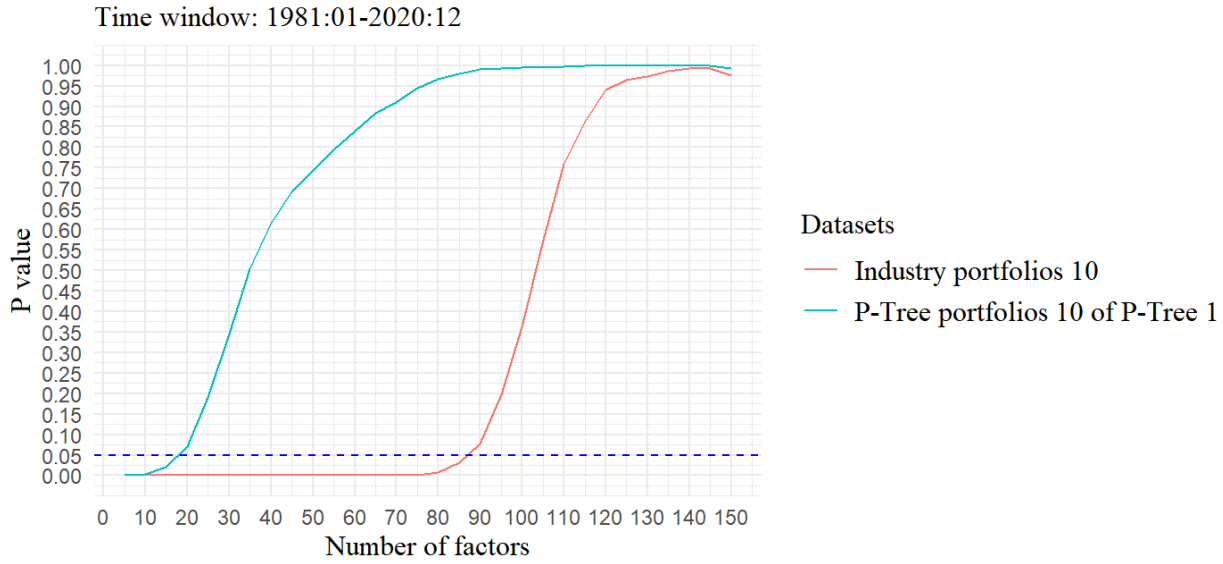


Panel B: FF6 Residuals

Figure 8: Industry and P-Tree portfolios of P-Tree 1: using FF residuals
This figure presents the test results using monthly data over 40 years, from January 1981 to December 2020. Before the test is conducted, all the x and y are residuals after taking a regression with respect to either the FF3 or FF6 model. Given $L = 153$ candidate characteristic-based factors, we plot the p-values for different hypothetical numbers of factors K, each corresponding to the null hypothesis $H_0(\kappa)$, where $\kappa = K/153$. The analysis uses two distinct datasets of test assets: 10 industry portfolios and 10 P-Tree portfolios of one single tree, i.e., P-Tree 1. In Panel A, we plot the results for residuals from FF3 model, and in Panel B, we plot the results for residuals from FF6 model. The blue dashed line indicates the 5% significance level.



Panel A: FF3 Residuals



Panel B: FF6 Residuals

Appendix

A Proof of Proposition 1

Proof. Consider the definition of q_j that

$$q_j(\cdot) = \begin{cases} (1-\kappa)p_0 & \text{for all } \tilde{\boldsymbol{\beta}}_j \text{ s.t. } x_j \leq x_j^*, \\ c_j p_{aj} & \text{for all } \tilde{\boldsymbol{\beta}}_j \text{ s.t. } x_j > x_j^*. \end{cases} \tag{A.1}$$

By definition,

$$\int q_j(b)\mathrm{d}b = (1-\kappa)\int_{x_j(b)\leq x_j^*} p_0(b)\mathrm{d}b + c_j \int_{x_j(b)>x_j^*} p_{aj}(b)\mathrm{d}b. \tag{A.2}$$

The condition $c_j p_{aj} = (1-\kappa)p_0$ when $x_j = x_j^*$ is satisfied iff $c_j = \frac{(1-\kappa)p_0(b)}{p_{aj}(b)}$ for $x_j(b) = x_j^*$.

Therefore

$$c_j = (1-\kappa)\det(I_N + \Omega_j)^{1/2}\exp(-\frac{1}{2}x_j^*). \tag{A.3}$$

Substitute this (A.3) into the integral (A.2) and we have

$$\frac{1}{(1-\kappa)}\int q_j(b)\mathrm{d}b = \int_{x_j(b)\leq x_j^*} p_0(b)\mathrm{d}b + \det(I_N+\Omega_j)^{1/2}\exp(-\frac{1}{2}x_j^*)\int_{x_j(b)>x_j^*} p_{aj}(b)\mathrm{d}b. \tag{A.4}$$

Differentiating the RHS of (A.4) with respect to x_j^* gives the derivative

$$\int_{x_j(b)=x_j^*} p_0(b)\mathrm{d}b - \det(I_N+\Omega_j)^{1/2}\left(\exp(-\frac{1}{2}x_j^*)\int_{x_j(b)=x_j^*} p_{aj}(b)\mathrm{d}b + \frac{1}{2}\exp(-\frac{1}{2}x_j^*)\int_{x_j(b)>x_j^*} p_{aj}(b)\mathrm{d}b\right). \tag{A.5}$$

Observe that when $x_j(b) = x_j^*$, we have readily seen that $p_{aj}(b) = p_0(b)\det(I_N + \Omega_j)^{-1/2} exp(\frac{1}{2}x_j^*)$. Therefore, the derivative (A.5) can be simplified to

$$-\det(I_N + \Omega_j)^{1/2}\frac{1}{2}\exp(-\frac{1}{2}x_j^*)\int_{x_j(b)>x_j^*} p_{aj}(b)\mathrm{d}b < 0. \tag{A.6}$$

Therefore, when there is a well-defined q_j such that $\int q_j = 1$, (A.4) implies there is a unique x_j^* that solves the equation

$$\frac{1}{(1-\kappa)} = \int_{x_j(b)\le x_j^*} p_0(b)\mathrm{d}b + \det(I_N + \Omega_j)^{1/2}\exp(-\frac{1}{2}x_j^*)\int_{x_j(b)>x_j^*} p_{aj}(b)\mathrm{d}b. \tag{A.7}$$

The solution is unique because (A.6) is negative.

Moreover, observe that as $x_j^* \to \infty$, the RHS of (A.4) reaches the minimum of 1. As $x_j^* \to 0$, the RHS of (A.4) reaches maximum $\det(I_N + \Omega_j)^{1/2}$. Therefore, there exists a unique solution if and only if

$$1 \le (1-\kappa)\det(I_N + \Omega_j)^{1/2}. \tag{A.8}$$

This completes the first part of the proof. It follows that the constant x_j^* is solved by setting the RHS of (A.4)

$$\int_{x_j(b)\le x_j^*} p_0(b)\mathrm{d}b + \det(I_N + \Omega_j)^{1/2}\exp(-\frac{1}{2}x_j^*)\int_{x_j(b)>x_j^*} p_{aj}(b)\mathrm{d}b \tag{A.9}$$

$$= \int_{x_j(b)\le x_j^*} p_0(b)\mathrm{d}b + \det(I_N + \Omega_j)^{1/2}\exp(-\frac{1}{2}x_j^*)\int_{b'\Omega_j b>x_j^*} p_0(b)\mathrm{d}b \tag{A.10}$$

to $1/(1-\kappa)$. ■

B Proof of Theorem 2

Proof. By definition, for each $j = 1, \ldots, L$, the N-dimensional vector $\tilde{\boldsymbol{\beta}}_j$ is

$$\tilde{\boldsymbol{\beta}}_j \overset{\mathrm{d}}{=} \boldsymbol{e}_j + \frac{\sqrt{T}}{s_j} diag\left(\frac{1}{\sigma_1}, \ldots, \frac{1}{\sigma_N}\right) \boldsymbol{\beta}_j = \boldsymbol{e}_j + \boldsymbol{\delta}_j, \tag{B.1}$$

where $\boldsymbol{e}_j$ is as defined as in the statement of the theorem and

$$\boldsymbol{\delta}_j = \frac{\sqrt{T}}{s_j} diag\left(\frac{1}{\sigma_1}, \ldots, \frac{1}{\sigma_N}\right) \boldsymbol{\beta}_j. \tag{B.2}$$

Therefore, under the null hypothesis, $\boldsymbol{\delta}_j = z_j \boldsymbol{\gamma}_j$ where $\boldsymbol{\gamma}_j \sim F$ for some arbitrary distribution F.

Recall that the unweighted j-th term in S is $\left[(1-z_j)\min\{x_j(\boldsymbol{e}_j), x_j^*\} + z_j x_j^*\right]$, and the unweighted j-th term in the test statistic T is $\min\left\{x_j(\tilde{\boldsymbol{\beta}}_j), x_j^*\right\}$. We want to show that the former is always an upper bound of the latter under the null hypothesis. According to the null hypothesis, with probability $1-\kappa$ we have $z_j = 0$, and hence $\boldsymbol{\delta}_j = 0$. Therefore

$$\begin{aligned} \min\left\{x_j(\tilde{\boldsymbol{\beta}}_j), x_j^*\right\} &= \min\left\{x_j(\boldsymbol{e}_j + \boldsymbol{\delta}_j), x_j^*\right\} \\ &= \min\left\{x_j(\boldsymbol{e}_j) + x_j(\boldsymbol{\delta}_j) + 2\boldsymbol{e}_j'\left(I_N - \left[I_N + \Omega_j\right]^{-1}\right)\boldsymbol{\delta}_j, x_j^*\right\} \\ &= (1-z_j)\min\left\{x_j(\boldsymbol{e}_j), x_j^*\right\} + z_j x_j^*. \end{aligned} \tag{B.3}$$

On the other hand, with probability κ we have $z_i = 1$. In this case

$$\min\left\{x_j(\tilde{\boldsymbol{\beta}}_j), x_j^*\right\} \leq x_j^* = (1-z_j)\min\left\{x_j(\boldsymbol{e}_j), x_j^*\right\} + z_j x_j^*. \tag{B.4}$$

This shows that for each $j = 1, \ldots, L$, the j-th term in S is always weakly greater than that in T. we conclude that under $H_0(\kappa)$, $T \leq_{1st} S$.

To see the bound is tight, consider in $H_0(\kappa)$ a sequence of $\{F_k\}_{k\in\mathbb{N}} \subseteq \mathscr{F}$ that diverges to infinity: for all $k \in \mathbb{N}$, $F_k(\{b \in \mathbb{R}^N \text{ such that } ||b|| < k\}) = 0$. Consider the test statistic T_k under the special generating distribution in the null hypothesis that γ_j is drawn from F_k for all j. We claim that the distribution of T_k converges to that of S.

To see so, for each $j = 1, \ldots, L$, we have $z_j = 0$ with probability $1 - \kappa$. In this case, $\min\left\{x_j(\tilde{\beta}_j), x_j^*\right\} = (1 - z_j)\min\left\{x_j(\boldsymbol{e}_j), x_j^*\right\} + z_j x_j^*$. Therefore, the j-th term in T equals that in S. On the other hand, with complementary probability κ we have $z_j = 1$. In this case, we have the j-th term in the test statistic $\min\{x_j(\tilde{\beta}_j), x_j^*\} \to x_j^*$ in probability F_k as $k \to \infty$. This is because F_k diverges as $k \to \infty$, causing γ_j diverges, and so δ_j and β_j also diverge to infinity. This limit is exactly the j-th term in S, $(1 - z_j)\min\{x_j(\boldsymbol{e}_j), x_j^*\} + z_j x_j^* = x_j^*$. Therefore under F_k, the j-th term in T converges in probability to that in S for all $j = 1, \ldots, L$. We conclude that the statistic T_k converges in distribution to S along F_k. ■

References

Akaike, H. 1974. A new look at the statistical identification model. *IEEE Transactions on Automatic Control* 19:716–.

Azriel, D., and A. Schwartzman. 2015. The empirical distribution of a large number of correlated normal variables. *Journal of the American Statistical Association* 110:1217–28.

Barillas, F., and J. Shanken. 2018. Comparing asset pricing models. *Journal of Finance* 73:715–54.

Bryzgalova, S., J. Huang, and C. Julliard. 2023. Bayesian solutions for the factor zoo: We just ran two quadrillion models. *Journal of Finance* 78:487–557.

Cai, T. T., and J. Jin. 2010. Optimal rates of convergence for estimating the null density and proportion of nonnull effects in large-scale multiple testing. *The Annals of Statistics* 100–45.

Carpentier, A., and N. Verzelen. 2019. Adaptive estimation of the sparsity in the gaussian vector model. *The Annals of Statistics* 47:93–126.

———. 2021. Optimal sparsity testing in linear regression model. *Bernoulli* 27:727–50.

Chib, S., X. Zeng, and L. Zhao. 2020. On comparing asset pricing models. *Journal of Finance* 75:551–77.

Chib, S., L. Zhao, and G. Zhou. 2024. Winners from winners: A tale of risk factors. *Management Science* 70:396–414.

Cong, L. W., G. Feng, J. He, and X. He. 2023. Growing the efficient frontier on panel trees. *NBER Working Paper* .

Daniel, K., D. Hirshleifer, and L. Sun. 2020. Short-and long-horizon behavioral factors. *Review of Financial Studies* 33:1673–736.

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

———. 2015. A five-factor asset pricing model. *Journal of Financial Economics* 116:1–22.

———. 2018. Choosing factors. *Journal of Financial Economics* 128:234–52.

Fan, J., and R. Li. 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association* 96:1348–60.

Feng, G., S. Giglio, and D. Xiu. 2020. Taming the factor zoo: A test of new factors. *The Journal of Finance* 75:1327–70.

Giannone, D., M. Lenza, and G. E. Primiceri. 2021. Economic predictions with big data: The illusion of sparsity. *Econometrica* 89:2409–37.

Gibbons, M. R., S. A. Ross, and J. Shanken. 1989. A test of the efficiency of a given portfolio. *Econometrica* 29:1121–52.

Green, J., J. R. M. Hand, and X. F. Zhang. 2017. The characteristics that provide independent information about average u.s. monthly stock returns. *Review of Financial Studies* 30:4389–436.

He, A., D. Huang, J. Li, and G. Zhou. 2023. Shrinking factor dimension: A reduced-rank approach. *Management Science* 69:Pages 4973–5693.

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

———. 2020. Replicating anomalies. *Review of Financial Studies* 33:2019–133.

Hsu, D., S. M. Kakade, and T. Zhang. 2014. Random design analysis of ridge regression. *Foundations of Computational Mathematics* 14:569–601.

Huber, P. J. 2004. *Robust statistics*. John Wiley & Sons, Inc.

Jaynes, E. T. 1968. Prior probabilities. *IEEE Transactions on systems science and cybernetics* 4:227–41.

Jensen, T. I., B. Kelly, and L. H. Pedersen. 2023. Is there a replication crisis in finance? *Journal of Finance* 78:2465–518.

Jin, J. 2008. Proportion of non-zero normal means: universal oracle equivalences and uniformly consistent estimators. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 70:461–93.

Jin, J., and T. T. Cai. 2007. Estimating the null and the proportion of nonnull effects in large-scale multiple comparisons. *Journal of the American Statistical Association* 102:495–506.

Kozak, S., S. Nagel, and S. Santosh. 2020. Shrinking the cross-section. *Journal of Financial Economics* 135:271–92.

McLean, R. D., and J. Pontiff. 2016. Does academic research destroy stock return predictability? *Journal of Finance* 71:5–32. ISSN 1540–6261.

Mitchell, T. J., and J. J. Beauchamp. 1988. Bayesian variable selection in linear regression. *Journal of the american statistical association* 83:1023–32.

Novy-Marx, R., and M. Velikov. 2016. A taxonomy of anomalies and their trading costs. *Review of Financial Studies* 29:104–47.

Schwarz, G. 1978. Estimating the dimension of a model. *The annals of statistics* 461–4.

Sharpe, W. F. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance* 19:425–42.

Stambaugh, R. F., and Y. Yuan. 2017. Mispricing factors. *Review of Financial Studies* 30:1270–315.

Tian, M. 2021. Firm characteristics and empirical factor models: A model mining experiment. *The Review of Financial Studies* 34:6087–125.

Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 58:267–88.

Zou, H., and T. Hastie. 2005. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 67:301–20.