Do Common Factors Really Explain the Cross-Section of Stock Returns?

Alejandro Lopez-Lira[1] and Nikolai Roussanov[2]

[1] *Warrington College of Business, University of Florida*

[2] *The Wharton School, University of Pennsylvania and NBER* *

First version: May 2020; This version: January 2021

Abstract

We study the empirical ability of stock characteristics to predict excess returns and document challenges to the notion of a trade-off between systematic risk and expected returns. First, we measure individual stocks' exposures to all common latent factors using a novel high-dimensional method. These latent factors appear to earn negligible risk premia despite explaining virtually all of the common time-series variation in stock returns. Next, we use machine learning methods to construct out-of-sample forecasts of stock returns based on a wide range of characteristics. A zero-cost beta-neutral portfolio that exploits this predictability but hedges all undiversifiable risk delivers a Sharpe ratio above one with no correlation with any systematic factor, thus rejecting the central prediction of the arbitrage pricing theory.

Key words: Asset Pricing, Arbitrage Pricing Theory, Machine Learning, Factor Models, Cross-Section of Returns

*We are grateful for helpful comments and suggestions from Jules van Binsbergen, Sophia Hua, Alexander Belyakov, Xiao Han, Andreas Brøgger, Shrihari Santosh (discussant), Stefano Giglio (discussant), and multiple seminar participants at the NBER Summer Institute 2021, the 3rd Future of Financial Information Conference, the Arison School of Business, IDC, The Wharton School, and BI Norwegian Business School. Emails: alejandro.lopez-lira@warrington.ufl.edu, nroussan@wharton.upenn.edu

The central insight of asset pricing theory is that only systematic risk should be rewarded with an average return in excess of the risk-free rate. In particular, the arbitrage pricing theory (APT) of Ross (1976) posits that certain securities earn higher expected returns than others only because they are more exposed to common (i.e., undiversifiable) risk factors. Conversely, the expected excess returns of portfolios that are hedged against all systematic risk should be zero as long as these portfolios are well-diversified, i.e., the number of securities is large enough for idiosyncratic shocks to average out (Chamberlain and Rothschild (1983)). This logic underlies most traditional multi-factor models that attempt to explain the cross-section of stock returns (E. F. Fama and K. R. French (1993), Cochrane (2005)).[1] In this paper, we provide new empirical evidence that challenges this interpretation.

Our empirical approach overcomes the main obstacle to directly testing the APT: identifying the common factors. We use a novel method to measure individual stocks' exposure to every common latent factor in the time series of stock returns that relies on a singular value decomposition of the return matrix. Importantly, we circumvent the need to estimate the entire covariance matrix of stock returns, which is typically not feasible given the large size of the cross-section of assets relative to the relatively short length of their time series available to the econometrician. We then construct an out-of-sample measure of expected returns for each stock based on a broad set of public and firm-specific signals using machine learning methods trained on past data, which help to avoid in-sample overfitting and data snooping (although, of

1. In contrast, multi-factor models that rely on either investor optimization (such as the intertemporal CAPM of Merton (1973) and Breeden (1979)) or optimal firm decisions (such as investment-based models of Cochrane (1996) and Hou, Xue, and Zhang (2015)), rather than mere absence of near-arbitrage opportunities, do not require that priced factors explain substantial amount of common variation in asset returns over time.

course, similar results obtain when using simpler linear models to forecast returns).

Do common factors driving the time-series variation in returns earn compensation for this undiversifiable risk? Since the systematic factors are latent, we use the estimates of factor loadings to construct factor-mimicking portfolios ("pure-plays"), extending the cross-sectional regression approach described by Fama (1976). These factor-mimicking portfolios explain a considerable fraction of the time-series variation of stock returns, despite a substantial amount of idiosyncratic risk (e.g., 50 factors corresponding to the 50 largest eigenvalues of the covariance matrix together explain just over 50 percent of the total variance of individual stocks' excess returns). However, all of these latent factors carry either negligible or zero price of risk on average. Consequently, they explain almost none of the cross-sectional variation in average excess returns. Furthermore, the latent factors carry a high variance by design, implying very low Sharpe ratios.

These findings are especially problematic when combining them with cross-sectional return predictability. We build pure-play portfolios using both factor loadings and our measure of the expected returns. This procedure allows us to construct a trading strategy that exploits the characteristics for the maximum out-of-sample fit that is effectively orthogonal to all common risk factors. We show that this "beta-netural" strategy remains profitable even after all of the ex-ante systematic risk exposure has been captured by the former. Thus, hedging portfolios based on the out-of-sample predictors of excess return against all systematic risk reduces their volatility without reducing their average returns, hence increasing their Sharpe ratios, in sharp contrast to the main prediction of the arbitrage pricing theory.

On the "bright" side, consistent with some of the existing literature, we observe a decline in the portfolios' performance after the 2000s, driven by the decrease in return predictability. On the not-so-bright side, the beta-neutral "arbitrage" portfolio is still a superior investment, with a Sharpe ratio higher than the market's and beating its unhedged counterpart both in terms of standard deviation and returns. However, the finding is especially troublesome considering that the hedged portfolios' theoretical excess return and Sharpe ratios should be exactly zero.

1 Related Literature

Our paper builds on and combines three notable strands of the cross-sectional asset pricing literature: estimation of latent drivers of common variation in returns ("risk factors"), predicting returns using stock characteristics ("anomalies"), and disentangling the roles of characteristics versus covariances with risk factors in driving expected returns.

Roll and Ross (1980), Chamberlain and Rothschild (1983), Connor and Korajczyk (1986), Connor and Korajczyk (1988), Shukla and Trzcinka (1990), and Pelger (2020), among others, show that under several different sets of assumptions that support the APT, the factor realizations and the factor loadings can be recovered from the covariance matrix. We build on their findings and note that the latent factor betas for every individual stock can also be obtained using the return matrix utilizing a singular value decomposition. We join Kim and Korajczyk (2018), Chen, Connor, and Korajczyk (2018), Pukthuanthong and Roll (2020) in noting the benefits of employing a large number of assets relative to the time period to estimate asset pricing relationships.

Fama and MacBeth (1973), Jacobs and Levy (1988), Lewellen (2015), Gu, Kelly, and Xiu (2020), and Freyberger, Neuhierl, and Weber (2020), among many others, provide ample evidence that cross-sectional returns are predictable, and machine learning enhances these predictions.[2] The APT logic has come under attack as carefully selected portfolios designed to exploit this characteristics-based predictability appear to defy risk-based explanations. There are two main arguments in defense of APT and risk-based models, more generally. First, maybe we have not found the right risk factors. Second, perhaps the carefully selected portfolios are a product of data snooping. We combine high-dimensional beta estimation and return prediction to address these two critics and test APT's implications. We take advantage of the existing literature on expected returns and use random forest regression to forecast returns in our main analysis.[3] Random forest regression is a standard non-linear and non-parametric ensemble method that averages multiple forecasts from (potentially) weak predictors.[4] As such, it is designed to provide an out-of-sample prediction that is less sensitive to data snooping than traditional forecasting methods.

We further complement the literature on return predictability by combining expected returns and high dimensional covariance estimation using high-frequency data. We then construct beta-neutral portfolios using pure-plays and show that the resulting hedged portfolios produce the same (if not higher) expected and realized returns while hedging all systematic risk exposure (in the APT sense). Thus, our paper

2. See especially Gu, Kelly, and Xiu (2020) for an excellent survey of standard methods. See Alti and Titman (2019) for a dynamic model that justifies the return predictability.

3. We also consider linear models as well as a Bayesian framework that combines variable selection with shrinkage: Spike-and-slab regression. See Giannone, Lenza, and Primiceri (2017) and Bianchi, Büchner, and Tamoni (2019) for recent applications in finance and economics of the latter.

4. See Binsbergen, Han, and Lopez-Lira (2020) for a thorough overview of random forest regression applied to earnings forecasting.

also relates to the debate of characteristics versus covariances. Daniel and Titman (1997) — and later Back, Kapadia, and Ostdiek (2015) and Kirby (2019) — show that the characteristics drive the risk-premium for the Fama–French three-factor model: "It is the characteristics rather than the covariance structure of returns that appear to explain the cross-sectional variation in stock returns." Complementing the previous findings, E. F. Fama and K. R. French (2020) find that using the cross-section regression approach of Fama and MacBeth (1973) to construct cross-section factors, those factors provide a better description of the cross-section of returns. Building on this observation, Daniel et al. (2020) show that we can use this finding to improve reduced-form portfolios' performance.

We complement their research in three critical ways. (i) Instead of building portfolios using characteristics, we build portfolios sorted on expected returns, summarizing all of the predictive information of multiple characteristics, following the machine-learning and prediction literature. (ii) We estimate all the latent factors' loadings and hedge against all latent factors instead of specifying an ex-ante multi-factor model and hedging each portfolio against one beta. (iii) We focus on testing the APT by forming portfolios sorted on expected returns with zero exposure against all risk factors that drive the covariance.

We also complement the recent research of Giglio and Xiu (2021) and Giglio, Xiu, and Zhang (2021) who show that standard estimators of risk premia in linear asset pricing models are biased if some priced factors are omitted. They argue for using a three-pass procedure that involves PCA and supervised PCA, respectively. Related to their results, we show that the statistical significance of the magnitude of the 'risk premia' of portfolios sorted on expected returns increases when restricting the loadings on the latent factors to be zero since the standard deviation decreases substantially.

Lettau and Pelger (2020) find that imposing a cross-sectional restriction that expected excess returns are linear in factor loadings while estimating latent factors improves their asset pricing performance. This result is not surprising given the common factor structure documented for characteristics-based portfolios, e.g. Kozak, Nagel, and Santosh (2018). Despite the seeming differences, we argue that our results are compatible with theirs, since they implicitly find a trade-off between the cross-sectional fit and the time-series fit.

In fact, Kelly, Pruitt, and Su (2019) also document an implicit trade-off between the model's ability to explain time-series and cross-sectional variation. The standard principal components explain the time-series variation significantly better, but fail entirely at explaining the cross-sectional variation. In contrast, their instrumented principal components explain better the cross-sectional variation at the cost of capturing less of the time-series variation.[5] We explain this finding in detail by noting that the latent factors are not priced and the cross-sectional variation seems to be primarily driven by return predictability unrelated to the latent factor exposures.

Kozak, Nagel, and Santosh (2018) argue that for "typical test asset portfolios, their return covariance structure essentially dictates that the first few principal components must explain the cross-section of expected returns. Otherwise, near-

5. Kelly, Pruitt, and Su (2019) use 'total R2' as their preferred measure of time-series variation. Total R2 is defined as Total $R^2 = 1 - \frac{\sum_{i,t}(r_{i,t+1} - \hat{\beta_{i,t}} f_{t+1})^2}{\sum_{i,t} r_{i,t+1}^2}$: "The total R2 thus includes the explained variation due to contemporaneous factor realizations and dynamic factor exposures, aggregated over all assets and time periods." We confirm in the online appendix that the results also apply when using the average time-series R2: $\frac{1}{N}\sum_i \frac{\sum_t (r_{i,t+1} - \hat{\beta}_i f_{t+1})^2}{\sum_t r_{i,t+1}^2}$.

arbitrage opportunities would exist." In our application, we use individual stocks instead of portfolios, and, as a result, many principal components drive the covariance structure. Nevertheless, we can hedge them, which results in portfolios with very high Sharpe ratios.

Cooper et al. (2021) apply PCA to portfolios sorted on CAPM anomalies and note that their resulting statistical factors produce competitive results when compared against common reduced form factor models. We explain the differences noting that the portfolios sorted on CAPM anomalies are selected on their ability to generate in-sample return spreads. When using the whole cross-section without selecting the best performing sorts we find the latent factors are not priced.

Kim, Korajczyk, and Neuhierl (2021) propose a new methodology for forming arbitrage portfolios that utilize the information contained in firm characteristics for both abnormal returns and factor loadings. First, they demean the returns using rolling windows and project them into the span of characteristics. Then they apply standard principal component analysis to the resulting matrix of characteristic-based portfolios (see Chen, Roussanov, and Wang (2021) for a refinement of this approach that accounts for the nonlinear relationship between characteristics and covariances). In contrast, we use high-frequency data to estimate the latent factor loadings directly from the matrix of returns without resorting to any preliminary dimension reduction. Hence, we can estimate and hedge all the relevant systematic factors, and we are not restricted to specifying the factor loadings as (linear) functions of a specific set of characteristics. Importantly, as argued above, without hedging most of the systematic variation, it is infeasible to test the approximate APT.

2 Theoretical Framework

We follow closely the setup in Ross (1976) and the recent exposition in Pelger (2020). The excess returns of the individual securities follow

$$r^e_{t+1} = \alpha_{i,t} + \beta'_{i,t} f_{t+1} + \epsilon_{i,t+1}, \tag{1}$$

where ϵ_{t+1} is mean zero and sufficiently uncorrelated to permit the law of large numbers to hold cross-sectionally. As pointed out by Hansen and Richard (1987), assuming that factor loadings are constant, i.e. $\beta_{i,t} = \beta_i$, is not innocuous, since the conditional mean-variance efficient portfolio is not necessarily unconditionally mean-variance efficient. We follow Lewellen and Nagel (2006), and Pelger (2020) and assume the covariance structre, and hence vector of betas, is stable within a short time window, but otherwise allow it to vary over time.[6] Without loss of generality, the elements of f_{t+1} are uncorrelated. APT implies that $\alpha_{i,t} = 0$. Roll and Ross (1980), Chamberlain and Rothschild (1983), Connor and Korajczyk (1986), Connor and Korajczyk (1988), Shukla and Trzcinka (1990), and Pelger (2020), among others, show under different sets of assumptions that the factor realizations and the factor loadings can be recovered from the covariance matrix: The factor loadings correspond to the eigenvectors of the covariance matrix, and the factor realizations to the projection of the assets into the eigenvectors.

There are two fundamental empirical challenges in testing the APT. First, identifying common factors requires estimating covariance matrices with a much larger

6. We use one year as the time period over which betas are approximately constant for our main analysis, but our results using one month are similar.

dimension than the length of the time series of stock returns. As a result, researchers are often forced to use portfolios sorted on selected characteristics of assets as proxies for common factors. Second, the econometrician does not observe expected returns, who must resort to using variables that have shown a historical relationship with average returns. Our approach aims to address both of these challenges.

2.1 Beta Estimation

We exploit the fact that the beta loadings with respect to the latent factors are the right singular vectors of the (compact) singular value decomposition of the return matrix, which allows us to estimate individual stock's conditional betas with respect to all the latent statistical factors driving the common time-series variation while circumventing the covariance estimation. We use daily data to estimate conditional factor loadings in the spirit of Lewellen and Nagel (2006).

At month t, let $\mathbf{R}_t$ be the *demeaned* matrix of size $T_t \times N_t$ with the past daily excess returns, where T_t is the number of days of trading in the past twelve month period, typically 252, and N_t is the number of stocks in the cross-section. We allow for a time varying number of stocks. We demean the returns using the rolling average return across time and stocks, since individual stock means are estimated poorly at the daily level (Merton (1980)). We winsorize this return matrix at the 1% level to remove outliers. We drop the time subscripts for the remaining of the section for notational ease. We consider the case where $N >> T$ which implies the rank of the return matrix $\mathbf{R}$ is equal to T.

Using the singular value decomposition, we get $\mathbf{R} = \mathbf{U_1 S_1 V_1}^\top$, where $\mathbf{U}_1, \mathbf{V}_1$ are orthogonal matrices with size $T \times T$, and $N \times N$ respectively. $\mathbf{S_1}$ is a (rectangular) diagonal matrix of size $T \times N$ that contains the singular values in decreasing order.

Since the elements off the diagonal of $\mathbf{S_1}$ are zero, we can write it as $\mathbf{S_1} = [\mathbf{S}, \mathbf{0}]$ with $\mathbf{S_1}$ a matrix of size $T \times T$ and instead focus on the compact singular value decomposition: $\mathbf{R} = \mathbf{U}\mathbf{S}\mathbf{V}^\top$. $\mathbf{S}$ is a $T \times T$ diagonal matrix that contains the singular values in descending order. $\mathbf{U}$ is a $T \times T$ orthogonal matrix. $\mathbf{V}$ is a $N \times T$ matrix with the property that: $\mathbf{V}^\top \mathbf{V} = \mathbf{I}_T$, and $\mathbf{I}$ is the identity matrix of size $T \times T$.

Let $\mathbf{C} = \frac{1}{T}\mathbf{R}^\top\mathbf{R}$, with size $N \times N$, denote the empirical covariance matrix of the daily returns, since $N >> T$, the covariance matrix is singular (the rank of $\mathbf{C}$ is at most $T - 1$). Then

$$\mathbf{C} = \frac{1}{T}\mathbf{R}^\top\mathbf{R} = \frac{1}{T}\mathbf{V}\ \mathbf{S^2}\ \mathbf{V}^\top = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^\top, \tag{2}$$

where the second equality is the eigenvalue decomposition of the covariance matrix.

The eigenvectors of the covariance matrix, contained in the columns of $\mathbf{V}$, are usually called the principal directions, and we show that they correspond to the betas of the returns with respect to the principal components. The principal loadings are defined as $\mathbf{V}\mathbf{S}/\sqrt{T}$. The principal components, the latent factor realizations, are given by $\mathbf{U}\mathbf{S} = \mathbf{R}\mathbf{V}$. The variance of each principal component is contained in the diagonal matrix $\mathbf{\Lambda} = \frac{1}{T}\mathbf{S^2}$:

$$\frac{1}{T}(\mathbf{U}\mathbf{S})^\top\mathbf{U}\mathbf{S} = \mathbf{\Lambda}, \tag{3}$$

and they correspond to the eigenvalues, which are equal to the normalized square singular values, $\lambda_k = \frac{s_k^2}{T}$.

Furthermore, the covariance of the returns with respect to the principal components is given by the following relationship:

$$\frac{1}{T}\mathbf{R}^{\top}(\mathbf{US}) = \frac{1}{T}\mathbf{VSU}^{\top}\mathbf{US} = \mathbf{V}\frac{\mathbf{S}^2}{T} = \mathbf{V\Lambda}. \tag{4}$$

Hence, the betas, in the usual sense, are contained in the columns of the $\mathbf{V}$ matrix. Using the singular value decomposition, we measure individual stocks' covariances (betas) with respect to every latent factor driving the cross-sectional covariance of returns, thus circumventing the need to estimate the entire covariance matrix, which is infeasible given the large size of the cross-section relative to the length of the time series.

The eigenvectors, and hence the betas, are orthogonal, but not (cross-sectionally) uncorrelated. The projections of the data into the space generated by the eigenvectors, the latent factors, are (time-series wise) uncorrelated.

To avoid confusion with time subscripts, we will define $K \equiv T$, so that the return matrix is of size $K \times N$.

2.2 Normalization of the Principal Components

For ease of exposition we normalize the weights of the eigenvectors to give the first factor a natural portfolio interpretation. Let $\omega_k = 1/sum(V_1)$ where V_1 is the first eigenvector, provided the sum is non zero, which is empirically the case. Since the k principal component is given by $PC_k = \mathbf{R}^{\top}V_k$, Where V_k is the k^{th} column of the $\mathbf{V}$ matrix, we define $w_k = V_k\omega_k$, and we normalize the factor k as: $f_k = \mathbf{R}^{\top}V_k\omega_k = \mathbf{R}^{\top}w_k$. We also define $\beta_k = V_k\omega_k^{-1}$ and the variance of the re-scaled factor is given by $\sigma_k^2 = \lambda_k\omega_k^2 = \frac{(s_k\omega_k)^2}{T}$.

The factors are latent, and are only normalized up to sign. We normalize each eigenvector to have positive cross-sectional mean at every period for consistency. This normalization allows us to retain the usual interpretation of

$$\frac{\lambda_i}{\sum_{i=1}^{n} \lambda_i}, \tag{5}$$

as the fraction of total return variance explained by the given factor.

2.3 Time-varying Beta Estimation

Let $\boldsymbol{\beta}_t$ the matrix that contains the betas (normalized eigenvectors) of size $N \times K$ and β_k is the vector of β_i, k, t for every stock index by i at a given time t (with the time-index omitted in the vector for notation easiness) for given factor k. Since the betas (and the factor realizations) are latent variables, their estimate can only be obtained ex-post, i.e., once the return is known at the end of the month. Hence, forming portfolios based on these betas is infeasible in real-time. However, because the betas are fairly stable over short time horizons, we can construct a "forecast" of $\boldsymbol{\beta}_t$ that we call $\hat{\boldsymbol{\beta}}_t$, using the beta estimates obtained in the previous periods.

We use a rolling panel linear model to "forecast" the realization of the betas using each stock beta's previous twelve lags. We use twelve lags since we are using twelve-month rolling windows to estimate the betas every month, which causes a high level of autocorrelation due to the overlapping windows. We forecast the beta for each factor separately, with a typical model of the form:

$$\beta_{i,k,t} = a_k + \sum_{j=1}^{12} \lambda_{t-j} \beta_{i,k,t-j} + u_{i,k,t}, \tag{6}$$

Every month, we run a regression of beta on its 12 lags using a 36-month window, which, together with the data used to construct the lagged betas, amounts to 48 months of data. We forecast the next month's beta using the coefficients from the rolling regression of monthly so as not to subject our estimation to look-ahead bias. In addition, we consider alternative window lengths of 12 months, 60 months, 120 months, and an expanding window approach. Results are reported in Table 12. When more than 36 months are used for training, the R^2 becomes stable. Furthermore, the table shows that higher-order betas are harder to predict, which coincides with them driving less of the time-series variation of the cross-section of returns.

2.4 Latent Factor Portfolios

Since the factors that we consider are latent, we need to project them onto the return space in order to study their empirical asset pricing properties. We construct factor-mimicking portfolios so that a portfolio tracking the k-th latent factor has a beta of one with respect to the k factor and zero otherwise. Let w be the vector of portfolio weights. We first consider zero-cost portfolios. The portfolio tracking each latent factor k solves the following problem, which leads to a maximally diversified portfolios in the sense that avoids extreme positions (either long or short) in individual stocks:

$$\underset{w}{\text{minimize}} \; \frac{1}{2} w' \Omega w, \tag{7}$$

$$s.t. \; w'\iota = 0, \tag{8}$$

$$s.t.\ w'\beta_k = 1, \tag{9}$$

$$s.t.\ w'\beta_j = 0\ \forall j \neq k, \tag{10}$$

with $\Omega_{ii} = \frac{1}{mktcap_i}$ if value-weighting is desired (and identity matrix for equal weighting). The returns of the portfolio k closely correspond to the coefficients of β_k in a Fama-Macbeth regression of realized returns on a constant and the betas as the regressors, except the portfolio weights are normalized.

We additionally consider factors with weight equal one, where the problem looks the same as above, except that

$$s.t.\ w'\iota = 1. \tag{11}$$

In the later case, the factors will be heavily correlated with each other, and with the market portfolio.

2.5 Expected Returns

We now consider the problem of forecasting returns at time t + 1 using only a (strict) subset of the information set available at time t. In practice, this amounts to modeling the conditional expectation as a (possibly non-linear) function of characteristics available at time t.

$$E[r_{i,t+1}|c_{i,t}] = f(c_{i,t}) \equiv \mu_{i,t} \tag{12}$$

For the main part of the paper, we use random forest regressions to forecast future returns. Random forest regression is a non-linear and non-parametric ensemble method that averages multiple forecasts from (potentially) weak predictors. Thus, the ultimate prediction is superior to a forecast following from any one individual predictor (Breiman 2001). We discuss the algorithm comprehensively in the Appendix. We train the algorithm using 60-months rolling windows, which is analogous to a linear rolling regression forecast. For robustness, we also consider a Bayesian framework, spike-and-slab regression (Ishwaran and Rao (2005)), as well as standard linear regressions and find similar results.

2.6 Pure Play Portfolios

Having obtained each stock's individual betas and expected returns, we can form hedge portfolios that deliver a pre-specified level of exposure to a particular latent factor while being orthogonal to all of the other factors, or delivering a desired conditional expected return. As such these are "pure plays" on particular factor exposures (or expected returns). Our construction builds on Fama (1976) interpretation of the slopes of the Fama-MacBeth cross-sectional regression (see also Back, Kapadia, and Ostdiek (2015), Gilje, Ready, and Roussanov (2016), Kirby (2019) and Lopez-Lira (2020)).[7] We omit the time subscripts for ease of exposition. We first describe the general approach and then consider its specific applications.

Let w be the portfolio weights. We collect the values of μ^i, the expected returns at time t constructed using the characteristics, and build portfolios with a given

7. See Kelly, Malamud, and Pedersen (2020) for a different and novel framework that instead of using pure-plays, uses eigenvectors of a prediction matrix to construct zero-beta portfolios.

expected return $w'\mu = \mu_0$. Furthermore, we use the betas with respect to the common factors and target specific levels of covariances $w'\beta_k = \beta_{k,0}$, where β_k is the vector of coefficients of the projection of returns into the k-th principal component and $\beta_{k,0}$ is the target value, usually zero.

The portfolio weights solve the following problem, which leads to a maximally diversified portfolio in the sense that avoids extreme positions (either long or short) in individual stocks, by giving them as even a weight as possible, subject to the constraints:

$$\underset{w}{\text{minimize}} \ \frac{1}{2} w' \Omega w, \tag{13}$$

where Ω is a weighting matrix, for example to make the portfolios value weighted.[8] The maximization is subject to portfolios being zero cost,

$$s.t. \ w'\iota = 0, \tag{14}$$

delivering a pre-specified expected return,

$$s.t. \ w'\mu = \mu_0, \tag{15}$$

and, most importantly, providing the desired factor exposures,

$$s.t. \ w'\beta_k = \beta_{0,k} \ \forall k = 1...\hat{k}. \tag{16}$$

8. A diagonal Ω with, $\Omega_{ii} = \frac{1}{mktcap_i}$, where $mktcap_i$ is the market capitalization for stock i results in the value weighted portfolio in the absence of beta restrictions and expected stocks restrictions. For the main analysis, we follow Kirby (2019) and omit stocks that are not liquid (stocks whose market capitalization is below the 20th percentile of the NYSE as in E. F. Fama and K. R. French (2008) and run an unweighted minimization problem but we also show results using the market capitalization as the weight.

We collect the restrictions in a vector $a_0 = [0, \mu_0, \beta_0]'$ of size $(2 + \hat{k}) \times 1$ and the characteristics in a matrix $A = [\iota, \mu, B]$ of size $n \times (2+\hat{k})$ where $\hat{k} \leq K$ is the number of betas to hedge and B is the matrix that contain those betas, of size $n \times \hat{k}$, so that the problem can be written compactly as

$$\underset{w}{\text{minimize}}\ \frac{1}{2} w'\Omega w \tag{17}$$

$$\text{subject to } A'w = a_0. \tag{18}$$

The Lagrangian is given by:

$$\frac{1}{2} w'\Omega w + \lambda'(a_0 - A'w), \tag{19}$$

with solution

$$w = \Omega^{-1} A (A'\Omega^{-1} A)^{-1} a_0. \tag{20}$$

When we use an identity matrix as the weighting matrix we end up with:

$$w = A(A'A)^{-1} a_0. \tag{21}$$

These weights are very similar to those generated by using a Fama-Macbeth Cross-sectional Regression (FM) of returns on both expected return predictors and factor betas, with the difference that in the FM regression approach we get $w_{k,FM} = A(A'A)^{-1} e_k$, where e_k is a standard basis vector in $\mathbb{R}^{\hat{k}}$ with a one in the k-th position and zeros in the rest.

Notice that when $\beta_{0,k} = 0\ \forall k = 1...\hat{k}$ then $a_0 = [0, \mu_0, \mathbf{0}_{\hat{k}\times 1}]' = \mu_0 [0, 1, \mathbf{0}_{\hat{k}\times 1}]' = \mu_0 e_2$, where again e_2 is a standard basis vector in $\mathbb{R}^{\hat{k}}$ with a one in the second position and zeros in the rest. Hence it follows that the weights can be written as:

$$w = \mu_0(\Omega^{-1}A(A'\Omega^{-1}A)^{-1}e_2). \tag{22}$$

Notice the weights are linear functions of μ_0. Instead of choosing an arbitrary level of μ_0, we normalize the weights so that the zero cost portfolio has the sum of the absolute value of its weights equal to two, $\sum_i |w_i|/2 = 1$, for comparability with traditional long-short portfolios. The normalization, naturally, does not affect the Sharpe ratio.

2.6.1 Unhedged and Hedged Portfolios Exploiting Expected Returns

In order to construct a zero cost portfolio that exploits return predictability but does not attempt to eliminate any of the systematic risk we can remove the beta constraint (16) and the portfolio problem specializes to the following:

$$\underset{w}{\text{minimize}} \; \frac{1}{2} w'\Omega w, \tag{23}$$

$$s.t. \; w'\iota = 0, \tag{24}$$

$$s.t. \; w'\mu = \mu_0, \tag{25}$$

with $\Omega_{ii} = \frac{1}{mktcap_i}$ for value weighting or identity for equal weights.

Since the solution is, again, linear in μ_0, the level of targeted expected returns, instead of specifying an ex-ante level we normalize the weights so that the sum of the absolute value of the weights equal to two, $\sum_i |w_i|/2 = 1$ for comparability with traditional long-short portfolios. The returns of the portfolio closely correspond to the coefficients in a Fama-Macbeth regression of realized returns on a constant and the predicted return as the sole regressor, except the portfolio weights are normalized.

In order to construct a long-short beta-neutral portfolio, we want to hedge the beta exposure and hence the portfolio problem is the following:

$$\underset{w}{\text{minimize}} \ \frac{1}{2} w' \Omega w, \tag{26}$$

$$s.t. \ w' \iota = 0, \tag{27}$$

$$s.t. \ w' \mu = \mu_0, \tag{28}$$

$$s.t. \ w' \beta_k = 0 \ \forall k = 1, ..., \hat{K}, \tag{29}$$

with $\Omega_{ii} = \frac{1}{mktcap_i}$ if value-weighting is desired (and identity matrix for equal weighting). And $\hat{K}$ the number of betas to hedge.

Since the solution is, once again, linear in μ_0, the level of targeted expected returns, instead of specifying an ex-ante level we normalize the weights so that the sum of the absolute value of the weights equal to two, $\sum_i |w_i|/2 = 1$ for comparability with traditional long-short portfolios. The returns of the portfolio closely correspond to the coefficients of the predicted return in a Fama-Macbeth regression of realized returns on a constant, the predicted return, and the betas as the regressors, except the portfolio weights are normalized.

3 Empirical Analysis

3.1 Data

We use the Center for Research in Security Prices (CRSP) daily stock files to measure individual stocks' betas as described in the previous section. For return prediction, we use the 62 characteristics for predicting returns constructed exactly as in Freyberger, Neuhierl, and Weber (2020). We build the characteristics using CRSP and Standard and Poor's Compustat. We follow Weber (2018) in determining when the balance-sheet data is available, typically in June of the year following the fiscal year-end.

We omit stocks whose size falls below the 20th percentile of the NYSE following E. F. Fama and K. R. French (2008) and Kirby (2019) to avoid any liquidity concerns. Our sample period is from July 1965 until June 2014, and we use the first ten years as a training period of the first rolling forecast. Hence, we start our primary analysis from 1974. We make sure not to use any forward-looking information, so all of our results are out-of-sample by design. We use monthly returns for the remainder of the analysis.

3.2 Latent Factors

We begin by estimating the betas with respect to the latent factors as described in Section 2.1. For each month we estimate the factor loadings using the previous year's return observations. Figure 1 plots the cumulative share of variance explained by the first $\hat{k}$ principal components (from largest to smallest), $\sum_{i=1}^{\hat{k}} \frac{\lambda_i}{\sum_{i=1}^{K} \lambda_i}$ for a particular sample month, January 1999 (the choice is arbitrary but the results are very similar for all months).

[Insert Figure 1 about here]

Figure 1 shows the very high-dimensional nature of the cross-section of stock returns in a typical year. The first five largest principal components explain about 20% of the total variance of returns. We need around 50 factors to capture 50% of the variation, 100 factors to capture 75% of the variation, and 200 factors to capture 99% of the variation. The previous result is despite the well-known fact that PCA provides the (in-sample) solution for which the (time-series) variation explained is the highest for a given number of factors.

[Insert Figure 2 about here]

Because we are interested in portfolio formation, instead of using the in-sample betas, which are measurable at the end of the period, we 'forecast' them using a rolling panel regression. Reassuringly, the fitted betas are almost as good as the in-sample betas in explaining the time-series variation. For example, Figure 2 shows the cross-sectional average of the time-series R2 in a one-year rolling regression of individual stocks into the latent factors using either the actual betas or the predicted betas as a function of the number of factors.

[Insert Figure 3 about here]

Figure 3 shows the time-varying cumulative percentage of the variance explained by the principal components. We can see that there is substantial degree of variation in the share of total variance explained by the first ten principal components, ranging from 10% to 40%, but it appears stationary over time. Notably, there is a lot less variation in the fraction that is explained by the first 100 PC, which varies around 80% throughout most of our sample.

3.3 Factor Risk Premiums

Since the factors are latent, we construct factor-mimicking portfolio using pure-plays as described in Section 2.4. We first consider zero-weight portfolios. For a given factor j, each stock's weight in this portfolio is positive if it has an above average beta on the principal component, and negative otherwise. Each portfolio has a beta of 1 with respect to its own factor, $\sum_i w_{it}^j \beta_{it}^j = 1$ and zero with respect to other factors, $\sum_i w_{it}^k \beta_{it}^k = 0, k \neq j$. We normalize them as described in the previous section. We also consider unit weight portfolios, which have weights that sum up to one, in order to mimic the behavior of portfolios that are long equities, e.g. the market portfolio. Note that technically these portfolios are still zero-weight, since we are considering stock returns in excess of the risk free rate.

[**Insert Figure 4 about here**]

Table 1 depicts the main properties of returns on the zero-weight factor mimicking portfolios corresponding to the first five largest principal components. None of them carry a (positive) risk-premium on average, with Sharpe ratios that are very close to zero, e.g. 0.10 for the first principal component and 0.19 on the third (which is the largest in terms of average return). The table also shows the average return, and standard deviation of the latent factors. The discrepancy between the average return of the first factor and that of the market portfolio comes from the zero-weight nature of the projection, as we can see from table 4, which reports the correlation matrix of the first five factor with the Fama-French 6 factors (including momentum). The first factor has a correlation of 0.65 with the market portfolio (as well as a positive correlation with SMB), the second factor has a correlation of −0.44 with the market (consistent with its somewhat negative average return), and all the others have essentially zero correlations with the market. The first latent factor is thus similar to a long-short portfolio that exploits variation in market betas. It is well known that the relationship between market beta and average returns is often flat (or even inverted) in the cross-section of stock returns, which is consistent with our evidence of a very small correlation for bearing systematic risk.[9] Table 2 reports the correlation between the zero-weight latent factors, the market portfolio and the Fama–French five factors.

It is instructive to compare these results with those for unit-weight portfolios, which more closely resemble that of a typical investor who is not able to take on short positions on stocks. Table 3 presents descriptive statistics for the first five factors. We see that now all of the five factor-mimicking portfolios earn substantial and statistically significant average returns, with Sharpe ratios between 0.5 and 0.65, as well as higher volatility than their zero-weight portfolios above. The correlation with the market portfolio is 0.95 for the first factor, which now closely resembles

9. See Black (1972), Blume (1975), E. Fama and K. French (1992), Baker, Bradley, and Wurgler (2011), and Frazzini and Pedersen (2014), among others.

a long portfolio of all equities, but is also positive and high, around 0.5 for the remaining four portfolios (notably smaller for second factor though, which had a large negative correlation in its zero-weight form, but now positive at 0.35. Interestingly, its correlation with SMB also flips from being strongly negative (at -0.49) to slightly positive.

Yet the correlation structure of these factor mimicking portfolios is otherwise largely unchanged. The first factor is positively correlated with SMB and negatively with the other Fama-French factors, including momentum, both in the zero- and unit-weight versions. The second factor is negatively correlated with momentum in both cases. It is strongly positively correlated with HML, RMW, and CMA in the zero-weight version, but most of these correlations go down and almost disappear in the zero-weight version, while the other factors become more negatively correlated to these factors but not strongly. In any case, the correlation structure of the unit-weight factors beyond the first one or two is not particularly informative, since the unit-weight constraint pushes them too far away from zeroing in on the independent sources of common variation. This is evidenced by the high cross-correlations of the unit-weight portfolios, despite the fact that they are supposed to mimic orthogonal factors. The zero-weight portfolios, in turn, largely fulfill their role, with correlations close to zero, especially beyond the first two factors. The remaining correlations reflect the unconditional nature of the estimated correlation matrix, since zero conditional covariances can translate into nonzero unconditional ones if returns are predictable and conditional expected returns are correlated (the SVD estimation above is done on rolling windows and thus approximates conditional betas).

Figure 4 displays cumulative returns to investing in the zero-weight factor mimicking portfolios over time compared to the CRSP value-weighted market return and rolling over the risk-free bond. The returns on all the factors are much closer to the latter than the former. Comparing this to the performance of unit-weight portfolios in Figure 5 we see that the latter much more closely resemble the market index, some outperforming and others underperforming it over different time periods.

[**Insert Table 1 about here**]

[**Insert Table 2 about here**]

3.4 Hedged and Unhedged Portfolios

We now consider the portfolios that are constructed to take advantage of the optimal predictors of conditional excess returns as described in Section 2.6.1. In constructing the hedged "Beta-Neutral" portfolio we consider the first 50 factors but results remain similar with 25 or 100 factors. The fact that the latent factors are identified only up to rotation is unimportant when we hedge them, since once a factor is hedged, any rotation is hedged as well, i.e., the null-space of implied by the betas is invariant to rotations.

Table 5 Panel A shows our main result. We find that the beta-neutral (i.e., hedged) portfolio produces a significant average excess return similar to its long-short (i.e., unhedged) counterpart (around .7% per month or 8.5% annualized) while typically doubling its annualized Sharpe ratio (1.51 vs 0.75). For comparison, the average market excess return is 0.52 % per month and its annualized Sharpe ratio is

0.39 We remark that this result is in complete opposition to APT's prediction, since both the average return of the hedged portfolio and its the Sharpe ratio should be zero. The return of the unhedged portfolio is not necessarily a problem for risk-based models such as the APT. The complications result from the return patterns of the hedged portfolio.

According to APT, all of the risk premium comes from factor exposure. We find exactly the reverse result: the latent factors are not rewarded at all, and there is a large unexplained excess return. To make matters more troublesome, the common factors drive most of the time-series variation and hedging all systematic risk significantly reduces the portfolio's variance, greatly increasing its Sharpe ratios; it even somewhat increases its average return in the latter half of the sample.

[**Insert Table 5 about here**]

Figure 6 shows the Sharpe ratio as a function of the number of factors for the hedged beta-neutral portfolio. The Sharpe ratios are annualized by multiplying by the square root of twelve. We can observe the Sharpe ratios stabilize between 25 and 100 factors. While the results remain similar afterwards, the computational burden increases and becomes potentially unstable, since, for example, when hedging 200 factors, we need to invert a 200×200 matrix.

[**Insert Figure 6 about here**]

Figure 7 shows the cumulative performance of the market, the unhedged portfolio and the hedged portfolio. There is a marked difference in the return patterns before and after the 2000s. Figures 8 and 9 show in the detail the periods before and after the 2000s.

[Insert Figure 7 about here]

Table 5 Panel B shows the result of the period 1974–1999. The average return of the long-short portfolio is 1.12% per month, and its Sharpe ratio is 1.49. For comparison, the average return of the beta-neutral portfolio is around 1% per month, and its Sharpe ratio is a substantial 2.32, which is very far from APT's prediction of zero. The market's average excess return and Sharpe ratio are 0.65% per month and 0.48, respectively.

[Insert Figure 8 about here]

Table 5 Panel C shows the result of the period 2000–2014. The long-short portfolio's average return is negative and equal to -0.06% per month, and its Sharpe ratio is -0.04. We conjecture that the variables that had good forecasting power before the in the pre-2000 period must have lost its forecasting power, perhaps because of improved market efficiency. However, the beta-neutral portfolio's average return remains positive and around 0.25% per month, and its Sharpe is 0.44. For comparison, the market's average excess return is 0.28% per month and its Sharpe ratio is 0.21. In short, we observe a decline in return predictability, consistent with the argument of Mclean and Pontiff (2016) that many "anomalies" have been "discovered" over the years, and presumably subsequently driven out due to the rise of quantitative (and factor-based!) investment management. Yet our evidence that the hedged portfolio's Sharpe ratio is not only positive but higher than the market's even in the post-2000 sample (albeit much smaller than pre-2000) is still quite problematic for the risk-based paradigm of asset pricing.

[Insert Figure 9 about here]

3.4.1 Characteristics and Covariances

At first glance our results might seem to be in contradiction with the recent literature, which argues that characteristics are proxies for factor betas (e.g., Kelly, Pruitt, and Su (2019) and Lettau and Pelger (2020)). In fact, our results are consistent with the evidence documented in those studies and the seeming differences stem from the inherent ability of characteristics based models to explain the cross-section of returns, albeit at a cost of explaining less of the time-series variation.

First, Kelly, Pruitt, and Su (2019) document an implicit trade-off between the model's ability to explain time-series and cross-sectional variation. For example, they show that the standard principal components explain the time-series variation significantly better than their instrumented principal components (IPCAs) with a 'Total' R^2 of 33.8% compared to 19% for their IPCAs. In contrast, their instrumented principal components explain better the cross-sectional variation with a 'predictive' R^2 of 0.7 %, whereas the PCAs fail entirely at explaining the cross-sectional variation, as evidenced by a negative predictive R2.[10]

Second, Lettau and Pelger (2020) find that imposing a cross-sectional restriction that expected excess returns are linear in factor loadings while estimating latent factors improves their asset pricing performance. However, they also find a trade-off

10. Total $R^2 = 1 - \frac{\sum_{i,t}(r_{i,t+1} - \hat{\beta_{i,t}} f_{t+1})^2}{\sum_{i,t} r^2_{i,t+1}}$, Predictive $R^2 = 1 - \frac{\sum_{i,t}(r_{i,t+1} - \hat{\beta_{i,t}} E[f_{t+1}])^2}{\sum_{i,t} r^2_{i,t+1}}$

between explaining the time-series variation and the cross-sectional variation, in fact it is inherent to their estimation approach.[11]

Of course, the classic APT does not contemplate such a trade-off. Kozak, Nagel, and Santosh (2018) argue that commonality in terms of time-series comovement of returns with factors is necessary in the absence of arbitrage opportunities (or infinite Sharpe ratios). In fact, priced risk factors should be the high-order principal components, i.e. the ones corresponding to the largest eigenvalues of the covariance matrix of returns. Thus, giving up on the time-series fit in order to improve on the cross-section makes the resulting factors difficult to interpret. Our findings reflect this difficulty: we show that stock expected returns that are functions of characteristics are *not* correlated to the largest common factors in the sense of driving the time-series variation of returns. Uncomfortably, these factors are not useful for explaining cross-sectional differences in mean returns, in sharp opposition to APT's main prediction.

3.4.2 Cross-Sectional Regressions

It is natural to look at the previous results through the lens of Fama-MacBeth cross-sectional regressions. We consider predictive regression of realized returns on our estimates of conditional expected returns every period. We expect our forecast to have a positive coefficient and be statistically significant. If we had the perfect forecast, the value of this coefficient would be unity. APT's prediction is that when adding

11. In detail, their minimization problem is:

$$\mathrm{RP-PCA:}\ \ \widehat{\mathrm{F}}_{\mathrm{RP}}, \widehat{\Lambda}_{\mathrm{RP}} = \underset{\Lambda,\mathrm{F}}{\operatorname{argmin}} \quad \underbrace{\frac{1}{NT}\sum_{n=1}^{N}\sum_{t=1}^{T}(X_{nt} - \mathrm{F}_t\,\Lambda_n^\top)^2}_{\text{unexplained TS variation}} + \gamma\,\underbrace{\frac{1}{N}\sum_{n=1}^{N}\left(\overline{\mathrm{X}}_n - \overline{\mathrm{F}}\,\Lambda_n^\top\right)^2}_{\text{XS pricing error}},$$

and $\gamma > 0$ is the parameter characterizing the trade-off.

the covariance estimates to the predictive regression, the return predictability coming from our forecast should decline, since according to APT, return predictability comes only from factor exposures.

We consider regressions of the form:

$$r_{i,t+1} = a_t + b_t x_{it} + u_{r,i,t}, \tag{30}$$

and

$$r_{i,t+1} = a_t + b_t x_{it} + \sum_{j=1}^{k} \lambda_t \beta_{it}^{j} + \varepsilon_{r,i,t}, \tag{31}$$

where x_{it} is the predicted return and β_{it}^{j} the exposure of the i-th stock to the j-th factor.

We do not control for common characteristics since they are used to form the predicted returns and consequently are correlated with it. We omit stocks whose size falls below the 20th percentile of the NYSE following E. F. Fama and K. R. French (2008) and Kirby (2019) to avoid any concern of illiquidity.

[Insert Table 6 about here]

[Insert Table 7 about here]

Tables 6 and 7 show the cross-sectional regression estimates. Table 6 shows unweighted regressions and Table 7 shows weighted regressions by the market capitalization. As expected, the coefficient corresponding to the predicted return remains statistically significant in almost all specifications. The exception is during the 2000-20014 period, consistent with the time-series descriptive statistics and showing the decline in return predictability. Since we are measuring expected returns with error, however, the estimate suffers from the usual attenuation bias, and so should be greater than zero but less than one, as we find in the data. In agreement with the descriptive statistics and in contrast to APT's prediction, the predicted return's coefficient increases its statistical significance when factor betas are added as controls.

3.5 Time-Series Regressions

Given the failure of latent common factors in explaining the expected returns predicted by stock characteristics, we turn to popular factor models considered in empirical asset pricing literature. We test whether the long-Short (unhedged) and the beta-neutral (hedged) strategy returns are spanned by the CAPM, the Fama–French Five-Factor model E. F. Fama and K. R. French (2015), the q5 model of Hou et al. (2021) and the mispricing factors model of Stambaugh and Yuan (2017). Tables 8 and 9 report the results. The only "traditional" factor that our trading strategy significantly loads on is momentum - both the unhedged and the hedged portfolios display a significantly positive betas, suggesting that momentum returns are not compensation for undiversifiable risk. Interestingly, the *hedged* portfolio also loads *negatively* on RMW portfolio, suggesting that the latter does relate to common sources of risk in stock returns but its average return has the "wrong" sign.

Needless to say, both portfolios have significantly positive alphas with respect to all of the models that we consider. Both the long-short (unhedged) and the beta-neutral (hedged) strategy also positively load on the expected growth factor of Hou et al. (2021) and on the PERF factor of Stambaugh and Yuan (2017). Interestingly, the alphas of the unhedged strategy are positive but not statistically significant against both of these models, while the hedged portfolio is clearly not spanned by either, with a statistically significant alpha of about 45 basis points per month with respect to both. This suggests that at least some of the factors utilized by these models capture important sources of common variation in stock returns as reflected in their covariance matrix, and thus hedged out by our construction of the beta neutral portfolio.

[**Insert Table 8 about here**]

[**Insert Table 9 about here**]

4 Conclusion

APT has had a remarkable impact on the literature of cross-sectional returns and multi-factor models. Nevertheless, testing APT has been generally complicated since tests using portfolios face the critic that maybe we haven't find the right factors or the portfolios incur in data snooping.

This paper combines high-dimensional beta estimation, return prediction, and pure plays to address these two critics and test APT's implications. First, we use a novel approach to high-dimensional beta estimation to get individual stock's time-varying betas with respect to all the latent statistical factors driving the covariance. Then, we combine the betas with ex-ante forecasts of returns and use pure plays to construct zero-beta portfolios sorted on expected returns.

We overwhelmingly reject APT's main theoretical prediction that the zero-beta portfolios' excess returns and Sharpe ratios should be zero. We attribute the inability of APT to explain the zero-beta portfolios to the fact that none of the latent factors have a non-negligible price of risk on average; consequently, hedging all risk exposure is not only harmless but also desirable. Hedging all exposures, however, seems problematic from a general equilibrium perspective.

References

Alti, Aydoğan, and Sheridan Titman. 2019. "A Dynamic Model of Characteristic-Based Return Predictability." *Journal of Finance* 74, no. 6 (December): 3187–3216. ISSN: 15406261. https://doi.org/10.1111/jofi.12839. https://onlinelibrary.wiley.com/doi/full/10.1111/jofi.12839%20https://onlinelibrary.wiley.com/doi/abs/10.1111/jofi.12839%20https://onlinelibrary.wiley.com/doi/10.1111/jofi.12839.

Back, Kerry, Nishad Kapadia, and Barbara Ostdiek. 2015. "Testing Factor Models on Characteristic and Covariance Pure Plays." *Working Paper,* https://ssrn.com/abstract=2621696.

Baker, Malcolm, Brendan Bradley, and Jeffrey Wurgler. 2011. "Benchmarks as limits to arbitrage: Understanding the low-volatility anomaly." *Financial Analysts Journal* 67:40–54.

Bianchi, Daniele, Matthias Büchner, and Andrea Tamoni. 2019. "What Matters When? Time-Varying Sparsity in Expected Returns." *Working Paper,* https://ssrn.com/abstract=3438754.

Binsbergen, Jules H. van, Xiao Han, and Alejandro Lopez-Lira. 2020. *Man vs. Machine Learning: The Term Structure of Earnings Expectations and Conditional Biases.* Technical report, Working Paper Series 27843. National Bureau of Economic Research. https://doi.org/10.3386/w27843. http://www.nber.org/papers/w27843.

Black, Fischer. 1972. "Capital market equilibrium with restricted borrowing." *The Journal of Business* 45:444–455.

Blume, Marshall. 1975. "BETAS AND THEIR REGRESSION TENDENCIES." *The Journal of Finance* 30 (3): 785–795. https://doi.org/https://doi.org/10.1111/j.1540-6261.1975.tb01850.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.1975.tb01850.x. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1975.tb01850.x.

Breeden, Douglas T. 1979. "An intertemporal asset pricing model with stochastic consumption and investment opportunities." *Journal of Financial Economics* 7 (3): 265–296. ISSN: 0304-405X. https://doi.org/10.1016/0304-405X(79)90016-3.

Breiman, Leo. 2001. "Random Forests." *Mach. Learn.* (USA) 45 (1): 5–32. ISSN: 0885-6125. https://doi.org/10.1023/A:1010933404324. https://doi.org/10.1023/A:1010933404324.

Chamberlain, Gary, and Michael Rothschild. 1983. "Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets." *Econometrica* 51 (5): 1281–1304. ISSN: 00129682, 14680262. http://www.jstor.org/stable/1912275.

Chen, Qihui, Nikolai Roussanov, and Xiaoliang Wang. 2021. "Semiparametric Conditional Factor Models: Estimation and Inference, with Qihui Chen and Xiaoliang Wang." *working paper.*

Chen, Zhuo, Gregory Connor, and Robert A Korajczyk. 2018. *A performance comparison of large-n factor estimators,* 1. https://doi.org/10.1093/rapstu/rax017. https://academic.oup.com/raps/article/8/1/153/3858092.

Cochrane, John H. 1996. "A cross-sectional test of an investment-based asset pricing model." *Journal of Political Economy* 104 (3): 572–621. ISSN: 00223808. https://doi.org/10.1086/262034. https://www.journals.uchicago.edu/doi/abs/10.1086/262034.

———. 2005. *Asset Pricing: Revised Edition.* Princeton University Press. ISBN: 9781400829132. https://books.google.co.uk/books?id=20pmeMaKNwsC.

Connor, Gregory, and Robert A Korajczyk. 1986. "Performance measurement with the arbitrage pricing theory. A new framework for analysis." *Journal of Financial Economics* 15 (3): 373–394. ISSN: 0304405X. https://doi.org/10.1016/0304-405X(86)90027-9.

———. 1988. "Risk and return in an equilibrium APT. Application of a new test methodology." *Journal of Financial Economics* 21 (2): 255–289. ISSN: 0304405X. https://doi.org/10.1016/0304-405X(88)90062-1.

Cooper, Ilan, Liang Ma, Paulo Maio, and Dennis Philip. 2021. "Multifactor Models and Their Consistency with the APT." *The Review of Asset Pricing Studies* 11 (2): 402–444. ISSN: 2045-9920. https://doi.org/10.1093/rapstu/raaa024. https://academic.oup.com/raps/article/11/2/402/6050855.

Daniel, Kent, Lira Mota, Simon Rottke, and Tano Santos. 2020. "The Cross-Section of Risk and Returns." *The Review of Financial Studies* 33 (5): 1927–1979. ISSN: 0893-9454. https://doi.org/10.1093/rfs/hhaa021. https://doi.org/10.1093/rfs/hhaa021.

Daniel, Kent, and Sheridan Titman. 1997. "Evidence on the Characteristics of Cross Sectional Variation in Stock Returns." *The Journal of Finance* 52 (1): 1–33. https://doi.org/https://doi.org/10.1111/j.1540-6261.1997.tb03806.x. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1997.tb03806.x.

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

Fama, Eugene F. 1976. *Foundations of Finance.* Basic Books.

Fama, Eugene F, and Kenneth R French. 1993. "Common risk factors in the returns on stocks and bonds." *Journal of Financial Economics* 33 (1): 3–56. ISSN: 0304-405X. https://doi.org/https://doi.org/10.1016/0304-405X(93)90023-5. http://www.sciencedirect.com/science/article/pii/0304405X93900235.

———. 2008. "Average Returns, B/M, and Share Issues." *The Journal of Finance* 63 (6): 2971–2995. https://doi.org/10.1111/j.1540-6261.2008.01418.x. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.2008.01418.x.

———. 2015. "A five-factor asset pricing model." *Journal of Financial Economics* 116 (1): 1–22. ISSN: 0304-405X. https://doi.org/http://dx.doi.org/10.1016/j.jfineco.2014.10.010. http://www.sciencedirect.com/science/article/pii/S0304405X14002323.

———. 2020. "Comparing Cross-Section and Time-Series Factor Models." Edited by Andrew Karolyi. *The Review of Financial Studies* 33 (5): 1891–1926. ISSN: 0893-9454. https://doi.org/10.1093/rfs/hhz089. https://academic.oup.com/rfs/article/33/5/1891/5555879.

Fama, Eugene F, and James D MacBeth. 1973. “Risk, Return, and Equilibrium: Empirical Tests.” *Journal of Political Economy* 81 (3): 607–636. ISSN: 00223808, 1537534X. http://www.jstor.org/stable/1831028.

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

Freyberger, Joachim, Andreas Neuhierl, and Michael Weber. 2020. “Dissecting Characteristics Nonparametrically.” *The Review of Financial Studies* 33 (5): 2326–2377. ISSN: 0893-9454. https://doi.org/10.1093/rfs/hhz123. https://doi.org/10.1093/rfs/hhz123.

Giannone, Domenico, Michele Lenza, and Giorgio E Primiceri. 2017. “Economic Predictions with Big Data: The Illusion Of Sparsity.” *CEPR Discussion Papers,* https://ideas.repec.org/p/cpr/ceprdp/12256.html.

Giglio, Stefano, and Dacheng Xiu. 2021. “Asset pricing with omitted factors.” *Journal of Political Economy* 129 (7): 1947–1990. ISSN: 1537534X. https://doi.org/10.1086/714090. https://www.journals.uchicago.edu/doi/abs/10.1086/714090.

Giglio, Stefano, Dacheng Xiu, and Dake Zhang. 2021. “Test Assets and Weak Factors.” *SSRN Electronic Journal,* https://doi.org/10.2139/ssrn.3768081. https://papers.ssrn.com/abstract=3768081.

Gilje, Erik, Robert Ready, and Nikolai Roussanov. 2016. “Fracking, drilling, and asset pricing: Estimating the economic benefits of the shale revolution.” *National Bureau of Economic Research working paper.*

Gu, Shihao, Bryan Kelly, and Dacheng Xiu. 2020. "Empirical Asset Pricing via Machine Learning." *The Review of Financial Studies* 33 (5): 2223–2273. ISSN: 0893-9454. https://doi.org/10.1093/rfs/hhaa009. https://doi.org/10.1093/rfs/hhaa009.

Hansen, Lars, and Scott F. Richard. 1987. "The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models." *Econometrica* 55:587–613.

Hou, Kewei, Haitao Mo, Chen Xue, and Lu Zhang. 2021. "An Augmented q-Factor Model with Expected Growth*." *Review of Finance* 25, no. 1 (February): 1–41. ISSN: 1572-3097. https://doi.org/10.1093/rof/rfaa004. https://academic.oup.com/rof/article/25/1/1/5727769.

Hou, Kewei, Chen Xue, and Lu Zhang. 2015. "Digesting anomalies: An investment approach." *Review of Financial Studies* 28 (3): 650–705. ISSN: 0893-9454. https://doi.org/10.1093/rfs/hhu068. https://academic.oup.com/rfs/article-lookup/doi/10.1093/rfs/hhu068.

Ishwaran, Hemant. 2015. "The Effect of Splitting on Random Forests." *Mach. Learn. (USA)* 99 (1): 75–118. ISSN: 0885-6125. https://doi.org/10.1007/s10994-014-5451-2. https://doi.org/10.1007/s10994-014-5451-2.

Ishwaran, Hemant, and J Sunil Rao. 2005. "Spike and slab variable selection: Frequentist and Bayesian strategies." *Ann. Statist.* 33 (2): 730–773. https://doi.org/10.1214/009053604000001147. https://doi.org/10.1214/009053604000001147.

Jacobs, Bruce I, and Kenneth N Levy. 1988. "Disentangling Equity Return Regularities: New Insights and Investment Opportunities." *Financial Analysts Journal* 44 (3). ISSN: 0015-198X. https://doi.org/10.2469/faj.v44.n3.18.

Kelly, Bryan T, Semyon Malamud, and Lasse H Pedersen. 2020. *Principal Portfolios.* Technical report, Working Paper Series 27388. National Bureau of Economic Research. https://doi.org/10.3386/w27388. http://www.nber.org/papers/w27388.

Kelly, Bryan T, Seth Pruitt, and Yinan Su. 2019. "Characteristics are covariances: A unified model of risk and return." *Journal of Financial Economics* 134 (3): 501–524. ISSN: 0304-405X. https://doi.org/https://doi.org/10.1016/j.jfineco.2019.05.001. http://www.sciencedirect.com/science/article/pii/S0304405X19301151.

Kim, Soohun, and Robert A Korajczyk. 2018. "Large Sample Estimators of the Stochastic Discount Factor." *SSRN Electronic Journal,* ISSN: 1556-5068. https://doi.org/10.2139/ssrn.3131274. https://papers.ssrn.com/abstract=3131274.

Kim, Soohun, Robert A Korajczyk, and Andreas Neuhierl. 2021. "Arbitrage Portfolios." *The Review of Financial Studies* 34 (6): 2813–2856. ISSN: 0893-9454. https://doi.org/10.1093/rfs/hhaa102. https://academic.oup.com/rfs/article/34/6/2813/5902462.

Kirby, Chris. 2019. "Firm Characteristics, Cross-Sectional Regression Estimates, and Asset Pricing Tests." *The Review of Asset Pricing Studies,* ISSN: 2045-9920. https://doi.org/10.1093/rapstu/raz005. https://doi.org/10.1093/rapstu/raz005.

Kozak, SERHIY, Stefan Nagel, and SHRIHARI Santosh. 2018. "Interpreting Factor Models." *The Journal of Finance* 73 (3): 1183–1223. ISSN: 15406261. https://doi.org/10.1111/jofi.12612. https://onlinelibrary.wiley.com/doi/abs/10.1111/jofi.12612%20https://onlinelibrary-wiley-com.ezproxy.library.bi.no/doi/full/10.1111/jofi.12612%20https://onlinelibrary-wiley-com.ezproxy.library.bi.no/doi/abs/10.1111/jofi.12612%20https://onlinelibrary-wiley-com.e.

Lettau, Martin, and Markus Pelger. 2020. "Factors That Fit the Time Series and Cross-Section of Stock Returns." Edited by Stijn Van Nieuwerburgh. *The Review of Financial Studies* 33 (5): 2274–2325. ISSN: 0893-9454. https://doi.org/10.1093/rfs/hhaa020. https://academic.oup.com/rfs/article/33/5/2274/5756219.

Lewellen, Jonathan. 2015. "The Cross-section of Expected Stock Returns." *Critical Finance Review* 4 (1): 1–44. ISSN: 2164-5744. https://doi.org/10.1561/104.00000024. http://dx.doi.org/10.1561/104.00000024.

Lewellen, Jonathan, and Stefan Nagel. 2006. "The conditional CAPM does not explain asset-pricing anomalies." *Journal of Financial Economics* 82 (2): 289–314. ISSN: 0304-405X. https://doi.org/https://doi.org/10.1016/j.jfineco.2005.05.012. http://www.sciencedirect.com/science/article/pii/S0304405X06001371.

Lopez-Lira, Alejandro. 2020. "Risk factors that matter: Textual analysis of risk disclosures for the cross-section of returns." *working paper.*

Mclean, R. David, and Jeffrey Pontiff. 2016. "Does Academic Research Destroy Stock Return Predictability?" *The Journal of Finance* 71 (1): 5–32. https://doi.org/https://doi.org/10.1111/jofi.12365. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jofi.12365. https://onlinelibrary.wiley.com/doi/abs/10.1111/jofi.12365.

Merton, Robert C. 1973. "An Intertemporal Capital Asset Pricing Model." *Econometrica* 41 (5): 867–887. ISSN: 00129682, 14680262. http://www.jstor.org/stable/1913811.

———. 1980. "On estimating the expected return on the market: An exploratory investigation." *Journal of Financial Economics* 8 (4): 323–361. ISSN: 0304-405X. https://doi.org/https://doi.org/10.1016/0304-405X(80)90007-0. http://www.sciencedirect.com/science/article/pii/0304405X80900070.

Nembrini, Stefano, Inke R König, and Marvin N Wright. 2018. "The revival of the Gini importance?" *Bioinformatics* 34 (21): 3711–3718. ISSN: 1367-4803. https://doi.org/10.1093/bioinformatics/bty373. https://doi.org/10.1093/bioinformatics/bty373.

Pelger, Markus. 2020. "Understanding Systematic Risk: A High-Frequency Approach." *The Journal of Finance* 75 (4): 2179–2220. https://doi.org/https://doi.org/10.1111/jofi.12898. https://onlinelibrary.wiley.com/doi/abs/10.1111/jofi.12898.

Pukthuanthong, Kuntara, and Richard W Roll. 2020. "An Agnostic and Practically Useful Estimator of the Stochastic Discount Factor." *SSRN Electronic Journal,* ISSN: 1556-5068. https://doi.org/10.2139/ssrn.3503974. https://papers.ssrn.com/abstract=3503974.

Roll, Richard, and Stephen A Ross. 1980. "An Empirical Investigation of the Arbitrage Pricing Theory." *The Journal of Finance* 35 (5): 1073–1103. ISSN: 00221082, 15406261. http://www.jstor.org/stable/2327087.

Ross, Stephen A. 1976. "The arbitrage theory of capital asset pricing." *Journal of Economic Theory* 13 (3): 341–360. ISSN: 0022-0531. https://doi.org/https://doi.org/10.1016/0022-0531(76)90046-6. http://www.sciencedirect.com/science/article/pii/0022053176900466.

Shukla, Ravi, and Charles Trzcinka. 1990. "Sequential Tests of the Arbitrage Pricing Theory: A Comparison of Principal Components and Maximum Likelihood Factors." *The Journal of Finance* 45 (5): 1541–1564. ISSN: 00221082, 15406261. http://www.jstor.org/stable/2328749.

Stambaugh, Robert F., and Yu Yuan. 2017. "Mispricing factors." *Review of Financial Studies* 30, no. 4 (April): 1270–1315. ISSN: 0893-9454. https://doi.org/10.1093/RFS/HHW107. https://academic.oup.com/rfs/article/30/4/1270/2965095.

Weber, Michael. 2018. "Cash flow duration and the term structure of equity returns." *Journal of Financial Economics* 128 (3): 486–503. ISSN: 0304-405X. https://doi.org/https://doi.org/10.1016/j.jfineco.2018.03.003. http://www.sciencedirect.com/science/article/pii/S0304405X18300667.

5 Tables

Table 1: Descriptive Statistics of the Zero-Cost Latent Factors: 1974–2014

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Mean	0.12	-0.05	0.09	-0.01	-0.06
t-statistic	0.61	-0.42	1.14	-0.14	-0.74
Std. dev	4.21	2.61	1.63	2.00	1.59
Sharpe ratio	0.10	-0.07	0.19	-0.02	-0.12

This table reports the descriptive statistics of the time-series of monthly excess-returns (in percent) for the first five latent common factors. The projection portfolio weights are zero-cost and have a beta of one with its respective factor and zero otherwise. The mean is the monthly arithmetic average of excess returns. The standard deviation is calculated monthly. The Sharpe ratios are annualized by multiplying by the square root of twelve. We omit stocks whose size falls below the 20th percentile of the NYSE following E. F. Fama and K. R. French (2008) and Kirby (2019) to avoid any concerns about liquidity.

Table 2: Correlation Matrix of the Zero-Cost Latent Factors and the Fama–French Five Factors plus Momentum: 1974–2014

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Factor 1	1.00	-0.36	0.11	0.02	0.09
Factor 2	-0.36	1.00	-0.12	-0.14	0.02
Factor 3	0.11	-0.12	1.00	-0.11	-0.11
Factor 4	0.02	-0.14	-0.11	1.00	-0.07
Factor 5	0.09	0.02	-0.11	-0.07	1.00
Mkt-RF	0.66	-0.44	0.02	-0.00	0.05
SMB	0.18	-0.49	0.24	0.15	-0.04
HML	-0.29	0.44	-0.10	-0.03	-0.03
RMW	-0.16	0.35	0.01	-0.06	-0.03
CMA	-0.36	0.39	-0.13	0.03	-0.04
Mom	-0.11	-0.29	-0.03	0.15	-0.03

This table reports the correlation of monthly excess-returns (in percent) for the first five latent common factors with the Fama–French five factors plus momentum. The projection portfolio weights are zero-cost and have a beta of one with its respective factor and zero otherwise. The mean is the monthly arithmetic average of excess returns. The standard deviation is calculated monthly. The Sharpe ratios are annualized by multiplying by the square root of twelve. We omit stocks whose size falls below the 20th percentile of the NYSE following E. F. Fama and K. R. French (2008) and Kirby (2019) to avoid any concerns about liquidity.

Table 3: Descriptive Statistics of the Unit-Cost Latent Factors: 1974–2014

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Mean	0.85	0.67	0.81	0.71	0.67
t-statistic	3.14	3.47	3.82	3.21	3.21
Std. dev	5.62	4.03	4.44	4.61	4.33
Sharpe ratio	0.52	0.58	0.64	0.53	0.53

This table reports the descriptive statistics of the time-series of monthly excess-returns (in percent) for the first five latent common factors. The projection portfolio weights are unit-cost and have a beta of one with its respective factor and zero otherwise. The mean is the monthly arithmetic average of excess returns. The standard deviation is calculated monthly. The Sharpe ratios are annualized by multiplying by the square root of twelve. We omit stocks whose size falls below the 20th percentile of the NYSE following E. F. Fama and K. R. French (2008) and Kirby (2019) to avoid any concerns about liquidity.

Table 4: Correlation Matrix of the Unit-Cost Latent Factors and the Fama–French Five Factors plus Momentum: 1974–2014

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Factor 1	1.00	0.34	0.65	0.61	0.64
Factor 2	0.34	1.00	0.71	0.68	0.74
Factor 3	0.65	0.71	1.00	0.82	0.85
Factor 4	0.61	0.68	0.82	1.00	0.83
Factor 5	0.64	0.74	0.85	0.83	1.00
Mkt-RF	0.95	0.35	0.58	0.55	0.61
SMB	0.45	0.12	0.48	0.45	0.39
HML	-0.28	0.20	-0.12	-0.09	-0.10
RMW	-0.32	-0.05	-0.25	-0.27	-0.27
CMA	-0.38	0.10	-0.18	-0.12	-0.15
Mom	-0.13	-0.25	-0.07	0.00	-0.07

This table reports the correlation of monthly excess-returns (in percent) for the first five latent common factors with the Fama–French five factors plus momentum. The projection portfolio weights are unit-cost and have a beta of one with its respective factor and zero otherwise. The mean is the monthly arithmetic average of excess returns. The standard deviation is calculated monthly. The Sharpe ratios are annualized by multiplying by the square root of twelve. We omit stocks whose size falls below the 20th percentile of the NYSE following E. F. Fama and K. R. French (2008) and Kirby (2019) to avoid any concerns about liquidity.

Table 5: Descriptive Statistics of the Portfolios: 1974–2014

Panel A: 1974–2014

	Market	Long-short	Beta-neutral
Mean	0.52	0.72	0.74
Std. dev	4.65	3.34	1.70
Sharpe ratio	0.39	0.75	1.51

Panel B: 1974–1999

	Market	Long-short	Beta-neutral
Mean	0.65	1.12	0.99
Std. dev	4.64	2.61	1.48
Sharpe ratio	0.48	1.49	2.32

Panel C: 2000–2014

	Market	Long-short	Beta-neutral
Mean	0.28	-0.06	0.25
Std. dev	4.67	4.31	1.97
Sharpe ratio	0.21	-0.04	0.44

This table reports the descriptive statistics of the time-series of monthly excess-returns (in percent) for the market, the long short portfolio sorted on expected returns, and the beta-neutral long-short portfolio. The later two return series are constructed using pure plays as described in the main text. The mean is the monthly arithmetic average of excess returns. The standard deviation is calculated monthly. The Sharpe ratios are annualized by multiplying by the square root of twelve. We omit stocks whose size falls below the 20th percentile of the NYSE following E. F. Fama and K. R. French (2008) and Kirby (2019) to avoid any concerns about liquidity.

Table 6: Fama–MacBeth Predictive Cross-Sectional Regressions

$$r^e_{i,t+1} = a_t + b_t x_{it} + \sum_{j=1}^{k} \lambda_t \beta^j_{it},$$

	Panel A: 1974–2014		Panel B: 1974–1999		Panel C: 2000–2014	
	(1)	(2)	(1)	(2)	(1)	(2)
Intercept	0.01	0.01	0.01	0.01	0.01	0.01
	(3.57)	(4.37)	(2.98)	(3.27)	(1.97)	(2.92)
predicted return	0.26	0.27	0.36	0.36	0.08	0.11
	(7.23)	(11.61)	(8.38)	(12.31)	(1.17)	(2.92)
BetaF1		−0.00		0.00		−0.00
		(−0.16)		(0.52)		(−0.73)
BetaF2		−0.00		−0.00		0.00
		(−0.35)		(−0.51)		(0.09)
BetaF3		0.00		0.00		0.00
		(1.62)		(0.10)		(2.40)
BetaF4		−0.00		−0.00		−0.00
		(−0.73)		(−0.42)		(−0.64)
BetaF5		0.00		−0.00		0.00
		(0.02)		(−0.18)		(0.31)
R2	0.01	0.09	0.01	0.08	0.01	0.11

Notes: This table reports the Fama-MacBeth cross-sectional regressions of monthly stocks' excess returns on the return forecast in the periods 1974–2014, 1974–1999, 2000–2014. "BetaFj" denotes the beta with respect to the j-th latent factor. We omit stocks whose size falls below the 20th percentile of the NYSE following E. F. Fama and K. R. French (2008) and Kirby (2019) to avoid any concerns about liquidity. $x_i t$ denotes the predicted return of stock i at time t. (1) and (2) report the regression results with and without control variables, respectively. The controls are limited to the risk exposures with respect to the latent factors, since the predicted return is constructed with the usual characteristics. We control for ten exposures, but present only the first five. All of the coefficients in the corresponding to the exposures insignificant. We report the time-series average of slope coefficients associated with Fama-MacBeth t-statistics (in parentheses).

Table 7: Weighted Fama–MacBeth Predictive Cross-Sectional Regressions

$$r^e_{i,t+1} = a_t + b_t x_{it} + \sum_{j=1}^{k} \lambda_t \beta^j_{it},$$

	Panel A: 1974–2014		Panel B: 1974–1999		Panel C: 2000–2014	
	(1)	(2)	(1)	(2)	(1)	(2)
Intercept	0.00	0.00	0.00	0.00	0.00	0.01
	(1.54)	(1.47)	(1.65)	(0.68)	(0.47)	(1.41)
predicted return	0.28	0.33	0.30	0.40	0.24	0.22
	(4.73)	(7.92)	(4.93)	(9.51)	(2.06)	(2.62)
BetaF1		0.00		0.00		−0.00
		(0.14)		(0.96)		(−0.67)
BetaF2		−0.00		−0.00		0.00
		(−0.02)		(−0.68)		(0.84)
BetaF3		0.00		−0.00		0.00
		(1.19)		(−0.42)		(1.99)
BetaF4		−0.00		0.00		−0.00
		(−0.25)		(0.07)		(−0.43)
BetaF5		0.00		−0.00		0.00
		(1.66)		(−0.30)		(2.57)
R2	0.01	0.09	0.01	0.08	0.01	0.11

Notes: This table reports the Fama-MacBeth cross-sectional regressions weighted by market capitalization of monthly stocks' excess returns on the return forecast in the periods 1974–2014, 1974–1999, 2000–2014. "BetaFj" denotes the beta with respect to the j-th latent factor. We omit stocks whose size falls below the 20th percentile of the NYSE following E. F. Fama and K. R. French (2008) and Kirby (2019) to avoid any concerns about liquidity. $x_i t$ denotes the predicted return of stock i at time t. (1) and (2) report the regression results with and without control variables, respectively. The controls are limited to the risk exposures with respect to the latent factors, since the predicted return is constructed with the usual characteristics. We control for ten exposures, but present only the first five. All of the coefficients in the corresponding to the exposures insignificant. We report the time-series average of slope coefficients associated with Fama-MacBeth t-statistics (in parentheses).

Table 8: Time Series Regression of the Long-Short and the Hedged Portfolio against the CAPM and Fama–French Five Factors plus Momentum

$$Portfolio_t = \alpha + \sum_{i=1}^{5} \beta_i F_{i,t} + \epsilon_t$$

	Long-Short	Beta-Neutral	Long-Short	Beta-Neutral
Intercept	0.75^{***}	0.76^{***}	0.53^{***}	0.69^{***}
	(4.41)	(8.85)	(2.95)	(8.44)
Mkt-RF	-0.06	-0.05^{**}	-0.04	-0.04
	(-1.20)	(-2.35)	(-0.94)	(-1.54)
SMB			0.10	-0.00
			(1.23)	(-0.04)
HML			0.01	-0.02
			(0.14)	(-0.36)
RMW			-0.19	-0.15^{*}
			(-1.36)	(-1.95)
CMA			-0.00	0.07
			(-0.00)	(0.99)
Mom			0.40^{***}	0.15^{***}
			(5.10)	(5.18)
Adj. R^2	0.01	0.02	0.28	0.19
Num. obs.	493	493	493	493

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Notes: This table reports the regression of monthly stock returns (in percent) of the long-short portfolio and beta-neutral portfolio on the CAPM and the Fama–French five-factor model. The t-statistic are shown in parenthesis. The sample period is 1974 to 2014. We omit stocks whose size falls below the 20th percentile of the NYSE following E. F. Fama and K. R. French (2008) and Kirby (2019) to avoid any concerns about liquidity. Returns of the Fama-French five-factor model (FF5) come from Kenneth's French website. *p<0.1; **p<0.05; ***p<0.01.

Table 9: Time Series Regression of the Long-Short and the Hedged Portfolio against the q5 and the Mispricing Model

$$Portfolio_t = \alpha + \sum_{i=1}^{5} \beta_i F_{i,t} + \epsilon_t$$

	Long-Short	Beta-Neutral	Long-Short	Beta-Neutral
Intercept	0.30	0.59***	0.37*	0.63***
	(1.43)	(6.85)	(1.85)	(6.47)
Mkt-RF	−0.02	−0.03	0.03	−0.01
	(−0.38)	(−0.94)	(0.59)	(−0.42)
R_ME	0.21	0.06		
	(1.79)	(0.87)		
R_IA	−0.03	0.03		
	(−0.20)	(0.49)		
R_ROE	0.13	0.01		
	(0.77)	(0.12)		
R_EG	0.38***	0.16**		
	(2.52)	(2.17)		
SMB			0.15	0.03
			(1.41)	(0.48)
MGMT			0.12	0.07*
			(1.41)	(1.96)
PERF			0.32***	0.09***
			(5.65)	(4.80)
Adj. R^2	0.07	0.04	0.13	0.05
Num. obs.	493	493	493	493

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Notes: This table reports the regression of monthly stock returns (in percent) of the long-short portfolio and beta-neutral portfolio on the q5 factors from Hou et al. (2021) and the m4 factors from Stambaugh and Yuan (2017). The t-statistic are shown in parenthesis. The sample period is 1974 to 2014. We omit stocks whose size falls below the 20th percentile of the NYSE following E. F. Fama and K. R. French (2008) and Kirby (2019) to avoid any concerns about liquidity. Returns of the Fama-French five-factor model (FF5) come from Kenneth's French website. *p<0.1; **p<0.05; ***p<0.01.

Table 10: Time Series Regression of the Value-Weighted Long-Short and the Value-Weighted Hedged Portfolio against the CAPM and Fama–French Five Factors plus Momentum

$$Portfolio_t = \alpha + \sum_{i=1}^{5} \beta_i F_{i,t} + \epsilon_t$$

	Long-Short	Beta-Neutral	Long-Short	Beta-Neutral
Intercept	0.60^{***}	0.55^{***}	0.42^{**}	0.52^{***}
	(3.95)	(7.77)	(2.59)	(7.17)
Mkt-RF	−0.03	−0.03	−0.04	-0.04^{*}
	(−0.56)	(−1.81)	(−1.05)	(−2.02)
SMB			0.08	0.01
			(1.35)	(0.34)
HML			0.07	0.01
			(0.93)	(0.29)
RMW			-0.25^{*}	−0.13
			(−1.98)	(−1.91)
CMA			−0.24	−0.06
			(−1.88)	(−1.19)
Mom			0.48^{***}	0.14^{***}
			(6.14)	(4.92)
Adj. R^2	−0.00	0.01	0.36	0.18
Num. obs.	493	493	493	493

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Notes: This table reports the regression of monthly stock returns (in percent) of the value-weighted long-short portfolio and the value-weighted beta-neutral portfolio on the CAPM and the Fama–French five-factor model. The t-statistic are shown in parenthesis. The sample period is 1974 to 2014. *p<0.1; **p<0.05; ***p<0.01.

Table 11: Time Series Regression of the Value-Weighted Long-Short and the Value-Weighted Hedged Portfolio against the q5 and the Mispricing Model

$$Portfolio_t = \alpha + \sum_{i=1}^{5} \beta_i F_{i,t} + \epsilon_t$$

	Long-Short	Beta-Neutral	Long-Short	Beta-Neutral
Intercept	0.28	0.45^{***}	0.25	0.46^{***}
	(1.33)	(5.84)	(1.49)	(5.82)
Mkt-RF	−0.02	−0.03	0.04	−0.01
	(−0.44)	(−1.30)	(1.02)	(−0.43)
R_ME	0.18	0.05		
	(1.71)	(1.06)		
R_IA	−0.22	−0.07		
	(−0.96)	(−0.92)		
R_ROE	0.13	−0.01		
	(0.75)	(−0.09)		
R_EG	0.35^{*}	0.14^{*}		
	(2.19)	(1.98)		
SMB			0.13	0.03
			(1.29)	(0.64)
MGMT			0.04	0.02
			(0.41)	(0.68)
PERF			0.37^{***}	0.09^{***}
			(5.21)	(3.66)
Adj. R^2	0.05	0.03	0.16	0.05
Num. obs.	493	493	493	493

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Notes: This table reports the regression of monthly stock returns (in percent) of the value-weighted long-short portfolio and the value-weighted beta-neutral portfolio on the factors from Hou et al. (2021) and the m4 factors from Stambaugh and Yuan (2017). The t-statistic are shown in parenthesis. The sample period is 1974 to 2014. We omit stocks whose size falls below the 20th percentile of the NYSE following E. F. Fama and K. R. French (2008) and Kirby (2019) to avoid any concerns about liquidity. Returns of the Fama-French five-factor model (FF5) come from Kenneth's French website. *p<0.1; **p<0.05; ***p<0.01.

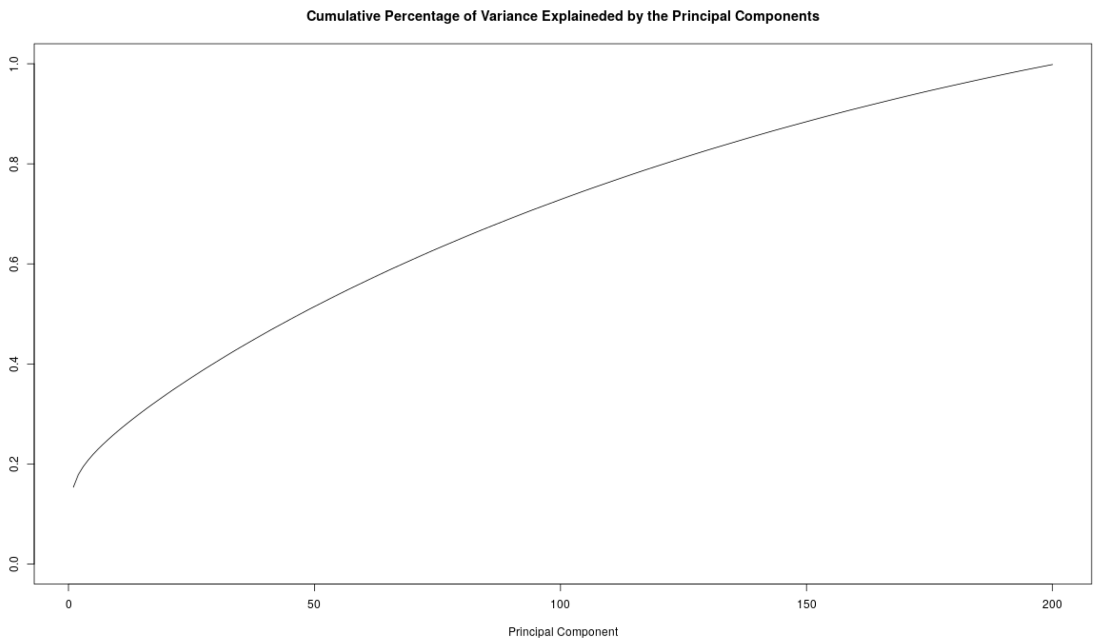
Table 12: Out-of-Sample R^2 from Predicting Latent Realized Betas for Different Models

Beta Number	Naive	12 Months	36 Months	60 Months	120 Months	Expanding
1	0.97	0.97	0.98	0.98	0.98	0.97
2	0.90	0.85	0.90	0.91	0.91	0.91
3	0.66	0.60	0.69	0.71	0.74	0.71
4	0.45	0.42	0.54	0.57	0.61	0.57
5	0.29	0.36	0.45	0.49	0.54	0.48
6	0.17	0.39	0.44	0.46	0.51	0.42
7	0.11	0.34	0.40	0.43	0.48	0.39
8	-0.09	0.27	0.32	0.34	0.39	0.30
9	-0.22	0.23	0.27	0.30	0.35	0.25
10	-0.30	0.18	0.24	0.27	0.32	0.23
Average	0.29	0.46	0.52	0.54	0.58	0.52

Notes: This table reports the out-of-sample R^2 when predicting the monthly realized betas using different models. Beta Number refers to which latent factor the beta corresponds to and Average refers to the arithmetic average of the latent factors' R^2. Naive is the model were the forecast is the last observation (random walk). N Months, where N is 12, 36, 60 or 120, correspond to a model using rolling panel regressions with 12 lags of the form $\beta_{i,p,t} = a_p + \sum_{j=1}^{12} \lambda_{t-j} \beta_{i,p,t-j} + u_{i,p,t}$. and where the regression is run using N (+12) months of data and Expanding corresponds to the same rolling panel regressions using all the available data up until the point where the model is trained. The out-of-sample R^2 is first calculated monthly for each latent beta forecast with the usual formula for a latent factor beta number p at time t: $R^2_{p,t} = 1 - \frac{\sum_i (\beta_{i,p,t} - \beta^f_{i,p,t})}{\beta_{i,p,t} - \bar{\beta}_{p,t}}$, where $\bar{\beta}_{p,t}$ is the cross-sectional average at time t for the latent factor beta p. The sample period is 1974 to 2014.

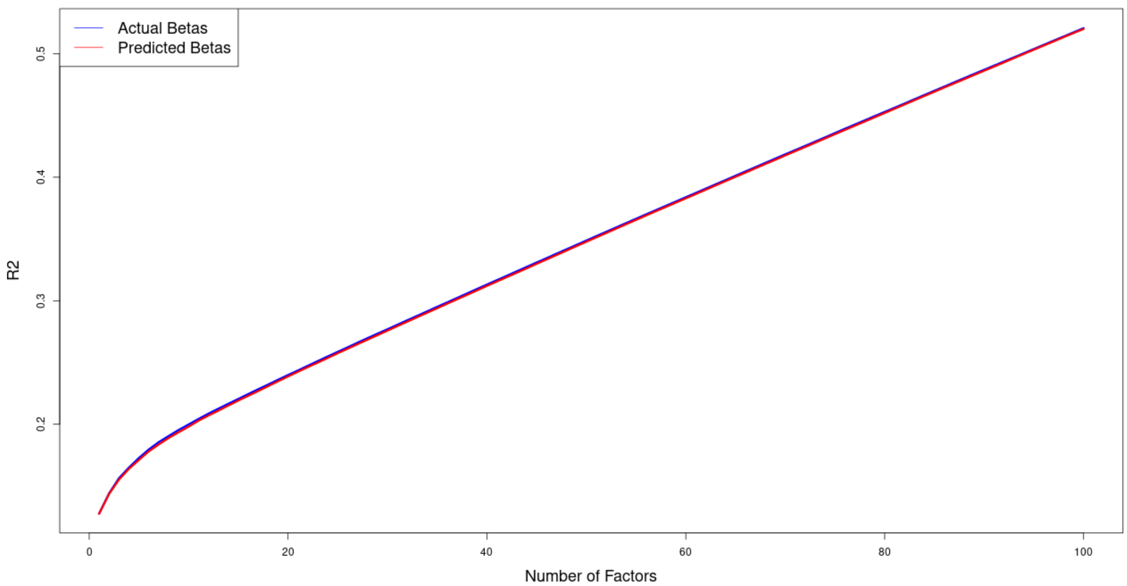
6 Figures

Figure 1:



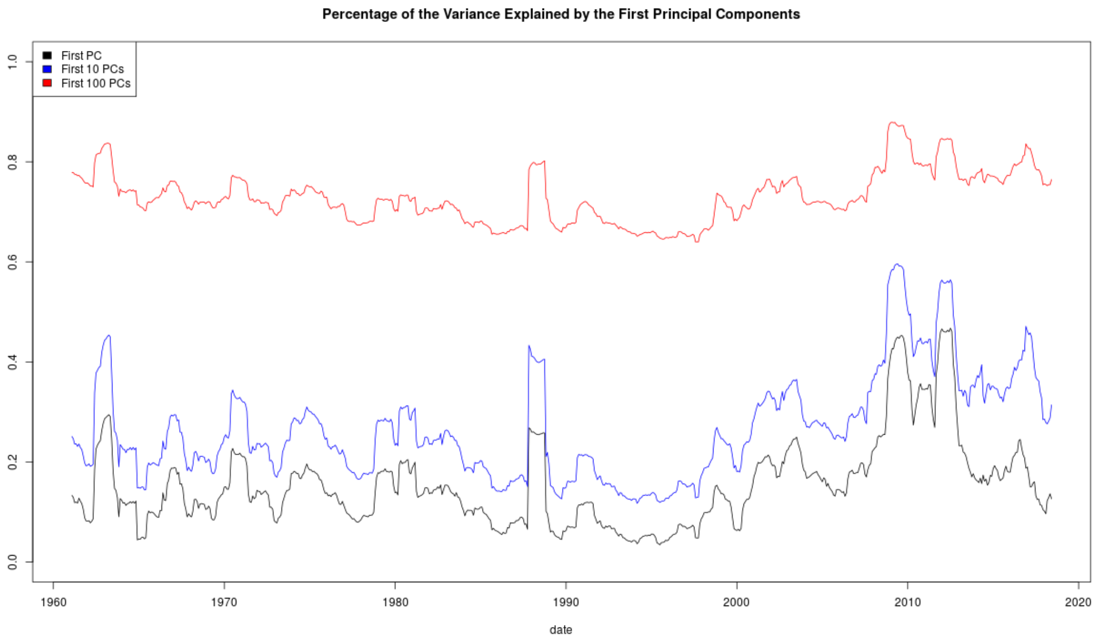
The figure shows the percentage of the variance explained by each of the principal components of the covariance matrix. The figure uses as en example January 1999 but the results are similar for any other period. We omit stocks whose size falls below the 20th percentile of the NYSE to avoid any concerns about liquidity.

Figure 2: Average Time-Series R2 of Stock Returns against Reconstructed Latent Factors



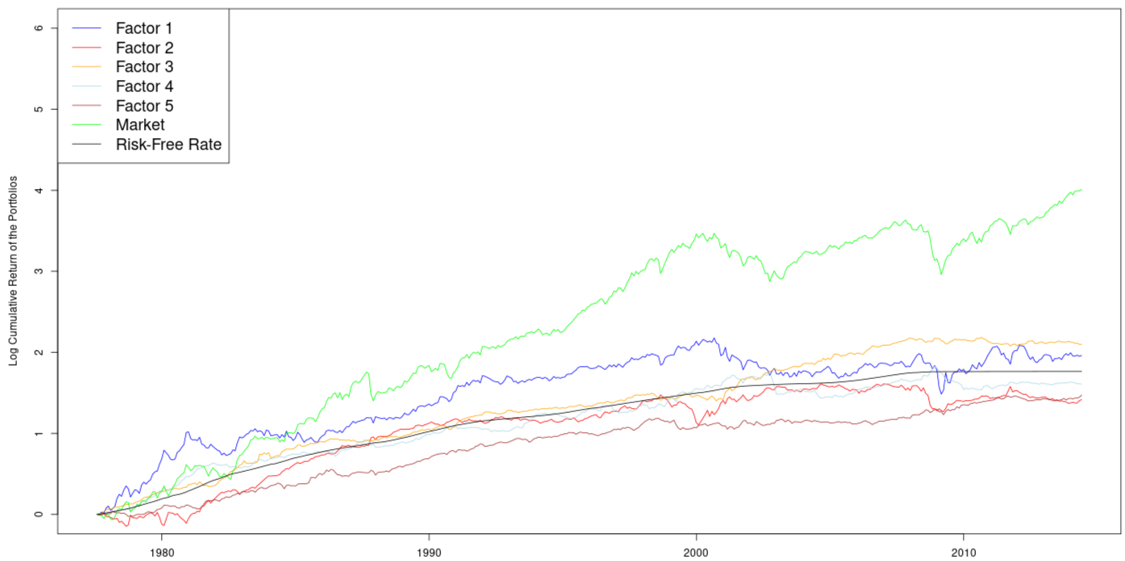
The figure shows the cross-sectional average of the time-series R2 in a regression of individual stocks into the latent factors using either the actual betas or the predicted betas as a function of the number of factors. The regression is performed in yearly rolling windows. The factors are re-estimated every month using previous year's returns. The sample period covers 1974–2014. We omit stocks whose size falls below the 20th percentile of the NYSE to avoid any concerns about liquidity.

Figure 3:



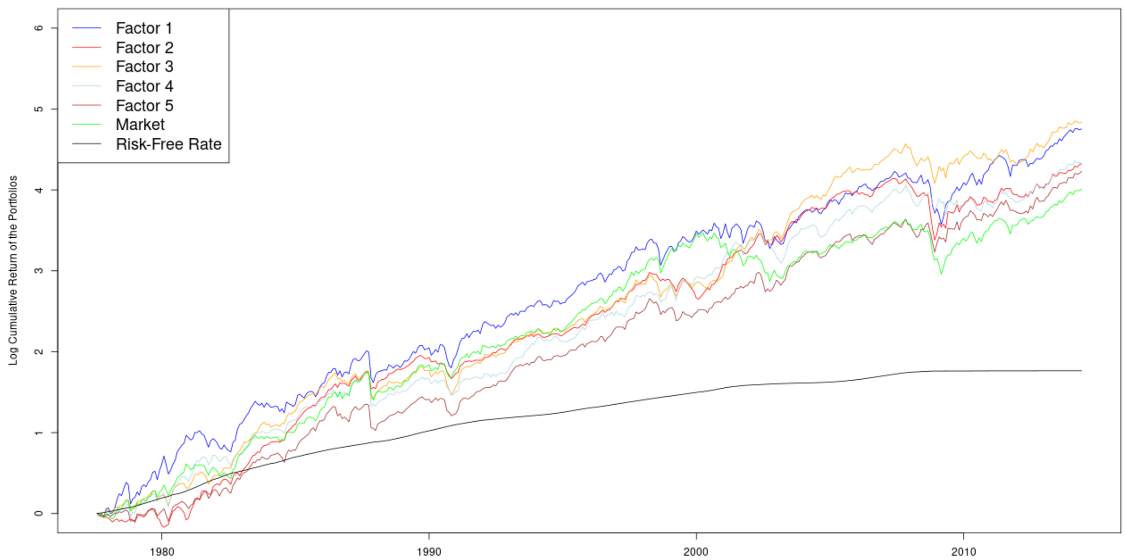
The figure shows the time-varying percentage of the variance explained by each of the principal components of the covariance matrix. The factors are re-estimated every month using previous year's returns. The sample period covers 1974–2014. We omit stocks whose size falls below the 20th percentile of the NYSE to avoid any concerns about liquidity.

Figure 4: Performance of the zero-weight Latent Factors: 1974–2014



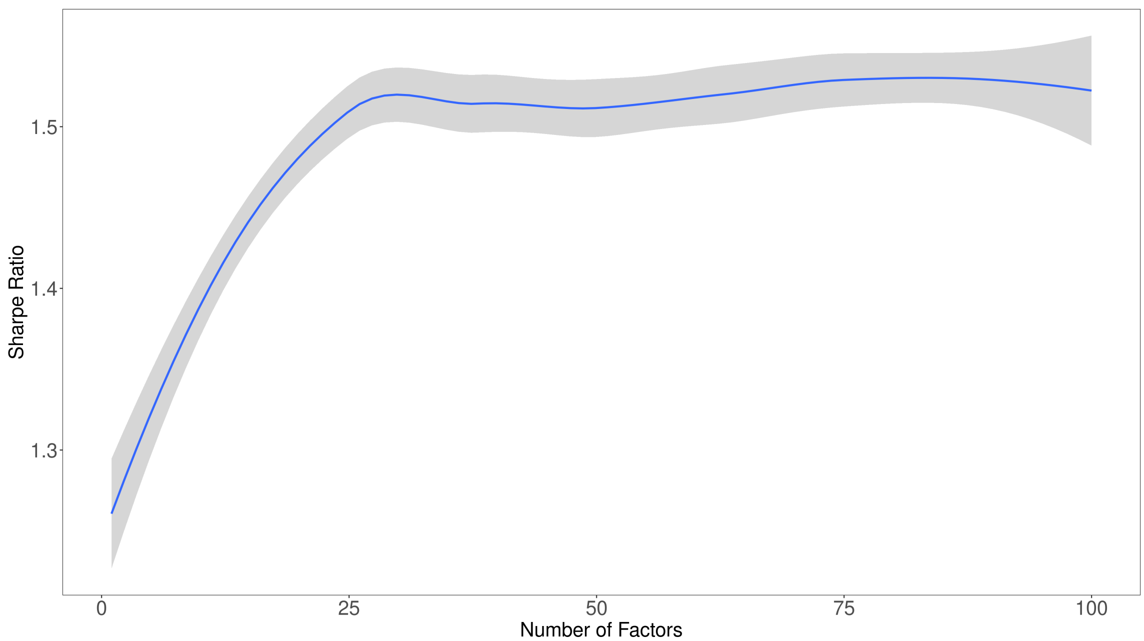
The figure shows the cumulative returns (in logs) of the market portfolio (green line), the risk-free rate (black line), and the projections into the return space of the latent common factors. The projection portfolio weights are zero-weight and have a beta of one with its respective factor and zero otherwise. We add the risk-free rate to the plot of the zero-weight portfolios, for comparability with the market, and assuming that the margin would be invested at the risk-free rate. The factors are projected using pure plays. The sample period covers 1974–2014. Returns of the market portfolio and the risk-free rate come from Kenneth's French website. We omit stocks whose size falls below the 20th percentile of the NYSE following E. F. Fama and K. R. French (2008) and Kirby (2019) to avoid any concerns about liquidity.

Figure 5: Performance of the unit-weight Latent Factors: 1974–2014



The figure shows the cumulative returns (in logs) of the market portfolio (green line), the risk-free rate (black line), and the projections into the return space of the latent common factors. The projection portfolio weights are unit-weight and have a beta of one with its respective factor and zero otherwise. The factors are projected using pure plays. The sample period covers 1974–2014. Returns of the market portfolio and the risk-free rate come from Kenneth's French website. We omit stocks whose size falls below the 20th percentile of the NYSE following E. F. Fama and K. R. French (2008) and Kirby (2019) to avoid any concerns about liquidity.

Figure 6: Sharpe Ratio as a Function of the Number of Factors Hedged

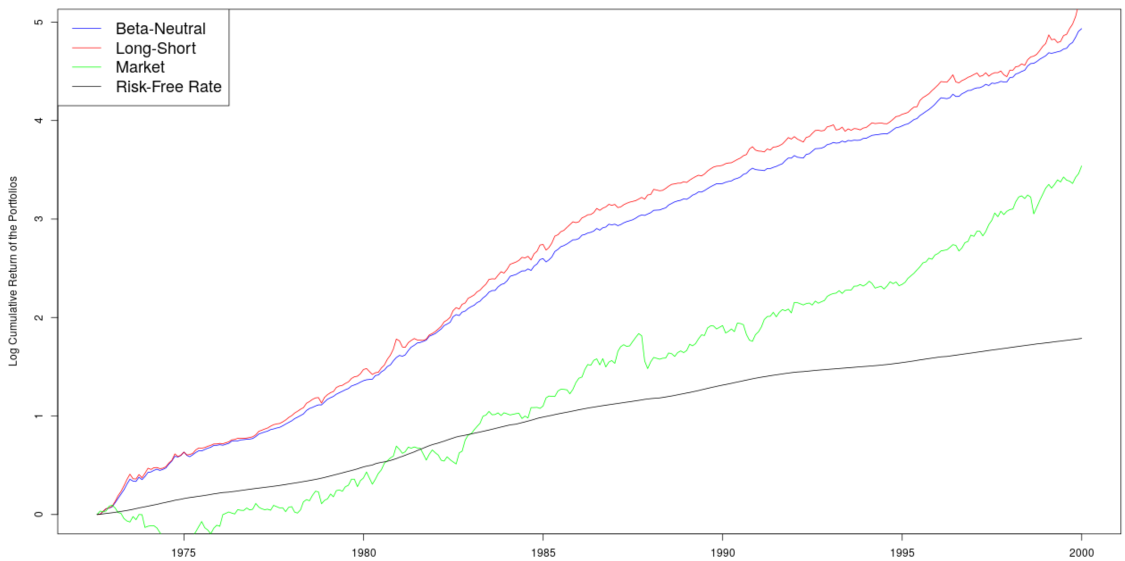


The figure shows the Sharpe ratio as a function of the number of factors for the hedged beta-neutral portfolio constructed using pure plays as described in the main text. The Sharpe ratios are annualized by multiplying by the square root of twelve. We omit stocks whose size falls below the 20th percentile of the NYSE following E. F. Fama and K. R. French (2008) and Kirby (2019) to avoid any concerns about liquidity. The sample period covers 1974–2014.

Figure 7: Cumulative performance of the portfolios: 1974–2014

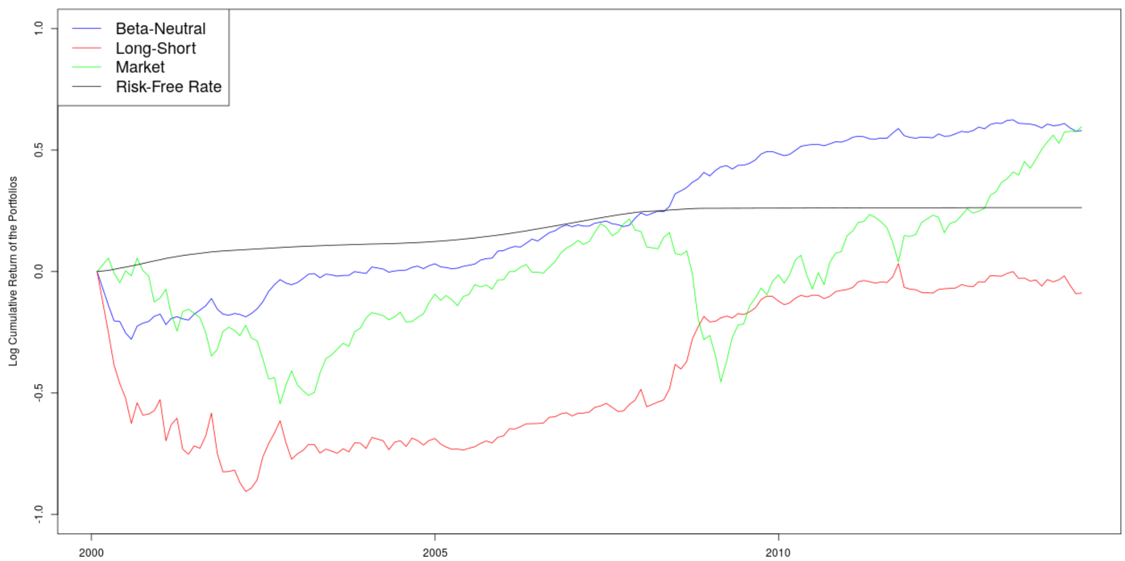
The figure shows the cumulative returns (in logs) of the market portfolio (green line), the risk-free rate (black line), the long-short portfolio sorted on expected returns (red line), and its hedged counterpart, the beta-neutral long-short portfolio (blue line). We add the risk-free rate to the plot of the zero-weight portfolios, for comparability with the market, and assuming that the margin would be invested at the risk-free rate. The later two return series are constructed using pure plays as described in the main text. According to APT, the blue line should equal the gray line. The sample period covers 1974–2014. Returns of the market portfolio and the risk-free rate come from Kenneth's French website. We omit stocks whose size falls below the 20th percentile of the NYSE following E. F. Fama and K. R. French (2008) and Kirby (2019) to avoid any concerns about liquidity.

Figure 8: Performance of the portfolios: 1974–1999



The figure shows the cumulative returns (in logs) of the market portfolio (green line), the risk-free rate (black line), the long-short portfolio sorted on expected returns (red line), and its hedged counterpart, the beta-neutral long-short portfolio (blue line). We add the risk-free rate to the plot of the zero-weight portfolios, for comparability with the market, and assuming that the margin would be invested at the risk-free rate. The later two return series are constructed using pure plays as described in the main text. According to APT, the blue line should equal the gray line. The sample period covers 1974–1999. Returns of the market portfolio and the risk-free rate come from Kenneth's French website. We omit stocks whose size falls below the 20th percentile of the NYSE following E. F. Fama and K. R. French (2008) and Kirby (2019) to avoid any concerns about liquidity.

Figure 9: Performance of the portfolios: 2000–2014



The figure shows the cumulative returns (in logs) of the market portfolio (green line), the risk-free rate (black line), the long-short portfolio sorted on expected returns (red line), and its hedged counterpart, the beta-neutral long-short portfolio (blue line). We add the risk-free rate to the plot of the zero-weight portfolios, for comparability with the market, and assuming that the margin would be invested at the risk-free rate. The later two return series are constructed using pure plays as described in the main text. According to APT, the blue line should equal the gray line. The sample period covers 1974–1999. Returns of the market portfolio and the risk-free rate come from Kenneth's French website. We omit stocks whose size falls below the 20th percentile of the NYSE following E. F. Fama and K. R. French (2008) and Kirby (2019) to avoid any concerns about liquidity.

A Appendix

A.1 Random Forest and Return Forecasts

In this study, we use random forest regressions to forecast future returns. Random forest regression is a non-linear and non-parametric ensemble method that averages multiple forecasts from (potentially) weak predictors and is asymptotically unbiased and able to approximate any function. The ultimate forecast is superior to a forecast following from any individual predictor (Breiman 2001).

We train the random forest model using data from the most recent 60 months and forecast returns in the following period using only the information available at the current time, analogous to a rolling regression forecast. The forecasts are therefore out-of-sample by design. The resulting forecasting regression is:

$$E[r_{i,t+1}|c_{i,t}] \approx f(c_{i,t}) \equiv \mu_{i,t}, \tag{32}$$

where f denotes the random forest model using data from the most recent periods and $c_{i,t}$ denotes the vector of firm characteristics.

For the data, we use the 62 characteristics for predicting returns constructed exactly as in Freyberger, Neuhierl, and Weber (2020). We standardize the characteristics month by month by cross-sectionally demeaning and dividing by the standard deviation. We replace missing values for each characteristics month by month with the median value in that month. We use 500 trees for the random forest regression and a maximum depth of 5. We explain the algorithm itself in detail in this subsection following closely Binsbergen, Han, and Lopez-Lira (2020).

The building blocks for random forest regression are decision trees with a flowchart structure in which the data are recursively split into non-intersecting regions. At each step, the algorithm splits the data choosing the variable and threshold that best minimizes the mean squared error when the average value of the variable to be forecasted is used as the prediction. Decision trees contain two fundamental substructures: *decision nodes* by which the data are split, and *leaves* that represent the outcomes. At the leaves, the forecast is a constant local model equal to the average for that region.

A decision tree model's goal is to partition the data to make optimal constant predictions in each partition (or subspace). Consequently, decision trees are fully non-parametric and allow for arbitrary non-linear interactions. The only parameter for training a decision tree model is the depth, i.e., the maximum length of the path from a root node to leaves. The larger the depth, the more complex the tree, and the more likely it will overfit the data.[12]

More formally, the decision tree model forecast is constant over a disjoint number of regions R_m:

$$\hat{y} = f(x) = \sum_m c_m I_{\{x \in R_m\}}, \tag{33}$$

$$c_m = \frac{1}{N_m} \sum_{\{y_i : x_i \in R_m\}} y_i, \tag{34}$$

and each region is chosen by forming rectangular hyper-regions in the space of

12. The standard approach to decrease the risk of overfitting is to stop the algorithm whenever the next split would result in a sample size smaller than a predetermined size, which is usually five observations for the regression model. This sample threshold is called the *minimum node size*.

the predictors:

$$R_m = \{x_i \in \bigtimes_{i \in I} X_i : x_i \leq k_i^m\}, \tag{35}$$

where $\bigtimes$ denotes a Cartesian product, I is the number of predictors. Thus, each predictor x_i can take values in the set X_i.

The algorithm minimizes the mean squared error numerically to best approximate the conditional expectation by choosing the variables and thresholds, and hence the regions R_m in a greedy fashion. Because of their non-parametric nature and flexibility, decision tree models are prone to overfitting when the depth is large. The most common solution is to use an ensemble of many decision trees with shorter depth: random forest regression models.

Random forest regression models are an ensemble of decision trees that bootstrap the predictions of different decision trees. Each tree is trained on a random sample, usually drawn with replacement. Instead of considering all predictors, decision trees are modified so that they use a strict random subset of features at each node to render the individual decision trees' predictions less correlated.[13] The final prediction of a random forest model is obtained by averaging each decision tree's predictions.

Random forest regressions provide a natural measure of the importance of each variable, the so-called *impurity importance* (Ishwaran 2015). The impurity importance for variable X_i is the sum of all mean squared error decreases of all nodes in the forest at which a split on X_i has been used, normalized by the number of trees. The

13. The algorithm allows a fixed set of variables always to be considered at each split. More generally, the algorithm enables us to specify the probability for each predictor to be considered at each partition.

impurity importance measure can be biased, and we use the correction of Nembrini, König, and Wright 2018 to address this well-known concern. Finally, we normalize the features' importance of each variable as percentages for ease of interpretation.

There are two main parameters in the random forest algorithm: (1) the number of decision trees and (2) the depth of the decision trees. Since the random forest is a bootstrapping procedure, a high number of decision trees are recommended. Notwithstanding computational time, there is no theoretical downside to using more trees. That said, performance tends to plateau following a large number of trees. The depth of each decision tree determines the overall complexity of the model. More complex models usually over-fit. Nevertheless, because of the inherent randomization, random forests are resilient to over-fitting in a wide variety of circumstances.

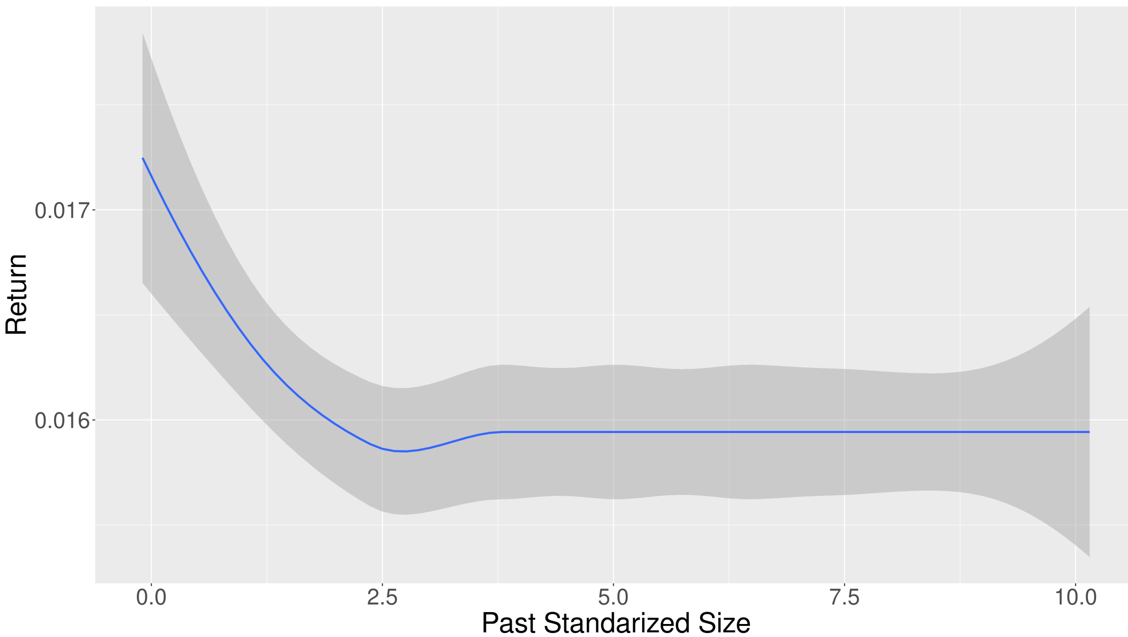
While random forest regressions are non-parametric, we can interpret them using partial dependence plots (PDPs). PDPs explain how features influence the predictions. They display the average marginal effect on the forecast for each value of variable x_s w. PDPs show the model predicts on average when each data instance has a fixed value for that feature. While a disadvantage is that the averages calculated for the partial dependence plot may include very unlikely data points, we include confidence intervals in the figures to address the uncertainty. Formally they are defined as:

$$\hat{f}_{x_S}(x_S) = \frac{1}{n}\sum_{i=1}^{n} \hat{f}(x_S, x_C^{(i)}) \approx E_{x_C}\left[\hat{f}(x_S, x_C)\right] = \int \hat{f}(x_S, x_C) d\mathbb{P}(x_C), \tag{36}$$

where x_S is the variable of interest, and x_C are the other variables. We show an example of a PDPs in Figure 10 .

A.2 Appendix Figures

Figure 10: Predicted Return as a Non-Linear Function of Past Size



Notes: The figure plots the partial dependence plot of one-month-ahead realized returns on size. The partial dependence plot is calculated from a random forest regression of the linear errors on the dependent variables used in Freyberger, Neuhierl, and Weber (2020).