Journal Of Investment Management, Vol. 20, No. 1, (2022), pp. 70–89

www.joim.com

CHARACTERISTIC-BASED RETURNS: ALPHA OR SMART BETA?*

Soohun Kim[a], *Robert A. Korajczyk*[b] *and Andreas Neuhierl*[c]

We propose new methodology to construct arbitrage portfolios by utilizing information contained in firm characteristics for both abnormal returns and betas (and, therefore, smart-beta risk premiums). Our methodology gives maximal weight to risk-based interpretations of characteristics' predictive power before any attribution to abnormal returns. The method allows the explanatory power of a characteristic for both alpha and beta to ebb and flow. This feature is particularly important when we expect that profit opportunities may be arbitraged away by investors. We apply the methodology to a large panel of U.S. stock returns from 1965 to 2018. Empirically, characteristics have time-varying explanatory power for both factor betas and alpha. We find that the arbitrage portfolio has (statistically and economically) significant alpha and annualized Sharpe ratios ranging from 1.31 to 1.66.

Many variables have shown some ability to predict the cross-section of asset returns (Harvey, 2016). This predictive power could be due to (1) their ability to predict the cross-section of systematic risk (beta); (2) their ability to predict asset mispricing (alpha); and (3) spurious cross-sectional relations due to overfitting (data snooping). A substantive debate has evolved regarding whether the evidence suggests that the predictability is due to predicting the cross-section of factor betas or the cross-section of

*This manuscript is based on Kim *et al.* (2021), which was originally published in the Review of Financial Studies by Oxford University Press. We thank Ananth Madhavan, an anonymous referee, and seminar participants at the Vanguard Investment Management Group for their helpful comments.
[a]College of Business, Korea Advanced Institute of Science and Technology, Seoul, 02455, South Korea.
[b]Kellogg School of Management, Northwestern University, IL 60208, USA.
[c]Olin Business School, Washington University in St. Louis, St. Louis, MO 63130, USA.

alpha. Many characteristics are used as the basis to build factor portfolios, e.g., Fama and French (1993, 2015); Hou *et al.* (2015), and multi-factor models seem to have had much success in explaining the cross-section of returns, suggesting that the predictability was from the characteristics' ability to explain beta. That is, the characteristic spread portfolios seem to be smart beta portfolios.

However, Daniel and Titman (1997) argue that it is difficult to disentangle a purely characteristic-based model (in which characteristics only predict alpha) from a risk-based model (in which characteristics only predict beta), because the characteristics and factor loadings in the characteristic-sorted portfolios are collinear. Their influential approach to disentangling the beta vs. alpha explanations uses double-sorts of assets into portfolios based on lagged beta estimates and firm characteristics. Returns on dollar-neutral portfolios made of long and short legs with similar beta exposure but different levels of the characteristics are designed to measure the pure returns to the characteristics (alpha). Similarly, returns on long–short portfolios made of portfolios with similar levels of the characteristic but different levels of beta exposure are designed to measure the pure risk premium. They find significant characteristic-based returns, controlling for betas, but not for beta-based returns, controlling for characteristics. These results suggest that the characteristics are predicting alpha rather than the returns to smart beta portfolios.

Despite its great intuitive appeal, the double-sorting procedure can also lead to wrong conclusions. One issue with the double-sorting procedure arises when the true betas are cross-sectionally related to firm characteristics. Lagged, regression-based estimates of systematic risk are often very noisy, and potentially stale, estimates of the true systematic risk. This may lead to the characteristics predicting returns, holding estimated betas constant, not because the characteristics predict abnormal returns, but because the characteristics are better predictors of true betas (Ferson and Harvey, 1997; Berk, 2000) than are the regression estimates, which are known to be relatively imprecise. Furthermore, the issue of staleness of the estimates is somewhat inescapable because the estimates in academic studies are typically backward-looking functions of unconditional covariances and variances. For example, leverage in a firm's capital structure implies that equity betas are time-varying and that time-series changes in equity betas will be related to changes in the firm's leverage. Since changes in firm size, book-to-market equity ratio, and the firm's past price movements are all correlated with leverage changes, commonly used characteristics (such as market capitalization, book-to-market equity ratios, and momentum) might help predict conditional betas, over and above the predictive power of unconditional betas. In addition to the issue of staleness, double sorting has the disadvantages that the approach handles one characteristic at a time and, hence, is unable to analyze many characteristics simultaneously and that sorting into portfolios may mask important variation in returns relative to using individual assets.

We propose a new methodology, which is an extension of the projected principal components procedure analysis (PPCA) of Fan *et al.* (2016). Unlike the double-sorting procedure, the estimator can accommodate many characteristics simultaneously; can use either individual assets or portfolios; and conditions systematic risk estimates on current values of firm characteristics. Thus, the method addresses the issues raised above. Our procedure gives characteristics, maximal explanatory power for risk premia before we attribute any explanatory power to alphas. We project time-series demeaned asset returns (which eliminates alpha) onto the characteristics. We then estimate the relation between factor betas and characteristics by applying principal component analysis (PCA) to the projected returns. Given the estimated systematic factor loading function, we extract the relation between alpha and the characteristics through a simple constrained ordinary least squares (OLS) regression.

Our procedure brings together the long-standing "fundamental beta" literature (e.g., Rosenberg and McKibben, 1973) and the latent factor model literature. The rolling regression procedure, often used in the academic literature, is not commonly used by practitioners for performance attribution. One obvious reason is that managed portfolios often have portfolio weights and, hence, betas that vary over time. The "fundamental beta" literature (e.g., Rosenberg and McKibben (1973)) models betas as functions of current firm characteristics. This has led to an array of tools available to portfolio managers that accommodate shifting betas due to both changing characteristics and to changing portfolios weights. We model betas and alphas as functions of observable characteristics at recent lags.

In a similar spirit as our work Ang, Madhavan, and Sobczyk (2017) and Madhavan, Sobczyk, and Ang (2021) modify the active/passive performance measures of Lo (2008) and Hsu, Kalesnik, and Myers (2010) to accommodate betas that are functions of firm characteristics. They use a prespecified factor model of traded factors that are exchange traded funds (ETFs) and assume that asset betas are functions of firm characteristics defined by the creators of the ETFs. However, the goals of the two approaches differ. They wish to measure performance of managed portfolios that may be run by managers with both factor timing skills and unspecified security selection skills. We desire to determine if the characteristics allow us to generate alpha that is distinct from beta premiums and, if so, to figure out portfolio weights to implement the strategy.

To illustrate the possible confusion of beta for alpha when using historical estimates of beta and to highlight the advantage of our approach over the double-sorting method, we simulate a simple economy in which the Capital Asset Pricing Model (CAPM) holds. Alphas, or abnormal returns, are identically zero, but the true underlying betas are functions, cross-sectionally, of a single firm characteristic. We perform month-by-month rolling sorts of assets based on OLS estimates of market betas (estimated over the previous 60 months) and the characteristic. We report average returns of double-sorted portfolios in Table 1 (details about the simulation are in the table legend). Although the true return-generating process has no alpha, the return differences of the high-minus-low characteristic portfolios (reported in the last row) are statistically significant, while the return differences of the high-minus-low estimated beta portfolios (reported in the last column) are insignificant. The double-sorting procedure indicates a strong relation between the characteristic and alpha in an economy in which no such relation exists. That is, the double-sorting procedure mischaracterizes beta as alpha. In contrast, when we apply our procedure to this economy, we find that the relation between abnormal returns and the characteristic is insignificantly different from zero. Our procedure is not fooled by the fact that the characteristic is related to the true beta.

When there does exist a relation between alpha and characteristics, one can use our method to construct an arbitrage portfolio (one that loads on alpha but has zero beta exposure) that exploits such a relation. Our arbitrage portfolio weights are proportional to the estimated alpha function, in the spirit of Treynor and Black (1973). We apply the procedure to U.S. stock return data using the characteristics data set of Freyberger *et al.* (2020), updated to December 2018. In the baseline implementation, we use 12 months of data to

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

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

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

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

estimate the weights of the arbitrage portfolio and then hold the portfolio for one month (our results are robust to using alternative-length estimation periods). We then roll the estimation forward by one period and repeat the process. Therefore, we obtain portfolio returns that are out-of-sample relative to the estimation period, in the sense that the arbitrage portfolio weights for period t only use information from periods prior to t. The arbitrage portfolio has (statistically and economically) significant alphas relative to several popular asset pricing models and annualized Sharpe ratios ranging from 1.31 to 1.66 (depending on the number of latent systematic factors we estimate).

Data snooping would suggest that our arbitrage portfolio's alpha should decline over time. The same might be true if sources of alpha are arbitraged away. We say "might be true" because the procedure might be able to switch out of dying sources of alpha into currently active sources of alpha. Thus, the demise of a particular source of alpha does not imply the arbitrage portfolio's alpha declines. We test for a trend in alpha over our sample period. We do find a slight downward trend, but the trend is statistically insignificant and economically inconsequential.

Our approach allows us to make a number of contributions to empirical asset pricing. First, we provide useful guidance in portfolio construction for investors who want to eliminate exposure to the common risks and focus on exploiting the mispricing of traded securities. Second, we address, in a unified manner, the question of "betas vs. characteristics" in a statistical factor pricing model, thereby not assuming a "true" factor model. This question has long been an issue since Fama and French (1993) and Daniel and Titman (1997). The "arbitrage" notion in our arbitrage portfolios is that we are constructing portfolios that hedge out the systematic risk associated with firm characteristics. In the simulated economy above, there are no arbitrage opportunities, and our procedure applied to those data correctly finds no evidence of arbitrage opportunities. Our procedure can separately identify risk (beta) and mispricing (alpha). The empirical results using U.S. stock return data imply that the cross-sectional predictability, found in the literature, is not solely due to smart beta effects but also due to mispricing effects (alpha).

1 The Model

We assume that there exist a large number of securities indexed by $i = 1, \ldots, N$, and the return-generating processes (e.g., alphas and betas) for those individual securities are stable for short blocks of time $t = 1, \ldots, T$. We allow the return-generating process to change across time periods. The return-generating process of each individual security follows a K-factor model in which the factors are unobservable, latent factors. In particular, the excess return of i-th asset at time t is generated by a factor model,

$$R_{i,t} = \alpha_i + \boldsymbol{\beta}_i'\mathbf{f}_t + e_{i,t}, \quad i = 1, \ldots, N \quad \text{and} \quad t = 1, \ldots, T, \tag{1}$$

where $\boldsymbol{\beta}_i = [\beta_{i,1} \cdots \beta_{i,K}]'$ is the $(K \times 1)$ factor loadings of the i-th asset, $\mathbf{f}_t$ is the $(K \times 1)$ systematic factor realization (plus risk premium) in period t, and $e_{i,t}$ is the zero-mean idiosyncratic residual return of asset i at time t. The mispricing of asset i is α_i. We use $\mathbf{0}_m$, $\mathbf{1}_m$, and $\mathbf{0}_{m\times l}$ to denote the $(m \times 1)$ vectors of zeros and ones and the $(m \times l)$ matrix of zeros, respectively. The return-generating process of Equation (1) is expressed compactly in matrix form as:

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

where the (i, t) element of the $(N \times T)$ matrix $\mathbf{R}$ is $R_{i,t}$, respectively, $\boldsymbol{\alpha}$ is the $(N \times 1)$ vector of $[\alpha_1 \cdots \alpha_N]'$, the i-th row of the $(N \times K)$ matrix $\mathbf{B}$ is $\boldsymbol{\beta}_i'$, the t-th row of the $(T \times K)$ matrix $\mathbf{F}$ is

$\mathbf{f}_t' = [f_{1,t} \cdots f_{K,t}]$, and the (i, t) element of the $(N \times T)$ matrix $\mathbf{E}$ is $e_{i,t}$.

Our approach is an extension of the projected principal components analysis (PPCA) procedure of Fan *et al.* (2016). They assume that alphas are zero and allow the factor loading matrix, $\mathbf{B}$, to be a nonparametric function of firm characteristics and estimate the model with the restriction that mispricing is zero. We allow both the mispricing, $\boldsymbol{\alpha}$, and the systematic risk, $\mathbf{B}$, to be functions of asset-specific characteristics. Let $\mathbf{x}_i = [x_{i,1} \cdots x_{i,L}]'$ be the $(L \times 1)$ vector of the characteristics associated with stock i. Define the $(N \times L)$ matrix of $\mathbf{X}$, the i-th row of which is $\mathbf{x}_i'$. We assume the following structure for $\boldsymbol{\alpha}$ and $\mathbf{B}$:

$$\boldsymbol{\alpha} = \mathbf{G}_\alpha(\mathbf{X}) + \Gamma_\alpha$$

$$\mathbf{B} = \mathbf{G}_\beta(\mathbf{X}) + \Gamma_\beta,$$

where the $(N \times 1)$ vector, Γ_α, and the $(N \times K)$ matrix, Γ_β, are cross-sectionally orthogonal to the characteristic space of $\mathbf{X}$. We call $\mathbf{G}_\alpha(\mathbf{X})$ the "mispricing function" and $\mathbf{G}_\beta(\mathbf{X})$ the "factor loading function." Γ_α and Γ_β represent the sources of alpha and beta that are not related to the characteristics, $\mathbf{X}$. While the mispricing function, $\mathbf{G}_\alpha(\mathbf{X})$ and factor loading function, $\mathbf{G}_\beta(\mathbf{X})$, can be consistently estimated in the large N/small T setting used here, consistent estimates of Γ_α and Γ_β are not available. Therefore, our procedure does not attempt to exploit the gammas, just their orthogonality to the characteristics. For ease of exposition, we will work with a linear factor loading function and also a linear mispricing function. There are a number of ways in which one could incorporate non-linearity into the mispricing and factor loading functions. As an example, we can extend $\mathbf{X}$ to be a large set of characteristics, possibly containing suitable polynomials of some underlying characteristics, $\mathbf{X}^*$. While we report results for linear specifications here, Kim *et al.* (2021) find that a parametric nonlinear specification using $\mathbf{X}^*$ leads to larger estimated alphas. Li *et al.* (2021) take a nonparametric approach to incorporating non-linearity. They also find that the alpha function varies through time, but is generally insignificant. We then rewrite the return-generating process (Equation (2)) as follows:

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

First, we can learn about alpha and beta through $\mathbf{G}_\alpha(\mathbf{X})$ and $\mathbf{G}_\beta(\mathbf{X})$ even when data are over short horizon (such as a year) by exploiting the large cross-section of assets. This is a strong advantage over other factor extraction methods requiring large time series or high-frequency observations. Second, our rolling estimation of Equation (3) enables us to study the *temporal* relation of characteristics to risk or mispricing. By estimating Equation (3) over rolling-windows, we can learn about the dynamics of $\mathbf{G}_\alpha(\mathbf{X})$ and $\mathbf{G}_\beta(\mathbf{X})$. For example, a characteristic could have a strong relation with alpha for a while and then dissipate. Lastly, we do not need to necessarily have all important characteristics for risk and mispricing. Since any information in missing characteristics is captured by Γ_α and Γ_β, our model already incorporates the possibility of missing characteristics. If some important characteristics are missing, we may lose some precision but will not generate spurious alpha.

The Arbitrage Pricing Theory (APT; Ross, 1976) implies that the sum of squared pricing errors is finite, so that the average sum of squared pricing errors goes to zero as N increases, $\frac{1}{N}\boldsymbol{\alpha}'\boldsymbol{\alpha} \to 0$. Thus, the APT implies that $\frac{\mathbf{G}_\alpha(\mathbf{X})'\mathbf{G}_\alpha(\mathbf{X})}{N} \to 0$, because $0 \le \frac{\mathbf{G}_\alpha(\mathbf{X})'\mathbf{G}_\alpha(\mathbf{X})}{N} \le \frac{1}{N}\boldsymbol{\alpha}'\boldsymbol{\alpha}$, since $\frac{1}{N}\boldsymbol{\alpha}'\boldsymbol{\alpha}$ also involves $\frac{1}{N}\boldsymbol{\Gamma}_{\boldsymbol{\alpha}}'\boldsymbol{\Gamma}_{\boldsymbol{\alpha}}$. Allowing for significant mispricing of assets implies that the cross-sectional average of the squared mispricing function $\mathbf{G}_\alpha(\mathbf{X})$, $\frac{\mathbf{G}_\alpha(\mathbf{X})'\mathbf{G}_\alpha(\mathbf{X})}{N}$, may be nonzero and converges (as N increases) to some constant, $\frac{\mathbf{G}_\alpha(\mathbf{X})'\mathbf{G}_\alpha(\mathbf{X})}{N} \to \delta \ge 0$. We do not assume that δ is constant across time, but do not use a time index for the sake of simplicity.

This assumption does not imply that characteristics capture all potential mispricing. Mispricing orthogonal to the characteristics is reflected in Γ_α. The main objective of this paper is to provide a method to detect the relation between $\mathbf{X}$ and $\boldsymbol{\alpha}$ while also allowing the characteristics to predict differences in systematic risk across assets. Using the relation between $\mathbf{X}$ and both $\boldsymbol{\alpha}$ and $\mathbf{B}$ allows us to form portfolios that yield alpha, if it is positive-while hedging out the systematic risk associated with the firm characteristics.

We make the natural assumption that alpha and beta are uncorrelated in the cross-section (in essence, any existing correlation is attributed to beta). Also, by definition, the components of alpha $\mathbf{G}_\alpha(\mathbf{X})$ and beta $\mathbf{G}_\beta(\mathbf{X})$ explained by the characteristics are orthogonal to those unexplained by the characteristics, Γ_α and Γ_β. A full set of technical assumptions are available in Kim *et al.* (2021). As mention above, we also will assume that the mispricing function is linear: $\mathbf{G}_\alpha(\mathbf{X}) = \mathbf{X}\theta_\alpha$, where θ_α is an $L \times 1$ vector.

1.1 Methodology[1]

Our procedure first (cross-sectionally) projects demeaned returns onto the firm-specific characteristics. The factor loading function is then estimated by applying a standard principal components procedure to the projected returns. Fan *et al.* (2016) show that the estimated factor loading function converges to the true factor loading function as the cross-sectional sample increases, even for small time-series samples. We extend the PPCA estimator to not only estimate factor loadings, but also the mispricing function, a case not covered in Fan *et al.* (2016). This allows us to implement the procedure using rolling blocks of data to estimate portfolio weights for the next month. It also allows for time variation in factor risk premia and predicted alphas.

To motivate the procedure, consider a one factor model with two characteristics, one driving alpha, $\mathbf{X}_\alpha = \alpha$, and the other driving beta, $\mathbf{X}_\beta = \beta$. Subtracting the time series mean eliminates $\mathbf{X}_\alpha$. The projected demeaned returns, $\widehat{\mathbf{R}}$, obtained from projecting the characteristics onto the demeaned returns converges (in N) to $\mathbf{X}_\beta\mathbf{F}^{*\prime}$ where $\mathbf{F}^*$ is the vector of factor realizations, minus the time series mean factor realization. The first eigenvector of $\frac{\widehat{\mathbf{R}}\widehat{\mathbf{R}}'}{N}$, in this example, gives us $\mathbf{X}_\beta\mathbf{G}_\beta(\mathbf{X})$. Using this estimate of the factor loading function we estimate the alpha by regressing mean return on the characteristics with the constraint that the portfolios beta is zero.

We achieve the goal of constructing an arbitrage portfolio in three steps. In the first step, we subtract the time-series mean from returns and obtain an estimator of $\mathbf{G}_\beta(\mathbf{X})$ by applying Asymptotic Principal Components (APC) to demeaned projected returns (Connor and Korajczyk (1986)). By demeaning the returns, we focus purely on systematic risk not on expected returns or realized premia. In the second step, we estimate the mispricing function, $\mathbf{X}\theta_\alpha$, by regressing (in the cross-section) average returns on the characteristic space orthogonal to the estimated $\mathbf{G}_\beta(\mathbf{X})$ from the first step. Although the average returns contain both mispricing and risk premiums from systematic risks, we extract the information about the mispricing by imposing orthogonality to the systematic risks. In the third step, we use the estimated alpha, $\mathbf{X}\hat{\theta}_\alpha$, to construct asset investment weights for the arbitrage portfolio. We outline our approach verbally here, include more technical material in the Appendix, and refer the reader to Kim *et al.* (2021) for all proofs.

Step 1: The first step of our procedure is the estimation of factor exposures, $\mathbf{G}_\beta(\mathbf{X})$. The observed returns in Equation (3) are driven both by alpha and beta. We eliminate the effect of alpha by time series demeaning the observed returns. We then project the demeaned returns of onto the span of the characteristics. The demeaned and projected returns, $\widehat{\mathbf{R}}$, converge, with large N, to the systematic component of returns (beta times demeaned factor realizations).

Step 2: We estimate factor exposures by applying standard principal component analysis to $\widehat{\mathbf{R}}$. The k-th column of the estimated factor loading function, $\widehat{\mathbf{G}}_\beta(\mathbf{X})$, is $\sqrt{N}$ times the eigenvector of $\frac{\widehat{\mathbf{R}}\widehat{\mathbf{R}}'}{N}$ corresponding to the k-th largest eigenvalue of $\frac{\widehat{\mathbf{R}}\widehat{\mathbf{R}}'}{N}$. From Connor and Korajczyk (1986) we know that the eigenvectors of $\frac{\widehat{\mathbf{R}}\widehat{\mathbf{R}}'}{N}$ yield estimates of the factor betas while the eigenvectors of $\frac{\widehat{\mathbf{R}}'\widehat{\mathbf{R}}}{N}$ yield estimates of the realized latent factors. In Kim *et al.* (2021), we show that $\widehat{\mathbf{G}}_\beta(\mathbf{X})$ converges to the true factor loading function as N increases, for fixed values of T.

Step 3: To estimate the mispricing component θ_α, where $\mathbf{G}_\alpha(\mathbf{X}) = \mathbf{X}\theta_\alpha$, we calculate the time series mean of $\mathbf{R}$ and denote the $(N \times 1)$ vector of average returns, $\overline{\mathbf{R}}$. Our objective is to extract θ_α from $\overline{\mathbf{R}}$. Note that simply projecting $\overline{\mathbf{R}}$ to the linear span of $\mathbf{X}$ does not work because $\overline{\mathbf{R}}$ contains not only $\mathbf{X}\theta_\alpha$ but $\mathbf{G}_\beta(\mathbf{X})\overline{\mathbf{F}}$. That is, projecting $\overline{\mathbf{R}}$ onto $\mathbf{X}$ confounds the cross-sectional predictability of returns due to mispricing with the predictability of returns due to factor risk premia. Hence, we project $\overline{\mathbf{R}}$ to the linear space spanned by $\mathbf{X}$ that is orthogonal to $\widehat{\mathbf{G}}_\beta(\mathbf{X})$. We estimate θ_α from a constrained cross-sectional OLS regression, where the constraint is that the portfolio's factor betas are zero:

$$\widehat{\boldsymbol{\theta}}_\alpha = \arg\min_{\boldsymbol{\theta}} \left(\overline{\mathbf{R}} - \mathbf{X}\boldsymbol{\theta}\right)' \left(\overline{\mathbf{R}} - \mathbf{X}\boldsymbol{\theta}\right)$$
$$\text{subject to } \widehat{\mathbf{G}}_\beta(\mathbf{X})'\mathbf{X}\boldsymbol{\theta} = \mathbf{0}_K. \quad (4)$$

As N increases, the estimated mispricing function converges to the true function, $\hat{\theta}_\alpha \xrightarrow{p} \theta_\alpha$. This regression is a conventional ordinary least square problem with linear equality constraints and the closed form solution is easily obtained.[2]

Step 4: Finally, we construct an arbitrage portfolio that exploits estimated mispricing information in characteristics. Consider the estimated portfolio weights, $\widehat{\mathbf{w}} = \frac{1}{N}\mathbf{X}\widehat{\theta}_\alpha$. Under our assumptions, as N increases, $\widehat{\mathbf{w}}\mathbf{R} \xrightarrow{p} \delta\mathbf{1}'_T$, where $\delta = \lim_{N\to\infty} \frac{\mathbf{G}_\alpha(\mathbf{X})'\mathbf{G}_\alpha(\mathbf{X})}{N}$. Our arbitrage portfolio weights are proportional to the estimated alpha function. This is very much in the spirit of Treynor and Black (1973) in which an asset's weight is proportional to alpha and inversely proportional to idiosyncratic risk. Our large N/small T gives us consistent estimates of alpha (due to large N) but not of idiosyncratic risk (due to small T). Thus, we only exploit the alpha component. It may be possible to improve on our results by incorporating exposures to idiosyncratic risk in the portfolio weights.

The above result is the punchline of this paper: an investor can consistently recover the arbitrage profits, should they exist (that is, should $\delta > 0$), as the number of securities in the cross-section grows large. Since our estimator does not require large T, we can estimate $\mathbf{w}$ over short rolling samples. In Kim *et al.* (2021) we subject our procedure to extensive simulation tests and it performs well. We now turn to the question of whether our approach uncovers any evidence of alpha in the U.S. stock market.

2 Empirical Application

2.1 Data

The asset-level characteristic data are the same as in Freyberger *et al.* (2020), extended to December 2018. The assets studied are U.S. equities from

Table 2 Firm characteristics by category.

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

This is a reproduction of Table 1 in Freyberger *et al.* (2020). It lists the characteristics we consider in our empirical analysis by category. We refer to their online Appendix for a precise definition of these variables and their construction in conventional data set (CRSP, Compustat). The sample period is January 1965 to December 2018.

the Center for Research in Security Prices (CRSP) monthly file. As is common in the literature, we limit the analysis to U.S. firms' common equity trading on NYSE, Amex, or Nasdaq. Accounting data are obtained from Compustat. Table 2 provides an overview of the characteristics used for estimation of the mispricing function and the factor loading function.

To alleviate potential concerns about survivorship bias, which may arise because of backfilling by the data vendor, we require that a firm have at least two years of data in Compustat. Our sample period is from 1965 through 2018. We use 12 months for the estimation period (24- and 36-month estimation periods yield similar results). For the full sample, we have approximately 1.75 million firm/month observations in our analysis. The appendix in Freyberger *et al.* (2020) contains a detailed description of the construction of the characteristic data.

2.2 Estimation

In the baseline model, we assume that the factor loading function and the mispricing function are linear in the characteristics. For example, $\widehat{\theta}_{\alpha,12}$ is estimated with stock returns over $t = 1, \ldots, 12$, and the characteristics observed at time 0, $\mathbf{X}_0$. The portfolio weight for the portfolio held from time 12 to time 13 is calculated as $\widehat{\mathbf{w}}_{12} = \frac{1}{N}\mathbf{X}_{12}\widehat{\theta}_{\alpha,12}$. The return of the arbitrage portfolio is measured in the following month, $t = 13$ is $\widehat{\mathbf{w}}'_{12}\mathbf{R}_{.,13}$, where $\mathbf{R}_{.,13}$ is the column vector of asset returns in month 13. We repeat this process month by month until December 2018. In order to make the results comparable in scale to common equity factors, we scale the portfolio weights so that the in-sample (over the estimation period) standard deviation is 20% per year. Changing this scaling, for example to 10%, will change the estimates of alpha but will not change *t*-statistics or the estimated Sharpe ratio of the strategy.

2.3 Performance of the arbitrage portfolio

Table 3 shows the summary statistics for returns of the arbitrage portfolio assuming different numbers of latent factors (eigenvectors). From Table 3 we see that the returns and Sharpe ratios increase with the number of eigenvectors until about six eigenvectors. Employing more than six eigenvectors does not seem to materially harm the properties of the portfolio, but there also does not seem to be an improvement in any performance metric. Overall, the Sharpe ratios are very high, ranging from 1.31 to 1.66. The increase in Sharpe ratios with increasing number of eigenvectors is driven by increasing means, not decreasing standard deviations, because the standard deviation is always normalized to be 20%, in-sample. The out-of-sample standard deviation is close to, but slightly smaller than, the in-sample standard deviation. The table also displays the maximum drawdown, which ranges between 20.1% and 38.5%. These drawdown numbers are relatively moderate compared to the maximum drawdowns of common factors over the same time period. The four factors in Fama–French–Carhart model have maximum drawdowns of 55.68% (market factor), 55.04% (size factor), 40.92% (value factor) and 57.31% (momentum factor) over our sample period. In addition, skewness, kurtosis, and the best and worst month are reported in Table 3.

If our latent factor approach somehow missed an important systematic factor the arbitrage portfolio's performance could be driven by high exposures to common risk factors rather than mispricing. As a check for this, we run a time-series regression of the arbitrage portfolio's returns onto common risk factors.[3] In Table 4 we report the risk-adjusted returns of the arbitrage portfolio (with six estimated factors) relative to the CAPM (column 1), the Fama and French (1993) three-factor model (column 2), the Fama–French three-factor model augmented with the Carhart

Table 3 Portfolio performance statistics.

# Eigenvectors	Mean (%)	Standard deviation (%)	Sharpe ratio	Skewness	Kurtosis	Maximum drawdown	Worst month (%)	Best month (%)
1	20.08	14.70	1.37	1.24	8.30	22.37	−18.77	29.46
2	24.84	17.51	1.42	0.53	6.07	23.16	−22.44	30.06
3	24.02	14.66	1.64	1.28	10.23	20.91	−19.89	34.05
4	27.54	16.56	1.66	1.07	6.71	22.21	−19.61	30.66
5	28.71	17.94	1.60	1.10	7.11	20.08	−20.08	36.16
6	29.48	18.42	1.60	1.29	8.67	20.84	−19.99	41.88
7	30.13	18.31	1.65	1.34	9.04	21.92	−20.21	42.82
8	29.67	19.84	1.50	1.26	10.38	27.92	−26.02	42.74
9	28.75	17.74	1.62	1.32	8.66	27.05	−20.43	36.56
10	24.93	19.02	1.31	0.29	15.72	38.52	−38.52	41.19

This table reports annualized percentage means, annualized percentage standard deviations, annualized Sharpe ratios, skewness, kurtosis, the maximum draw down, and the best and worst month returns. The arbitrage portfolio with one through ten eigenvectors is estimated every month using the steps outlined in Section 2. The sample period is January 1968 to December 2018.

(1997) momentum factor (column 3), the Fama and French (2015) five-factor model (column 4), the Fama–French five-factor model augmented with the momentum factor (column 5), the Hou *et al.* (2015) four-factor model (column 6) and the HXZ model augmented with the momentum factor (column 7). Neither the arbitrage portfolio's returns nor the benchmark portfolios' returns in Table 4 incorporate the costs of trading as well as the costs of borrowing shares for the short side of the portfolios.

We limit our discussion to the case in which we extract six factors (six eigenvectors). The results for all other cases are contained in the online Appendix to Kim *et al.* (2021). In Table 4 we can see that the alpha (or the intercept in the time-series regression) is fairly consistent across various asset pricing models and exceed 2% per month for all models. Although our arbitrage portfolio has significant exposures to some factors, the adjusted R^2 is fairly low with the minimum of 0.00 and the maximum of 0.23. We illustrate the relation between out-of-sample alpha and the number of eigenvectors used in the estimator in Figure 1. There are two interesting aspects of the figure. One is the cross-model variation in alpha. The other is the hump shape to the alpha versus number of factors plots.

To highlight the dispersion in alphas, we note that the returns on the arbitrage portfolio does not change across the various graphs in Figure 1. What does change is the benchmark risk/return model. Generally, a model with fewer smart-beta factors will attribute to alpha the component of returns that the additional smart beta factors attribute to the risk premiums on the additional factors. That is, adding more factors generally reduces the estimated alpha. This is true as long as the arbitrage portfolio has non-negative beta relative to the additional factor. Figure 1 shows that adding a momentum factor (UMD) always leads to a smaller estimated alpha. Also, the Fama–French three-factor model has smaller estimated alpha than the CAPM since the CAPM attributes to alpha the returns that the FF3 model calls size and value premiums. The one anomaly is the Fama–French five-factor model, which has the highest estimated alphas. This appears to be

Table 4 Risk-adjusted returns with six eigenvectors.

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

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

This table reports alphas (%/month) and factor loadings on the factors by the Fama and French (1993, 2015), Carhart (1997), and the q-factor model (HXZ4) by Hou *et al.* (2015). The arbitrage portfolio with six eigenvectors is estimated every month using the steps outlined in Section 2. Newey and West (1987) standard errors are given in parentheses. The sample period is January 1968 to December 2018.

due to the fact that the arbitrage portfolio's only significant factor loading in FF5 (on the Robust Minus Week (RMW) factor) is negative.

The alpha plots have a hump shape and decrease after approximately seven eigenvectors. Given that we use only 12 monthly observations in the estimation, we attribute the deterioration to the overfitting during the estimation period. For example, with $N = 2{,}000$ assets and $T = 12$ months, we have 12,000 observations per parameter for a one-factor model and 36 observations per parameter with a ten-factor model. Adding more latent factors always improves in-sample performance. However, the arbitrage portfolio returns are out-of-sample returns. Overfitting in the estimation stage leads to worse performance out of sample.

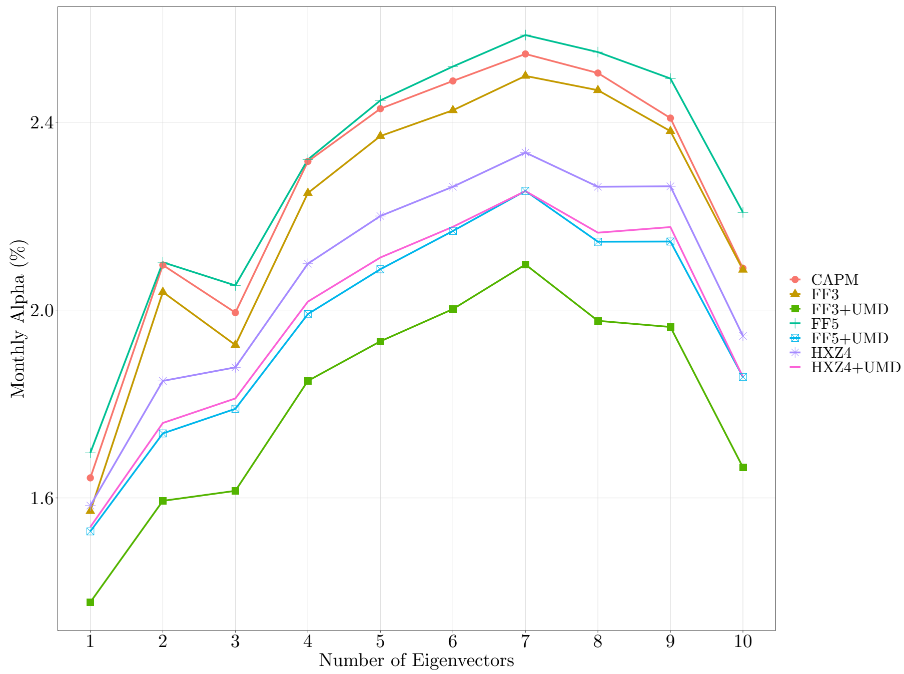

Figure 1 Alpha for varying the number of eigenvectors.

This figure shows the monthly alpha of the arbitrage portfolio against the CAPM, the Fama–French three- and five-factor models, the Hou,Xue, Zhang four-factor model and their "momentum augmented" versions for one through ten eigenvectors. The sample period is from January 1968 to December 2018.

2.4 Properties of the arbitrage portfolio

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

2.4.1 Time-series properties

To develop further intuition about the performance of the arbitrage portfolio, we explore its time-series properties more closely. In Figure 2 we plot the cumulative return. The arbitrage portfolio did not have a negative return (for a full calendar year) during the recent financial crises. Overall, the returns are positive in 49 out of 51 calendar years. Also, the arbitrage portfolio does not have significantly different returns during NBER recessions vs. other periods. For a regression of the portfolio return on a constant and an NBER recession indicator, i.e., $r_t = a + b \times \text{NBER}_t + \varepsilon_t$, we obtain point estimates of $\widehat{a} = 2.36$ (significant at the 1% level) and $\widehat{b} = 0.71$, with a p-value of 0.26. This strongly suggests that the portfolio returns are not systematically related to the business cycle.

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

$$r_t = a + b \times t^{\gamma} + \varepsilon_t. \tag{5}$$

We estimate the model using non-linear least squares. The point estimates are $\widehat{a} = 5.22$, $\widehat{b} = -0.11$, $\widehat{\gamma} = 0.57$. Only the intercept is significant at conventional levels, with a p-value of less than 0.01 (p-values of $\widehat{b}$ and $\widehat{\gamma}$ are 0.75

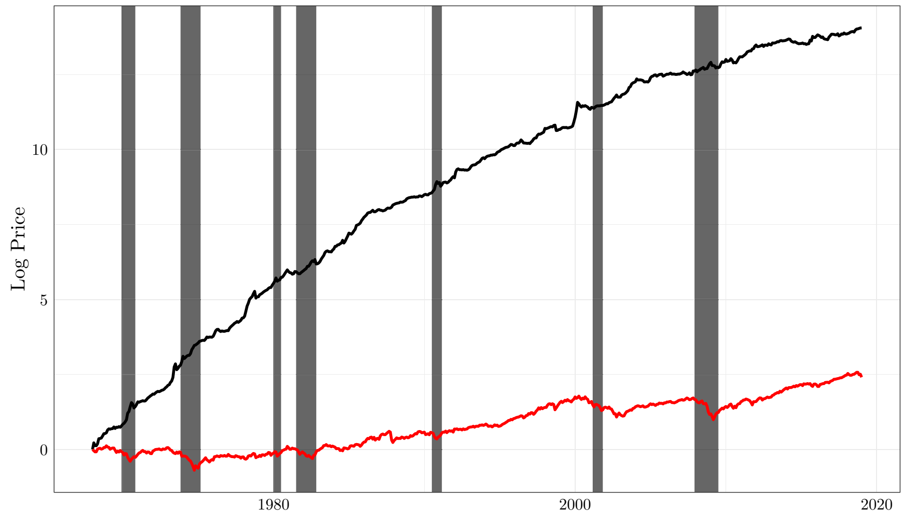

Figure 2 Price path and returns of the arbitrage portfolio.

The top panel of the figure shows the logarithmic price path (i.e., the cumulative returns) of the arbitrage portfolio (using six eigenvectors) in black line and the market portfolio in red line. The areas shaded in gray depict NBER recessions. The lower panel shows the yearly returns of the arbitrage portfolio (with six eigenvectors). The sample period is January 1968 to December 2018.

and 0.18, respectively). A possibly undesirable feature of this specification is that it does not rule out arbitrarily negative returns in the limit. It seems reasonable to restrict the model to only allow returns to be zero in the limit since arbitrage activities and data mining should lead to zero alpha in the long run. One easy way to achieve this is restrict the intercept to be zero and require a positive value for b in this case, we estimate $\widehat{b} = 10.67$ and $\widehat{\gamma} = -0.28$. This specification suggests a mild decay in excess returns and predicts the returns to reach less than 1%

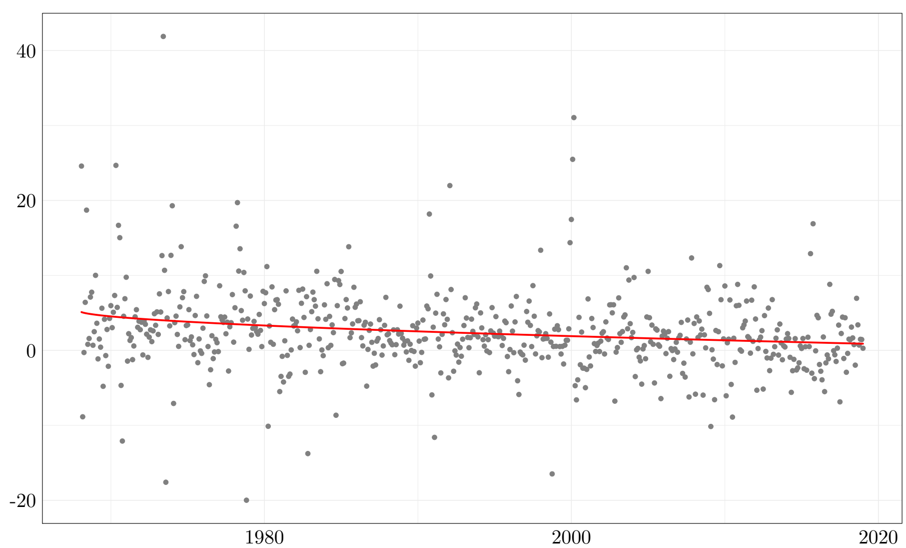

Figure 3 Monthly returns of the arbitrage portfolio 1968–2018.

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

per month in approximately 5,000 months. We plot the monthly returns of the arbitrage portfolio and the trend estimated from this specification in Figure 3. Both trend specifications confirm that the excess returns appear not to diminish significantly over time. This finding is important in the context of the work of McLean and Pontiff (2016) and Linnainmaa and Roberts (2018), who document that many anomalies have become significantly weaker post publication. While it is possible that data snooping will lead to reduced future performance of the arbitrage portfolio, many of the predictive characteristics are the result of research done decades ago. We conclude that the significant average excess returns are at least partially due to mispricing of assets.

2.4.2 *Time variation in characteristic strength for beta and alpha*

The heatmaps for the relations between characteristics and both beta and alpha, in Figures 4 and 5, show how the weights on the different characteristics change from period to period. This

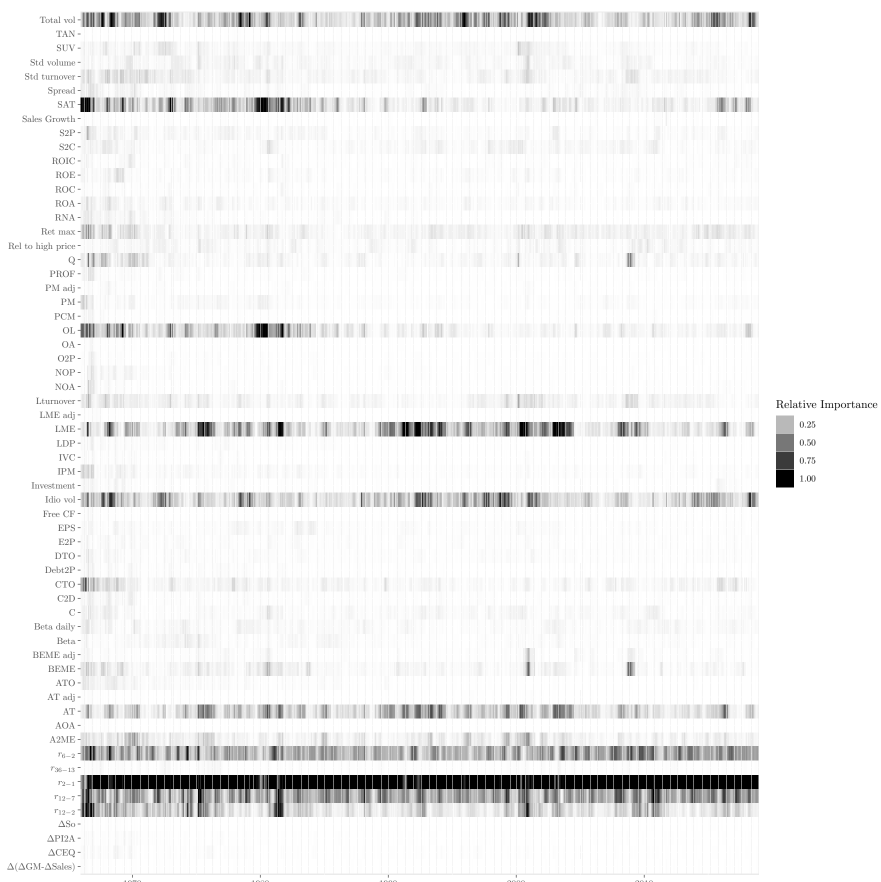

Figure 4 Beta heatmap.

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

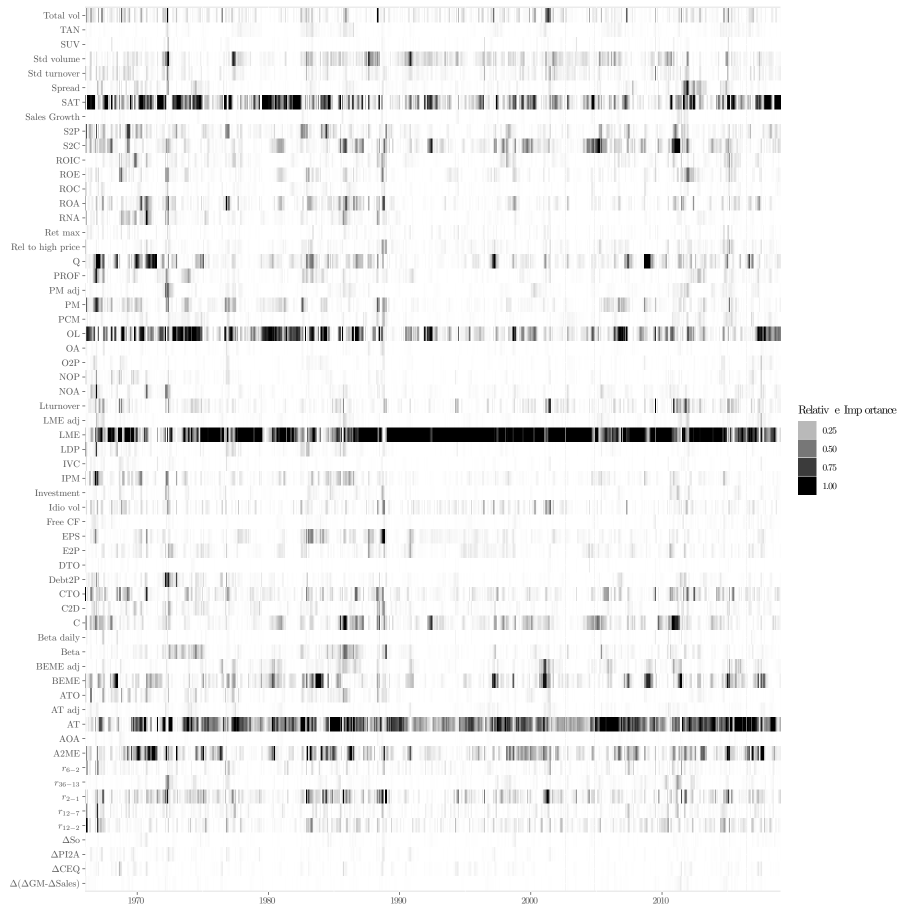

Figure 5 Alpha heatmap.

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

suggests that an advantage of the rolling estimation of θ_{α} is an important component in the performance of the arbitrage portfolio. Characteristics that have the most consistent relation with factor beta include momentum-related characteristics (e.g., r_{2-1}), market capitalization (LME), and total volatility. The first two might reflect a leverage effect in betas as well as the well-known result that small-capitalization firms tend to have higher betas. Characteristics that have a more transitory relation with betas are operating leverage (OL) and idiosyncratic volatility.

Characteristics that have the most consistent relation with alpha include market capitalization (LME), sales to assets (SAT), and total assets (AT). Characteristics that have a more transitory relation with alphas are Tobin's Q (Q) and operating leverage (OL). Interestingly, the "value" characteristic, measured by book-to-market equity (BEME), has very transitory

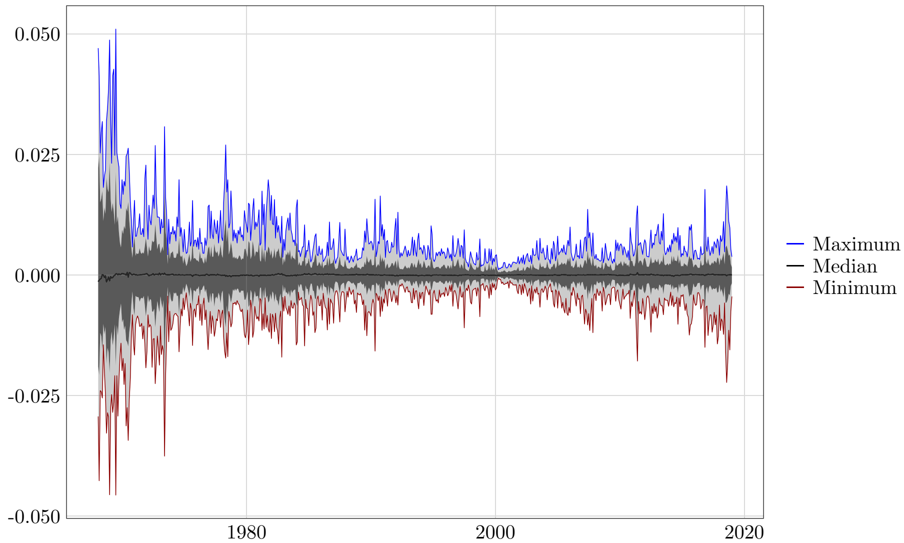

Figure 6 Portfolio weights.

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

explanatory power for alpha, when controlling for the other characteristics.

As can be seen from the heatmaps, a number of these characteristics seem to explain neither alpha nor beta. We have chosen to avoid pre-test biases by utilizing the full set of 61 characteristics. It may be possible to improve on our performance by utilizing variable selection methods to cull useless characteristics, such as those in Freyberger *et al.* (2020) and Li *et al.* (2021).

2.4.3 Portfolio weights

The theory does not impose much in the way of limits or discipline on the portfolio weights of the arbitrage portfolio. In the implementation, we scale the portfolio weights such that the in-sample standard deviation of the arbitrage portfolio is 20% annualized. We de-mean the characteristics so that the resulting portfolio weights of the arbitrage portfolio sum to zero so the portfolio is, by construction, "dollar-neutral." However, we do not impose any constraints on the largest (smallest) position within the portfolio. A potential concern is that the portfolio allocates an unrealistically large amount into individual assets. In Figure 6, we plot the median, minimum, maximum as well as the 5% and 95% quantile of the weights in each month over the sample period from January 1968 to December 2018. The largest weight to a single asset (in absolute value) is approximately 5.1%. In later parts of the sample, when the number of stocks is larger, the weights are considerably smaller, with the largest weights often being less than 1% in absolute value. One could easily change the magnitude of the asset weights by changing the in-sample volatility target used in forming the portfolio weights.

3 Conclusion

We propose new methodology to simultaneously recover returns on "smart-beta" portfolios, estimate assets' conditional factor loadings, estimate conditional alphas using firm-level characteristics, and construct portfolio weights for the arbitrage portfolio. Our methodology has several advantages, relative to the method of double sorting on past estimated betas and lagged characteristics, which is frequently used in the literature. Our approach is more in the spirit of the fundamental beta literature beginning with Rosenberg and McKibben (1973). The methodology only requires a large cross-section and can accommodate a short time span. This allows us to incorporate factor momentum (e.g., Arnott *et al.* (2019); Gupta and Kelly (2019)), or anomaly decay (for example, due to attempts to trade on the anomaly) into a real-time portfolio construction process. Our approach parses the ability of firm characteristics to explain the cross-section of returns into risk and mispricing components.

When applied to U.S. equities over the period from 1968 to 2018, we find that characteristics carry significant information about mispricing despite giving maximal explanatory power to the risk loadings of the statistical factor model. Alphas against popular factor models exceed 2% per month (when we hedge exposure to six latent factors). While some characteristics consistently predict betas, their predictive power for alphas is more transitory. This highlights the importance of allowing the relation between mispricing and characteristics to vary over time.

A separate, unresolved, question in the literature is whether ESG characteristics can predict alpha. Our procedure may be useful in shedding light on this topic.

A. Appendix

Step 1: The first step of our procedure is the estimation of factor exposures, $\mathbf{G}_\beta(\mathbf{X})$. The observed returns in (3) are driven both by alpha and beta. We eliminate the effect of alpha by time series demeaning the observed returns:

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

where $\mathbf{J}_T = \left(\mathbf{I}_T - \frac{1}{T}\mathbf{1}_T\mathbf{1}'_T\right)$.[4] We then project the demeaned returns of Equation (A.1) on the (linear) span of the characteristics by premultiplying by the projection matrix $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Then, we get

$$\begin{aligned}\widehat{\mathbf{R}} \equiv \mathbf{P}\mathbf{R}\mathbf{J}_T &= \mathbf{P}\mathbf{G}_\beta(\mathbf{X})\mathbf{F}'\mathbf{J}_T \\ &\quad + \mathbf{P}\Gamma_\beta\mathbf{F}'\mathbf{J}_T + \mathbf{P}\mathbf{E}\mathbf{J}_T. \quad \text{(A.2)}\end{aligned}$$

Note that $\mathbf{P}\mathbf{G}_\beta(\mathbf{X}) = \mathbf{G}_\beta(\mathbf{X})$, since $\mathbf{G}_\beta(\mathbf{X})$ is already in the linear span of $\mathbf{X}$. The orthogonality of Γ_β and $\mathbf{X}$ make $\mathbf{P}\Gamma_\beta$ and $\mathbf{PE}$ negligible for large N. Hence, $\widehat{\mathbf{R}} = \mathbf{P}\mathbf{R}\mathbf{J}_T \approx \mathbf{G}_\beta(\mathbf{X})\mathbf{F}'\mathbf{J}_T$ with large N, i.e., the projected demeanded returns are only related to risk and not mispricing.

Step 2: We estimate factor exposures by applying standard principal component analysis to $\widehat{\mathbf{R}}$, as in Fan *et al.* (2016). The k-th column of $\widehat{\mathbf{G}}_\beta(\mathbf{X})$ is $\sqrt{N}$ times the eigenvector of $\frac{\widehat{\mathbf{R}}\widehat{\mathbf{R}}'}{N}$ corresponding to the k-th largest eigenvalue of $\frac{\widehat{\mathbf{R}}\widehat{\mathbf{R}}'}{N}$, where $\widehat{\mathbf{R}}$ is given by Equation (A.2). From Connor and Korajczyk (1986) we know that the eigenvectors of $\frac{\widehat{\mathbf{R}}\widehat{\mathbf{R}}'}{N}$ yield estimates of the factor betas, while the eigenvectors of $\frac{\widehat{\mathbf{R}}'\widehat{\mathbf{R}}}{N}$ yield estimates of the realized latent factors. In Kim *et al.* (2021), we show that $\widehat{\mathbf{G}}_\beta(\mathbf{X})$ converges to the true factor loading function as N increases, for fixed values of T.

Step 3: To estimate the mispricing component θ_α, where $\mathbf{G}_\alpha(\mathbf{X}) = \mathbf{X}\theta_\alpha$, we calculate the time series mean of $\mathbf{R}$ and denote the $(N \times 1)$ vector of average returns, $\frac{1}{T}\mathbf{R}\mathbf{1}_T = \overline{\mathbf{R}}$.

$$\overline{\mathbf{R}} = \mathbf{X}\theta_\alpha + \Gamma_\alpha + (\mathbf{G}_\beta(\mathbf{X}) + \Gamma_\beta)\overline{\mathbf{F}} + \overline{\mathbf{E}}. \quad \text{(A.3)}$$

Our objective is to extract θ_α from $\overline{\mathbf{R}}$. Note that simply projecting $\overline{\mathbf{R}}$ to the linear span of $\mathbf{X}$ does not work because $\overline{\mathbf{R}}$ contains not only $\mathbf{X}\theta_\alpha$ but also $\mathbf{G}_\beta(\mathbf{X})\overline{\mathbf{F}}$. That is, projecting $\overline{\mathbf{R}}$ onto $\mathbf{X}$ confounds the cross-sectional predictability of returns due to mispricing with the predictability of returns due to factor risk premia. Hence, we project $\overline{\mathbf{R}}$ to the linear space spanned by $\mathbf{X}$ that is orthogonal to $\widehat{\mathbf{G}}_\beta(\mathbf{X})$. We estimate θ_α from a constrained cross-sectional OLS regression:

$$\widehat{\boldsymbol{\theta}}_\alpha = \arg\min_{\boldsymbol{\theta}} (\overline{\mathbf{R}} - \mathbf{X}\boldsymbol{\theta})'(\overline{\mathbf{R}} - \mathbf{X}\boldsymbol{\theta}) \quad \text{subject to } \widehat{\mathbf{G}}_\beta(\mathbf{X})'\mathbf{X}\boldsymbol{\theta} = \mathbf{0}_K. \quad \text{(A.4)}$$

As N increases, the estimated mispricing function converges to the true function, $\hat{\theta}_\alpha \xrightarrow{p} \theta_\alpha$. This regression is a conventional ordinary least square problem with linear equality constraints and the closed-form solution is easily obtained.[5]

Step 4: Finally, we construct an arbitrage portfolio that exploits estimated mispricing information in characteristics. Consider the estimated portfolio weights, $\widehat{\mathbf{w}} = \frac{1}{N}\mathbf{X}\widehat{\theta}_\alpha$. Under our assumptions, as N increases, $\widehat{\mathbf{w}}\mathbf{R} \xrightarrow{p} \delta\mathbf{1}'_T$, where $\delta = \lim_{N\to\infty} \frac{\mathbf{G}_\alpha(\mathbf{X})'\mathbf{G}_\alpha(\mathbf{X})}{N}$. Our arbitrage portfolio weights are proportional to the estimated alpha function.

Notes

1 The code for our procedure is available at https://github.com/aneuhierl/arbitrage-portfolios (in R and Python).

2 The constrained minimization of (4) yields the closed form solution of $\widehat{\boldsymbol{\theta}}_\alpha = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\overline{\mathbf{R}} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\widehat{\mathbf{G}}_\beta(\mathbf{X})\left(\widehat{\mathbf{G}}_\beta(\mathbf{X})'\widehat{\mathbf{G}}_\beta(\mathbf{X})\right)^{-1}\widehat{\mathbf{G}}_\beta(\mathbf{X})'\overline{\mathbf{R}}$.

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

4 Note that $\mathbf{R}\mathbf{J}_T$ demeans the return in a time-series dimension because $\mathbf{R}\mathbf{J}_T = \mathbf{R} - \left(\frac{1}{T}\mathbf{R}\mathbf{1}_T\right)\mathbf{1}'_T$.

5 The constrained minimization of (A.4) yields the closed form solution of $\widehat{\boldsymbol{\theta}}_\alpha = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\overline{\mathbf{R}} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\widehat{\mathbf{G}}_\beta(\mathbf{X})(\widehat{\mathbf{G}}_\beta(\mathbf{X})'\widehat{\mathbf{G}}_\beta(\mathbf{X}))^{-1}\widehat{\mathbf{G}}_\beta(\mathbf{X})'\overline{\mathbf{R}}$.

References

Ang, A., Madhavan, A., and Sobczyk, A. (2017). "Estimating time-varying factor exposures," *Financial Analysts Journal* **73**, 41–54.

Arnott, R. D., Clements, M., Kalesnik, V., and Linnainmaa, J. T. (2019). "Factor Momentum," Available at SSRN 3116974.

Berk, J. B. (2000). "Sorting Out Sorts," *Journal of Finance* **55**, 407–427.

Carhart, M. M. (1997). "On Persistence in Mutual Fund Performance," *Journal of Finance* **52**, 57–82.

Connor, G. and Korajczyk, R. A. (1986). "Performance Measurement with the Arbitrage Pricing Theory: A New Framework for Analysis," *Journal of Financial Economics* **15**, 373–394.

Daniel, K. and Titman, S. (1997). "Evidence on the Characteristics of Cross Sectional Variation in Stock Returns," *Journal of Finance* **52**, 1–33.

Fama, E. F. and French, K. R. (1993). "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics* **33**, 3–56.

Fama, E. F. and French, K. R. (2015). "A Five-Factor Asset Pricing Model," *Journal of Financial Economics* **116**, 1–22.

Fan, J., Liao, Y., and Wang, W. (2016). "Projected Principal Component Analysis in Factor Models," *Annals of Statistics* **44**, 219–254.

Ferson, W. E. and Harvey, C. R. (1997). "Fundamental Determinants of National Equity Market Returns: A Perspective on Conditional Asset Pricing," *Journal of Banking & Finance* **21**, 1625–1665.

Freyberger, J., Neuhierl, A., and Weber, M. (2020). "Dissecting Characteristics Nonparametrically," *Review of Financial Studies* **33**, 2326–2377.

Gupta, T. and Kelly, B. (2019). "Factor Momentum Everywhere," *Journal of Portfolio Management* **45**, 13–36.

Harvey, C. R., Liu, Y., and Zhu, H. (2016). "... and the Cross-Section of Expected Returns," *Review of Financial Studies* **29**, 5–68.

Hou, K., Xue, C., and Zhang, L. (2015). "Digesting Anomalies: An Investment Approach," *Review of Financial Studies* **28**, 660–705.

Hsu, J. C., Kalesnik, V., and Myers, B. W. (2010). "Performance Attribution: Measuring Dynamic Allocation Skill," *Financial Analysts Journal* **66**, 17–26.

Kim, S., Korajczyk, R. A., and Neuhierl, A. (2021). "Arbitrage Portfolios," *Review of Financial Studies* **34**, 2813–2856.

Li, S., Linton, O. B., and Ge, S. (2021). "Dynamic Peer Groups of Arbitrage Characteristics," Available at SSRN 3638105.

Linnainmaa, J. T. and Roberts, M. R. (2018). "The History of the Cross-Section of Stock Returns," *Review of Financial Studies* **31**, 2606–2649.

Lo, A. W. (2008). "Where do Alphas Come From?: A Measure of the Value of Active Investment Management," *Journal of Investment Management* **6**, 1–29.

Madhavan, A., Sobczyk, A., and Ang, A. (2021). 11Toward esg alpha: Analyzing esg exposures through a factor lens," *Financial Analysts Journal* **77**, 69–88.

McLean, D. R. and Pontiff, J. (2016). "Does Academic Research Destroy Return Predictability," *Journal of Finance* **71**, 5–32.

Newey, W. K. and West, K. D. (1987). "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation consistent Covariance Matrix," *Econometrica* **55**, 703–708.

Rosenberg, B. and McKibben, W. (1973). "The Prediction of Systematic and Specific Risk in Common Stocks," *Journal of Financial and Quantitative Analysis* **8**, 317–333.

Ross, S. A. (1976). "The Arbitrage Theory of Capital Asset Pricing," *Journal of Economic Theory* **13**, 341–60.

Treynor, J. L. and Black, F. (1973). "How to Use Security Analysis to Improve Portfolio Selection," *The Journal of Business* **46**, 66–86.

Keywords: Arbitrage; characteristic-based model; smart beta; factor pricing

JEL Classification Codes: G12, C58, C38