Aggregation of Information About the Cross Section of Stock Returns: A Latent Variable Approach

Nathaniel Light
American University of Sharjah

Denys Maslov
Moody's Analytics

Oleg Rytchkov
Temple University

We propose a new approach for estimating expected returns on individual stocks from a large number of firm characteristics. We treat expected returns as latent variables and apply the partial least squares (PLS) estimator that filters them out from the characteristics under an assumption that the characteristics are linked to expected returns through one or few common latent factors. The estimates of expected returns constructed by our approach from 26 firm characteristics generate a wide cross-sectional dispersion of realized returns and outperform estimates obtained by alternative techniques. Our results also provide evidence of commonality in asset pricing anomalies. (*JEL* G12, C58)

Received June 26, 2014; editorial decision May 15, 2016 by Editor Leonid Kogan.

Any rational asset pricing theory implies that expected stock returns admit a beta representation, where betas are computed with respect to a discount factor (Cochrane 2005). However, the nature of the discount factor remains elusive despite several decades of active academic research: many initially promising models such as the CAPM, CCAPM, and Fama-French 3-factor model cannot

---

We thank Leonid Kogan (the editor) and two anonymous referees for numerous valuable suggestions. We also thank Naresh Bansal, Sudipta Basu, Nina Boyarchenko, Tarun Chordia, Serdar Dinc, Andres Donangelo, Bruno Feunou, John Griffin, Travis Johnson, Revansiddha Khanapure, Soohun Kim, Shimon Kogan, Cheng-Few Lee, Jonathan Lewellen, Igor Makarov, Anna Obizhaeva, Pavel Savor, Elizaveta Shevyakhova, Clemens Sialm, Stathis Tompaidis, as well as seminar participants at Fordham University, Rutgers University, and Temple University and participants of the 2013 Midwest Finance Association Meeting, 2013 Financial Management Association Meeting, 2014 Financial Intermediation Research Society Conference, 2014 Northern Finance Association Meeting, 9th Annual Mid-Atlantic Research Conference in Finance, and 2016 UAE Quantitative Research Symposium for helpful comments. All remaining errors are our own. Supplementary data can be found on *The Review of Financial Studies* web site. Send correspondence to Oleg Rytchkov, Fox School of Business, Temple University, 1801 Liacouras Walk, 423 Alter Hall (006-01), Philadelphia, PA 19122; telephone: (215) 204-4146. E-mail: rytchkov@temple.edu.

doi:10.1093/rfs/hhw102 Advance Access publication February 6, 2017

fully explain the cross section of stock returns. Instead, there is a large and still growing body of evidence that expected returns tend to line up with various firm characteristics.[1] Such patterns, known as asset pricing anomalies, have received much attention and now seem ubiquitous. Subrahmanyam (2010) claims that more than 50 variables are correlated with future stock returns in the cross section. Harvey, Liu, and Zhu (2016) catalogue 314 different factors, although some of them are highly correlated. Green, Hand, and Zhang (2013) conduct an extensive search for "return predictive signals" in the accounting and finance literature and find 330 of them that have been reported.

A large number of variables related to expected returns poses an important question: how to aggregate the information contained in various firm characteristics and construct the most precise estimates of expected returns on individual stocks? Such estimates would improve optimal portfolios (e.g., Treynor and Black 1973), provide a characteristic-based cost of capital as well as a benchmark for portfolio performance evaluation (e.g., Chan, Dimmock, and Lakonishok 2009), allow researchers to form basis assets that would increase the power of asset pricing tests (e.g., Haugen and Baker 1996), facilitate the study of the properties of expected returns in the cross section (e.g., Lewellen 2015), and help to identify possible mispricing (e.g., Stambaugh, Yu, and Yuan 2015; Cao and Han 2016).

In this paper, we propose a novel approach for aggregating information about future stock returns from a large number of firm characteristics. The approach is based on two main premises. First, we explicitly acknowledge that expected returns on individual stocks are unobservable to an econometrician and consider them as latent variables. Second, we assume that unobservable expected returns are related to observable firm characteristics through one or few latent factors (in the case of one latent factor, the factor itself can be identified with expected returns). The assumption is quite general and may hold in a variety of settings. For example, it holds when expected returns are described by a rational asset pricing model, in which case the latent factors coincide with unobservable betas. This interpretation of the latent factors is consistent with a large amount of literature that establishes explicit theoretical links between various firm characteristics and betas and shows that many characteristics can predict future returns even when expected returns are described by a rational asset pricing model.[2] Alternatively, the latent factors can be associated with frictions such as illiquidity costs, which are related to expected returns as in Acharya and

[1] Subrahmanyam (2010) and Goyal (2012) review the literature on cross-sectional predictors of stock returns.

[2] In particular, expected returns have been related to size (e.g., Berk 1995; Berk, Green, and Naik 1999; Gomes, Kogan, and Zhang 2003; Carlson, Fisher, and Giammarino 2004), book-to-market ratio (e.g., Berk, Green, and Naik 1999; Gomes, Kogan, and Zhang 2003; Carlson, Fisher, and Giammarino 2004; Zhang 2005), price-earnings ratio (e.g., Kogan and Papanikolaou 2013), leverage (e.g, Garlappi and Yan 2011), investments (e.g., Cochrane 1996; Li, Livdan, and Zhang 2009; Kogan and Papanikolaou 2013), equity issuance (e.g., Li, Livdan, and Zhang 2009), accruals (e.g., Wu, Zhang, and Zhang 2010), past returns (e.g., Berk, Green, and Naik 1999; Johnson 2002; Liu and Zhang 2014), financial distress (e.g., Garlappi and Yan 2011), and idiosyncratic volatility (e.g., Kogan and Papanikolaou 2013; Babenko, Boguth, and Tserlukevich 2016).

Pedersen (2005), or margin requirements, which are related to expected returns as in Gârleanu and Pedersen (2011). Our framework is also valid when the latent factors describe sentiment reflected in multiple firm characteristics.[3]

Under our assumptions, the problem of an econometrician is to construct the best estimate of latent expected returns on individual stocks using firm characteristics as signals. The key insight of our paper is that this problem can be naturally solved within the partial least squares (PLS) framework. PLS is a statistical technique that extracts a common factor from a set of predictors such that it has the highest covariance with the forecasted variable while not necessarily being the main source of common variation in the predictors.[4] In the case with only one returns-related factor in the characteristic space, the PLS-based estimation of expected returns can be implemented as a sequence of two OLS regressions. In the first step, realized returns at time t are regressed on each lagged standardized characteristic individually (these are cross-sectional regressions that use all stocks for which the characteristic is available). The obtained slopes, whose number is equal to the number of characteristics, are used in the second step: for each individual stock, all firm characteristics available at time t are regressed on the slopes obtained in the first step. We prove that the slopes from the second-step regressions are proportional to cross-sectionally demeaned expected returns on individual stocks at time t when the number of stocks and characteristics is sufficiently large. Note that the PLS procedure is purely cross-sectional as it uses data from only two time periods. However, if the relations between characteristics and expected returns are stable over time, the precision of the PLS estimates can be increased by using averages of the first-step regression slopes obtained in the current and previous periods instead of their most recent values in the second-step regressions. We average the slopes over all previous periods in the baseline specification and consider alternative averaging schemes as a robustness check.

Although the assumption that all characteristics are proxies for the same latent variable (expected returns) is relatively general, it still can be unduly restrictive. For example, it excludes an important situation in which a multifactor asset pricing model holds and firm characteristics are proxies for various linear combinations of multiple betas. To accommodate such cases, we consider an extension of the model with several unobservable returns-related factors in the characteristics. Note that as in the 1-factor case, the validity of a multifactor asset pricing model is only one of many possible reasons why the information about expected returns contained in the characteristics can be summarized by a few latent factors. Because in the presence of multiple factors, not only the factor values but also their relative weights should be estimated to find expected returns, the multifactor PLS-based procedure is more complicated

[3] Kozak, Nagel, and Santosh (2015) and Stambaugh and Yuan (2016) provide arguments that even in a sentiment-driven economy the cross section of expected stock returns is described by a low-dimensional factor model.

[4] Vinzi et al. (2010) review partial least squares methods and their applications.

than its 1-factor analog and contains 3 steps that closely follow those of the 3-pass regression filter (3PRF) developed in Kelly and Pruitt (2015).

We apply the proposed estimation procedure to 26 firm characteristics that are known to be related to stock returns and construct a new variable *AFER* (the acronym stands for aggregate filtered expected returns) that aggregates information from all characteristics and estimates expected returns on individual stocks. To measure the ability of *AFER* to predict returns, we sort stocks according to *AFER*, form decile portfolios, compute future realized equal- and value-weighted returns on the portfolios, and find the average difference in returns on the top and bottom portfolios (hedge returns). Our results show that *AFER* is indeed aligned with expected returns in the cross section: the returns on the *AFER* portfolios monotonically increase with the portfolio number, and in the baseline specification the monthly hedge returns are 2.06% and 1.56% with *t*-statistics of 9.31 and 5.45 for equal- and value-weighted returns, respectively.[5] Notably, the hedge returns on *AFER* exceed those produced by individual characteristics.

This result has two implications. First, it is already remarkable that *AFER* has at least some predictive power for future returns. Because our procedure estimates a common returns-related factor in the characteristics, the dispersion of returns on the *AFER* portfolios implies that such a factor exists and there is commonality in the information about returns contained in various characteristics.[6] Second, because the dispersion of returns produced by *AFER* is larger than that produced by individual characteristics, we conclude that the characteristics contain different information about future returns. In particular, each characteristic can be viewed as a noisy signal about expected returns, and PLS successfully aggregates those signals and filters out the noise contained in them.

To illustrate the advantages of the PLS-based aggregation, we compare *AFER* with the estimates of expected returns obtained by several alternative techniques such as the factor analysis, principal component analysis (PCA), Fama-MacBeth regression, and the averaging of stock ranks with respect to individual characteristics. Each of these approaches has its own shortcomings compared with PLS (we discuss them in detail in Section 3). In particular, the factor analysis, PCA, and the averaging of stock ranks pick up common variation in the characteristics, which can be produced by various factors unrelated to expected stock returns. The fitted values from the Fama-MacBeth regression can be less informative than *AFER* is because the regression estimates ignore the assumed

[5] In our analysis, we exclude stocks with prices below \$5. When those stocks are included in the sample, the hedge returns on *AFER* are 3.16 (*t*-statistic 15.14) and 1.49 (*t*-statistic 5.84) for equal- and value-weighted portfolios, respectively.

[6] This conclusion echoes the results of Avramov et al. (2013), who relate many asset pricing anomalies to firm credit rating downgrades.

constraints on the factor structure of the characteristics. Moreover, the Fama-MacBeth regression estimates are likely to be particularly imprecise when the number of characteristics is large (that is, exactly when the PLS estimates are precise) because cross-sectional OLS regressions effectively involve the computation of the variance-covariance matrix of all regressors, whose size grows quadratically with the number of regressors. Our empirical analysis confirms that none of the alternative aggregation techniques performs better than the PLS-based approach, and that in the vast majority of cases *AFER* is more informative about future returns than the other predictors. Moreover, this result holds even for the largest and most liquid stocks from the top 10% of market capitalization.

Overall, we find that the *AFER*-produced dispersion of realized returns is statistically and economically significant and larger than that generated by either individual characteristics or alternative information aggregation techniques. Our results indicate that PLS is a promising statistical approach to estimation of expected returns from multiple firm characteristics. As such, it may help to understand the cross-sectional behavior of stock returns, as well as to develop profitable trading strategies if the variation in stock returns is at least partially explained by mispricing.

Our paper is related to several strands in the literature. First, our interpretation of expected returns as latent variables makes the paper close to the literature on the time-series predictability of aggregate stock returns by filtered expectations.[7] Specifically, our econometric technique belongs to the PLS framework adopted by Kelly and Pruitt (2013, 2015) and Huang et al. (2015). The latent variable approach is also used by Gibbons and Ferson (1985); Ferson (1990); and Ferson, Foerster, and Keim (1993), who aim to explain the time series of expected returns using a small number of expected risk premiums. In contrast to these works, we use individual firm characteristics for predicting the cross-sectional distribution of expected stock returns rather than the time series of stock returns.

Second, our paper contributes to the analysis of stock return predictability by multiple firm characteristics (e.g., Haugen and Baker 1996; Hanna and Ready 2005; Lewellen 2015; Green, Hand, and Zhang 2016). In particular, it extends the set of techniques that combine various accounting variables into informative signals about returns (e.g., Ou and Penman 1989; Abarbanell and Bushee 1998; Piotroski 2000; Mohanram 2005; Asness, Frazzini, and Pedersen 2014).

Third, our paper belongs to the growing literature that promotes a comprehensive approach to asset pricing anomalies. Fama and French (2008) examine the strength of 7 anomalies across size groups. Stambaugh, Yu, and

[7] An incomplete list of papers includes Conrad and Kaul (1988); Brandt and Kang (2004); Pástor and Stambaugh (2009); Binsbergen and Koijen (2010); Rytchkov (2012); and Piatti and Trojani (2015). Various expectations are modeled as unobservable state variables also in Hamilton (1985); Balke and Wohar (2002); and Rytchkov (2010), among others.

Yuan (2012) find that 11 anomalies appear to be stronger following periods of high investor sentiment. Avramov et al. (2013) document that many anomalies are concentrated in firms with low credit ratings and the profitability of those anomalies (except asset growth) derives from credit rating downgrades. Chordia, Subrahmanyam, and Tong (2014) reconsider 12 popular anomalies and argue that their returns diminished in recent years because of a decline in trading costs and an increase in trading activity. McLean and Pontiff (2016) compare pre- and postpublication returns on 97 anomalies and find that the profitability of the anomalies tends to decline after their publication. Novy-Marx and Velikov (2016) estimate after-trading-cost returns on 23 anomalies. Kogan and Tian (2015) explore how easy it is to explain 27 asset pricing anomalies by a 3-factor model in which 2 factors are return spreads produced by anomalous characteristics. Green, Hand, and Zhang (2016) consider 94 firm characteristics and find that 12 of them contain independent information about returns in the sample of non-microcap stocks. None of these papers focuses on the aggregation of information from a large number of firm characteristics.

1. Methodology

1.1 Baseline specification

Consider N stocks whose characteristics are observed in at least two time periods. The best predictor of returns on stock i at time t is expected return $\mu_{it} = E[R_{it+1}|\mathcal{F}_t]$, where $\mathcal{F}_t$ denotes all information available to market participants. By the definition of expectation, the realized return on stock i can be written as

$$R_{it+1} = \mu_{it} + \varepsilon_{it+1}, \tag{1}$$

where $E[\varepsilon_{it+1}|\mathcal{F}_t]=0$ and unexpected returns ε_{it+1}, $i=1,\dots,N$, are assumed to be independent from all variables in the information set $\mathcal{F}_t$, although ε_{it+1} and ε_{jt+1} for $i \neq j$ can be correlated. Econometricians lack some information contained in $\mathcal{F}_t$ and do not know μ_{it}. Instead, they observe firm characteristics X^a_{it}, $a=1,\dots,A$, such that $X^a_{it-s} \subset \mathcal{F}_t$ for $s \geq 0$. In practice, the characteristics can describe different aspects of a firm and be measured in incomparable units. Therefore, it is common to demean and standardize them, so we assume that at each time period the variables X^a_{it}, $a=1,\dots,A$, have zero cross-sectional means and unit variances.

The key assumption of our framework is that the latent vector μ_{it} is the only factor in the characteristic space related to future stock returns. Specifically, we assume that

$$X^a_{it} = \delta^a_t(\mu_{it} - \bar{\mu}_t) + u^a_{it}, \tag{2}$$

where δ^a_t measures the sensitivity of characteristic a to expected returns, and $\bar{\mu}_t$ is the cross-sectional average of expected returns at time t. By construction, the cross-sectional average of u^a_{it} is zero, but there is no restriction on its cross-characteristic mean. Note that if expected returns are described by an asset pricing model that admits a single-beta representation, the expected excess returns are proportional to betas with respect to a risk factor. Therefore, in this case Equation (2) can be viewed as a relation between characteristics and unobservable betas. Such a relation is in the spirit of the large theoretical literature that demonstrates how firm characteristics and betas are linked in a rational asset pricing framework (e.g., Kogan and Papanikolaou 2012). When betas are unobservable, the relations between characteristics and betas can explain the existence of asset pricing anomalies (e.g., Lin and Zhang 2012).[8] It should be emphasized that the validity of a 1-factor asset pricing model is a sufficient but not necessary condition for Equation (2) to hold. For example, our framework is valid even when the latent variables describe some sort of mispricing reflected in multiple firm characteristics.

For future convenience, we denote the sample cross-sectional variance and covariance as $\overline{Var}$ and $\overline{Cov}$, respectively, and reserve $\widetilde{Var}$ and $\widetilde{Cov}$ for the sample variance and covariance in the characteristic space. The standardization of the characteristics implies that $\overline{Var}(X_{it}^a)=1$ in each time period t. For the identifiability of expected returns, we make the following additional assumptions.

Assumption 1 (distribution of expected returns). In each period t,

$$\bar{\mu}_t = \frac{1}{N}\sum_{i=1}^{N}\mu_{it} \xrightarrow{p} \mu_t \quad \text{and} \quad \overline{Var}(\mu_{it}) = \frac{1}{N}\sum_{i=1}^{N}(\mu_{it}-\bar{\mu}_t)^2 \xrightarrow{p} V_t$$

as $N\to\infty$, where $V_t>0$.

Assumption 2 (distribution of characteristic loadings). In each period t,

$$\tilde{\delta}_t = \frac{1}{A}\sum_{a=1}^{A}\delta_t^a \xrightarrow{p} \delta_t \quad \text{and} \quad \widetilde{Var}(\delta_t^a) = \frac{1}{A}\sum_{a=1}^{A}(\delta_t^a-\tilde{\delta}_t)^2 \xrightarrow{p} \Lambda_{t,t}$$

as $A\to\infty$, where $\Lambda_{t,t}>0$. Also, for consecutive periods $t-1$ and t,

$$\widetilde{Cov}(\delta_{t-1}^a,\delta_t^a) = \frac{1}{A}\sum_{a=1}^{A}(\delta_{t-1}^a-\tilde{\delta}_{t-1})(\delta_t^a-\tilde{\delta}_t) \xrightarrow{p} \Lambda_{t-1,t}$$

as $A\to\infty$, where $\Lambda_{t-1,t}>0$.

8 True betas in conditional linear factor models are theoretically infeasible because they depend on the information available to investors but not to an econometrician (Hansen and Richard 1987). The estimation of betas is complicated by mismeasurement of risk factors (e.g., Black, Jensen, and Scholes 1972; Roll 1977) and time variation in betas (e.g., Harvey 1989; Ferson and Harvey 1991; Lewellen and Nagel 2006; Li and Yang 2011; Ang and Kristensen 2012).

Assumption 3 (orthogonality of errors and expected returns). In each period t and for each characteristic a, $a=1,\ldots,A$,

$$\overline{Cov}(\mu_{it},u_{it}^a)=\frac{1}{N}\sum_{i=1}^{N}(\mu_{it}-\bar{\mu}_t)(u_{it}^a-\bar{u}_t^a)\xrightarrow{p}0 \quad \text{as} \quad N\to\infty.$$

Assumption 4 (orthogonality of errors and loadings). In each period t and for each stock i, $i=1,\ldots,N$,

$$\widetilde{Cov}(\delta_{t-1}^a,u_{it}^a)=\frac{1}{A}\sum_{a=1}^{A}(\delta_{t-1}^a-\tilde{\delta}_{t-1})(u_{it}^a-\tilde{u}_{it})\xrightarrow{p}0 \quad \text{as} \quad A\to\infty.$$

Assumption 1 formalizes a natural condition that the cross-sectional distribution of expected stock returns has a finite mean and positive standard deviation. Similarly, Assumption 2 implies that the population distribution of the vector of characteristic loadings (δ_{t-1}^a δ_t^a) has a finite first moment and non-zero second moments. It can be interpreted as a condition that when the number of characteristics increases, their average informativeness stays the same because otherwise a disproportionately large number of useless characteristics would shift the distribution of δ_t toward zero and decrease $\Lambda_{t,t}$. Assumption 2 also requires that relative magnitudes of the characteristic slopes are statistically stable over time. Assumption 3 states that there is no cross-sectional relation between u_{it}^a and the individual expected stock returns. In other words, all information about the cross section of stock returns contained in the characteristics is summarized by μ_{it}. Assumption 4 implies that there is no systematic relation between the past sensitivity of each characteristic to expected returns and the part of the characteristic unrelated to expected returns.

It should be emphasized that our assumptions impose restrictions on neither the cross-sectional correlations of ε_{it} (all stocks are likely to have exposure to the market factor) nor the cross-characteristic or cross-sectional correlations of u_{it}^a. In particular, the cross-characteristic correlations between u's imply that X_{it}^a may have a complex correlation structure with multiple factors in the characteristic space where μ_{it} is only one of them.[9] The existence of returns-unrelated factors in the characteristics has an important consequence: a simple averaging of the characteristics or naive extraction of their common component by the factor analysis or PCA would mix returns-related and returns-unrelated factors, thereby providing inconsistent estimates of expected stock returns.

In the described framework, the main problem of an econometrician is to estimate μ_{it} by filtering them from the observable characteristics X_{it}^a. To solve

9 For example, in the simulation analysis conducted in the Online Appendix we assume that $X_{it}^a=\delta_t^a(\mu_{it}-\bar{\mu}_t)+f_{it}^u+\eta_{it}^a$, where f_{it}^u is a common factor in characteristics that is unrelated to expected returns.

this problem, we use a procedure that belongs to a family of partial least squares (PLS) algorithms. Initially promoted by Herman Wold in a series of publications (e.g., Wold 1975, 1982), the PLS framework is actively used in computational chemistry (e.g., Frank and Friedman 1993; Wold, Sjöström, and Eriksson 2001) and behavioral sciences (e.g., Vinzi et al. 2010). Recently, it has found applications in finance: PLS is employed by Kelly and Pruitt (2013) to extract market expectations from the cross section of valuation ratios and by Huang et al. (2015) to construct an investor sentiment index from several sentiment proxies. An extension of PLS to quantile regression is used by Giglio, Kelly, and Pruitt (2016) for the aggregation of 19 measures of systemic risk.

The main objective of PLS is the extraction of a common factor from a set of predictive variables that has the highest covariance with the predicted (target) variable. As in the words of Kelly and Pruitt (2015), PLS is a "disciplined" dimension reduction technique. In contrast to PCA and factor analysis, which also extract one or few factors that concisely describe the variability of data and correlations between predictors, respectively, PLS identifies a factor with the best ability to predict the target variable even though this factor may not be the most important source of common variation in the predictors. The PLS-based estimation of our model at time t can be implemented in two steps.

Step 1. Run separate cross-sectional regressions of R_{it}, $i=1,\ldots,N$, on each individual firm characteristic X^a_{it-1}, $i=1,\ldots,N$, for $a=1,\ldots,A$ and denote the obtained slopes as λ^a_t.

Step 2. For each firm i, $i=1,\ldots,N$, run a regression of X^a_{it} on λ^a_t, $a=1,\ldots,A$, and denote the obtained slopes as $\hat{\mu}_{it}$.

The steps of the procedure admit an intuitive interpretation. By running regressions of current returns on each lagged characteristic at Step 1, we effectively estimate the loadings of characteristics on expected returns in the previous time period (up to a scalar multiplicative factor), so λ^a_t can be viewed as a proxy for δ^a_{t-1}. When the loadings are stable over time (Assumption 2 holds), λ^a_t is also a proxy for δ^a_t. Running a regression of current characteristics of each stock on the estimated loadings at Step 2, we find the slope in Equation (2) (again, up to a multiplicative factor that is the same for all stocks), which coincides with the current demeaned expected return on the stock. Because λ^a_t and $\hat{\mu}_{it}$ are determined only by realized returns on stocks at time t and characteristics of firms at times t and $t-1$, the estimation procedure does not suffer from look-ahead bias, and all computations can be performed in real time.

Our estimation procedure is similar to the three-pass regression filter (3PRF) developed by Kelly and Pruitt (2015), which is a generalization of classic PLS, and it can be viewed as an implementation of 3PRF with 3 modifications. First, the objective of 3PRF is to find the best forecast for a single target variable y from a long history of multiple observables (at time t only y_{t+1} is predicted), whereas we estimate expected returns on multiple stocks at a given moment from multiple firm characteristics. Effectively, the time dimension in Kelly and Pruitt (2015) corresponds to our stock dimension and multiple time series of predictors in Kelly and Pruitt (2015) correspond to multiple firm characteristics in our approach. However, Kelly and Pruitt (2015) do not "forecast" past realizations of the variable y, whereas we estimate expected returns on all stocks. Second, 3PRF uses the same set of predictors to estimate factor loadings and to predict the target variable, whereas we use lagged characteristics to estimate the loadings but current characteristics to estimate expected returns. Moreover, we allow the loadings to vary over time. Third, we ignore the last (third) step of 3PRF because without additional assumptions, the scale of expected returns is unidentified, although the ranking of expected returns is asymptotically revealed after implementing only two steps. The following Proposition formalizes this result.

Proposition 1. If Assumptions 1–4 hold, the described 2-step procedure asymptotically reveals the cross section of expected stock returns up to an unobservable time-varying factor F_t, which is the same for all stocks, that is,

$$\operatorname*{plim}_{A\to\infty}\operatorname*{plim}_{N\to\infty}\hat{\mu}_{it}=F_t(\mu_{it}-\mu_t), \qquad \text{where} \qquad F_t=\frac{\Lambda_{t-1,t}}{\Lambda_{t-1,t-1}V_{t-1}},$$

and $\Lambda_{t-1,t}$, $\Lambda_{t,t}$, and V_{t-1} are defined in Assumptions 1 and 2.

Proof. See Appendix A. ■

Proposition 1 states that $\hat{\mu}_{it}$ are consistent estimates of demeaned expected stock returns up to a scaling factor. Thus, asymptotically the cross-sectional ranking of stocks based on $\hat{\mu}_{it}$ coincides with that based on μ_{it}, and sorting stocks into portfolios using $\hat{\mu}_{it}$ should deliver the highest dispersion of future portfolio returns. Proposition 1 is an analog of Theorem 1 in Kelly and Pruitt (2015), which establishes consistency of the 3PRF forecast.

As in Kelly and Pruitt (2015), the PLS-based estimates $\hat{\mu}_{it}$ admit a closed-form matrix representation. However, the 2-step procedure is more convenient when characteristic values are missing for some stocks. Indeed, in the first-step regressions the characteristics are used one by one, so missing values of one characteristic do not affect the estimates λ_t^a for other characteristics. Thus, the estimation of λ_t^a efficiently uses all available information. The second-step regression of all characteristics available for the given stock at the given moment on the slopes of those characteristics obtained at the first step is again a univariate regression with the number of observations equal to the number of characteristics. Missing values of characteristics reduce the sample size (and possibly the precision of inference) but do not invalidate the estimation of expected returns. The robustness of our procedure to missing observations distinguishes it from the Fama-MacBeth regression and factor analysis, which require the availability of all characteristics for all stocks.

To predict the ranking of returns at time $t+1$, our procedure uses only realized returns at time t and characteristics measured at times $t-1$ and t. However, in practice, firm characteristics and returns are observed in multiple periods, and our approach can be easily modified to take advantage of this additional information. Because λ_t^a is a proxy for δ_t^a, under a mild assumption that the relation between characteristics and expected returns is stable over time, the slope δ_t^a can be more precisely estimated by using averages of λ_s^a, $s \leq t$, instead of only the most recent λ_t^a. For example, when δ_t^a does not vary over time, it is natural to average λ_s^a from all previous periods and use the result at Step 2. To some extent, the averaging of λ_s^a is similar to averaging of cross-sectional regression slopes in the standard Fama-MacBeth approach.

There is another benefit from the averaging of λ_s^a. The assumed independence of unexpected returns ε_{it} from all variables at time $t-1$ implies that ε_{it} and expected returns μ_{it-1} are uncorrelated in the cross section. This assumption can be violated when the conditional volatility of returns is related to expected returns, the unexpected returns have a common factor (it could be related to macroeconomic shocks), and the cross-sectional dispersion of return volatilities is determined by heterogeneity of loadings on the common factor.[10] Indeed, assume that expected returns are positively related to volatilities and the realization of the factor in ε_{it} is positive. Then stocks with higher μ_{it-1} have higher volatilities, higher loadings on the common factor and, hence, higher realizations of ε_{it}. When ε_{it} and μ_{it-1} are correlated, the estimates of λ_t^a are contaminated by an additional term, which appears to be proportional to $\overline{Cov}(X_{it-1}^a, \varepsilon_{it})$ and makes the estimates of expected returns inconsistent. The time series averaging substantially alleviates this problem (and completely eliminates it as the number of periods tends to infinity) because ε_{it} enter the additional term linearly, and their time-series averages asymptotically converge to $E[\varepsilon_{it}]=0$. Effectively, the periods with positive and negative shocks to the common factor in ε_{it} are averaged away, and the consistency of the estimates is restored.

1.2 Extension: multiple returns-related factors in the characteristic space

In this section, we consider an extension of our baseline model in which characteristics are linked to expected returns through L common latent factors β_{lit}, $l=1,\dots,L$. Specifically, we assume that the expected return on stock i at time t is related to the latent factors as

$$\mu_{it} = \sum_{l=1}^{L} \beta_{lit} \gamma_l, \tag{3}$$

where γ_l, $l=1,\dots,L$, are loadings on various factors. As in the previous section, we stay agnostic about the nature of the latent factors, but when returns are

10 We thank Soohun Kim for pointing this out.

described by a rational asset pricing model, the latent factors β_{lit} can be identified with unobservable betas with respect to risk factors.[11] Introducing a $1\times L$ matrix of factor values B_{it} and an $L\times 1$ matrix of loadings Γ, the realized return on stock i at time $t+1$ can be written as

$$R_{it+1}=B_{it}\Gamma+\varepsilon_{it+1}. \tag{4}$$

Again, unexpected returns ε_{it+1}, $i=1,\dots,N$, are independent from all variables available at time t, $E[\varepsilon_{it+1}|\mathcal{F}_t]=0$, and the matrices B_{it} and Γ are unobservable to econometricians.

As in the baseline model, we assume that firm characteristics are aligned with stock returns because they are proxies for the unobservable latent factors: the demeaned and standardized characteristics X^a_{it}, $a=1,\dots,A$, of firm i at time t are related to B_{it} as

$$X^a_{it}=(B_{it}-\bar{B}_t)\Delta^a+u^a_{it}, \tag{5}$$

where Δ^a, $a=1,\dots,A$, are $L\times 1$ matrices that determine the sensitivity of characteristics to the factors, $\bar{B}_t$ is the cross-sectional average of the factors, and u^a_{it} are components of characteristics that are unrelated to stock returns and that can be arbitrary correlated among themselves. Note that because the characteristics do not contain information about the matrix Γ, it would be unidentifiable in period t if it were allowed to vary over time. Therefore, to get consistent estimates of expected returns, we assume that Γ is constant, so it can be estimated using characteristics and returns available in period $t-1$. Also, we assume that the sensitivities Δ^a are time-constant because otherwise it would be impossible to identify the relative scales of β_{lit}, $l=1,\dots,L$, from X^a_{it} and, therefore, to combine the estimated factor values in consistent estimates of expected returns. The assumption of constant characteristic loadings distinguishes the multifactor model from our baseline 1-factor specification, in which the slopes δ_t can change from period to period and, as a result, the single latent factor can be estimated only up to a multiplicative constant.

To estimate unobservable expected returns, we again employ an analog of 3PRF from Kelly and Pruitt (2015). The procedure in the multifactor case involves three types of variables: a realized target variable, which is R_{it} in our case; a set of predictors, which are firm characteristics X^a_{it} in our case; and a set of so-called proxy variables Z^l_{it}, $l=1,\dots,L$, whose construction is described below. As in the 1-factor case, we make several additional assumptions under which the procedure yields consistent estimates of expected returns.

Assumption 1' (distribution of latent factors). In each period t, $\overline{Var}(B_{it}) \xrightarrow{p} V_t$ as $N\to\infty$ and the matrix V_t is positive definite.

[11] In this case betas determine expected excess returns, not expected raw returns, but adding a constant to Equation (3) is inconsequential for our analysis because our procedure estimates only cross-sectionally demeaned expected returns.

Assumption 2' (distribution of characteristic loadings). $\widetilde{Var}(\Delta^a) \xrightarrow{p} \Lambda$ as $A \to \infty$ and the matrix Λ is positive definite.

Assumption 3' (orthogonality of errors and latent factors). In each period t and for each characteristic a, $a=1,\ldots,A$, $\overline{Cov}(B_{it}, u_{it}^a) \xrightarrow{p} 0$ as $N \to \infty$.

Assumption 4' (orthogonality of errors and loadings). In each period t and for each stock i, $i=1,\ldots,N$, $\widetilde{Cov}(\Delta^a, u_{it}^a) \xrightarrow{p} 0$ as $A \to \infty$.

The assumptions are similar to Assumptions 1–4 from Section 1.1 and correspond to Assumptions 2.1, 2.2, 4.3, and 4.1 in Kelly and Pruitt (2015).[12] The steps of the 3PRF procedure adopted to our framework are as follows.

Step 1. Run cross-sectional regressions of the characteristics X_{it-1}^a, $a=1,\ldots,A$, on the proxy variables Z_{it}^l, $l=1,\ldots,L$, and retain the $L \times 1$ matrices of slopes λ_t^a.

Step 2. For each stock i, $i=1,\ldots,N$, run regressions of X_{it-1}^a and X_{it}^a on λ_t^a in the characteristic space and denote the obtained matrices of slopes as $\hat{B}_{it-1}$ and $\hat{B}_{it}$.

Step 3. Run a cross-sectional regression of R_{it} on $\hat{B}_{it-1}$, $i=1,\ldots,N$, and denote the obtained matrix of slopes as $\hat{\Gamma}_t$. Compute an estimate of expected returns as $\hat{\mu}_{it} = \hat{B}_{it}\hat{\Gamma}_t$.

Steps 1 and 2 are similar to those in the 1-factor procedure. Indeed, when the realized return R_{it} is chosen to be the only proxy variable Z_{it}^1, the slopes λ_t^a are proportional to the covariances between individual characteristics and future returns, which are effectively estimated by cross-sectional regressions of realized returns on past characteristics at Step 1 of the 1-factor procedure. Also, as in the 1-factor model and in contrast to Kelly and Pruitt (2015), we use characteristics from time $t-1$ to estimate the factor loadings but those from time t to find betas. An additional step (Step 3) is required because now expected returns are determined by multiple latent factors, so not only the realizations of those factors should be estimated, but also a correct linear combination of them should be found. Note that in contrast to the 1-factor procedure, the multifactor procedure estimates $\hat{B}_i$ not only in period t but also in period $t-1$.

To construct the proxy variables Z_{it}^l, Kelly and Pruitt (2015) propose an iterative algorithm in which each subsequent proxy is the error of the 3PRF

[12] Kelly and Pruitt (2015) also impose several other technical conditions that are required for derivation of the 3PRF forecast standard errors.

forecast constructed using proxies obtained at the previous step, and the first proxy is set to be the target variable. We adjust this procedure to our framework and construct the variables Z_{it}^{l}, $l=1,\dots,L$, as follows. We start with $l=1$ and initialize the iteration as $Z_{it}^{1}=R_{it}$. At the iteration step, we implement Step 1 (find λ_t^a), Step 2 (find $\hat{B}_{it-1}$), and Step 3 (find $\hat{\Gamma}_t$), using Z_{it}^{k}, $k=1,\dots,l$, as proxies. A new proxy Z_{it}^{l+1} is computed as $Z_{it}^{l+1}=R_{it}-\hat{B}_{it-1}\hat{\Gamma}_t$. If $l+1=L$, the algorithm stops. Otherwise, we repeat Steps 1–3 with all proxies from the previous iterations and find the next proxy.

Theorem 1 in Kelly and Pruitt (2015) states that the 3PRF forecast is a consistent estimate of the infeasible best forecast that uses all information available at time t. Because our procedure can be viewed as 3PRF with relabeled dimensions and two sets of predictors (those from periods $t-1$ and t) instead of one, the theorem also applies to our setting and establishes that the vector of estimates $\hat{\mu}_{it}$ asymptotically reveals the demeaned expected stock returns.

Finally, note that as the 1-factor procedure, its multifactor version can benefit from the observability of returns and characteristics in multiple time periods. In particular, the errors-in-variables problem in the Step 2 regressions can be mitigated by averaging the estimates λ_s^a, $s\le t$, obtained at Step 1. Moreover, the precision of $\hat{\Gamma}_t$ at Step 3 can also be improved by averaging its estimates across previous time periods (in general, the time spans over which λ_s^a and $\hat{\Gamma}_s$ are averaged can be different), and we exploit this possibility in the empirical analysis as well. To ensure internal consistency of the estimation, we apply the same averaging schemes in the main estimation procedure and in the construction of the proxies.

2. Main Empirical Analysis

In this section, we illustrate how our aggregation technique works by applying it to 26 firm characteristics that are identified in previous studies as predictors of future stock returns.

2.1 Data

Our data come from the standard sources. Stock returns, stock prices, and numbers of shares outstanding are from CRSP monthly files, while accounting data are from Compustat Fundamentals annual files. We exclude financial firms (with the SIC codes between 6000 and 6999) and consider only NYSE, AMEX, and NASDAQ firms with common stocks (with SHRCD equal to 10 or 11). We take accounting data used for construction of characteristics in calendar year t from financial statements with the fiscal year end in year $t-1$. Returns are monthly stock returns with dividends and adjusted for delisting. To reduce the effect of transaction costs and other market frictions on our results, in each month, we exclude stocks that had prices below \$5 at the end of the previous month.

As signals about future stock returns, we use 26 variables that are associated with prominent asset pricing anomalies. These variables can be classified into six groups. The first group contains value-related characteristics such as size (S), book-to-market ratio (B/M), earnings-to-price ratio (E/P), and cash flow-to-price ratio (C/P). The second group includes returns-based variables such as market beta (B), momentum (MOM), long-term reversal (LTR), short-term reversal (STR), idiosyncratic volatility ($IdVol$), maximum daily return over the past month (MAX), and expected idiosyncratic skewness (EIS). The third group contains growth-related characteristics such as total asset growth (AG), abnormal capital investments (CI), investment growth (IG), investment-to-capital ratio (I/K), investment-to-assets ratio (I/A), accruals (ACC), net operating assets (NOA), net stock issues (NS), and composite stock issuance (ι). The fourth group contains returns on equity (ROE) and returns on assets (ROA), which measure firm profitability. The fifth group includes leverage (LV) and O-score (O), which are distress-related variables. The last group contains turnover (TO) and analysts' forecasts dispersion (D). The construction of each characteristic is described in Appendix B and it largely follows the first paper in which the variable is identified as related to stock returns.

2.2 Individual characteristics

We start our analysis with investigating whether the selected characteristics can individually predict the cross section of stock returns. For consistency with the subsequent analysis, we limit the sample to the period from 1970 to 2012.[13] Each month we sort all stocks with respect to each characteristic, form decile portfolios, and compute equal-weighted and value-weighted returns on them in the next month. The value-related characteristics, growth-related characteristics (except size), and leverage are revised once a year at the end of June. The portfolios based on returns on assets and O-score are rebalaced quarterly. The portfolios based on size, turnover, analysts' forecasts dispersion, and all returns-based variables are rebalanced at the end of each month.

We measure the ability of each characteristic to predict returns by the difference between the next month realized returns on top and bottom decile portfolios, which is referred to as hedge returns. Table 1 reports the average hedge returns on each characteristic as well as the sample standard deviations of hedge returns and their t-statistics. Overall, Table 1 confirms that the vast majority of the 26 characteristics are informative about future stock returns. Panel A shows that only three variables, S, B, and ROE, fail to produce statistically significant hedge returns on equal-weighted portfolios, whereas the absolute values of t-statistics of 21 variables exceed 3.[14] The variation in

13 Because of data availability, the sample starts in 1987 for expected idiosyncratic skewness, in 1975 for returns on assets, in 1976 for O-score, and in 1977 for analysts' forecasts dispersion.

14 Because the price and market capitalization are highly correlated in the cross section and we exclude low-priced stocks from the analysis, the inability of size to generate a statistically significant dispersion of returns is consistent with the evidence in Fama and French (2008) that the size anomaly is largely produced by microcap stocks.

Table 1
Individual characteristics

Panel A: equal-weighted portfolios

	S	B/M	E/P	C/P	B	MOM	LTR	STR	IdVol	MAX	EIS	ROE	ROA
Means	−0.16	0.98	0.84	0.95	−0.46	1.59	−0.55	−1.04	−1.02	−0.96	−0.65	−0.07	1.54
Stds	3.71	4.55	3.75	4.26	7.33	5.86	3.59	4.97	6.43	5.97	3.47	2.16	4.30
t-stats	−1.01	4.88	5.11	5.08	−1.43	6.15	−3.50	−4.76	−3.59	−3.65	−3.26	−0.69	7.63
	AG	**CI**	**IG**	**I/K**	**I/A**	**ACC**	**NOA**	**NS**	ι	**LV**	**O**	**TO**	**D**
Means	−0.73	−0.25	−0.43	−0.55	−0.68	−0.52	−0.60	−0.34	−0.70	0.70	−0.64	−0.56	−0.72
Stds	2.86	1.60	1.98	4.59	2.53	2.32	3.62	2.30	3.62	5.17	3.11	4.79	3.92
t-stats	−5.82	−3.54	−4.95	−2.75	−6.08	−5.15	−3.79	−3.40	−4.40	3.09	−4.33	−2.65	−3.83

Panel B: value-weighted portfolios

	S	B/M	E/P	C/P	B	MOM	LTR	STR	IdVol	MAX	EIS	ROE	ROA
Means	−0.17	0.58	0.62	0.49	−0.19	1.34	−0.26	−0.20	−0.99	−0.62	−0.73	−0.07	0.90
Stds	4.02	4.99	4.97	5.08	7.46	7.34	5.07	6.19	7.18	7.13	5.19	3.53	5.29
t-stats	−0.94	2.66	2.83	2.20	−0.57	4.16	−1.17	−0.72	−3.12	−1.98	−2.43	−0.44	3.62
	AG	**CI**	**IG**	**I/K**	**I/A**	**ACC**	**NOA**	**NS**	ι	**LV**	**O**	**TO**	**D**
Means	−0.37	−0.32	−0.29	−0.35	−0.49	−0.38	−0.28	−0.57	−0.59	0.51	−0.12	−0.27	−0.40
Stds	3.70	3.27	3.43	6.43	3.15	3.95	4.04	2.97	4.19	5.74	4.22	5.15	5.15
t-stats	−2.28	−2.23	−1.93	−1.25	−3.55	−2.17	−1.57	−4.35	−3.19	2.00	−0.58	−1.19	−1.61

This table reports time-series averages, standard deviations, and *t*-statistics of hedge returns defined as a difference in monthly returns on top and bottom decile portfolios formed individually by 26 firm characteristics. The decile portfolios are equal-weighted in panel A and value-weighted in panel B. All variables are named in Section 2.1 and their construction is described in Appendix B. Average returns and their standard deviations are reported in percentage points. The sample is from 1970 to 2012 for all characteristics except for *EIS*, *ROA*, *O*, and *D*, for which it starts in 1987, 1975, 1976, and 1977, respectively.

the *t*-statistics across the characteristics is driven by the dispersion in average hedge returns (they vary from 0.07% for ROE to 1.59% for momentum), as well as in the volatilities of hedge returns (they range from 1.60% for abnormal capital investments to 7.33% for beta). In the considered sample, the widest dispersion of returns is produced by momentum and returns on assets. The absolute values of average hedge returns on 20 characteristics exceed 50 basis points per month.

Consistent with the literature, panel B of Table 1 shows that the hedge returns tend to be smaller when computed for value-weighted portfolios and appear to be statistically insignificant not only for S, B, and ROE but also for 9 additional variables LTR, STR, MAX, IG, I/K, NOA, O, TO, and D. Note that the standard deviations are typically larger for the value-weighted portfolios, although for many characteristics the decrease in *t*-statistics is produced by lower expected returns rather than higher volatility of returns. In contrast to the other variables, expected idiosyncratic skewness EIS, abnormal capital investments CI, and net stock issues NS appear to have larger absolute values of hedge returns on value-weighted portfolios than on equal-weighted portfolios. The highest spread is again generated by momentum.

2.3 Filtered expected returns: main results

Next, we apply the procedure proposed in Section 1 to the same 26 characteristics and construct a variable that aggregates information on expected stock returns from all of them. We denote the new variable *AFER*, which stands for "aggregate filtered expected returns." In the baseline analysis, we assume that there is only one factor in the characteristic space that is related to expected returns, so the estimates of expected returns $\hat{\mu}_{it}$ are constructed in two steps: we first find the proxies λ_t^a for the slopes δ_t^a by running cross-sectional regressions of returns on lagged individual characteristics and then obtain $\hat{\mu}_{it}$ by running regressions of current stock characteristics on the estimated slopes.

As discussed in Section 1, the estimates of expected returns are likely to be more precise and less biased if the whole history of characteristics and returns, not only data from the periods $t-1$ and t, is used in the estimation. A natural way to incorporate the time-series dimension into the estimation procedure is to compute time-series averages of λ_s^a, $s \leq t$ and use them in the Step 2 regressions at time t instead of λ_t^a. The way that λ_s^a are averaged is an econometrician's choice, and we separately consider the versions of *AFER* based on the averages of λ_s^a over the past 5 years, past 10 years, and all previous months.[15]

Because in the early part of the sample period some characteristics are missing for many stocks, we augment the estimation procedure with a few additional conventions. First, we run the Step 1 regression for characteristic a in month t only if there are at least 10 stocks for which the characteristic is available. Otherwise, the slope λ_t^a is deemed missing.[16] Second, we run the Step 2 regression for stock i in month t only if there are at least 4 characteristics available for this stock in the given month. If the condition is violated, the expected return on stock i in month t is recorded as missing.

Our estimation procedure provides the variable *AFER*, which can be treated as a new firm characteristic. Because *AFER* is an estimate of demeaned expected returns only up to a multiplicative factor (the asymptotic limit of the factor is denoted by F_t in Proposition 1), the usual standard errors for it are undefined. To investigate the empirical relation between *AFER* and future stock returns, we adopt the same approach as we use for individual firm characteristics in Section 2.2. Specifically, each month, we sort stocks into decile portfolios according to *AFER*, compute realized equal-weighted and value-weighted returns on the portfolios, as well as the difference in returns on the top and bottom portfolios in the next month, and use the time series averages of returns, their standard errors, and the corresponding t-statistics to measure the cross-sectional dispersion of returns produced by *AFER* and test hypotheses about it. Note that because we sort stocks into portfolios, only the rank of a stock

15 In the Online Appendix, we consider another specification in which λ_t^a are assumed to follow autoregressive processes with characteristic-specific parameters and demonstrate that potential benefits of a more accurate modeling of λ_t^a are offset by additional estimation noise.

16 Practically, the characteristic is either missing completely for all stocks or available for more than 10 stocks.

according to *AFER* matters, so the identifiability of expected returns only up to a positive multiplicative factor is not an obstacle for our analysis. There is also another benefit from using portfolio sorts. The slope coefficients λ_t^a are noisy estimates of δ_t^a, so the Step 2 regression pre-asymptotically suffers from the errors-in-variables problem and, therefore, the estimates $\hat{\mu}_{it}$ are biased. Because only the ranking of estimated expected returns matters for portfolio sorts, our results are robust to any bias that shifts all expected returns by a constant or rescales them by a certain factor (the latter case includes the classic attenuation bias). Moreover, the error in the estimates of expected returns matters only when a stock is placed in a wrong portfolio, but this can happen only to stocks that are close to the decile boundary, and it is unlikely that many stocks are affected in this way.

The average returns on decile portfolios formed by various versions of *AFER*, as well as the volatilities of realized portfolio returns and their *t*-statistics, are reported in Table 2. It shows that even a 2-period version of *AFER* that uses only the most recent estimates of λ_t^a produces a statistically significant dispersion in stock returns: the hedge returns are 1.69% (*t*-statistic is 5.45) and 1.21% (*t*-statistic is 3.61) per month for equal-weighted and value-weighted portfolios, respectively. Moreover, the returns on the *AFER* portfolios tend to monotonically increase with the portfolio number. Table 2 also demonstrates that the average hedge returns and their *t*-statistics increase with the averaging horizon for λ_t^a. In particular, the strongest result is observed when λ_t^a are averaged over all previous months: in this case, the spread is 2.06% per month with the *t*-statistic of 9.31 for equal-weighted portfolios and 1.56% per month with the *t*-statistic of 5.45 for value-weighted portfolios. This result implies that, as discussed in Section 1.1, the averaging of λ_t^a over the longest possible period produces the most precise estimates of δ_t^a by diminishing the estimation noise. The version of *AFER* in which λ_t^a are averaged over all previous months is used as a default choice in the subsequent analysis.

Our results are nontrivial and have two implications. First, they imply that the characteristics indeed share the same returns-related latent factor, so there is commonality in the asset pricing anomalies associated with the considered characteristics. Second, the comparison of Tables 1 and 2 shows that the hedge returns on *AFER* with averaged λ_t^a exceed those on individual characteristics and have higher *t*-statistics, so high expected returns are generated by *AFER* without a commensurate increase in the volatility of returns. For example, the monthly volatilities of the *AFER* hedge returns in panel D of Table 2, which are 5.02% and 6.52% for equal-weighted and value-weighted portfolios, respectively, fall in the range of the volatilities of hedge returns on the individual characteristics reported in Table 1. Thus, our results imply that various firm characteristics indeed can be viewed as noisy signals that contain different information about expected returns.

Finally, to assess the hedge returns produced by *AFER* from the investment perspective, we examine the turnover of the *AFER* portfolios using two metrics.

Table 2
Returns on decile *AFER* portfolios

Panel A: no averaging of λ_t^a

	EW portfolios											VW portfolios										
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**(10-1)**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**(10-1)**
Means	−0.10	0.43	0.70	1.00	1.06	1.33	1.38	1.46	1.57	1.59	1.69	0.10	0.62	0.74	0.78	0.90	1.06	0.97	1.10	1.15	1.31	1.21
Stds	8.02	7.09	6.45	5.98	5.66	5.49	5.48	5.63	5.91	6.82	7.05	8.00	6.93	6.15	5.88	5.46	5.27	5.22	5.50	5.82	6.76	7.60
t-stats	−0.28	1.38	2.45	3.79	4.25	5.48	5.71	5.89	6.04	5.30	5.45	0.28	2.04	2.72	3.00	3.73	4.56	4.23	4.54	4.49	4.39	3.61

Panel B: averaging of λ_t^a over past 5 years

	EW portfolios											VW portfolios										
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**(10-1)**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**(10-1)**
Means	−0.30	0.49	0.82	1.07	1.15	1.32	1.39	1.41	1.48	1.58	1.88	−0.27	0.34	0.66	0.74	0.96	1.05	1.00	1.16	1.23	1.14	1.41
Stds	7.85	7.11	6.55	6.09	5.78	5.54	5.32	5.13	5.16	5.88	4.68	8.60	7.46	6.76	6.03	5.69	5.20	4.90	4.81	4.72	5.27	6.53
t-stats	−0.87	1.55	2.84	4.00	4.51	5.41	5.95	6.24	6.50	6.10	9.12	−0.72	1.05	2.21	2.79	3.83	4.57	4.65	5.47	5.93	4.91	4.91

Panel C: averaging of λ_t^a over past 10 years

	EW portfolios											VW portfolios										
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**(10-1)**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**(10-1)**
Means	−0.35	0.48	0.84	1.09	1.18	1.34	1.35	1.43	1.46	1.59	1.94	−0.17	0.24	0.49	0.79	0.99	0.94	1.02	1.03	1.06	1.20	1.37
Stds	8.19	7.35	6.72	6.20	5.83	5.45	5.13	4.90	4.81	5.58	4.49	8.84	7.56	6.69	6.14	5.49	5.10	4.56	4.31	4.17	4.79	6.51
t-stats	−0.97	1.47	2.84	3.99	4.61	5.58	5.95	6.62	6.89	6.49	9.82	−0.44	0.73	1.66	2.92	4.08	4.17	5.09	5.41	5.77	5.68	4.78

Panel D: averaging of λ_t^a over all previous months

	EW portfolios											VW portfolios										
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**(10-1)**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**(10-1)**
Means	−0.38	0.43	0.81	1.04	1.16	1.33	1.39	1.42	1.53	1.68	2.06	−0.28	0.27	0.55	0.85	0.95	1.00	1.06	1.14	1.10	1.28	1.56
Stds	8.49	7.60	6.91	6.23	5.84	5.42	5.09	4.81	4.64	5.16	5.02	8.93	7.58	6.58	5.85	5.36	4.93	4.47	4.14	4.20	4.94	6.52
t-stats	−1.02	1.29	2.65	3.81	4.50	5.58	6.21	6.69	7.48	7.38	9.31	−0.72	0.80	1.89	3.29	4.01	4.62	5.39	6.25	5.96	5.89	5.45

This table shows time-series averages, standard deviations, and *t*-statistics of monthly equal-weighted (EW) and value-weighted (VW) stock returns on decile portfolios formed by sorting firms on the aggregate filtered expected returns *AFER*. The columns (10–1) report the statistics for the difference in returns on the top and bottom portfolios. The variable *AFER* is constructed without time-series averaging of λ_t^a (in panel A), with averaging of λ_t^a over the most recent 5 years (in panel B), with averaging of λ_t^a over the most recent 10 years (in panel C), and with averaging of λ_t^a over all previous months (in panel D). The sample is from January 1970 to December 2012. The portfolios are rebalanced monthly. All returns and standard deviations are reported in percentage points.

First, we compute the average cross-sectional correlation of *AFER* in two subsequent periods and find that it is 0.68. It is lower than the correlations of the characteristics that are revised once a year such as the valuation ratios (for which the correlation is typically higher than 0.9) but much higher than the correlations of the characteristics that are substantially revised every month (it is 0.3 for *MAX* and 0.5 for *IdVol* and *D*). Second, we compute the average percentage of stocks that are assigned by *AFER* to the extreme decile portfolios in two subsequent months, and it appears to be 51% and 61% for the bottom and top portfolios, respectively. Again, the turnover is higher than implied by rarely revised characteristics: typically, more than 90% of stocks in the extreme portfolios formed by them are the same in two consecutive months. However, for the top-decile portfolios based on *MAX* and *IdVol* this number drops to 25% and 35%, respectively, so their turnover is much higher than the turnover of the *AFER* portfolios. Overall, the intermediate turnover of the *AFER* portfolios is a consequence of constructing *AFER* from both often and rarely revised characteristics.

2.4 Robustness tests

The previous section shows that *AFER* efficiently aggregates information from firm characteristics. In this section we explore whether this conclusion is sensitive to the details of how *AFER* is constructed. Unless stated otherwise, when computing *AFER*, we average λ_t^a from all previous months.

In the first modification, we change the cross-sectional regressions that we run at Step 1. In the baseline case, we use individual stocks. However, extreme realizations of characteristics and returns can disproportionately affect the estimated slopes λ_t^a and, as a result, decrease the precision of $\hat{\mu}_{it}$ obtained at Step 2. One possible remedy is to sort stocks into P portfolios with respect to each characteristic and then run a regression of average portfolio returns on average values of the characteristics in the portfolios.[17] We consider the cases $P=10$, $P=20$, $P=50$, and $P=100$. To ensure that each portfolio contains at least several stocks, the sample starts in January 1975. The hedge returns and their t-statistics produced by *AFER* in this case are presented in panel A of Table 3.

Our results show that regressions on portfolios at Step 1 produce almost the same decile hedge returns and their t-statistics as do regressions on individual stocks. For example, the spread for equal-weighted portfolios in the case $P=10$ is 1.90%, which is only slightly smaller than 2.06% reported in Table 2, and the results for a larger number of portfolios are even closer to our benchmark results. A similar pattern holds for value-weighted portfolios. Thus, either outliers do not distort the Step 1 regressions, or the robustness gain brought by portfolios

17 The idea to run cross-sectional regressions on portfolios goes back to Black, Jensen, and Scholes (1972) and Fama and MacBeth (1973), where portfolios are formed to mitigate the errors-in-variables problem in the estimated betas.

Table 3
Alternative specifications

Panel A: cross-sectional regressions on portfolios

	Hedge returns				t-stats			
	$P=10$	$P=20$	$P=50$	$P=100$	$P=10$	$P=20$	$P=50$	$P=100$
EW	1.90	1.98	2.04	2.02	9.66	9.64	9.46	9.26
VW	1.42	1.49	1.47	1.50	5.32	5.40	5.20	5.16

Panel B: alternative number of lags

	Hedge returns				t-stats			
	$K=1$	$K=3$	$K=6$	$K=12$	$K=1$	$K=3$	$K=6$	$K=12$
EW	2.06	2.06	1.85	1.68	9.31	8.73	7.51	6.71
VW	1.56	1.60	1.26	1.10	5.45	5.34	4.07	3.87

Panel C: nonlinear specification

	Hedge returns	t-stats
EW	2.22	10.82
VW	1.68	6.39

Panel D: risk-adjusted returns

	Hedge returns	t-stats
EW	2.11	14.43
VW	1.44	8.00

Panel E: 2-factor model

	Hedge returns				t-stats			
	no	5 years	10 years	all	no	5 years	10 years	all
EW	1.79	1.96	1.93	1.55	7.76	8.71	8.90	9.18
VW	1.45	1.38	1.37	1.06	5.22	5.20	5.05	4.62

This table shows the differences in monthly returns on top and bottom equal-weighted (EW) and value-weighted (VW) decile portfolios formed by *AFER* constructed in several alternative ways. In panel A the Step 1 cross-sectional regressions are run on P portfolios rather than on individual stocks. In panel B observables include K lags of each characteristic. In panel C the observables include the characteristics, as well as all possible quadratic terms constructed from them. In panel D *AFER* is constructed using stock returns adjusted for risk by the Fama-French three-factor model. In panels A–D the first step estimates λ_t^a are averaged over all previous months. In panel E expected returns are estimated using the 2-factor model, and the columns correspond to different time series averaging schemes for λ_t^a and $\hat{\Gamma}_t$. All returns are reported in percentage points. The sample is January 1970–December 2012.

is offset by the loss in efficiency produced by the decrease in the characteristic dispersion that results from portfolio formation (Ang, Liu, and Schwarz 2011).

In general, not only current characteristics but also their lags can contain information about future returns.[18] To allow for this possibility, we consider all characteristics from K previous months as signals about future returns, where $K=3$, $K=6$, and $K=12$. The baseline specification corresponds to $K=1$. Note that many variables are revised quarterly or annually, so 26 characteristics from K months yield less than $26K$ different signals. Overall, panel B of Table 3 demonstrates that the dispersion of returns produced by *AFER* tends to decrease with K. This effect is likely to be explained by the increase in the number of

[18] For example, Novy-Marx (2012) argues that momentum is largely driven by returns realized 7–12 months before portfolio formation.

weak signals about expected returns when distant lags of characteristics are used in the estimation. As a result, the dispersion of the characteristic loadings δ_t^a shrinks, and this makes the estimates $\hat{\mu}_{it}$ less precise.

The theoretical literature suggests that both betas and characteristics can predict stock returns even when a rational asset pricing model holds because characteristics can be nonlinear functions of expected returns (e.g., Lin and Zhang 2012). Motivated by this insight, we consider *AFER* constructed from 377 observables, which include 26 characteristics and all possible quadratic terms formed from them. The hedge returns produced by such *AFER* and presented in panel C of Table 3 appear to be slightly higher than those produced by the standard *AFER* (2.22% instead of 2.06% for equal-weighted portfolios and 1.68% instead of 1.56% for value-weighted portfolios). Thus, including the higher-order terms may help to obtain higher hedge returns.

So far we have constructed estimates for expected raw returns. However, many of the considered characteristics have become prominent owing to their ability to predict returns adjusted for risk by the Fama-French three-factor model. Therefore, we also consider a version of *AFER* that is built as a predictor of risk-adjusted returns. Following Brennan, Chordia, and Subrahmanyam (1998) and Avramov and Chordia (2006), we compute risk-adjusted returns $\tilde{r}_{it}$ on stock i in month t as

$$\tilde{r}_{it} = r_{it} - \beta_i^{MKT} \times MKT_t - \beta_i^{HML} \times HML_t - \beta_i^{SMB} \times SMB_t.$$

The individual stock betas β_i^{MKT}, β_i^{HML}, and β_i^{SMB} are estimated every month by regressing excess stock returns on a constant and the Fama-French factors MKT_t, HML_t, and SMB_t. In the regressions we use the previous 60 months of observations and require the availability of at least 24 months of return data. The risk-adjusted hedge returns produced by the risk-adjusted version of *AFER* are reported in panel D of Table 3. It shows that the hedge returns on the modified *AFER* are slightly higher (lower) for equal-weighted (value-weighted) portfolios than in the baseline case. In both cases they are statistically significant, and their t-statistics are higher than for raw returns.

Our baseline model assumes that there is only one returns-related factor in the characteristics. To investigate the sensitivity of our results to this assumption, we consider a 2-factor specification of the model, apply the estimation procedure described in Section 1.2 to the same characteristics, and construct a 2-factor version of *AFER*. Note that this extension entails a tradeoff. On the one hand, allowing for an additional factor may help to improve the fit of the model, decrease the estimation error, and produce more precise estimates of expected returns. On the other hand, the 2-factor model is more data demanding because now not only the returns-related factors but also their relative weights should be estimated. This may decrease the precision of the estimates, especially when the factors are highly correlated. Which effect dominates is an empirical question.

Table 4
Sensitivity of returns on *AFER* portfolios to individual characteristics

Panel A: equal-weighted portfolios

	S	B/M	E/P	C/P	B	MOM	LTR	STR	IdVol	MAX	EIS	ROE	ROA
Means	2.12	1.99	2.07	2.05	2.05	1.76	2.04	2.03	2.10	2.16	2.06	2.04	2.07
t-stats	9.43	9.41	9.41	9.31	9.72	7.32	9.20	8.32	10.29	10.02	9.26	9.14	9.36

	AG	CI	IG	I/K	I/A	ACC	NOA	NS	ι	LV	O	TO	D
Means	2.06	2.05	2.05	2.06	2.04	2.04	2.05	2.02	2.04	2.05	1.88	2.05	2.04
t-stats	9.30	9.46	9.23	9.15	9.09	9.23	9.34	9.51	9.15	9.30	9.00	9.31	9.27

Panel B: value-weighted portfolios

	S	B/M	E/P	C/P	B	MOM	LTR	STR	IdVol	MAX	EIS	ROE	ROA
Means	1.63	1.44	1.53	1.54	1.52	1.31	1.56	1.50	1.43	1.54	1.53	1.49	1.56
t-stats	5.82	5.00	5.26	5.35	5.46	4.29	5.40	4.92	5.35	5.49	5.37	5.08	5.45

	AG	CI	IG	I/K	I/A	ACC	NOA	NS	ι	LV	O	TO	D
Means	1.57	1.50	1.46	1.56	1.49	1.53	1.51	1.53	1.53	1.56	1.38	1.51	1.55
t-stats	5.45	5.34	5.06	5.42	5.15	5.39	5.31	5.29	5.35	5.43	4.82	5.31	5.41

This table reports average equal-weighted (panel A) and value-weighted (panel B) *AFER* hedge returns and their t-statistics. The hedge returns are defined as a difference in monthly returns on top and bottom decile *AFER* portfolios. All returns are reported in percentage points. The columns correspond to different versions of *AFER*: in each column, *AFER* is constructed using 25 firm characteristics from our standard set of 26 characteristics, *excluding* the characteristic indicated in the column name. In all columns, the Step 1 estimates λ_t^a are averaged over all previous months. The sample is from January 1970 to December 2012.

As mentioned in Section 1.2, the estimation of a multifactor model requires the choice of the averaging scheme for λ_t^a and $\hat{\Gamma}_t$. For consistency with Table 2, we consider 4 cases in which both λ_t^a and $\hat{\Gamma}_t$ are i) not averaged, ii) averaged over the past 5 years, iii) averaged over the past 10 years, and iv) averaged over all previous months. As before, we assess the ability of *AFER* to produce a dispersion in expected returns by the difference in average returns on the top and bottom decile *AFER* portfolios.

The results are reported in panel E of Table 3. It shows that the hedge returns in a 2-factor model tend to be smaller than those produced by 1-factor *AFER*. For example, the highest average hedge returns on equal- and value-weighted portfolios are 1.96% and 1.45%, respectively, whereas they are 2.06% and 1.56% in the baseline 1-factor model. An additional analysis reveals that betas of the 2 factors obtained at Step 2 are highly correlated, so the Step 3 regression suffers from a multicollinearity problem. This explains why adding a new factor to the baseline model does not help to improve its performance.

Although by construction PLS identifies the common component in characteristics that is related to future returns, there still may exist a concern that high hedge returns on the *AFER* portfolios are largely produced by one characteristic with a particulary strong relation to future returns. To show that this is not the case, we sequentially exclude each characteristic and construct the filtered expectations from the rest of them. The hedge returns produced by various combinations of 25 characteristics are reported in Table 4.

As expected, the hedge returns on the modified filtered expectations and their t-statistics are almost the same as in the baseline case with 26 characteristics, and this result holds for both equal-weighted and value-weighted returns. Among all characteristics, the largest effect is produced by the exclusion of momentum, but even in this case, the decrease in the hedge returns does not exceed 16%. For other characteristics the effect is much weaker. In particular, there is almost no sensitivity to excluding growth-related characteristics. Moreover, the exclusion of some characteristics such as size increases hedge returns. Thus, none of the individual characteristics is crucial for generating large *AFER* hedge returns.

2.5 Filtered expected returns in subsamples

Finally, we examine the dispersion of expected returns produced by *AFER* within subsamples determined by time period, firm capitalization, and idiosyncratic volatility of stock returns. Each partitioning is motivated by recent literature. Fama and French (2008) find that some characteristics such as size, book-to-market, profitability, and asset growth are related to returns only in a subsample of small stocks. Pontiff (1996, 2006) argues that anomalies are more likely to exist among stocks with high idiosyncratic volatility, and indeed the latter has been found to affect the relation between returns and book-to-market (Ali, Hwang, and Trombley 2003), accruals (Mashruwala, Rajgopal, and Shevlin 2006), and asset growth (Lipson, Mortal, and Schill 2011). The consideration of different time periods is motivated by weaker relations between characteristics and returns in the recent period (e.g., Horowitz, Loughran, and Savin 2000; Schwert 2003; Green, Hand, and Soliman 2011; Chordia, Subrahmanyam, and Tong 2014; McLean and Pontiff 2016).

To form size portfolios, we follow Fama and French (2008) and assign each stock to one of the 3 groups: microcaps, small stocks, and big stocks. The breakpoints are the 20th and 50th percentiles of the end-of-June market capitalization of the NYSE stocks. The idiosyncratic volatility is measured as the standard deviation of residuals in the time series regression of daily returns on the Fama-French three factors. We classify all stocks into three idiosyncratic volatility groups (low, medium, and high). As breakpoints, we use the 30th percentile and 70th percentile, that is, stocks with idiosyncratic volatility below the 30th percentile are in the low group, for example. Because the idiosyncratic volatility is measured on a monthly basis, we rebalance the volatility groups every month. As sample periods we consider January 1970–December 1995 (early sample) and January 1996–December 2012 (late sample).

Table 5 reports the differences in equal- and value-weighted returns on top- and bottom-decile portfolios formed by *AFER* within various subsamples. The panels correspond to subsamples based on time period, size, and idiosyncratic volatility. Our results indicate that *AFER* preserves its power to predict returns in all subsamples. In particular, the hedge returns are remarkably stable across time periods, although the *t*-statistics are lower in the late sample. On the one hand, this result implies that the cross-sectional predictability of returns by *AFER* is a real phenomenon, not a statistical fluke specific to a particular time period. On the other hand, the lower *t*-statistic is consistent with the evidence in

Table 5
Filtered expected returns in subsamples

	Sample periods					
	Hedge returns			t-stats		
	(1)	(2)	(3)	(1)	(2)	(3)
EW	2.06	2.15	1.92	9.31	10.12	4.21
VW	1.56	1.55	1.58	5.45	5.11	2.83

	Size portfolios						Idiosyncratic volatility portfolios					
	Hedge returns			t-stats			Hedge returns			t-stats		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
EW	2.52	1.75	1.15	11.31	6.81	4.56	0.95	1.94	2.66	6.30	10.37	11.68
VW	2.45	1.68	0.93	10.11	6.50	3.53	1.03	1.68	1.94	4.57	6.12	5.28

This table reports the average differences in monthly equal-weighted (EW) and value-weighted (VW) returns on top and bottom decile *AFER* portfolios and their t-statistics in various subsamples. To denote subsamples we use the following notation. Sample periods: (1) full sample (January 1970–December 2012), (2) early sample (January 1970–December 1995), (3) late sample (January 1996–December 2012); size portfolios: (1) microcap, (2) small, (3) large; idiosyncratic volatility portfolios: (1) low volatility, (2) medium volatility, (3) high volatility. In all panels *AFER* is constructed by averaging λ_t^a from all previous months. All returns are reported in percentage points.

the literature that it is more difficult to detect individual asset pricing anomalies in the recent years.

Table 5 also demonstrates that *AFER* has more power to predict returns on microcap stocks than on large stocks: it generates equal-weighted hedge returns of 2.52% per month on the sample of microcap stocks, but the returns drop to 1.15% on the sample of large stocks. A similar pattern is observed for value-weighted returns and t-statistics. Also, the dispersion of returns produced by *AFER* appears to be wider for high-volatility stocks (2.66% for equal-weighted returns) than for stocks with low volatility (0.95% for equal-weighted returns). These results are consistent with the studies that demonstrate the prevalence of anomalies in small and highly volatile stocks and they indicate that *AFER* inherits some properties of the characteristics from which it is constructed.

3. Alternative Aggregation Techniques

In this section we consider alternative techniques for the aggregation of information about expected stock returns and compare them with the proposed PLS-based approach. First, we use the factor analysis and PCA to extract common components from the characteristics and test their ability to predict stock returns. Second, we consider fitted values from the Fama-MacBeth regression of returns on characteristics as proxies for expected returns. Finally, we investigate whether the average percentile rank of stocks sorted with respect to multiple firm characteristics can forecast future returns.

Note that in contrast to PLS, the factor analysis, PCA, and Fama-MacBeth regression require the availability of all characteristics for each stock. However, many characteristic values are missing in the early part of the sample, and restricting the analysis to stocks for which all characteristics are available substantially reduces the sample size. The standard way to deal with incomplete observations is to impute missing values. Following the literature (e.g., Haugen and Baker 1996; Green, Hand, and Zhang 2016), we assign each missing characteristic its average value across all firms for which the characteristic is available in the same period. Also, we limit the sample to the period from January 1988 through December 2011, in which all 26 characteristics are available for at least a subsample of stocks.

3.1 Factor analysis

The objective of the factor analysis is to identify common sources of variation in a set of variables, so when applied to firm characteristics it can potentially reveal common information about future returns contained in them. However, the extracted factor is expected to be positively related to future returns only when individual characteristics that negatively predict returns are inverted, so the application of the factor analysis requires *ex ante* information on the relations between characteristics and returns. More importantly, because the factor analysis uses only the correlations among characteristics, it could reveal a factor that explains the commonality in characteristics but is silent about expected returns. In contrast, PLS by construction identifies only the common component in the characteristics that is related to future returns and ignores other common factors in the characteristic space. As a result, PLS and the factor analysis deliver different estimates for expected returns, and those provided by the factor analysis are likely to be poor predictors of returns unless the characteristics are described by a 1-factor model. This conclusion is illustrated by simulations conducted in the Online Appendix.

To compare the two approaches empirically, we repeat the analysis from Section 2.3, using the common factor in the characteristics at time t as a predictor for stock returns at time $t+1$. To estimate the factor values, we invert the signs of the characteristics that are negatively related to returns (all characteristics except B/M, E/P, C/P, MOM, ROA, and LV), find loadings $\lambda^a_{(f)t}$ of all characteristics X^a_{it} on a common factor, and then run a regression of characteristics X^a_{it} on the factor loadings.[19] The slopes in the second-step regression coincide with the factor values for stock i at time t. To ensure that the results of PLS and the factor analysis are comparable, we assume that there is only one factor in the characteristic space. Thus, the estimation of the factor values is similar to the estimation of μ_{it} in the PLS-based approach, but instead of the slopes λ^a_t from OLS regressions of returns on lagged individual characteristics, we use the loadings $\lambda^a_{(f)t}$ from the factor analysis in the second-step regressions. As in Section 2.3, we consider 4 specifications of the estimator: in the first specification only contemporaneously estimated loadings $\lambda^a_{(f)t}$ are

[19] The factor loadings are estimated using MLE. In a few months for which optimization yields a degenerate variance-covariance matrix of the factor model errors, the loadings are recorded as missing.

Table 6
Aggregation using alternative techniques

Panel A: factor analysis

	Hedge returns				*t*-stats			
	no	**5 years**	**10 years**	**all**	**no**	**5 years**	**10 years**	**all**
EW	0.23	0.83	0.90	0.91	0.45	2.22	2.30	2.42
VW	0.56	0.09	0.17	0.27	1.10	0.25	0.41	0.63

Panel B: principal component analysis (PCA)

	Hedge returns				*t*-stats			
	no	**5 years**	**10 years**	**all**	**no**	**5 years**	**10 years**	**all**
EW	−0.07	0.43	0.88	1.00	−0.17	1.23	2.94	3.27
VW	−0.64	0.04	0.27	0.36	−1.54	0.11	0.77	0.98

Panel C: Fama-MacBeth regression

	Hedge returns				*t*-stats			
	no	**5 years**	**10 years**	**all**	**no**	**5 years**	**10 years**	**all**
EW	1.68	1.64	1.65	1.79	3.54	6.01	6.22	6.89
VW	0.78	0.98	0.97	1.10	1.57	3.08	2.84	3.51

Panel D: average percentile rank

	Hedge returns	***t*-stats**
EW	1.36	7.02
VW	0.91	3.69

This table shows the differences in monthly equal-weighted (EW) and value-weighted (VW) returns on top- and bottom-decile portfolios formed by various alternative estimates of expected returns. In panels A and B the main factor in the characteristics and the first principal component of the characteristics are used as proxies for expected returns. In panel C the returns are predicted by fitted values from the Fama-MacBeth regression of returns on lagged characteristics. In panel D the average percentile rank in sorts with respect to multiple characteristics is used to predict returns. The columns in panels A–C indicate how $\lambda^a_{(f)t}$, $\lambda^a_{(PCA)t}$, and $\lambda_{(FM)t}$ are averaged over time (no averaging, over past 5 years, over past 10 years, and over all previous months). All returns are reported in percentage points. The sample is from January 1988 to December 2011 in panels A–C and from January 1970 to December 2012 in panel D.

used in the second-step regression and in the three others the estimates $\lambda^a_{(f)s}$, $s \leq t$, are averaged over particular time periods (past 5 years, past 10 years, or all previous months). Note that because the loadings of all characteristics are estimated simultaneously, they are sensitive to which characteristics are used in the estimation. As a result, the loadings in different time periods are comparable and can be averaged only if they are estimated from the same characteristics, and this explains why we limit the sample to the period in which all characteristics are available for at least a subsample of stocks.

Having estimated the factor values for all stocks in all periods, we construct equal- and value-weighted decile portfolios and, as in Section 2.3, compute average hedge returns and their *t*-statistics. The results are reported in panel A of Table 6. We observe that the common factor in the characteristics produces much smaller hedge returns than *AFER* does. Moreover, the results are sensitive to the weighting scheme: the hedge returns are positive, statistically significant in 3 out of 4 cases, and comparable to the hedge returns produced by individual characteristics for equal-weighted portfolios but are much smaller and statistically insignificant for value-weighted portfolios. Finally, we compute the differences in the hedge returns produced by the estimated factor and *AFER* and find that except for the case with no averaging of $\lambda^a_{(f)t}$ and λ^a_t, those differences are statistically significant. For example, when $\lambda^a_{(f)t}$ and λ^a_t are averaged over all previous months, the hedge returns on *AFER* and the factor are different with the t-statistics of 4.59 and 3.56 for equal- and value-weighted portfolios, respectively. Thus, our results confirm the superiority of the PLS-based approach over the factor analysis.

3.2 Principal component analysis (PCA)

An alternative way to extract common information from multiple variables is to use PCA, and next we investigate whether the first principal component of all firm characteristics can predict stock returns. The estimation procedure again contains two steps. In the first step, we use demeaned and standardized characteristics X^a_{it} to find the coefficients $\lambda^a_{(PCA)t}$, $a=1,\ldots,A$, of the first principal component. In the second step, we construct a potential predictor of returns as

$$\hat{\mu}_{(PCA)it}=\sum_{a=1}^{A}\left(\sum_{s\in\mathcal{T}}\lambda^a_{(PCA)s}\right)X^a_{it},$$

where for consistency with the other approaches, the coefficients $\lambda^a_{(PCA)s}$ are averaged over a particular time period $\mathcal{T}$ (past 5 years, past 10 years, or all previous months). As in the case of the factor analysis, we invert the signs of individual characteristics that negatively predict returns. Note that like the factor analysis, PCA aggregates characteristics using only covariances between them. Therefore, the first principal component is likely to be a poor predictor of returns in the presence of returns-unrelated common factors in the characteristic space. Kelly and Pruitt (2015) make the same argument in the time series context and demonstrate the superiority of 3PRF over PCA in the presence of irrelevant factors using Monte Carlo simulations.

To examine the performance of PCA empirically, we sort stocks on the variable $\hat{\mu}_{(PCA)it}$ every month, construct equal- and value-weighted decile portfolios, and compute statistics of hedge returns. Panel B of Table 6 demonstrates that the first principal component is a poor predictor of future returns: the value-weighted hedge returns, as well as equal-weighted hedge returns, when the averaging window for $\lambda^a_{(PCA)t}$ is short are statistically indistinguishable from zero. In the two cases when the hypothesis of zero hedge returns is rejected, the returns are much smaller than those produced by *AFER*. Formally, the hypothesis that the hedge returns generated by PCA and PLS are equal is reliably rejected in all specifications. For example, when $\lambda^a_{(PCA)t}$ and λ^a_t are averaged over all previous months, the corresponding t-statistics are 4.57 and 3.10 for equal- and value-weighted portfolios, respectively. Thus, PCA indeed underperforms PLS. This result implies that the characteristics share common returns-unrelated factors that contaminate the first principal component and hamper its ability to predict returns but that leave the quality of the PLS estimates unaffected.

3.3 Fama-MacBeth regression

A common approach to estimating expected returns is to use fitted values from the Fama-MacBeth regression of realized returns on multiple lagged firm characteristics (e.g., Haugen and Baker 1996; Chan, Dimmock, and Lakonishok 2009; Lewellen 2015). This approach also has several shortcomings compared with PLS. First, the regression-based estimates of expected returns are likely to be less efficient than the PLS-based estimates because PLS assumes that there is only 1 factor in the characteristic space that is related to expected returns, and this assumption puts additional constraints on the estimation. Those constraints for the case of 3PRF are explicitly stated by Kelly and Pruitt (2015), who show how 3PRF can be derived as a solution to a constrained least-squares problem.

Second, the regression estimates of expected returns are imprecise and even nonexistent when the number of characteristics is large. The coefficients $\lambda_{(FM)t} = (\frac{1}{N} X'_{t-1} X_{t-1})^{-1} (\frac{1}{N} X'_{t-1} R_t)$ of cross-sectional OLS regressions require the computation of the sample variance-covariance matrices of all regressors (the term $\frac{1}{N} X'_{t-1} X_{t-1}$), whose size grows with the number of characteristics as A^2. When A is large, it becomes an imprecise estimate of the population matrix and even non-invertible when there are more characteristics than stocks.[20] In contrast, PLS requires the estimation of the loadings δ_t^a, not the variance-covariance matrices of all characteristic. As a result, it outperforms the Fama-MacBeth regression when the number of characteristics is large. This conclusion is illustrated in the Online Appendix, which reports simulated performance of both methods for different numbers of stocks and characteristics.

Third, the Fama-MacBeth regression can suffer from the multicollinearity problem. This is another detrimental effect of the term $\frac{1}{N} X'_{t-1} X_{t-1}$. Therefore, a researcher should make a judgement about which of the highly correlated characteristics is the most informative about returns and include only that characteristic in the regression. PLS does not have this limitation and can simultaneously handle multiple highly correlated signals about returns.

Along with these shortcomings, the Fama-MacBeth regression has one advantage over the PLS-based aggregation. Because it does not rely on the factor structure of characteristics, it is more robust and may deliver consistent estimates of expected returns when the assumptions of PLS are violated. In particular, the PLS-based aggregation can be inferior to aggregation by the Fama-MacBeth regression when individual characteristics contain information about future returns that cannot be summarized by the common component μ_{it}.

[20] The techniques designed to deal with a large number of regressors include sliced inverse regression (e.g., Li 1991), reduced rank regression (e.g., Anderson 1951), and sparse regression models (e.g., Huang, Horowitz, and Ma 2008; Belloni, Chernozhukov, and Hansen 2013). The methods of forecasting a single time series when there are many predictors are reviewed by Stock and Watson (2006).

To compare the empirical performances of these two approaches, we run the Fama-MacBeth regression of returns on lagged characteristics in each month and construct estimates of expected returns as fitted values. As in Section 2.3, we consider 4 types of fitted values. In the first specification, only contemporaneous regression slopes $\lambda_{(FM)t}$ are used for computing fitted values, and in the 3 others, the regression slopes are averaged over particular time periods (past 5 years, past 10 years, or all previous months). The means and t-statistics of hedge returns on decile portfolios formed on the fitted values are presented in panel C of Table 6.

Our results show that the Fama-MacBeth regression provides better proxies for expected returns than the factor analysis and PCA. In particular, the hedge returns on the fitted values appear to be close to the equal-weighted hedge returns produced by *AFER* without averaging of λ_t^a. Nevertheless, the hedge returns tend to be smaller than those on *AFER* in all specifications for value-weighted portfolios, and all t-statistics are lower than their analogs from Table 2. Thus, the Fama-MacBeth regression does not produce better estimates of expected returns than does the PLS-based approach. This is another piece of evidence in favor of our assumption that the characteristics share a common returns-related component, which is estimated by PLS.

3.4 Rank-based approach

Finally, we compare PLS with a simple heuristic alternative approach that is based on the average ranks of stocks with respect to multiple characteristics (e.g., Stambaugh, Yu, and Yuan 2015). Specifically, in each month t, we sort stocks by each characteristic and find their percentile ranks defined for each stock as the rank divided by the number of stocks for which the given characteristic is available. To ensure that high ranks of stocks correspond to high returns, we reverse the order of stocks in the sorts with respect to those characteristics that are negatively related to returns. The average of the percentile ranks of a stock with respect to each characteristic is another proxy for its expected returns. Note that in contrast to the other considered alternative aggregation techniques, this approach works well even when only a subset of characteristics is available for a stock. Therefore, our sample period is from January 1970 to December 2012, and we do not impute missing characteristic values.[21]

The rank-based approach has its own shortcomings. First, similar to the factor analysis and PCA, the average rank reveals a common component in the characteristics that could explain their correlation (high average rank implies high individual characteristic ranks of the particular stock) but that may be unrelated to future returns. Second, the average rank is likely to provide inefficient estimates of expected returns because it ignores information from

[21] The imputation would shift the average percentile ranks toward the median and weaken their forecasting power. Our unreported results confirm this conclusion.

characteristic values and effectively presumes that only ranks of stocks are related to future returns. Third, the averaging of ranks implies that there is *ex ante* information on whether the relation between each characteristic and returns is positive or negative.[22]

Panel D of Table 6 reports the means and *t*-statistics of hedge returns produced by average percentile ranks. Note that in contrast to other aggregation techniques discussed above, the average percentile rank method implies that all characteristics equally contribute to expected returns in all time periods, so the weight of ranks with respect to each characteristic is $1/A$, where A is the number of characteristics. Thus, the method cannot be adjusted to include the time dimension in the same way as we adjust the PLS, PCA, factor analysis, and OLS regression approaches. As a result, we report only 1 set of hedge returns, not 4 sets as for the other aggregation techniques.

Overall, the average percentile ranks produce a statistically significant dispersion of returns. However, the hedge returns are lower than those on *AFER* even without the averaging of λ_t^a (1.36% instead of 1.69% for equal-weighted portfolios and 0.91% instead of 1.21% for value-weighted portfolios). As demonstrated in Table 2, the possibility to estimate the relation between each characteristic and returns more precisely by averaging λ_t^a from previous periods further increases the advantage of PLS: the hedge returns produced by *AFER* with λ_t^a averaged over all previous months are 2.06% (1.56%) for equal-weighted (value-weighted) portfolios. They are statistically different from hedge returns produced by the average percentile rank with *t*-statistics of 7.09 and 3.79, respectively. Thus, PLS dominates the average percentile rank approach not only theoretically but also empirically.

3.5 Large-cap stocks

Finally, we demonstrate that the PLS-based aggregation technique outperforms alternatives even when the sample contains only the most liquid large-cap stocks. In each month, we estimate expected returns by the same methods (PLS, PCA, factor analysis, Fama-MacBeth regression, and average percentile ranks) but using only stocks in the top 10% of market capitalization. Note that in contrast to Section 2.5, in which the hedge returns are computed for various subsamples of stocks but all stocks are used to construct *AFER*, now a subsample of stocks is used both to construct *AFER* and to test its relation to future returns.[23] The hedge returns produced by the five techniques in this subsample are reported in Table 7. To save space, for the first four techniques we consider only specifications without time series averaging of the Step 1 coefficients and with averaging them over all previous months.

[22] Novy-Marx (2016) demonstrates the existence of the overfitting bias that results from signing characteristics so that each of them predicts positive in-sample returns.

[23] The sample of stocks from the top 10% of market capitalization is smaller than the sample of large stocks in Section 2.5. The former contains on average 296 stocks, whereas the latter contains on average 795 stocks.

Table 7
PLS and alternative aggregation techniques in the subsample of large-cap stocks

	PLS(no)	PLS(all)	FA(no)	FA(all)	PCA(no)	PCA(all)	FM(no)	FM(all)	APR
Hedge returns									
EW	1.00	1.03	−0.09	−0.06	−0.07	0.24	0.82	0.45	0.77
VW	0.87	0.71	0.02	−0.26	−0.22	0.26	0.63	0.40	0.65
***t*-stats**									
	PLS(no)	**PLS(all)**	**FA(no)**	**FA(all)**	**PCA(no)**	**PCA(all)**	**FM(no)**	**FM(all)**	**APR**
EW	2.83	3.65	−0.15	−0.12	−0.14	0.49	1.83	1.35	3.06
VW	2.71	2.88	0.04	−0.53	−0.44	0.61	1.42	1.13	2.68

This table shows the differences in monthly equal-weighted (EW) and value-weighted (VW) returns on top- and bottom-decile portfolios formed by various estimates of expected returns using only stocks in the top 10% of market capitalization. PLS, FA, PCA, FM, and APR stand for estimates of expected returns obtained using PLS, factor analysis, PCA, Fama-MacBeth regression, and average percentile ranks, respectively. The indicator in brackets shows whether λ_t^a, $\lambda_{(f)t}^a$, $\lambda_{(PCA)t}^a$, and $\lambda_{(FM)t}$ are not averaged at all (no) or averaged over all previous months (all). All returns are reported in percentage points. The sample is from January 1988 to December 2011 for FA, PCA, and FM, and from January 1970 to December 2012 for PLS and APR.

Table 7 provides several observations. First, both equal-weighted and value-weighted hedge returns produced by all techniques are smaller than those reported in Table 6. In particular, the factor analysis, PCA, and Fama-MacBeth regression fail to produce a statistically significant dispersion of realized returns. This result is consistent with the fact that many individual anomalies are weaker on the large-cap stocks and it echoes the subsample results from Table 5. Second, both versions of *AFER* still predict future returns: even without averaging λ_t^a the hedge returns are 1% per month (0.87% per month) for equal-weighted (value-weighted) portfolios, which are statistically different from zero. The average percentile rank is also capable of delivering statistically significant hedge returns in this subsample, but they are lower than those produced by *AFER*. Thus, PLS again dominates the alternatives. One more interesting result of Table 7 is that the averaging of the Step 1 estimates λ_t^a does not help to improve the quality of the estimator. It can be explained by less stable relations between characteristics and expected returns in the subsample of large stocks.[24] Thus, the averaging of λ_t^a only contaminates its most recent and relevant values with noise.

4. Conclusion

The analysis of asset pricing anomalies helps to understand the cross section of expected stock returns. However, it is complicated by the large variety of firm characteristics that can predict returns. In this paper, we propose a simple, powerful, and easily implementable PLS-based approach to aggregating information about returns contained in those characteristics. Thus, our paper

[24] The average cross-sectional correlation between the PLS Step 1 estimates λ_t^a and λ_{t-1}^a is 0.16 for all stocks but only 0.07 for the large cap stocks.

is a step toward addressing the challenge posed by the existence of multiple anomalies and highlighted by John Cochrane in his AFA 2011 Presidential Address (Cochrane 2011), where he argues that to address many questions on the relation between firm characteristics and expected stock returns "… in the zoo of new variables … we will have to use different methods."

Our paper makes both methodological and empirical contributions. On the methodological side, we propose a new approach to the aggregation of information from a large number of firm characteristics that explicitly exploits the idea that characteristics can be viewed as signals about unobservable expected returns. On the empirical side, we find that the hedge returns produced by expected returns filtered from 26 characteristics are large, highly statistically significant, and exceed the hedge returns produced by each characteristic individually. This result suggests that there is commonality in the considered asset pricing anomalies, and individual characteristics are likely to be diverse signals about expected returns (otherwise, the hedge returns on our estimates would be comparable to those on individual characteristics).

We apply our aggregation technique to the U.S. market. However, owing to its numerous advantages such as efficient use of characteristics when their number is comparable to the number of stocks, robustness to missing observations, and ability to handle highly correlated characteristics, the PLS-based approach may be particularly valuable for the analysis of foreign markets with a small number of stocks and scarce information about firm fundamentals. The application of our methodology to other markets is an interesting direction for future research.

Appendix A. Proof of Proposition 1

The cross-sectional standardization of the characteristics implies that $\overline{Var}(X^a_{it-1})=1$ for all a, $a=1,\dots,A$. Hence, the slopes in the cross-sectional regressions of R_{it} on X^a_{it-1} obtained at Step 1 are

$$\lambda^a_t = \frac{\overline{Cov}(R_{it}, X_{it}-1^a)}{\overline{Var}(X^a_{it-1})} = \overline{Cov}(R_{it}, X^a_{it-1}).$$

Taking the first limit $N\to\infty$, we get

$$\overline{Cov}(R_{it}, X^a_{it-1}) = \overline{Cov}(\mu_{it-1}+\varepsilon_{it}, \delta^a_{t-1}(\mu_{it-1}-\bar{\mu}_{t-1})+u^a_{it-1}) \xrightarrow{p} \delta^a_{t-1}V_{t-1},$$

where we use Assumptions 1 and 3 along with the independence of ε_{it} from all variables available at time $t-1$.

At Step 2, the characteristics X^a_{it} are regressed on λ^a_t in the characteristic space for each stock, and the obtained slopes are

$$\hat{\mu}_{it} = \frac{\widetilde{Cov}(X^a_{it}, \lambda^a_t)}{\widetilde{Var}(\lambda^a_t)}.$$

Using the Slutsky's theorem,

$$\widetilde{Cov}(X^a_{it}, \lambda^a_t) = \widetilde{Cov}(\delta^a_t(\mu_{it}-\bar{\mu}_t)+u^a_{it}, \lambda^a_t) \xrightarrow[N\to\infty]{p} \widetilde{Cov}(\delta^a_t(\mu_{it}-\mu_t)+u^a_{it}, \delta^a_{t-1}V_{t-1})$$
$$= (\widetilde{Cov}(\delta^a_{t-1}, \delta^a_t)(\mu_{it}-\mu_t)+\widetilde{Cov}(u^a_{it}, \delta^a_{t-1}))V_{t-1},$$

$$\widetilde{Var}(\lambda_t^a) \xrightarrow[N\to\infty]{p} \widetilde{Var}(\delta_{t-1}^a V_{t-1}) = \widetilde{Var}(\delta_{t-1}^a) V_{t-1}^2.$$

Taking the second limit $A \to \infty$ and applying the rules for probability limits, we get

$$\operatorname*{plim}_{A\to\infty} \operatorname*{plim}_{N\to\infty} \hat{\mu}_{it} = \operatorname*{plim}_{A\to\infty} \operatorname*{plim}_{N\to\infty} \frac{\widetilde{Cov}(X_{it}^a, \lambda_t^a)}{\widetilde{Var}(\lambda_t^a)}$$

$$= \operatorname*{plim}_{A\to\infty} \frac{\widetilde{Cov}(\delta_{t-1}^a, \delta_t^a)(\mu_{it} - \mu_t) + \widetilde{Cov}(u_{it}^a, \delta_{t-1}^a)}{\widetilde{Var}(\delta_{t-1}^a) V_{t-1}} = \frac{\Lambda_{t-1,t}}{\Lambda_{t-1,t-1} V_{t-1}} (\mu_{it} - \mu_t).$$

The computation of the last limit uses Assumptions 2 and 4. After denoting the scalar factor $\Lambda_{t-1,t}/(\Lambda_{t-1,t-1}V_{t-1})$ as F_t, we get the statement of the proposition.

Appendix B. Characteristics

This Appendix describes the characteristics used in our analysis.

Size (*S*). Banz (1981) and Reinganum (1981) were the first to document that small stocks have higher returns than large stocks do. Following Fama and French (1992), S is defined as a logarithm of market capitalization. The latter is the product of the share price and the number of shares outstanding at the end of the previous month.

Book-to-market ratio (*B/M*). The value anomaly is a tendency of value stocks characterized by high ratios of fundamentals to price to have on average higher returns than growth stocks, i.e., the stocks with low fundamentals-to-price ratios (e.g., Basu 1977; Rosenberg, Reid, and Lanstein 1985). The book-to-market ratio B/M is one of the value indicators, and following Fama and French (1992), we compute it as a logarithm of the ratio of Book Value over Market Value. Market Value is equal to $CSHO \times PRCH\_C$, where $CSHO$ is the number of shares outstanding and $PRCH\_C$ is the stock price at the end of the calendar year $t-1$. Book Value is defined as $SEQ-(PSTKL$, or $PSTKRV$, or $PSTK$ in this order of availability) + $TXDITC$ (if not missing), where SEQ is stockholders' equity, $PSTKL$ is the preferred stock liquidating value, $PSTKRV$ is the preferred stock redemption value, $PSTK$ is the preferred stock par value, and $TXDITC$ is the balance sheet deferred taxes and investment tax credit. If SEQ is missing, $CEQ+PSTK$ is used, where CEQ is common equity. If all variables above are missing, $AT-LT$ is used, where AT is total assets, and LT is total liabilities.

Earnings-to-price ratio (*E/P*). This is another value indicator. Following Basu (1977, 1983), we define E/P as

$$E/P = \frac{IB}{ME},$$

where IB is the total earnings before extraordinary items and ME is the market value of equity. The latter is the product of the share price and the number of shares outstanding at the end of the previous year. Only firms with positive earnings are included in the sample.

Cash flow-to-price ratio (*C/P*). This is one more value indicator. Following Ball (1978), we define C/P as

$$C/P = \frac{IB+EDP+TXDI}{ME},$$

where IB is the total earnings before extraordinary items, EDP is the equity's share of depreciation, $TXDI$ is the deferred taxes (if available), and ME is the market value of equity described in the definition of E/P. $EDP = \frac{ME}{ME+AT-BE} DP$, where DP is the depreciation and amortization, AT is the total assets, and BE is the book value of equity, described in the definition of B/M. Only firms with positive earnings are included in the sample.

Market beta ($\boldsymbol{B}$). Following Black, Jensen, and Scholes (1972), B is defined as the slope in the daily time series regression

$$R_t - R_t^f = \alpha + \beta(R_t^M - R_t^f) + \varepsilon_t,$$

where R_t is the stock return, R_t^f is the short-term interest rate, and R_t^M is the market factor. B is estimated using daily data from the previous year for firms with at least 200 non-missing observations.

Momentum ($\boldsymbol{MOM}$). Jegadeesh and Titman (1993) document the ability of stocks with relatively high past returns (winners) to outperform stocks with relatively low past returns (losers). We construct the sorting variable MOM_t as a cumulative return over the previous thirteen months excluding the last month, that is,

$$MOM_t = \prod_{s=t-13}^{t-1} (1+R_s),$$

where R_s is the stock return in month s. MOM_t is recalculated for each stock every month. For consistency with the other characteristics, we deviate from Jegadeesh and Titman (1993) and interpret MOM_t as an anomalous characteristic: every month stocks are sorted on the most recent MOM_t, and all portfolios are rebalanced (in Jegadeesh and Titman (1993) only a part of the portfolio is revised every month).

Long-term reversal ($\boldsymbol{LTR}$). DeBondt and Thaler (1985) document that stock returns demonstrate a mean-reverting property at horizons longer than 1 year. Following this paper,

$$LTR_t = \prod_{s=t-61}^{t-14} (1+R_s),$$

where R_s is the stock return in month s. Months $t-13$ through $t-2$ are excluded to separate the long-term reversal effect from the Jegadeesh and Titman (1993) momentum effect.

Short-term reversal ($\boldsymbol{STR}$). Following Jegadeesh (1990), we define STR for each stock as the stock return in the previous month.

Idiosyncratic volatility ($\boldsymbol{IdVol}$). Ang et al. (2006) find that stocks with high idiosyncratic volatility have low future returns. Following this paper, we define *IdVol* as the standard deviation of residuals in the regression of daily CRSP returns on the daily realizations of the Fama-French three factors. In month t the idiosyncratic volatility is computed using daily data for the previous month, so *IdVol* is updated every month.

Maximum daily return over the past month ($\boldsymbol{MAX}$). This characteristic is suggested by Bali, Cakici, and Whitelaw (2011), who find that it negatively predicts returns.

Expected idiosyncratic skewness ($\boldsymbol{EIS}$). Boyer, Mitton, and Vorkink (2010) document that stocks with high expected idiosyncratic skewness have lower expected returns than stocks with low expected idiosyncratic skewness. We use data on expected idiosyncratic skewness provided on Brian Boyer's website (http://marriottschool.net/emp/boyer/Research/skewdata.html).

Return on equity ($\boldsymbol{ROE}$). The relation between this profitability measure and future returns is discussed in Haugen and Baker (1996), Fama and French (2006), and Fama and French (2008), among others. Following Fama and French (2008), we define *ROE* as

$$ROE = \frac{IB - DVP + TXDI}{BE},$$

where *IB* is the total earnings before extraordinary items, *DVP* is the preferred dividends (if available), *TXDI* is the deferred taxes (if available), and *BE* is the book value of equity, described in the definition of B/M. Only firms with positive earnings are included in the sample.

Returns on assets (ROA). This is another profitability measure whose relation to future returns is considered in Balakrishnan, Bartov, and Faurel (2010), Wang and Yu (2013), and Novy-Marx (2013). We define ROA_t as

$$ROA_t = \frac{IBQ_{t-1}}{ATQ_{t-2}},$$

where IBQ_t is income before extraordinary items in quarter t, and ATQ_t is total assets in quarter t. This characteristic is available starting from year 1975.

Total asset growth (AG). Cooper, Gulen, and Schill (2008) find a strong negative relation between firm's asset growth and its future stock returns. Following this paper, AG is defined as

$$AG_t = \frac{AT_{t-1} - AT_{t-2}}{AT_{t-2}},$$

where AT_{t-1} is total assets in the fiscal year ending in calendar year $t-1$. To reduce the effect of erroneous data entries, we winsorize the characteristic at the 1% level.

Abnormal capital investments (CI). The investments anomaly is the tendency of firms that recently experienced high capital investments to have low future stock returns. It is documented in Titman, Wei, and Xie (2004) and further explored in Polk and Sapienza (2009). Following Titman, Wei, and Xie (2004), we define abnormal capital investments as

$$CI_t = \frac{CE_{t-1}}{(CE_{t-2} + CE_{t-3} + CE_{t-4})/3} - 1,$$

where CE_t is the firm's capital expenditures scaled by its sales, that is, $CE_t = CAPX_t / SALE_t$. Similar to total asset growth, we winsorize abnormal capital investments at the 1% level.

Investment growth (IG). An alternative measure of investment growth is the change in the firm's capital expenditures $CAPX$. Following Xing (2008), we define IG as

$$IG_t = \frac{CAPX_t - CAPX_{t-1}}{CAPX_{t-1}}.$$

Investment-to-capital ratio (I/K). Another investments-based firm characteristic related to returns is the investment-to-capital ratio. Following Xing (2008), we define it as

$$I/K = \frac{CAPX}{PPENT},$$

where $CAPX$ is the capital expenditures and $PPENT$ is the property, plant, and equipment.

Investment-to-assets ratio (I/A). This measure of investments is proposed by Lyandres, Sun, and Zhang (2008). We define I/A as

$$(I/A)_t = \frac{INVT_{t-1} - INVT_{t-2} + PPEGT_{t-1} - PPEGT_{t-2}}{AT_{t-2}},$$

where $INVT_t$ is inventories, $PPEGT_t$ is gross property, plant, and equipment, and AT_{t-2} is total assets in fiscal year ending in calendar year $t-2$.

Accruals (ACC). Sloan (1996) discovers that firms with high accruals have lower future returns. Following this study, we define accruals as

$$ACC_t = \frac{(\Delta ACT_{t-1} - \Delta CHE_{t-1}) - (\Delta LCT_{t-1} - \Delta DLC_{t-1} - \Delta TXP_{t-1}) - DP_{t-1}}{AT_{t-2}},$$

where ACT_t is total current assets, CHE_t is cash and short-term investments, LCT_t is total current liabilities, DLC_t is debt in current liabilities, TXP_t is income taxes payable, DP_t is depreciation and amortization, AT_t is total assets. Each accounting variable X_t is from the annual report with the fiscal year ending in calendar year t and $\Delta X_t = X_t - X_{t-1}$.

Net operating assets (*NOA*). This characteristic is suggested by Hirshleifer et al. (2004), who find that it is negatively related to future stock returns. Following Hirshleifer et al. (2004), we define *NOA* as

$$NOA_t = \frac{\text{Operating Assets}_{t-1} - \text{Operating Liabilities}_{t-1}}{AT_{t-2}},$$

where

$$\text{Operating Assets}_t = AT_t - CHE_t,$$

$$\text{Operating Liabilities}_t = AT_t - DLC_t - DLTT_t - MIB_t - PSTK_t - CEQ_t.$$

Here AT_t is total assets, CHE_t is cash and short-term investments, DLC_t is debt in current liabilities, $DLTT_t$ is total long-term debt, MIB_t is minority interest, $PSTK_t$ is preferred stock, CEQ_t is total common equity.

Net stock issues (*NS*). This anomaly is highlighted in Fama and French (2008) and Pontiff and Woodgate (2008), who document that the change in the number of shares outstanding is negatively related to future stock returns. *NS* is defined as

$$NS_t = \log\left(\frac{SASO_{t-1}}{SASO_{t-2}}\right),$$

where $SASO_t$ is the split-adjusted shares outstanding from the fiscal year-end in year t computed as $SASO_t = CSHO_t \times AJEX_t$, where $CSHO_t$ is the shares outstanding and $AJEX_t$ is the cumulative adjustment factor. The characteristic is defined only for firms with non-zero values of *NS*.

Composite stock issuance (ι). This anomaly is documented in Daniel and Titman (2006). Composite stock issuance is defined as

$$\iota_t = \log\left(\frac{ME_{t-1}}{ME_{t-6}}\right) - r(t-6, t-1),$$

where ME_t is the market value of equity at the end of calendar year t and $r(t-\tau, t)$ is the cumulative log return between the last trading day of calendar year $t-\tau$ and the last trading day of calendar year t.

Leverage (*LV*). Bhandari (1988) finds that firms with high leverage have high returns. Following this paper, we measure leverage as

$$LV = \frac{AT - CEQ}{ME},$$

where *AT* and *CEQ* are book values of total assets and common equity, respectively, and *ME* is the market value of equity. ME is computed using monthly data corresponding to the end of the annual accounting period. We accommodate changes to the accounting period by allowing the data to be more than 12 months old, but we eliminate all firms whose most recent data is older than 18 months.

***O*-score**. *O*-score measures the likelihood of financial distress. Several studies (e.g., Dichev 1998; Griffin and Lemmon 2002; Campbell, Hilscher, and Szilagyi 2008) show that firms in financial distress have low stock returns. *O*-score is calculated using the model of Ohlson (1980) who defines it as

$$\begin{aligned} O\text{-}score_t = & -1.32 - 0.407\log(Size_t) + 6.03\,TLTA_t - 1.43\,WCTA_t + 0.076\,CLCA_t \\ & - 1.72\,OENEG_t - 2.37\,NITA_t - 1.83\,FUTL_t + 0.285\,INTWO_t - 0.521\,CHIN_t. \end{aligned}$$

The inputs of the formula are as follows. $Size_t = ATQ_t / CPI_t$ is total assets adjusted for inflation, where ATQ_t is total assets, and CPI_t is the consumer price index from the U.S. Bureau of Labor

Statistics. $TLTA_t = (DLCQ_t + DLTTQ_t)/ATQ_{t-1}$ is total liabilities divided by lagged total assets, where $DLCQ_t$ is debt in current liabilities, and $DLTTQ_t$ is total long-term debt. $WCTA_t = (ACTQ_t - LCTQ_t)/ATQ_{t-1}$ is working capital divided by lagged total assets, where $ACTQ_t$ is current assets, and $LCTQ_t$ is current liabilities. $CLCA_t = LCTQ_t/ACTQ_t$ is the ratio of current liabilities and current assets. $OENEG_t$ is 1 if total liabilities LTQ_t exceeds total assets and 0 otherwise. $NITA_t = NIQ_t/ATQ_{t-1}$, where NIQ_t is net income. $FUTL_t = PIQ_t/LTQ_{t-1}$, where PIQ_t is pretax income. $INTWO_t$ is 1 if net income NIQ_t was negative in the last 2 years and 0 otherwise. $CHIN_t = (NIQ_t - NIQ_{t-1})/(|NIQ_t| + |NIQ_{t-1}|)$ is level-adjusted change in net income. Each accounting variable X_t is from the quarterly report with the fiscal quarter ending in calendar month t, and portfolios are constructed at the end of month $t+3$ to allow for the late filing.

Turnover (TO). Datar, Naik, and Radcliffe (1998) and Lee and Swaminathan (2000) show that low turnover stocks earn higher returns than high turnover stocks. TO is defined as the moving average of daily turnover over the past 3–12 months, where daily turnover is the ratio of the number of shares traded each day to the number of shares outstanding at the end of the day. Following Lee and Swaminathan (2000), we compute turnover only for the NYSE and AMEX stocks.

Analysts' forecasts dispersion (D). Diether, Malloy, and Scherbina (2002) demonstrate that firms with higher dispersion in analysts' forecasts of future earnings have lower future returns. D is defined as the standard deviation of I/B/E/S next quarter earnings forecasts divided by the mean earnings forecast.

References

Abarbanell, J. S., and B. J. Bushee. 1998. Abnormal returns to a fundamental analysis strategy. *Accounting Review* 73:19–45.

Acharya, V. V., and L. H. Pedersen. 2005. Asset pricing with liquidity risk. *Journal of Financial Economics* 77:375–410.

Ali, A., L.-S. Hwang, and M. A. Trombley. 2003. Arbitrage risk and the book-to-market anomaly. *Journal of Financial Economics* 69:355–73.

Anderson, T. W. 1951. Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Annals of Mathematical Statistics* 22:327–51.

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang. 2006. The cross-section of volatility and expected returns. *Journal of Finance* 61:259–99.

Ang, A., and D. Kristensen. 2012. Testing conditional factor models. *Journal of Financial Economics* 106:132–56.

Ang, A., J. Liu, and K. Schwarz. 2011. Using stocks or portfolios in tests of factor models. Working Paper, Columbia Business School.

Asness, C. S., A. Frazzini, and L. H. Pedersen. 2014. Quality minus Junk. Working Paper, Copenhagen Business School.

Avramov, D., and T. Chordia. 2006. Asset pricing models and financial market anomalies. *Review of Financial Studies* 19:1001–40.

Avramov, D., T. Chordia, G. Jostova, and A. Philipov. 2013. Anomalies and financial distress. *Journal of Financial Economics* 108:139–59.

Babenko, I., O. Boguth, and Y. Tserlukevich. 2016. Idiosyncratic cash flows and systematic risk. *Journal of Finance* 71:425–56.

Balakrishnan, K., E. Bartov, and L. Faurel. 2010. Post loss/profit announcement drift. *Journal of Accounting and Economics* 50:20–41.

Bali, T. G., N. Cakici, and R. F. Whitelaw. 2011. Maxing out: Stocks as lotteries and the cross-section of expected returns. *Journal of Financial Economics* 99:427–46.

Balke, N. S., and M. E. Wohar. 2002. Low-frequency movements in stock prices: A state-space decomposition. *Review of Economics and Statistics* 84:649–67.

Ball, R. 1978. Anomalies in relationships between securities' yields and yield-surrogates. *Journal of Financial Economics* 6:103–26.

Banz, R. W. 1981. The relationship between return and market value of common stocks. *Journal of Financial Economics* 9:3–18.

Basu, S. 1977. Investment performance of common stocks in relation to their price-earnings ratios: A test of the Efficient Market Hypothesis. *Journal of Finance* 32:663–82.

Basu, S. 1983. The relationship between earnings' yield, market value and return for NYSE common stocks: Further evidence. *Journal of Financial Economics* 12:129–56.

Belloni, A., V. Chernozhukov, and C. Hansen. 2013. Inference for high-dimensional sparse econometric models. In *Advances in economics and econometrics, 10th World Congress of Econometric Society*, vol. 3, ed. D. Acemoglu, M. Arellano, and E. Dekel. 245–95. Cambridge: Cambridge University Press.

Berk, J. B. 1995. A critique of size-related anomalies. *Review of Financial Studies* 8:275–86.

Berk, J. B., R. C. Green, and V. Naik. 1999. Optimal investment, growth options, and security returns. *Journal of Finance* 54:1553–607.

Bhandari, L. C. 1988. Debt/equity ratio and expected common stock returns: Empirical evidence. *Journal of Finance* 43:507–28.

Binsbergen, J. H., and R. S. J. Koijen. 2010. Predictive regressions: A present-value approach. *Journal of Finance* 65:1439–71.

Black, F., M. C. Jensen, and M. Scholes. 1972. The Capital Asset Pricing Model: Some empirical tests. In *Studies in the theory of capital markets*, ed. M. C. Jensen. 79–121. New York: Praeger.

Boyer, B., T. Mitton, and K. Vorkink. 2010. Expected idiosyncratic skewness. *Review of Financial Studies* 23:169–202.

Brandt, M. W., and Q. Kang. 2004. On the relationship between the conditional mean and volatility of stock returns: A latent VAR approach. *Journal of Financial Economics* 72:217–57.

Brennan, M. J., T. Chordia, and A. Subrahmanyam. 1998. Alternative factor specifications, security characteristics, and the cross-section of expected stock returns. *Journal of Financial Economics* 49:345–73.

Campbell, J. Y., J. Hilscher, and J. Szilagyi. 2008. In search of distress risk. *Journal of Finance* 63:2899–939.

Cao, J., and B. Han. 2016. Idiosyncratic risk, costly arbitrage, and the cross-section of stock returns. *Journal of Banking and Finance*, forthcoming.

Carlson, M., A. Fisher, and R. Giammarino. 2004. Corporate investment and asset price dynamics: Implications for the cross-section of returns. *Journal of Finance* 59:2577–603.

Chan, L. K. C., S. G. Dimmock, and J. Lakonishok. 2009. Benchmarking money manager performance: Issues and evidence. *Review of Financial Studies* 22:4553–99.

Chordia, T., A. Subrahmanyam, and Q. Tong. 2014. Have capital market anomalies attenuated in the recent era of high liquidity and trading activity? *Journal of Accounting and Economics* 58:41–58.

Cochrane, J. H. 1996. A cross-sectional test of an investment-based asset pricing model. *Journal of Political Economy* 104:572–621.

Cochrane, J. H. 2005. *Asset pricing*. Princeton: Princeton University Press.

———. 2011. Presidential address: Discount rates. *Journal of Finance* 66:1047–108.

Conrad, J., and G. Kaul. 1988. Time-variation in expected returns. *Journal of Business* 61:409–25.

Cooper, M. J., H. Gulen, and M. J. Schill. 2008. Asset growth and the cross-section of stock returns. *Journal of Finance* 63:1609–51.

Daniel, K., and S. Titman. 2006. Market reactions to tangible and intangible information. *Journal of Finance* 61:1605–43.

Datar, V. T., N. Y. Naik, and R. Radcliffe. 1998. Liquidity and stock returns: An alternative test. *Journal of Financial Markets* 1:203–19.

DeBondt, W., and R. Thaler. 1985. Does the stock market overreact? *Journal of Finance* 40:793–805.

Dichev, I. D. 1998. Is the risk of bankruptcy a systematic risk? *Journal of Finance* 53:1131–47.

Diether, K. B., C. J. Malloy, and A. Scherbina. 2002. Differences of opinion and the cross section of stock returns. *Journal of Finance* 57:2113–41.

Fama, E. F., and K. R. French. 1992. The cross-section of expected stock returns. *Journal of Finance* 47:427–65.

———. 2006. Profitability, investment and average returns. *Journal of Financial Economics* 82:491–518.

———. 2008. Dissecting anomalies. *Journal of Finance* 63:1653–78.

Fama, E. F., and J. D. MacBeth. 1973. Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy* 81:607–36.

Ferson, W. E. 1990. Are the latent variables in time-varying expected returns compensation for consumption risk? *Journal of Finance* 45:397–429.

Ferson, W. E., S. R. Foerster, and D. B. Keim. 1993. General tests of latent variable models and mean-variance spanning. *Journal of Finance* 48:131–56.

Ferson, W. E., and C. R. Harvey. 1991. The variation of economic risk premiums. *Journal of Political Economy* 99:385–415.

Frank, I. E., and J. H. Friedman. 1993. A statistical view of some chemometrics regression tools. *Technometrics* 35:109–35.

Garlappi, L., and H. Yan. 2011. Financial distress and the cross-section of equity returns. *Journal of Finance* 66:789–822.

Gârleanu, N., and L. H. Pedersen. 2011. Margin-based asset pricing and deviations from the law of one price. *Review of Financial Studies* 24:1980–2022.

Gibbons, M. R., and W. Ferson. 1985. Testing asset pricing models with changing expectations and an unobservable market portfolio. *Journal of Financial Economics* 14:217–36.

Giglio, S., B. Kelly, and S. Pruitt. 2016. Systemic risk and the macroeconomy: An empirical evaluation. *Journal of Financial Economics* 119:457–71.

Gomes, J., L. Kogan, and L. Zhang. 2003. Equilibrium cross section of returns. *Journal of Political Economy* 111:693–732.

Goyal, A. 2012. Empirical cross-sectional asset pricing: A survey. *Financial Markets and Portfolio Management* 26:3–28.

Green, J., J. R. M. Hand, and M. T. Soliman. 2011. Going, going, gone? The apparent demise of the accruals anomaly. *Management Science* 57:797–816.

Green, J., J. R. M. Hand, and X. F. Zhang. 2013. The supraview of return predictive signals. *Review of Accounting Studies* 18:692–730.

———. 2016. The characteristics that provide independent information about average U.S. monthly stock returns. Working Paper, Pennsylvania State University.

Griffin, J. M., and M. L. Lemmon. 2002. Book–to–market equity, distress risk, and stock returns. *Journal of Finance* 57:2317–36.

Hamilton, J. D. 1985. Uncovering financial market expectations of inflation. *Journal of Political Economy* 93:1224–41.

Hanna, J. D., and M. J. Ready. 2005. Profitable predictability in the cross section of stock returns. *Journal of Financial Economics* 78:463–505.

Hansen, L. P., and S. F. Richard. 1987. The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models. *Econometrica* 55:587–613.

Harvey, C. R. 1989. Time-varying conditional covariances in tests of asset pricing models. *Journal of Financial Economics* 24:289–317.

Harvey, C. R., Y. Liu, and H. Zhu. 2016. . . . and the cross-section of expected returns. *Review of Financial Studies* 29:5–68.

Haugen, R. A., and N. L. Baker. 1996. Commonality in the determinants of expected stock returns. *Journal of Financial Economics* 41:401–39.

Hirshleifer, D., K. Hou, S. H. Teoh, and Y. Zhang. 2004. Do investors overvalue firms with bloated balance sheets? *Journal of Accounting and Economics* 38:297–331.

Horowitz, J. L., T. Loughran, and N. Savin. 2000. Three analyses of the firm size premium. *Journal of Empirical Finance* 7:143–53.

Huang, D., F. Jiang, J. Tu, and G. Zhou. 2015. Investor sentiment aligned: A powerful predictor of stock returns. *Review of Financial Studies* 28:791–837.

Huang, J., J. L. Horowitz, and S. Ma. 2008. Asymptotic properties of bridge estimators in sparse high-dimensional regression models. *Annals of Statistics* 36:587–613.

Jegadeesh, N. 1990. Evidence of predictable behavior of security returns. *Journal of Finance* 45:881–98.

Jegadeesh, N., and S. Titman. 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance* 48:65–91.

Johnson, T. C. 2002. Rational momentum effects. *Journal of Finance* 57:585–608.

Kelly, B., and S. Pruitt. 2013. Market expectations in the cross section of present values. *Journal of Finance* 68:1721–56.

———. 2015. The three-pass regression filter: A new approach to forecasting using many predictors. *Journal of Econometrics* 186:294–316.

Kogan, L., and D. Papanikolaou. 2012. Economic activity of firms and asset prices. *Annual Review of Financial Economics* 4:361–84.

———. 2013. Firm characteristics and stock returns: The role of investment-specific shocks. *Review of Financial Studies* 26:2718–59.

Kogan, L., and M. Tian. 2015. Firm characteristics and empirical factor models: A model-mining experiment. Working Paper, Federal Reserve Board.

Kozak, S., S. Nagel, and S. Santosh. 2015. Interpreting factor models. Working Paper, University of Michigan.

Lee, C., and B. Swaminathan. 2000. Price momentum and trading volume. *Journal of Finance* 55:2017–69.

Lewellen, J. 2015. The cross section of expected stock returns. *Critical Finance Review* 4:1–44.

Lewellen, J., and S. Nagel. 2006. The conditional CAPM does not explain asset-pricing anomalies. *Journal of Financial Economics* 82:289–314.

Li, E. X. N., D. Livdan, and L. Zhang. 2009. Anomalies. *Review of Financial Studies* 22:4301–34.

Li, K.-C. 1991. Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association* 86:316–27.

Li, Y., and L. Yang. 2011. Testing conditional factor models: A nonparametric approach. *Journal of Empirical Finance* 18:972–92.

Lin, X., and L. Zhang. 2012. The investment manifesto. *Journal of Monetary Economics* 60:351–66.

Lipson, M. L., S. Mortal, and M. J. Schill. 2011. On the scope and drivers of the asset growth effect. *Journal of Financial and Quantitative Analysis* 46:1651–82.

Liu, L. X., and L. Zhang. 2014. A neoclassical interpretation of momentum. *Journal of Monetary Economics* 67:109–28.

Lyandres, E., L. Sun, and L. Zhang. 2008. The new issues puzzle: Testing the investment-based explanation. *Review of Financial Studies* 21:2825–55.

Mashruwala, C., S. Rajgopal, and T. Shevlin. 2006. Why is the accrual anomaly not arbitraged away? The role of idiosyncratic risk and transaction costs. *Journal of Accounting and Economics* 42:3–33.

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

Mohanram, P. S. 2005. Separating winners from losers among low book-to-market stocks using financial statement analysis. *Review of Accounting Studies* 10:133–70.

Novy-Marx, R. 2012. Is momentum really momentum? *Journal of Financial Economics* 103:429–53.

———. 2013. The other side of value: The gross profitability premium. *Journal of Financial Economics* 108: 1–28.

———. 2016. Backtesting strategies based on multiple signals. Working Paper, University of Rochester.

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

Ohlson, J. A. 1980. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research* 18:109–31.

Ou, J. A., and S. H. Penman. 1989. Financial statement analysis and the prediction of stock returns. *Journal of Accounting and Economics* 11:295–329.

Pástor, Ľ., and R. F. Stambaugh. 2009. Predictive systems: Living with imperfect predictors. *Journal of Finance* 64:1583–628.

Piatti, I., and F. Trojani. 2015. Predictable risks and predictive regression in present-value models. Working Paper, University of Oxford.

Piotroski, J. D. 2000. Value investing: The use of historical financial statement information to separate winners from losers. *Journal of Accounting Research* 38:1–41.

Polk, C., and P. Sapienza. 2009. The stock market and corporate investment: A test of catering theory. *Review of Financial Studies* 22:187–217.

Pontiff, J. 1996. Costly arbitrage: Evidence from closed-end funds. *Quarterly Journal of Economics* 111:1135–51.

———. 2006. Costly arbitrage and the myth of idiosyncratic risk. *Journal of Accounting and Economics* 42: 35–52.

Pontiff, J., and A. Woodgate. 2008. Share issuance and cross-sectional returns. *Journal of Finance* 63:921–45.

Reinganum, M. R. 1981. Misspecification of capital asset pricing: Empirical anomalies based on earnings' yields and market values. *Journal of Financial Economics* 9:19–46.

Roll, R. 1977. A critique of the asset pricing theory's tests. Part I: On past and potential testability of the theory. *Journal of Financial Economics* 4:129–76.

Rosenberg, B., K. Reid, and R. Lanstein. 1985. Persuasive evidence of market inefficiency. *Journal of Portfolio Management* 11:9–16.

Rytchkov, O. 2010. Expected returns on value, growth, and HML. *Journal of Empirical Finance* 17:552–65.

Rytchkov, O. 2012. Filtering out expected dividends and expected returns. *Quarterly Journal of Finance* 2:1250012 (56 pages).

Schwert, G. W. 2003. Anomalies and market efficiency. In *Handbook of the economics of finance*, ed. G. Constantinides, M. Harris, and R. M. Stulz. 939–74. Amsterdam: North-Holland.

Sloan, R. G. 1996. Do stock prices fully reflect information in accruals and cash flows about future earnings? *Accounting Review* 71:289–315.

Stambaugh, R. F., J. Yu, and Y. Yuan. 2012. The short of it: Investor sentiment and anomalies. *Journal of Financial Economics* 104:288–302.

———. 2015. Arbitrage asymmetry and the idiosyncratic volatility puzzle. *Journal of Finance* 70:1903–48.

Stambaugh, R. F., and Y. Yuan. 2016. Mispricing factors. Working Paper, University of Pennsylvania.

Stock, J. H., and M. W. Watson. 2006. Forecasting with many predictors. In *Handbook of economic forecasting*, vol. 1, ed. G. Elliot, C. W. Granger, and A. Timmermann. 515–54. Amsterdam: North-Holland.

Subrahmanyam, A. 2010. The cross-section of expected stock returns: What have we learnt from the past twenty-five years of research? *European Financial Management* 16:27–42.

Titman, S., K. C. J. Wei, and F. Xie. 2004. Capital investments and stock returns. *Journal of Financial and Quantitative Analysis* 39:677–700.

Treynor, J. L., and F. Black. 1973. How to use security analysis to improve portfolio selection. *Journal of Business* 46:66–86.

Vinzi, V. E., W. W. Chin, J. Henseler, and H. Wang. 2010. *Handbook of partial least squares: Concepts, methods and applications*. Berlin: Springer-Verlag.

Wang, H., and J. Yu. 2013. Dissecting the profitability premium. Working Paper, University of Delaware.

Wold, H. 1975. Path models with latent variables: The NIPALS approach. In *Quantitative sociology: International perspectives on mathematical and statistical modeling*, ed. H. M. Blalock, A. Aganbegian, F. M. Borodkin, R. Boudon, and V. Cappecchi. 307–57. New York: Academic Press.

Wold, H. 1982. Soft modelling: The basic design and some extensions. In *Systems under indirect observation*, ed. K. G. Jöreskog and H. Wold. 1–54. Amsterdam: North-Holland.

Wold, S., M. Sjöström, and L. Eriksson. 2001. PLS-regression: A basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems* 58:109–30.

Wu, J. G., L. Zhang, and X. F. Zhang. 2010. The q-theory approach to understanding the accrual anomaly. *Journal of Accounting Research* 48:177–223.

Xing, Y. 2008. Interpreting the value effect through the Q-theory: An empirical investigation. *Review of Financial Studies* 21:1767–95.

Zhang, L. 2005. The value premium. *Journal of Finance* 60:67–103.