Out-of-Sample Performance of Factor Return Predictors*

Du Nguyen†

First Draft: June 2023
This Draft: July 2024

Abstract

In a factor timing context, recent studies have emphasized on developing techniques that reduce the factor dimension and demonstrated return predictability using only a few predictors with specific choice of estimation design. This focus inadvertently neglects the crucial issue of model instability that has been shown to plague the forecasting literature. Using almost a hundred equity factors and a broader set of predictor variables, I find that the forecasting performance of recent factor timing techniques is indeed sensitive to the choice of empirical design. Applying a variety of shrinkage methods on predictors and focusing on forecasting individual factors to better capture the dynamics between factor returns and predictive signals, I document robust evidence of out-of-sample predictability and more stable investment performance for factor timing strategies. The optimal timing portfolio has a 30% higher Sharpe ratio and generates more than twice the economic gains relative to the factor dimension-reduction approach.

Keywords: Factor portfolios, Factor timing, Return predictability

JEL classification: G10, G11, G17, C53

*I am grateful to Michael O'Doherty for insightful discussions and comments. I thank Chun-Wei Lin (discussant) and seminar participants at the 2024 SWFA annual meeting, and University of Missouri for helpful comments and suggestions.

†Du Nguyen is a Ph.D. candidate in finance in the Trulaske College of Business, University of Missouri, Columbia, MO 65211, USA. E-mail: du.nguyen@missouri.edu.

1 Introduction

Factor investing is one of the fastest-growing areas in the asset management industry, with about $1.45 trillion in U.S. factor exchange-traded products alone as of 2021 (Morningstar, 2022). There is a growing literature in asset pricing that attempts to predict factor returns, and thus to improve performance of factor investing with factor timing.[1] To assess whether a variable (e.g., prior one-year factor returns) has predictive ability, a typical specification is to regress the future return of a factor (e.g., the size factor) on the lagged variable

$$R_{t:t+h} = \beta_0 + \beta_1 X_t + \varepsilon_{t:t+h}, \tag{1}$$

where $\mathrm{R}_{t:t+h}$ is the h-period ahead excess return of the factor, and X_t is the lagged predictor variable. Table 1 provides a list of variables that have been shown to be good predictors, including investor sentiment (Stambaugh, Yu, and Yuan, 2012), aggregate mutual fund flows (Akbas, Armstrong, Sorescu, and Subrahmanyam, 2015), book-to-market ratio (Cohen, Polk, and Vuolteenaho, 2003, Kelly and Pruitt, 2013, Haddad et al., 2020, Baba-Yara, Boons, and Tamoni, 2021), industry-adjusted book-to-market ratio (Baba-Yara et al., 2021), issuer-repurchaser spread (Greenwood and Hanson, 2012), time-series momentum (Moskowitz, Ooi, and Pedersen, 2012), volatility (Moreira and Muir, 2017), factor momentum (Gupta and Kelly, 2019, Ehsani and Linnainmaa, 2022), and portfolio characteristics (Kelly et al., 2023, Kagkadis, Nolte, Nolte, and Vasilas, 2023).[2]

Similarly growing is the documentation of hundreds of equity factors (e.g., Cochrane, 2011 and McLean and Pontiff, 2016) that creates a challenge to provide credible evidence of predictability using Equation 1. This challenge has shifted recent studies to focus on developing methodologies that reduce the factor dimension for predictability (e.g., Haddad et al., 2020). As a result, these papers often illustrate the predictive ability of their techniques with a few predictor variables and specific choice of test factors and estimation designs. Because model instability is a critical issue that has plagued the financial forecasting literature (e.g., Goyal and Welch, 2008 and Goyal, Welch, and Zafirov, 2023), it is reasonable to question whether the forecasting performance documented in these studies extends to the wide range of factors and predictors featured in the factor timing literature, and whether it is robust to the choice of estimation design (e.g., Rossi and Inoue, 2012 and Rossi, 2013).[3]

I answer these questions by providing a comprehensive analysis of the predictive ability of nine variables for a broad sample of 92 equity factors. Using both predictive regressions and techniques that reduce the

[1] See, e.g., Haddad, Kozak, and Santosh (2020) and Kelly, Malamud, and Pedersen (2023).

[2] Favero, Melone, and Tamoni (2022), Huang, Song, and Xiang (2022), and Dong, Kang, and Peress (2022) are examples of working papers that explore other predictors of factor returns.

[3] There is also evidence from the asset management industry suggesting that the factor timing strategies is difficult to implement in real time due to outliers (e.g., Asness, 2016, Asness, Chandra, Ilmanen, and Israel, 2017, and Dai and Dong, 2023).

factor dimension, I find that no systematic evidence of return predictability for most variables under common estimation designs. I provide statistical evidence that this poor performance stems primarily from structural instability in forecasting models. Exploring several shrinkage techniques that combine predictive signals from all variables and focusing on predicting individual factor returns, I find robust evidence in favor of factor return predictability. This finding highlights the role of machine learning applications in improving factor timing strategies.

I first examine the predictive ability of nine variables for a broad sample of 92 equity factors with an emphasis on their real-time performance. Using univariate predictive regressions, I conduct a comprehensive out-of-sample analysis with a recursive estimation design. The main performance measure is Campbell and Thompson's (2008) out-of-sample R^2 (R^2_{OS}), which intuitively captures how good return forecasts from predictive regressions are compared to historical mean forecasts. Positive R^2_{OS} suggests that predictive regressions beat the historical average, and higher R^2_{OS} implies potentially larger economic value of factor timing. I also use total R^2_{OS} to evaluate the predictability across all factors (see, e.g., Gu, Kelly, and Xiu, 2020).

The results indicate no systematic evidence that factor returns are predictable out of sample. The median R^2_{OS} across 92 factors is negative for eight of nine variables. The total R^2_{OS} is either negative or almost zero for all but investor sentiment. Across nine predictors, only 5%-28% of factors are predictable with standard statistical significance.[4] The weak performance is pervasive across all categories. Momentum and investment factors exhibit almost no evidence of predictability. Favorable evidence among other factor groups is, if any, meager. These out-of-sample results highlight the weakness of forecasting factor returns with predictive regressions and motivate the development of other forecasting approaches.

Recently, Haddad, Kozak, and Santosh (2020) (hereafter HKS) propose an approach to predicting factor returns that avoids running univariate regressions for many factors. HKS focus on forecasting returns of the largest principal components (PC) that summarize the structure of factor returns, and infer individual factor predictability via exposure of each factor to these PCs. For out-of-sample tests, HKS adopt a split-sample estimation weighting scheme and find strong evidence of factor predictability. Using their exact empirical design, I confirm the main finding that book-to-market ratio has a strong predictive ability for a large number of factors.

Although this empirical choice shows an impressive result, the split-sample design is an unconventional choice in the literature. There are several econometric concerns that motivate the use of more common designs (e.g., recursive regressions), including estimation risk, structural breaks, and power of out-of-sample tests.

[4] Inoue and Kilian (2005) show that out-of-sample tests based on the mean squared prediction errors (MSPE)-normal statistic (Diebold and Mariano, 1995, West, 1996) have low power, and thus we should discount weak out-of-sample evidence in favor of strong in-sample results. The out-of-sample tests in this paper use the MSPE-adjusted statistic, which Clark and West (2007) show to have higher power.

Due to these concerns, I re-examine the performance of the PC portfolio approach with three estimation schemes (i.e., split-sample, recursive, and rolling-window). Combined with nine predictors, I conduct an extensive analysis for 27 cases in total.

I find no systematic evidence for factor return prediction under the more common estimation designs. The mean R^2_{OS} with book-to-market ratio as predictor reduces from 0.64% under the split-sample design to -0.11% and -0.31% under the recursive and rolling-window designs, respectively. The number of predictable factors with statistical significance also reduces by half. This deterioration in performance is pervasive across all variables. Across all 27 cases, the split-sample results are more likely to indicate stronger evidence of predictability, but such evidence either deteriorate or disappear under the common estimation schemes. For most predictors, less than a handful of factors are predictable with statistical significance at the 1% level.

Next I investigate whether the PC portfolio approach systematically improves performance of the optimal timing portfolio relative to the static factor investing portfolio. The optimal portfolio is constructed from the market factor and the first five PC portfolios. I find that while the PC portfolio approach improves the Sharpe ratio and utility gains of the optimal timing portfolio under the split-sample design, the incremental economic values from anomaly timing become marginal under the recursive and rolling-window schemes. Across nine predictors, the Sharpe ratio of the optimal timing portfolio is systematically lower than that of the static strategy. For utility gains, only the timing strategies based on factor momentum appear to outperform. I also find that increases in the variance of the implied stochastic discount factor (SDF) are typically more modest under the alternative estimation schemes. Except for factor momentum, the increase ranges between 11% to 45%, which is much less impressive than the 80% increase in the base case with book-to-market ratio. These results collectively question the ability of the PC portfolio approach to document time-varying factor premia and its implications for factor-based asset pricing model.

An economic explanation for the differences in the predictive performance across estimation designs is potential structural changes in the relation between factor returns and their predictors. I provide evidence that there exists several structural breaks in the predictive relations for three out of five largest PCs. Moreover, only a few factors maintain consistent predictive relations (i.e., correct predictive signs with statistical significance) in both halves of the sample at the individual factor level. I also examine another source of model instability, namely the time-varying factor loadings (i.e., the eigenvectors of the covariance matrix of factor returns). Intuitively, the factor loadings determine the exposure of factor portfolios to the PCs, and thus contribute to the individual factor predictability. I find statistical evidence that there exists at least one large structural change in the factor loadings during the sample period. Taken as a whole, substantial structural breaks create a difficulty for the PC portfolio approach to provide stable evidence of factor timing.

The results so far highlight structural changes as a critical hurdle to provide consistent evidence of factor return predictability out of sample. A simple solution to reduce the impact of structural breaks on forecast accuracy is combining forecasts from multiple signals (e.g., Diebold and Pauly, 1987). This approach also avoids the difficulty in estimating break points in real time, and the bias-variance trade-off in break estimation methods. I explore five common shrinkage techniques, including forecast combination, discount mean square of prediction errors, predictor average, principal component regression (PCR), and partial least squares (PLS). I find systematic evidence in favor of factor return predictability across all five methods. The mean R^2_{OS} is between 0.52% and 1.08%, and the total R^2_{OS} ranges between 0.60% and 1.30%. Advanced machine learning methods such as PCR and PLS show that more than half of factors are predictable with statistical significance at least at the 10% level. The shrinkage approach improves predictability across all categories. While simple methods such as forecast combination improves predictability for momentum and investment factors, machine learning methods show strong evidence of predictability for other factor categories. These results highlight the stable ability of the shrinkage approach in predicting a large set of anomaly portfolios.

The strong predictability implies potential economic gains for factor timing strategies. From the perspective of mean-variance investors, I use return forecasts from the shrinkage methods to estimate optimal allocations to the original factor portfolios, and evaluate factor timing portfolios in terms of their Sharpe ratios and Certainty Equivalent Returns (CER). The shrinkage approach routinely leads to substantial improvement for timing strategies. While the mean annualized Sharpe ratio across the pure factor investing strategies (i.e., allocations based on the historical averages) is only 0.11, it increases more than twice for timing strategies across five shrinkage methods. There are also substantial utility gains with factor timing under this approach. The mean annualized CER is between 2.39% and 4.00%, compared to an average of 0.40% with factor investing.

Finally, I examine the performance of the optimal timing portfolio under the shrinkage approach. To make a direct comparison with the PC portfolio approach, I adopt the exact procedure of the latter, but estimate expected returns of the PCs using the individual factor return forecasts from the shrinkage methods rather than using the PC predictors. The contributing weight of individual forecasts for each PC corresponds to their loadings on the PC. Following this modification, the optimal timing portfolio produces annualized Sharpe ratio between 0.74 to 0.94 across five methods, compared to an average of 0.72 under the original PC approach across nine predictors. In terms of utility gains, PCR and PLS methods deliver almost twice as large as CER. These methods also imply large variance of the implied SDF at 2.00 and 2.89, respectively, compared to the average 1.70 under the original PC approach.

This paper contributes to the literature on factor timing by documenting that the shrinkage approach provides conclusive evidence for factor return predictability. A recent approach developed by Kelly et al.

(2023) exploits the cross-predictability of a factor characteristic for other factor returns. Another paper by Kagkadis et al. (2023) follow HKS to forecast PC portfolio returns, but apply dimension-reduction techniques on PC portfolio characteristics to construct predictors. The approaches in these studies rely on estimation of predictability of principal portfolios, and it remains unclear whether its performance depends on the subjective choice of factors and number of principal components (e.g., Bessembinder, Burt, and Hrdlicka, 2022, Bessembinder, Burt, and Hrdlicka, 2024). The shrinkage approach avoids this by focusing on the predictability of individual factors. In this direction, a paper closely related to this paper is Neuhierl, Randl, Reschenhofer, and Zechner (2023). While their paper adopts only partial least squares regressions to assess the economic gain of timing individual factors, this paper shows that other shrinkage techniques can also improve factor return predictability.

This paper also adds to the literature on data snooping in return prediction. Starting with Goyal and Welch (2008), there has been a growing number of studies that invalidates the ability of predictor variables (e.g., Goyal et al., 2023, Cakici, Fieberg, Metko, and Zaremba, 2024) and new methods (e.g., Cakici, Fieberg, Neumaier, Poddig, and Zaremba, 2024) in timing the market factor. Interests from both investment industry and finance academia, combined with the factor zoo, have spurred the discovery of more variables and methods in forecasting anomaly returns. This paper complements recent papers that challenge the ability of some prior predictors in this literature (e.g., Novy-Marx, 2014, Fan, Li, Liao, and Liu, 2022) by providing a comprehensive assessment for a broad set of factors and predictors in out-of-sample settings. This paper also re-emphasizes the importance of common empirical exercises in documenting time-series predictability from the econometrics literature by showing that variations in estimation designs can produce different conclusions of predictability (Rossi, 2013).

Last, the evidence from this paper have important investment implications. There has been a debate whether factor timing is implementable in practice (e.g., Asness, 2016, Arnott, Beck, and Kalesnik, 2016). This paper sheds light on understanding where the disparity in conclusion may come from, and suggests that advanced machine learning methods are promising for the practical implementation of factor timing strategies.

2 Data and Methodology

I describe the sample of factor portfolios used in the empirical analyses in Section 2.1, the predictor variables in Section 2.2, and the methodology for forecasting evaluation in Section 2.3.

2.1 Factor portfolios

I consider non-overlapped factors from Hou, Xue, and Zhang (2015), Kelly, Pruitt, and Su (2019), and Haddad et al. (2020), and restrict the sample to factors that are based on a single, continuous sorting variable. To ensure sample consistency, I require that data on sorting variables is available from 1970.[5] This process identifies 92 characteristic variables for which I can construct factor portfolios and their predictors from January 1970 to December 2021.

I follow the convention in selecting stocks for the construction of factor portfolios. I consider the universe of firms covered by the Center for Research in Security Prices (CRSP) and the Compustat Fundamentals Annual (Compustat). I include only U.S. common stocks that are listed on NYSE, AMEX, and NASDAQ, and exclude utility and financial firms. To mitigate the impact of small stocks, I require stocks to have market price greater than $1 at the portfolio formation date. I collect the firm-level characteristics from the CRSP Monthly and Daily Stock Files and the Compustat Annual and Quarterly Files, and sort firms into ten value-weighted portfolios for each of the characteristic using the breakpoints from only NYSE firms. I use the CRSP monthly file to calculate the monthly return series for each decile portfolio. Each factor portfolio is constructed by taking a long (short) position in the decile that is expected to outperform (underperform) based on prior literature.

Since many factors use related characteristics, I follow the classification scheme of Hou et al. (2015) and Hou, Xue, and Zhang (2020) to classify each factor into one of six categories: momentum, value, investment, profitability, intangibles, and trading frictions. Table A1 provides additional detail on sample selection, factor categories, average return and Fama and French's (1993) 3-factor alpha for the factor portfolios.

2.2 Predictor variables

I consider predictors of anomaly returns from published studies in which the time-series data of a predictor is sufficiently long. This restriction mitigates concerns of data snooping and estimation risk, especially in out-of-sample settings (e.g., Rossi and Inoue, 2012). This process identifies nine predictors: book-to-market ratio (BM), industry-adjusted book-to-market ratio ($IND.BM$), issuer-repurchaser spread ($ISSREP$), prior one-moth factor return ($MOM1$), prior 12-month factor average returns ($MOM12$), volatility ($Volatility$), characteristic spread (CS), long-run reversal ($LRREV$), and investor sentiment ($SENT$). Except for $SENT$, the others are factor-specific. I briefly describe the construction of these predictors here, and leave the detail in Table A2 in the Appendix.

[5]Since data on the bid-ask spreads is sparse before 1980, I use Amihud's (2002) illiquidity measure as a substitution.

Book-to-market ratio and its industry-adjused version I follow Baba-Yara et al. (2021) to construct the book and market value for each firm. Following HKS, a portfolio's book-to-market ratio is defined as the sum of the book value relative to the total market value of all firms in that portfolio. The (net) book-to-market ratio of a factor portfolio is constructed as the difference in natural logarithm of the book-to-market ratio between the long and the short portfolios. For industry-adjusted book-to-market ratio of a firm, I subtract from the firm's book-to-market ratio the value-weighted average book-to-market ratio of the industry the firm belongs to. A portfolio's industry-adjusted book-to-market ratio is defined as the value-weighted average of the industry-adjusted book-to-market ratio of all firms in that portfolio.

Issuer-repurchaser spread I follow Greenwood and Hanson (2012) to construct *ISSREP* of a characteristic portfolio as the difference in the weighted average of the underlying characteristic between issuers and purchasers. Greenwood and Hanson (2012) define issuers (purchasers) as firms whose net stock issuance is greater than 10% (less than -0.5%). I follow Fama and French (2008) and define net stock issuance as the change in log split-adjusted shares outstanding from Compustat annual file.

Factor momentum I use Gupta and Kelly's (2019) prior one-month factor return for *MOM1*. For *MOM12*, I use the indicator version of prior 12-month average returns as Ehsani and Linnainmaa (2022) use in their main analysis (see their Table II). In particular, *MOM12* for a factor is an indicator variable that equals one if the factor portfolio's average return over the past 12 months is positive and zero otherwise.

Volatility Following Moreira and Muir (2017), I use the realized variance of daily factor returns in the prior month to estimate the factor's volatility. To mitigate skewness concerns, I scale the variable by the average of monthly variances up to the prior month, and take the natural logarithm.

Characteristic spread I transform each characteristic into a [-0.5,+0.5] interval and calculate a factor's characteristic spread as the difference in the value-weighted characteristic between the long and the short leg of the factor (Kelly et al., 2023).

Long-run reversal I calculate a factor's long-run reversal signal as cumulative returns over the past 5 years, scaled by the realized variance of the factor's returns over the same period (Moskowitz et al., 2012).

Investor sentiment For in-sample tests and out-of-sample tests that use a split-sample estimation design, I use Baker and Wurgler (2006) full-sample orthogonalized sentiment series. For out-of-sample tests that use either a recursive or a rolling-window estimation design, I follow Huang, Jiang, Tu, and Zhou (2015) to create a look-ahead bias-free version of the index. First, at each out-of-sample month $t+1$, I use the data on five individual components of investor sentiment (i.e., the close-end fund discount rate, the number of IPOs, the 12-month lagged first-day returns of IPOs, the 12-month lagged dividend premium, and the equity share in new issues) from July 1965 to month t, and standardize them to have mean of 0 and standard deviation of 1. Second, I obtain the orthogonalized version of these series by regressing each series on six macroeconomic variables as in Baker and Wurgler (2006), and retain the residuals from the regressions. I smooth each residual series with their six-month average values to mitigate outliers in each series. With predictive regressions and the PC portfolio approach, I use the first principal component of the five individual series as predictor. For all shrinkage methods, I use all five individual measures as predictors. I obtain data on the full-sample orthogonalized sentiment index, its original five components, and six macroeconomic variables from Jeffrey Wurgler's webpage.[6]

2.3 Methodology

2.3.1 Forecasting methods

Predictive regressions Under the assumption of time-varying expected returns, I obtain predictive coefficient estimates by running the time-series regression

$$R_{t:t+1} = \beta_0 + \beta_1 X_t + \varepsilon_{t:t+1}, \tag{2}$$

where $R_{t:t+1}$ is the one-period ahead excess return of the factor, and X_t is the lagged predictor variable. A one-month ahead return forecast formed at time t for a factor i is

$$\hat{R}_{i,t+1} = \hat{\beta}_{i0} + \hat{\beta}_{i1} X_t, \tag{3}$$

where $\hat{\beta}_{i0}$ and $\hat{\beta}_{i1}$ are predictive coefficient estimates.

The PC portfolio approach Under the assumption that excess factor returns have a linear latent factor specification, a factor portfolio i's return can be represented as follows

$$R_{i,t+1} = \sum_{k=1}^{K} \omega_i^k PC_{t+1}^k + \varepsilon_{i,t+1}, \tag{4}$$

where PC_{t+1}^k is the kth PC in month $t+1$, and ω_i^k is the loading of factor i on the kth PC. In empirical tests, I follow HKS and set $K = 5$. I estimate the PCs and factor loadings with the principal component estimation. The data I use for estimation depending on the forecasting estimation design. For example, if the design is recursive window, I recursively re-estimate the PC returns and factor loadings in every month.

[6]The data, updated until June 2022, is available at https://pages.stern.nyu.edu/~jwurgler/. I thank Jeffrey Wurgler for making the data available.

For factor-specific predictor variables, I use the eigenvectors obtained from the PC estimation to construct the PC predictors. For *SENT*, I use the sentiment index to predict one-month ahead PC returns. I estimate predictive coefficients for the PCs as follows:

$$PC_{t+1}^{k} = \lambda_0^k + \lambda_1^k X_t + \epsilon_{t+1}^k. \tag{5}$$

It follows that a return forecast formed in month t for the kth PC portfolio is $\hat{\lambda}_0^k + \hat{\lambda}_1^k X_t$, where $\hat{\lambda}_0^k$ and $\hat{\lambda}_1^k$ are estimated from Equation 5. I obtain a return forecast formed at time t for factor i as

$$\hat{R}_{i,t+1} = \sum_{k=1}^{5} \hat{\omega}_i^k \widehat{PC}_{t+1}^{k}, \tag{6}$$

where $\hat{\omega}_i^k$ is the estimated loading of factor i on the kth PC, and $\widehat{PC}_{t+1}^{k}$ is a one-month ahead forecast of the kth PC.

The shrinkage approach I use five shrinkage techniques, including forecast combination (*FC*), discount mean square of prediction errors (*DMSPE*), predictor average (*Average*), principal component regression (*PCR*), and partial least squares (*PLS*). *FC* forecast is the equal-weighted average of forecasts from univariate predictive regressions. For each factor, I obtain individual forecasts by running the predictive regression with each predictor separately, then take the arithmetic mean of the individual forecasts. Similar to *FC*, *DMSPE* forecast is the weighted average of individual forecasts from univariate predictive regressions; however, forecasts that have lower prediction errors over a holdout period have greater weight. For *Average*, I first take the arithmetic mean of all (standardized) predictors to construct a single variable. For *PCR*, I first extract the first principal component from the set of predictors. Both *Average* and *PCR* then use the resulting variables to predict factor returns in univariate predictive regressions. For *PLS*, I construct a target-relevant series from all variables, and use this series to predict factor returns with univariate predictive regressions. For all shrinkage methods, I perform the estimations using only real-time data to avoid look-ahead bias.

2.3.2 Performance evaluation

I assess the statistical accuracy of each forecasting approach via mean squared of prediction errors (MSPE). Denote $\hat{\epsilon}_{i,t+1|t}^{m}$ and $\hat{\epsilon}_{i,t+1|t}^{PM}$ as the conditional forecast error of factor i in month $t+1$ under approach m and the prevailing mean, respectively. The sample MSPE is calculated as

$$\widehat{MSPE}_i^{m} = \frac{1}{T} \sum_{t=0}^{T} \hat{\epsilon}_{i,t+1|t}^{2,m}, \tag{7}$$

where T is the number of out-of-sample observations. I adopt Clark and West's (2007) procedure to test the null hypothesis that the MSPE under the prevailing mean benchmark (i.e., $\widehat{MSPE}_i^{PM}$) is less than or equal to the MSPE under approach m. A rejection indicates that forecasts made by method m are statistically better than the historical mean.

I also calculate and report the Campbell and Thompson's (2008) R^2_{OS} for each factor (subscript suppressed) as

$$R^2_{OS} = 1 - \frac{\widehat{MSPE}^m}{\widehat{MSFE}^{PM}}. \tag{8}$$

To assess the predictive ability across all factors, I follow Gu et al. (2020) and extend Equation 8 to calculate total R^2 as follows[7]

$$\text{Total } R^2_{OS} = 1 - \frac{\sum_i \widehat{MSPE}_i^m}{\sum_i \widehat{MSPE}_i^{PM}}. \tag{9}$$

3 Predicting factor returns with predictive regression approach

In this section I analyze the out-of-sample predictive ability of the variables for each factor using the conventional predictive regression approach. I adopt a recursive estimation design to evaluate out-of-sample evidence. For each factor and predictor, I use the first half of the sample (1970:01 to 1995:12) as the initial training period to estimate the parameters of the predictive regression model, namely β_0 and β_1 from Equation 2, for the forecasts of 1996:01. I expand the estimation window each month to construct forecasts for the rest of the sample. The out-of-sample evaluation period is from 1996:01 to 2021:12. The expanding-window estimation design allows me to use all past information for estimation; hence, alleviates estimation risk. Moreover, the long evaluation period mitigates the concern of low power in the out-of-sample tests (Inoue and Kilian, 2005).

I present the out-of-sample results in Table 2. Panel A reports the summary statistics of R^2_{OS}s, the total R^2_{OS}, and the number of R^2_{OS} that is statistically significant at least at the 10% level. The out-of-sample predictive ability of most predictors is weak. Column (1) shows that except for investor sentiment, the mean R^2_{OS} is either negative or marginally positive. The median R^2_{OS} in Column (3) is negative for all predictors but *SENT*, suggesting that the predictive regression model for a median factor performs worse than the historical mean model. Column (5) provides further evidence that out-of-sample predictability as a whole is weak. Column (6) reports the number of predictions that yield positive R^2_{OS}, and those that are statistically significant (p-value less than 10%). Across the predictors, the number is low, ranging from 5 to 26 out of 92 factors. For example, the number of predictable factors using *BM* is only 23, less than a third of the

[7] Gu et al. (2020) and Gu, Kelly, and Xiu (2021) calculate the total R^2_{OS} for individual stocks. Their calculation is different from Equation 9 in that they assume an unconditional expected return of zero for all stocks. In my setting, this assumption implies comparing forecasts made under method m to a zero-return forecast. However, Gu et al. (2020) highlight that this assumption is only suitable for individual stocks, and is not reasonable for long-short portfolios.

factor sample. These results suggests that the forecasts made by most predictors perform no better than the prevailing mean forecasts.

Next I examine the out-of-sample evidence across six factor categories. Panel B reports the total R^2_{OS} and the number of statistically significant R^2_{OS} (p-value less than 10%) for all six categories. There is weak evidence of factor return predictability across all factor categories, with only a few exceptions. Value and volatility appear to predict momentum factors, but the number of predictable factors in this group remains less than half. While value factors appear to be predictable with the 12-month factor momentum, profitability factors are strongly predictable with the one-month version. For all factor categories except momentum, *SENT* appears to be good predictor. Nevertheless, the number of predictable factors using investor sentiment remains low, suggesting that a few strong predictable factors are driving the result.

In summary, the out-of-sample results show that the predictive regression approach does not systematically provide evidence in favor of factor return predictability. In the next section, I analyze whether the principal-component portfolio approach can provide stronger evidence.

4 Predicting factor returns with principal component portfolio approach

Haddad et al. (2020) estimate expected factor returns through the forecasts of the largest principal components of the factors. HKS show that a few large components is sufficient to explain majority of variances in the cross-section of expected factor returns. Subsequently, return forecasts for individual factors based on the forecasts of the dominant components are less noisy. Using the book-to-market ratio as predictor and a split-sample estimation design, HKS demonstrate a strong out-of-sample forecasting performance for a set of 50 equity factors.

In Section 4.1 I use my broader sample and HKS's estimation design to confirm the outstanding performance of the PC portfolio approach. In Section 4.2 I show that this performance disappears when other estimation designs common in the literature are employed. I provide some potential explanations in Section 4.4.

4.1 The PC portfolio approach in predicting factor returns

I begin by describing in detail the empirical choices used in HKS. The original study constructs a set of 50 factor portfolios from 1974:01 to 2017:12. Table A1 in the Appendix provides more detail on the identity of these factors. The first step in the PC portfolio approach requires identifying the dominant equity components that summarize the cross-section of expected factor returns. It also requires estimating the loading of each factor on these common components. HKS obtain these estimates using data from the first half of the sample. Based on the factor structure (Equation 4), they use the PC estimators to estimate the first five PCs, and the factor loadings on these components for the full sample. Using the eigenvectors obtained from this process and the factor-specific book-to-market ratios, they construct the book-to-market ratio for each PC.[8] Using only the first half of the data on the PC returns and their predictors, HKS estimate the predictive parameters in Equation 5, and obtain forecasts of the PC returns for the second half of their sample. Given the PC return forecasts and factor loadings on the PCs, the authors obtain return forecasts for each factor (Equation 6). HKS use the second half of their sample for out-of-sample evaluation.

I adopt the above split-sample estimation design for my sample, and summarize the predictive performance of the approach in Table 3. The first half of the sample is from 1970:01 to 1995:12. Panel A reports the predictive performance for five PCs using their own book-to-market ratios. The first two rows report the predictive coefficient estimates and their corresponding t-statistics. Three out of five PCs produce correct predictive sign, and the second PC is predictable with statistical significance (t-statistic = 2.74). The next two rows present the full-sample R^2 and the R^2_{OS}. The first two PCs have positive R^2_{OS} and demonstrate significant out-of-sample predictability with R^2_{OS} over 1%.

The next step in the PC portfolio approach is to obtain return forecasts for individual factors through their loadings on the five PCs, and evaluate the out-of-sample predictive performance. Panel B of Table 3 reports the predictive performance across factors using the conventional predictive regression approach and the PC portfolio approach. Columns (1) and (2) provide the distribution statistics of R^2_{OS}. On the one hand, the mean (median) R^2_{OS} under the conventional approach is −0.10% (0.03%), confirming the method's poor predictive performance documented in Section 3. On the other hand, the PC portfolio method produces substantially higher R^2_{OS}. The mean (median) under this approach is 0.64% (0.99%). The total R^2_{OS} in Column (3) under the PC portfolio method is 1.15%, compared to 0.23% under the conventional method.[9] The results suggest that the PC portfolio approach provides evidence in favor of factor return predictability. In Column (4) I report the number of factors that have positive and statistically significant R^2_{OS}. The number of predictable factors with statistical significance at the 5% level increases more than twice from 21 under the traditional approach to 50 under the PC portfolio approach. This statistic confirms the superior performance of the PC approach method in predicting factor returns.

[8] HKS market adjust and rescale both the PC returns and their predictors. In their Table 4 (Row 12) they show that the overall predictive performance in terms of the total R^2_{OS} is even stronger without these adjustments. To minimize unnecessary data adjustments and maintain consistency through all forecasting approaches, I do not market adjust and rescale the data in my sample.

[9] HKS show that the mean (median) R^2_{OS} among their 50 factors under the PC portfolio approach is 1.00% (0.70%). The total R^2_{OS} in their sample is 0.93%. My sample includes 44 factors from HKS's sample, and does not include the other 6 factors due to data availability from 1970. When I repeat the test for only 44 factors, the mean, the median, and the total R^2_{OS} is 0.60%, 0.74%, and 0.93%, respectively.

The evidence from the PC portfolio approach is impactful because they resolve the difficulty of recent studies that use the book-to-market ratio to document factor return predictability (e.g., Asness, 2016 and Asness et al., 2017). However, I argue in the next section that the estimation design adopted in the original study is unconventional in the forecasting literature. I show that the implementation of other common estimation schemes can change the favorable conclusion of factor return prediction.

4.2 Estimation designs and performance of the PC portfolio approach

There are ex ante econometric concerns with the split-sample estimation scheme. First, this scheme is not desirable to mitigate estimation risk as we expect longer training windows are more likely to reduce estimation error. Given that estimation error is always a challenge for real-time forecasting exercises, it is widely believed among empiricists that more data is preferable. Second, it is not obvious how the split-sample estimation scheme biases return forecasts when there are structural breaks in the relation between future returns and their predictors.[10] For example, Pesaran and Timmermann (2002) suggest that the rolling-window estimation design is more appropriate if structural changes are frequent and large. Third, the inference of predictability is affected by the power of out-of-sample tests, which might differ across estimation designs. For instance, McCracken (2007) show that the recursive design is generally most powerful and the split-sample scheme is powerful only in particular cases.[11] Because of these econometric concerns, it is common in practice to report forecasting results for several estimation weighting designs (Rossi, 2013).[12]

To assess how the PC portfolio approach performs under different estimation designs, I re-perform the tests in Section 4.1, but adopt both an expanding-widow and a rolling-window design. For direct comparison with the split-sample results, I use data from the first half to estimate the initial set of all relevant parameters for both designs, including the eigenvectors, the PC returns, the factor loadings on the PCs, and the parameters from the predictive regressions. I use these estimates to construct return forecasts for 1996:01. I repeat this procedure each month until the end of the sample period to obtain forecasts for the rest of the sample. In the expanding-window design, I use all data up to the month that I re-estimate the parameters. In the rolling window design, I use a rolling window equal to the initial training period.

[10] An example of a potential structural change in the predictive relation is that value spreads (e.g., the book-to-market ratio) are no longer informative about the future performance of the value anomaly in recent decades (e.g., Asness, 2016). Another pontential source of structural instability is time-varying factor premia. For example, Fama and French (2021) document a difference in the premium of the value factor before and after 1991, or McLean and Pontiff (2016) show that the factor premia on average are lower out of sample and post publication.

[11] McCracken's (2007) results are based on the tests that use the MSPE-normal statistic (Diebold and Mariano, 1995, West, 1996). The out-of-sample test in this paper uses the MSPE-adjusted statistic, which Clark and West (2007) show to have higher power for rolling and recursive regressions. However, it is unclear about the power of this test under the split-sample estimation scheme.

[12] I also use a reverse-sample design to conduct an "out-of-sample" test. In particular, I use the second half of the sample for estimation, and the first half of the sample for evaluation. Table A3 in the Appendix summarizes results from this test that uses BM as predictor. The general conclusion about the performance of the PC portfolio approach across estimation designs in later analyses does not change materially.

With three estimation designs and nine predictors, I examine the out-of-sample predictive performance of the PC portfolio approach for 27 cases. The base case uses *BM* as predictor and the split-sample design for estimation. Figure 1 summarizes the distribution of the R^2_{OS} for all cases. The black, red, and blue bar colors indicate results for the split-sample, the expanding-window, and the rolling-window design, respectively. Panels A, B, and C show the mean, the median, and the total R^2_{OS} across 92 factors for each case, respectively.

First, I investigate the results using *BM* as predictor. Panel A reveals a remarkable difference in the mean R^2_{OS} between the split-sample scheme and the other two schemes. Compared to the mean of 0.64% under the split-sample scheme, the statistic is negative using the expanding- (-0.11%) and rolling-window (-0.31%) estimators. The result suggests that return forecasts for a factor that use the value spread is worse than the prevailing mean forecasts on average. Panel B confirms this conclusion as the median R^2_{OS} is 0.07% and -0.23% under the recursive and rolling-window schemes, respectively. The total R^2_{OS} reported in Panel C reduces more than three times from the split-sample scheme, further suggesting that the evidence in favor of factor return predictability using book valuation is unstable.

The large gap in the predictive performance between the split-sample scheme and the other two schemes is not exclusive to *BM* as predictor. Across all 27 cases, the split-sample results are more likely to indicate stronger evidence of predictability, but such evidence either deteriorate or disappear under the recursive and rolling-window designs. Among nine predictors, only *MOM1* and *SENT* appear to indicate some evidence of factor return predictability. However, the variation of evidence is still large across the schemes: the median R^2_{OS} for *MOM1* and *SENT* is positive under the recursive design but is either marginally positive or negative under the rolling-window scheme. Figure 2 shows the number of significant predictions from the out-of-sample tests across three estimation designs. Panel A shows that the results for the R^2_{OS} tests at the 5% level of statistical significance. When the predictor is value spread, the number of predictable factors reduces by half from 50 under the split-sample scheme to 25 under the rolling-window scheme. The deteriorating performance is more pronounced with the number of significant predictions at the 1% level. For most predictors, only less than a handful of factors are predictable with strong statistical evidence.

4.3 Performance of the optimal timing portfolio and variance of the stochastic discount factor

HKS show that strong factor predictability implies better performance for the optimal timing portfolio relative to its static version. Because the optimal portfolio is equivalent to the stochastic discount factor (SDF), they also show that the variance of the implied SDF is substantially larger, challenging theoretical models to explain this property in the cross-section of asset returns.

I investigate how the optimal timing portfolio performs and the implied SDF varies under the PC portfolio approach across the three estimation designs. The optimal timing portfolio is constructed based on the market factor and the first five PC portfolios. To focus on the incremental benefit of timing anomaly portfolios for each of nine predictors, I use historical means for market forecasts (i.e., no timing for the market factor), and use the PC portfolio approach to make expected return forecasts for the PCs. The optimal weight for this anomaly timing (AT) strategy is estimated as

$$\hat{\omega}_{\mathrm{AT},t} = (\gamma \hat{\mathbf{\Sigma}}_t)^{-1}[\hat{R}_{MKTRF,t+1}, \hat{PC}^1_{t+1}, \ldots, \hat{PC}^5_{t+1}]', \tag{10}$$

where is $\hat{\mathbf{\Sigma}}_t$ the estimated covariance matrix of forecast errors, $\hat{R}_{MKTRF,t+1}$ is the historical mean of the market factor, $\hat{PC}^k_{t+1}$ is the kth PC return forecast, and γ is a risk aversion parameter equal to one. I estimate the conditional variance of the SDF m_{t+1} as

$$\mathrm{var}_t(\hat{m}_{t+1}) = \hat{\omega}'_{\mathrm{AT},t} \hat{\mathbf{\Sigma}}_t \hat{\omega}_{\mathrm{AT},t}. \tag{11}$$

Figure 3 shows the relative performance of the optimal AT portfolio to the optimal factor investing portfolio in which return forecasts are historical means. Panel A (B) shows the incremental change of the annualized Sharpe ratio (Certainty Equivalent Return) in percentage. Under the split-sample design, anomaly timing with *BM* increases the economic value of the optimal portfolio, especially for the utility gains. However, the gains decrease substantially under the other two estimation schemes. Relative to anomaly investing, anomaly timing produces lower Sharpe ratio for the optimal portfolio with almost no differences in utility gains. Across nine predictors, the Sharpe ratio of the optimal AT portfolio is systematically lower than that of the static strategy. For utility gains, only factor momentum appears to yield consistently higher value for either recursive or rolling-window design.

Panel C shows the average of estimated conditional variance of the SDF. With *BM* as predictor, the average SDF variance increases by 80% under the split-sample design, consistent with HKS's argument that incorporating conditional information increases the variance of the SDF to a great extent. However, the increase reduces to about 40% under the other two estimation designs. Across nine predictors with the exception of factor momentum, the increases in the average of SDF variance are typically less than 40% across designs.

Overall, the analyses show that when common estimation schemes are employed, the evidence of factor return predictability is significantly less impressive and even non-existent in most cases. The performance of the optimal timing strategy is also not systematically better than the pure factor investing strategy. Since it is uncommon for empiricists and practitioners to adopt a half-split estimation design in testing predictive ability of different models, the results show that models that adopt the PC portfolio approach are not stable. I discuss potential reasons for the instable performance of the PC portfolio approach in the next section.

4.4 Explanations for instable out-of-sample performance

The recursive and rolling regressions perform worse than the split-sample regressions for two potential reasons: (i) estimation risk in the regressions of the PC returns on their predictors, and (ii) structural breaks in the relations between future factor returns and their predictors.

Estimation risk is a prominent concern in the forecasting literature of asset returns (see, e.g., Stambaugh, 1999). An easy solution to mitigate this risk is to have more data for estimation, which aims at reducing the variance of forecast errors. However, the gain in sample size can be negative if additional information has large variance or noise (e.g., Meng and Xie, 2014). The intuition is that the OLS estimates of the predictive parameters equally weight observations with large and small noise; hence, can produce less precise forecasts. In this context, it is possible that the predictors are more noisy in the second half of the sample.[13] This in turn increases the estimation risk under the designs that use data in this period. While this explanation might be reasonable for some predictors and factors, it does not appear to be a satisfactory reason for the systematic weaker performance across all predictors and factors.

A more reasonable explanation is potential structural changes in the relation between the PC returns and their predictors. In Figure A3 I examine the general predictive performance of the PC portfolio approach at each out-of-sample month during the evaluation period. For brevity and focus, I show the results for *BM* (Panel A), *MOM1* (Panel B), and *VOL* (Panel C). I plot the standardized cumulative sum squares of errors (SSE) difference, which is a rough proxy for total R^2_{OS} at each out-of-sample month for three designs. The solid, dashed, and dotted line indicate the split-sample, recursive, and rolling-window estimation design, respectively. The shaded areas indicate NBER-dated recessions. For *BM* in Panel A, the standardized cumulative SSE difference is almost the same across the designs from the beginning of the out-of-sample period (1996:01) until early 2000. However, it appears that the dot-com bubble and the followed 2001 financial crisis had strong impact on the return forecasts across three estimation schemes. While the split-sample regressions produce better forecasts than the historical mean benchmark, the recursive and rolling regressions yield worse forecasts. Interestingly, the cumulative SSE difference lines are almost parallel after the 2001 financial crisis. Therefore, it is possible that the lack of modelling the structural changes around 2000 for the recursive and rolling regressions leads to the disparity in performance across these models. A similar pattern is observable for *VOL* in Panel C. For *MOM1* in Panel B, structural breaks appear to happen during the 2008 financial crisis and the COVID-19 crisis.

[13] Gonçalves and Leonard (2023), for example, show that book value is not a good proxy for fundamental information in recent decades. An implication is that the book-to-market ratio can be a noisy predictor of future returns in more recent periods.

Structural changes in the relation between the PC returns and their predictors can stem from the underlying relation among the individual factors. To investigate this possibility, I perform individual in-sample predictive regressions for each factor and predictor for two sub-periods and record how many predictive relations are stable in both halves of the sample. I evaluate the stability based on consistency in predictive sign and statistical significance. Table A4 in the Appendix reports the results. Column (7) shows the number of predictions that is consistent from the first to the second half across nine predictors. For most cases, less than a handful of factors that are predictable with standard statistical evidence in both sub-samples. For example, *BM* only predicts seven out of 92 factors consistently for the two periods. These results imply that the predictive relations among individual factors are subject to structural breaks that might affect the stability of factor timing models.

Next I use Bai and Perron's (1998) and Bai and Perron's (2003) tests to formally investigate structural changes in the predictive relation between the PC portfolio returns and their predictors. For brevity, I use the book-to-market ratio as predictor. Table A5 Panel A in the Appendix summarizes results from these tests for the predictive regressions. I find statistical evidence that three out of five PC portfolios exhibit structural changes in their relation with *BM*. PC1 and PC4 have two breaks, and PC2 have one break during the sample period. Consistent with the observation that cumulative SSE difference has dramatic changes in the early 2000s, the structural break tests show that PC1, PC2, and PC4 have a break between 2000 and 2001. The predictive coefficient estimates also experience significant difference before and after this break date. For example, the predictive estimate for PC2 is negative and statistically insignificant between 1988:11 and 2000:08, but is positive and statistically significant afterward.

Under the PC portfolio approach, the individual factor predictability relies not only on the predictability of the PC returns, but also on factor loadings on the PCs. Therefore, I use another structural break test, developed by Chen, Dolado, and Gonzalo (2014), to examine whether the factor loadings (or the eigenvectors of the covariance of factor returns) are time-varying. Chen et al. (2014) develop a simple test that uses the *supWald* statistic to detect one big unknown structural break in factor loadings of large factor models. The procedure regresses the first PC on the other PCs, and tests the null hypothesis that there is no structural changes in the factor loadings via the relations among the PCs.[14] Table A5 Panel B in the Appendix summarizes results from these tests. I find statistical evidence that there is a big structural change in the factor loadings among the PCs in 2000:05, which is consistent with my earlier findings that structural breaks in the predictive regressions are prominent in the early 2000s. In the context of factor timing, the existence of

[14] By construction, the PCs are orthogonal to each other so the regression of the first PC on the other PCs is not meaningful when the full-sample data is used. However, the coefficient estimates from the regression can be different from 0 in subsamples if there are breaks.

large structural breaks creates a difficulty for a stable performance of forecasting models that use individual predictors.

The discussion so far highlights structural changes as a critical hurdle in predicting factor returns for the PC portfolio approach. A clear solution is to incorporate the forecasting of structural breaks into the forecasting process. However, this exercise is not always implementable due to the difficulty of estimating break points in real time and the bias-variance trade-off in break estimation methods (see, e.g., Pesaran and Timmermann, 2005 or Boot and Pick, 2020). A simpler solution to reduce the impact of structural changes on forecast accuracy is combining forecasts (e.g., Diebold and Pauly, 1987). In the next section I examine the evidence of factor return predictability using different shrinkage methods that combine forecasts from all predictors.

5 Predicting factor returns with shrinkage methods

5.1 Summary of out-of-sample predictive performance

First I examine the predictive ability of five shrinkage methods, including forecast combination (*FC*), discount mean square of prediction errors (*DMSPE*), predictor average (*Average*), principal component regression (*PCR*), and partial least squares (*PLS*).[15] The set of predictors include *Book-to-market ratio*, *Industry-adjusted book-to-market ratio*, *Issuer-repurchaser spread*, *One-month momentum*, *12-month momentum*, *Volatility*, *Reversal*, *Characteristic spread*, and five components of *Sentiment index*. I adopt the same expanding-window design for the out-of-sample tests from previous sections.

For *FC*, *Average*, *PCR*, and *PLS*, I use an initial training window equal to the first half of the sample. For *DMSPE* that requires holdout periods, I use an initial training window from 1970:01 to 1985:12, and the subsequent 120 months as the holdout period. I expand the training window each month to estimate forecasts for the rest of the sample. The out-of-sample evaluation period is from 1996:01 to 2021:12.

Table 4 Panel A reports the summary statistics for R^2_{OS}, total R^2_{OS}, and the number of significant predictions for five methods. For comparison, Panel B reports the lowest and the highest values of the same statistics across all predictors for the conventional predictive regression and the PC portfolio approach using the same estimation design. Column (1) in Panel A shows that the mean R^2_{OS} for the shrinkage methods is positive and relatively high for all shrinkage methods, ranging from 0.52% to 1.08%. The median under the *PCR* method is 1.08%, about more than 50% higher than that of the highest value under the PC portfolio approach. For simple combination methods (*FC* and *DMSPE*), the 25th percentile is about 0.15%, suggesting

[15] Advanced machine learning methods are also potential. However, such methods are more suitable for forecasting exercises that involve a relatively large number of predictors compared to the number of observations. Given that my tests use less than ten predictors and the main interest is on the relative performance of the shrinkage methods, I adopt the most common methods among shrinkage techniques.

that at least three-quarters of 92 factors have forecasts better than the prevailing mean forecasts. For the predictive performance of these methods as a whole, both *PCR* and *PLS* perform equally well with total R^2_{OS} at 1.30% and 0.91%, respectively. These statistics represent a substantial improvement over the highest from the predictive regression (0.68%) and the PC portfolio approach (0.63%). The number of significant predictions also improves using most of the shrinkage methods. Column (6) shows that 81 out of 92 factors have forecasts better than the prevailing mean forecasts under the *FC* and *DMSPE* methods. Except for *Average*, the results show that more than half of the factors have positive R^2_{OS} that are statistically significant at least at the 10% level.

5.2 Predictive performance across individual factors

To examine which factors are more (less) likely to be predictable, Figure 4 shows the R^2_{OS} for 92 factors using the shrinkage methods. Each plot shows the R^2_{OS} in percent for factors in one of six categories. The six colors indicate the six ranges of R^2_{OS} among all factors with darker red (green) indicating lower (higher) R^2_{OS}. The dominance of the green color suggests that the shrinkage methods generally provide evidence in favor of factor return predictability for all factor categories.

Among six momentum factors, intermediate momentum is predictable with high R^2_{OS} under most methods. Industry momentum, prior 6- and 12-month momentum appear to exhibit weak predictability. Among value factors, 15 out of 20 factors are predictable by all methods. The value factor is highly predictable with R^2_{OS} up to 1.14% under the *PCR* method. This result suggests that real-time investors can benefit from timing the value factor even when its unconditional average return is low (Fama and French, 2021). The profitability factors appear to be strongly predictable for most shrinkage methods with R^2_{OS} greater than 0.50%. I also observe a similar pattern for factors in the intangibles and trading frictions groups, although the R^2_{OS} is slightly lower on average. The *PCR* and *PLS* methods do not seem to provide strong evidence of predictability for investment factors. However, the simple combination methods (*FC* and *DMSPE*) still show that 12 out of 18 factors have positive R^2_{OS}. Overall, the results show the shrinkage methods generally improve the predictability for all factor categories.

5.3 Direct comparison across forecasting approaches

To further highlight the outperformance of the shrinkage methods over the conventional and the PC portfolio approach, I directly compare how many factors have higher R^2_{OS} under the five shrinkage methods. Panels A and B in Table 5 report the results in comparison to the traditional and the PC portfolio approach, respectively. To offer evidence of statistical significance, I report in brackets the p-value from a one-sided binomial test with the null hypothesis that the proportion is less than half. The results from Panel A show that the number of factors with higher R^2_{OS} under the shrinkage methods is substantially higher compared to the conventional predictive regression, ranging from 44 to 85. The proportion of such factors is statistically greater than 50%, with only a few exceptions. In comparison with the PC portfolio approach, the proportion is still statistically significant for most individual predictors. The PC portfolio approach that uses *BM* underperforms all shrinkage methods. Although the PC portfolio approach appears to perform better for *MOM1* and *SENT*, the *PCR* method still yields higher R^2_{OS} for more than 50% of the factors.

5.4 Economic value of factor timing

In this section I examine the economic value of factor return predictability through the investment performance of timed factor strategies. I also use forecast returns obtained from the shrinkage methods to improve the performance of the PC portfolio approach.

5.4.1 Sharpe ratio

I start with the Sharpe ratio of timed strategies for individual factor portfolios. The main interest is on the performance of strategies in which I use the shrinkage methods to predict factor returns. For comparison, I also present the results that use the conventional and PC portfolio approaches. I adopt the same recursive estimation design from the previous sections to evaluate the performance of timed factor portfolios. For each out-of-sample month $t+1$, I obtain a factor portfolio's real-time weight at month t as

$$w^m_{i,t+1} = \frac{1}{\gamma}\frac{\hat{R}^m_{i,t+1}}{\hat{\sigma}^2_{t+1}}, \tag{12}$$

where $w^m_{i,t+1}$ is the weight of factor portfolio i using forecasting method m for month $t+1$, $\hat{R}^m_{i,t+1}$ is the return forecast of factor portfolio i using forecasting method m for month $t+1$, $\hat{\sigma}^2_{i,t+1}$ is the sample variance of factor portfolio i for month $t+1$, and γ is a risk aversion parameter. I use all data on factor returns up to month t to estimate $\hat{\sigma}^2_{i,t+1}$. I construct the portfolio return for month $t+1$ as $w^m_{i,t+1}R_{i,t+1}$, where $R_{i,t+1}$ is the actual return of factor portfolio i for month $t+1$. The outcome of this procedure is a time series of monthly returns for each timed strategy for the second half of the sample. I set the risk aversion parameter γ to one. To mitigate estimation risk that can yield fluctuating portfolio weights over time, I impose a leverage constraint that the absolute weight on timed portfolios is less than or equal to two. This constraint on portfolio positions also takes into account the implementability of these strategies in practice.

Table 6 reports the distribution of annualized Sharpe ratio from timed factor strategies that use return forecasts from all forecasting methods. Panel A shows the results for the shrinkage methods. To set a base case, I examine the Sharpe ratio from pure factor investing in which optimal weights use the historical means as return forecasts. Row (1) from Panel A shows that the mean Sharpe ratio across 92 factor investing strategies is 0.11. The median is only 0.09, suggesting that almost half of strategies do not have positive returns on average during the evaluation period. The subsequent rows report the performance across five shrinkage methods. The *FC* and *DMSPE* methods produce almost the same distribution of the Sharpe ratio. The mean Sharpe ratio increases more than twice from the factor investing strategies to 0.23. The *Average* method has slightly weaker performance but all distribution statistics still improve compared to the original factor strategies. For the *PCR* and *PLS* methods, the Sharpe ratio improves substantially. For example, the median R^2_{OS} using the *PCR* method increases by more than three times to 0.28.

Next I compare the benefits of factor timing between the shrinkage methods and the first two approaches. Panel B presents the mean and the median Sharpe ratios of the timed strategies using six predictors under the conventional predictive regressions, and the PC portfolio approach. Given that two valuation predictors *BM* and *IND.BM* have weak predictive ability under both forecasting approaches, it is not surprising that their mean Sharpe ratio is only higher than that of factor investing by a small margin. There are higher gains from predicting factor portfolio returns when *MOM1*, *MOM12*, and *SENT* are predictor. However, the results show that there is a lack of substantial difference in the performance between the conventional and the PC portfolio approach. For instance, the mean R^2_{OS} under the PC portfolio method using *MOM1* as predictor is 0.24, a small improvement from 0.22 under the conventional approach. More importantly, most of these strategies underperform advanced shrinkage methods (*PCR* and *PLS*). These results show that the strong predictive ability of the shrinkage methods can translate into significant economic gains for factor timing strategies.

To identify timed factor strategies that have high (low) Sharpe ratio, Figure A4 in the Appendix shows the annualized Sharpe ratio for 92 factors across five shrinkage methods. For each method, I sort the Sharpe ratios into deciles and use different colors from dark red to dark green to indicate their Sharpe ratio magnitude. There are differences in the Sharpe ratio among common factors from leading asset pricing models. While the timed size (SIZE) and gross profitability (PROF) strategies have relatively high Sharpe ratio, the value (VALUE), investment (INV), and 12-month momentum (MOM12) strategies have more modest gains from timing exercises.

5.4.2 Certainty Equivalent Return

Another measure to gauge the economic value of timed factor strategies is Certainty Equivalent Return (CER). I measure the CER for strategy i during the evaluation period as

$$CER_i = \bar{R}_i - \frac{\gamma}{2}\hat{\sigma}_i^2, \tag{13}$$

where $\bar{R}_i$ and $\hat{\sigma}_i^2$ are the average return and sample variance of the timed strategy i, respectively. I adopt the same parameter choices from the previous section by setting the risk aversion parameter γ to one and imposing a leverage constraint on portfolio allocations of two.

Table 7 reports the distribution of annualized CER in percent from timed factor strategies. Panel A shows the results when the forecasting methods are the shrinkage techniques. The first row presents the summary statistics for the factor investing strategies. The mean CER is 0.40%, suggesting that the utility gains from factor investing are low on average. Nevertheless, 25% of the strategies have high utility gains as the 75th percentile is 2.98%. The subsequent rows show that the CER improves significantly when I adopt the shrinkage methods to forecast factor returns. The mean CER is between 2.00% and 4.00%. The utility gains increase by a large magnitude at the higher end of the distribution. For example, 25% of the factors have annualized CER greater than 5.00%.

The utility gains under the shrinkage methods are also substantially higher than that under both the conventional and PC portfolio approaches. Panel B presents the mean and median CER of the timed strategies using these two methods. Consistent with the poor Sharpe ratios from the previous section, most predictors yield low CER. Only *MOM1* and *SENT* produce relatively high mean CER using the PC portfolio approach, at 2.56% and 1.49%, respectively. Nevertheless, these economic gains are substantially lower than those from the *PCR* and *PLS* methods. For example, the mean CER from the *PLS* method is more than twice with investor sentiment as predictor.

Figure A5 in the Appendix shows the annualized CER for 92 factors across five shrinkage methods. Again for each method, I sort the CERs into deciles and use different colors from dark red to dark green to indicate their magnitude. I find that a positive relationship between CER among common factor strategies and their R^2_{OS} and Sharpe ratios. For example, the high R^2_{OS} and Sharpe ratios for the timed size and gross profitability factors turn into large economic value, while the value and timed 12-month momentum strategy have subtle economic gains.

5.4.3 Optimal timing portfolio and the stochastic discount factor

Previous analyses suggest that the shrinkage methods improve factor predictability over the PC portfolio approach. I examine whether this improvement has any implications for the performance of the optimal timing portfolio and the variance of the implied SDF. In particular, I use the return forecasts of individual factors from the shrinkage methods to estimate the expected returns for the PC portfolios, and construct the optimal anomaly timing portfolio as in Section 4.3. To highlight the improvement in the PC return forecasts, I adopt a homogeneous covariance matrix assumption. Consequently, the optimal weight is estimated as

$$\hat{\omega}_{\text{AT},t} = (\gamma\hat{\mathbf{\Sigma}})^{-1}[\hat{R}_{MKTRF,t+1}, \hat{PC}^{k}_{t+1}, \ldots, \hat{PC}^{k}_{t+1}]', \tag{14}$$

where $\hat{\mathbf{\Sigma}}$ is the estimated covariance matrix of forecast errors using the first half of the sample with historical means as return forecasts, $\hat{R}_{MKTRF,t+1}$ is the historical mean of the market factor, $\hat{PC}^{k}_{t+1}$ is the kth PC return forecast, and γ is a risk aversion parameter equal to one.

Table 8 shows the performance of the optimal anomaly timing portfolio under this approach. For comparison, the first row shows the average performance under the PC portfolio approach across nine predictors, and the subsequent rows show the performance for five shrinkage methods. Column (1) shows that the Sharpe ratio of the optimal timing portfolio improves when using return forecasts from shrinkage methods. Simple methods such as *FC* and *DMSPE* can improve the Sharpe ratio by about 30%. Column (2) shows that this improvised method of estimating PC's expected returns delivers substantial information ratio. There is also evidence of larger economic gains in Column (3). Finally, the variance of the implied SDF varies across methods with advanced techniques such as *PLS* producing more volatile SDF.

5.5 Robustness check

5.5.1 Choice of estimation designs

To address concerns that different training windows or sample-split points can affect the out-of-sample predictive performance (e.g., Inoue, Jin, and Rossi, 2017), I repeat the out-of-sample tests using the shrinkage methods with two choices of initial training periods: 240 and 360 months. Table A6 Panels A and B report the results. Using 240 months as the initial training sample does not change the original results significantly. The mean R^2_{OS} is between 0.65% and 1.15%. The total R^2_{OS} can be as high as 1.35%. The predictive performance slightly decreases when I use 360 months as the initial training window. However, the evidence in favor of factor return predictability remains clear.

I also perform the out-of-sample tests using the rolling-window estimation scheme. I use the first half of the data as the initial training sample, and use this window as the rolling window. The predictive ability is strong for the *FC* and *PCR* methods. For example, the total R^2_{OS} under the *FC* and *PCR* methods is 0.75% and 0.82%, respectively. The *PLS* method exhibits larger decrease in performance. Nevertheless, the total R^2_{OS} remains high at 0.34%.

5.5.2 Variable exclusion and inclusion

The predictive ability of the shrinkage approach can rely on a few predictor variables. To assess how each of six predictors affects the predictive performance, I repeat the out-of-sample tests in Section 5.2 but exclude a predictor each time. I adopt the same expanding window estimation design from Section 5.2. Table A7 in the Appendix reports the results. Panel A reports the mean and the total R^2_{OS}, while Panels B and C report the annualized mean (median) Sharpe ratios and CER, respectively. When I remove all five components of *SENT*, the *PCR* and the *PLS* methods have weaker performance. These results suggest that *SENT* has a significant contribution to predicting factor returns. Nevertheless, simple combination methods (*FC* and *DMSPE*) still provide strong evidence. For example, the total R^2_{OS} under the *FC* method when removing *SENT* is still 0.66%, and the mean Sharpe ratio across timed strategies is 0.23.

Finally, I examine whether macroeconomic variables affect the performance of the shrinkage methods. I use 14 economic variables from Goyal and Welch's (2008) study.[16] I repeat the out-of-sample tests using the same expanding window estimation design by adding one of 14 variables to the set of the original predictors each time. I also examine an extreme case when I add all 14 variables. Table A8 reports the mean and the median R^2_{OS} across the cases for all five shrinkage methods. Most results do not change significantly from the base case. However, there is a considerable decrease when I add all 14 variables to my original set of predictors. For example, the total R^2_{OS} under the *FC* method reduces from 0.60% in the base case to 0.42%. This result implies that the macroeconomic variables bring noise to the forecasting of factor returns. Nevertheless, I still find strong evidence in favor of predictability from the more advanced techniques such as *PCR*.

6 Conclusions

The predictability of factor returns has strong implications for both asset pricing research and industry applications, leading to a growing search among academics to find variables that predict factor returns. Most studies use the conventional predictive regression approach to document evidence of predictability, while recent papers avoid predicting many factors by focusing on predicting PC portfolio returns. Regardless of the approach, most studies differ in their choices about test factor, predictor variable, sample period, return horizon, and estimation design. Compounded by concerns of consistent out-of-sample performance, results from prior studies create a difficulty to draw a broad conclusion about the predictive ability of prior variables.

Using a broad sample of 92 equity factors and nine prominent predictors of factor returns, I do not find systematic evidence in favor of factor return predictability. The out-of-sample predictive ability is weak and

[16] I collect data on these variables from https://sites.google.com/view/agoyal145. I thank Amit Goyal for making the data available.

even non-existent for most predictors under the conventional predictive regression approach. The evidence remains weak under the PC portfolio approach in which factor returns share common structures. I attribute the poor out-of-sample performance of these two approaches to model instabilities, and find evidence consistent with this explanation.

I explore several shrinkage techniques that combine signals from all predictors to mitigate the impact of structural instability. I find more consistent evidence in favor of predictability across the methods. I also find that the predictability exists in all factor categories. Moreover, the strong predictability routinely translates into significant improvement in the performance of factor timing portfolios. Taken as a whole, the shrinkage approach offers a conclusive evidence for factor timing.

References

Akbas, F., Armstrong, W. J., Sorescu, S., and Subrahmanyam, A. (2015). Smart money, dumb money, and capital market anomalies. *Journal of Financial Economics*, *118*(2), 355–382.

Amihud, Y. (2002). Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets*, 31–56.

Arnott, R. D., Beck, N., and Kalesnik, V. (2016). *Timing 'Smart Beta' Strategies? Of Course! Buy Low, Sell High!* (SSRN Scholarly Paper No. ID 3040956). Rochester, NY: Social Science Research Network. doi: 10.2139/ssrn.3040956

Asness, C. S. (2016). The siren song of factor timing. *Journal of Portfolio Management*, *42*(5), 1–6.

Asness, C. S., Chandra, S., Ilmanen, A., and Israel, R. (2017). Contrarian factor timing is deceptively difficult. *Journal of Portfolio Management*, *43*(5), 72–87.

Baba-Yara, F., Boons, M., and Tamoni, A. (2021). Value return predictability across asset classes and commonalities in risk premia. *Review of Finance*, *25*(2), 449–484.

Bai, J., and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, 47–78.

Bai, J., and Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, *18*(1), 1–22.

Baker, M., and Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *Journal of Finance*, *61*(4), 1645–1680.

Bessembinder, H., Burt, A., and Hrdlicka, C. M. (2022). Factor returns and out-of-sample alphas: Factor construction matters. *Working Paper*.

Bessembinder, H., Burt, A., and Hrdlicka, C. M. (2024). More factors matter and factors matter more than you might think: The role of time variation in factor premia. *Working Paper*.

Boot, T., and Pick, A. (2020). Does modeling a structural break improve forecast accuracy? *Journal of Econometrics*, *215*(1), 35–59.

Cakici, N., Fieberg, C., Metko, D., and Zaremba, A. (2024). Do anomalies really predict market returns? new data and new evidence. *Review of Finance*, *28*(1), 1–44.

Cakici, N., Fieberg, C., Neumaier, T., Poddig, T., and Zaremba, A. (2024). Pockets of predictability: A replication. *Journal of Finance, forthcoming*.

Campbell, J. Y., and Thompson, S. B. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies*, *21*(4), 1509–1531.

Chen, L., Dolado, J. J., and Gonzalo, J. (2014). Detecting big structural breaks in large factor models. *Journal of Econometrics*, *180*(1), 30–48.

Clark, T. E., and West, K. D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, *138*(1), 291–311.

Cochrane, J. H. (2011). Presidential address: Discount rates. *Journal of Finance*, *66*(4), 1047–1108.

Cohen, R. B., Polk, C., and Vuolteenaho, T. (2003). The value spread. *Journal of Finance*, *58*(2), 609–641.

Da, Z., Liu, Q., and Schaumburg, E. (2014). A closer look at the short-term return reversal. *Management Science*, *60*(3), 658–674.

Dai, W., and Dong, A. (2023). Another look at timing the equity premiums. *Working Paper*.

Diebold, F. X., and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, *13*, 253–263.

Diebold, F. X., and Pauly, P. (1987). Structural change and the combination of forecasts. *Journal of Forecasting*, *6*(1), 21–40.

Dong, X., Kang, N., and Peress, J. (2022). Fast and slow arbitrage: The predictive power of capital flows for factor returns.

Working Paper.

Ehsani, S., and Linnainmaa, J. T. (2022). Factor momentum and the momentum factor. *Journal of Finance*, *77*(3), 1877–1919.

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

Fama, E. F., and French, K. R. (2008). Dissecting anomalies. *Journal of Finance*, *63*(4), 1653–1678.

Fama, E. F., and French, K. R. (2021). The value premium. *The Review of Asset Pricing Studies*, *11*(1), 105–121.

Fan, M., Li, Y., Liao, M., and Liu, J. (2022). A reexamination of factor momentum: How strong is it? *Financial Review*, *57*(3), 585–615.

Favero, C. A., Melone, A., and Tamoni, A. (2022). Factor models with drifting prices. *Working Paper*.

Gonçalves, A. S., and Leonard, G. (2023). The fundamental-to-market ratio and the value premium decline. *Journal of Financial Economics*, *147*(2), 382–405.

Goyal, A., and Welch, I. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, *21*(4), 1455–1508.

Goyal, A., Welch, I., and Zafirov, A. (2023). *A Comprehensive 2022 Look at the Empirical Performance of Equity Premium Prediction.*

Greenwood, R., and Hanson, S. G. (2012). Share issuance and factor timing. *Journal of Finance*, *67*(2), 761–798.

Gu, S., Kelly, B. T., and Xiu, D. (2020). Empirical asset pricing via machine learning. *Review of Financial Studies*, *33*(5), 2223–2273.

Gu, S., Kelly, B. T., and Xiu, D. (2021). Autoencoder asset pricing models. *Journal of Econometrics*, *222*(1), 429–450.

Gupta, T., and Kelly, B. T. (2019). Factor momentum everywhere. *Journal of Portfolio Management*, *45*(3), 13–36.

Haddad, V., Kozak, S., and Santosh, S. (2020). Factor timing. *Review of Financial Studies*, *33*(5), 1980–2018.

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

Hou, K., Xue, C., and Zhang, L. (2020). Replicating anomalies. *Review of Financial Studies*, *33*(5), 2019–2133.

Huang, S., Song, Y., and Xiang, H. (2022). Noise trading and asset pricing factors. *Working Paper*.

Inoue, A., Jin, L., and Rossi, B. (2017). Rolling window selection for out-of-sample forecasting with time-varying parameters. *Journal of Econometrics*, *196*(1), 55–67.

Inoue, A., and Kilian, L. (2005). In-sample or out-of-sample tests of predictability: Which one should we use? *Econometric Reviews*, *23*(4), 371–402.

Kagkadis, A., Nolte, I., Nolte, S., and Vasilas, N. (2023). Factor timing with portfolio characteristics. *Review of Asset Pricing Studies*.

Kelly, B. T., Malamud, S., and Pedersen, L. H. (2023). Principal portfolios. *Journal of Finance*.

Kelly, B. T., and Pruitt, S. (2013). Market expectations in the cross-section of present values. *Journal of Finance*, *68*(5), 1721–1756.

Kelly, B. T., Pruitt, S., and Su, Y. (2019). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics*, *134*(3), 501–524.

McCracken, M. W. (2007). Asymptotics for out of sample tests of granger causality. *Journal of Econometrics*, *140*(2), 719–752.

McLean, R. D., and Pontiff, J. (2016). Does academic research destroy stock return predictability. *Journal of Finance*, *71*(1), 5–32.

Meng, X. L., and Xie, X. (2014). I got more data, my model is more refined, but my estimator is getting worse! am i just dumb? *Econometric Reviews*, *33*(1-4), 218–250.

Moreira, A., and Muir, T. (2017). Volatility-managed portfolios. *Journal of Finance*, *72*(4), 1611–1644.

Morningstar. (2022). A Global Guide to Strategic-Beta Exchange-Traded Products. Technical Report.

Moskowitz, T. J., Ooi, Y. H., and Pedersen, L. H. (2012). Time series momentum. *Journal of financial economics*, *104*(2), 228–250.

Neuhierl, A., Randl, O., Reschenhofer, C., and Zechner, J. (2023). Timing the factor zoo. *Working Paper*.

Newey, W. K., and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, *55*(3), 703–708.

Novy-Marx, R. (2014). Predicting anomaly performance with politics, the weather, global warming, sunspots, and the stars. *Journal of Financial Economics*, *112*(2), 137–146.

Pesaran, M. H., and Timmermann, A. (2002). Market timing and return prediction under model instability. *Journal of Empirical Finance*, *9*(5), 495–510.

Pesaran, M. H., and Timmermann, A. (2005). Small sample properties of forecasts from autoregressive models under structural breaks. *Journal of Econometrics*, *129*(1-2), 183–217.

Rossi, B. (2013). Chapter 21 - advances in forecasting under instability. In G. Elliott and A. Timmermann (Eds.), *Handbook of economic forecasting* (Vol. 2, pp. 1203–1324). Elsevier.

Rossi, B., and Inoue, A. (2012). Out-of-sample forecast tests robust to the choice of window size. *Journal of Business & Economic Statistics*, *30*(3), 432–453.

Stambaugh, R. F. (1999). Predictive regressions. *Journal of Financial Economics*, *54*(3).

Stambaugh, R. F., Yu, J., and Yuan, Y. (2012). The short of it: Investor sentiment and anomalies. *Journal of Financial Economics*, *104*(2), 288–302.

West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica*, 1067–1084.

Figure 1. Factor prediction with principal component portfolio approach: Out-of-sample $\mathbf{R^2}$
This figure summarizes the distribution of the R^2_{OS} from the out-of-sample tests of factor returns under the principal component (PC) portfolio approach. The sample period is from 1970:01 to 2021:12. The number of factors is 92. Additional detail on the factors is in Table A1 in the Appendix. The nine predictors are *Book-to-market ratio (BM)*, *Industry-adjusted book-to-market ratio (IND.BM)*, *Issuer-repurchaser spread (ISSREP)*, *One-month momentum (MOM1)*, *12-month momentum (MOM12)*, *Volatility (VOL)*, *Characteristic spread (CS)*, *Long-run reversal (LRREV)*, and *Sentiment index (SENT)*. For *SENT*, I reconstruct the series for each period from its underlying five components using the same recursive design of the out-of-sample tests. Predictor construction detail is described in Table A2 in the Appendix. I obtain a PC-based return forecast for factor i as

$$\hat{R}_{i,t+1} = \sum_{k=1}^{5} \hat{\omega}^k_{i,t+1} \widehat{PC}^k_{t+1},$$

where $\hat{R}_{i,t+1}$ and $\widehat{PC}^k_{t+1}$ are the excess return forecasts of factor portfolio i and PC portfolio k in month $t+1$, respectively. $\hat{\omega}^k_{i,k,t+1}$ is the loading of factor portfolio i on PC portfolio k from the PC estimation. Panels A, B, and C show the mean, the median, and the total R^2_{OS} across 92 factors, respectively. The forecasting designs are split-sample (black bar), expanding-window (red bar), and rolling-window (blue bar). For all designs, I use an initial training window equal to the first half of the sample (1970:01 - 1995:12). For the split-sample design, I use the parameter estimates to construct forecasts for the second half. For expanding- (rolling-)window design, I expand (roll) the training window each month to update parameter estimates and construct forecasts for the rest of the sample. The rolling window is equal to the initial training window. The out-of-sample evaluation period is from 1996:01 to 2021:12. I calculate Campbell and Thompson's (2008) R^2_{OS} using Equation 8. The total R^2_{OS} is calculated using Equation 9. I use Clark and West's (2007) procedure for the R^2_{OS} tests.

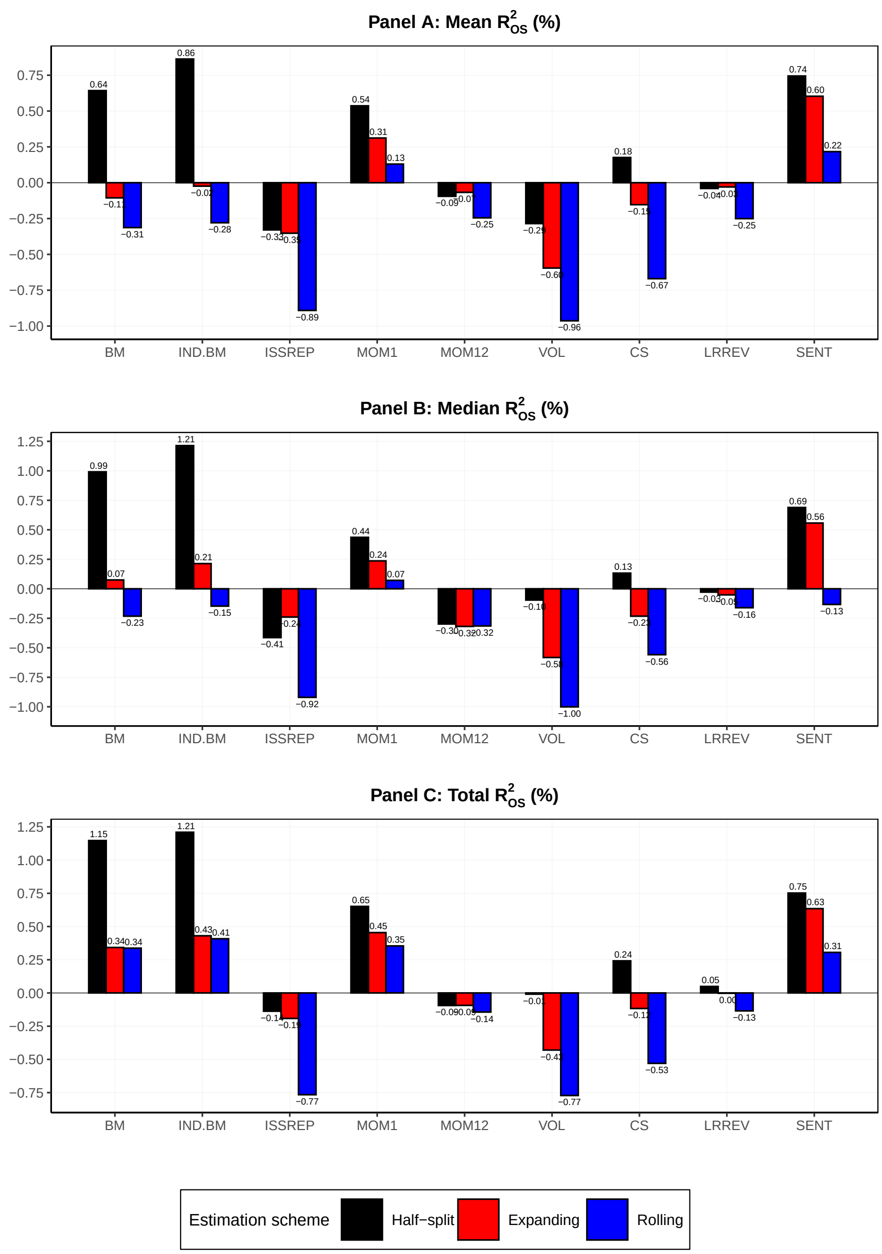

Figure 2. Factor prediction with principal component portfolio approach: Number of significant predictions
This figure shows the number of significant predictions from the out-of-sample tests of factor returns under the principal component (PC) portfolio approach. The sample period is from 1970:01 to 2021:12. The number of factors is 92. Additional detail on the factors is in Table A1 in the Appendix. The nine predictors are *Book-to-market ratio (BM)*, *Industry-adjusted book-to-market ratio (IND.BM)*, *Issuer-repurchaser spread (ISSREP)*, *One-month momentum (MOM1)*, *12-month momentum (MOM12)*, *Volatility (VOL)*, *Characteristic spread (CS)*, *Long-run reversal (LRREV)*, and *Sentiment index (SENT)*. For *SENT*, I reconstruct the series for each period from its underlying five components using the same recursive design of the out-of-sample tests. Predictor construction detail is described in Table A2 in the Appendix. I obtain a PC-based return forecast for factor i as

$$\hat{R}_{i,t+1} = \sum_{k=1}^{5} \hat{\omega}^{k}_{i,t+1} \widehat{PC}^{k}_{t+1},$$

where $\hat{R}_{i,t+1}$ and $\widehat{PC}^{k}_{t+1}$ are the excess return forecasts of factor portfolio i and PC portfolio k in month $t+1$, respectively. $\hat{\omega}^{k}_{i,k,t+1}$ is the loading of factor portfolio i on PC portfolio k from the PC estimation. Panels A and B show the number of factors with significant predictions based on R^2_{OS} at 5% and 1% levels, respectively. The forecasting designs are split-sample (black bar), expanding-window (red bar), and rolling-window (blue bar). For all designs, I use an initial training window equal to the first half of the sample (1970:01 - 1995:12). For the split-sample design, I use the parameter estimates to construct forecasts for the second half. For expanding- (rolling-)window design, I expand (roll) the training window each month to update parameter estimates and construct forecasts for the rest of the sample. The rolling window is equal to the initial training window. The out-of-sample evaluation period is from 1996:01 to 2021:12. I calculate Campbell and Thompson's (2008) R^2_{OS} using Equation 8. I use Clark and West's (2007) procedure for the R^2_{OS} tests.

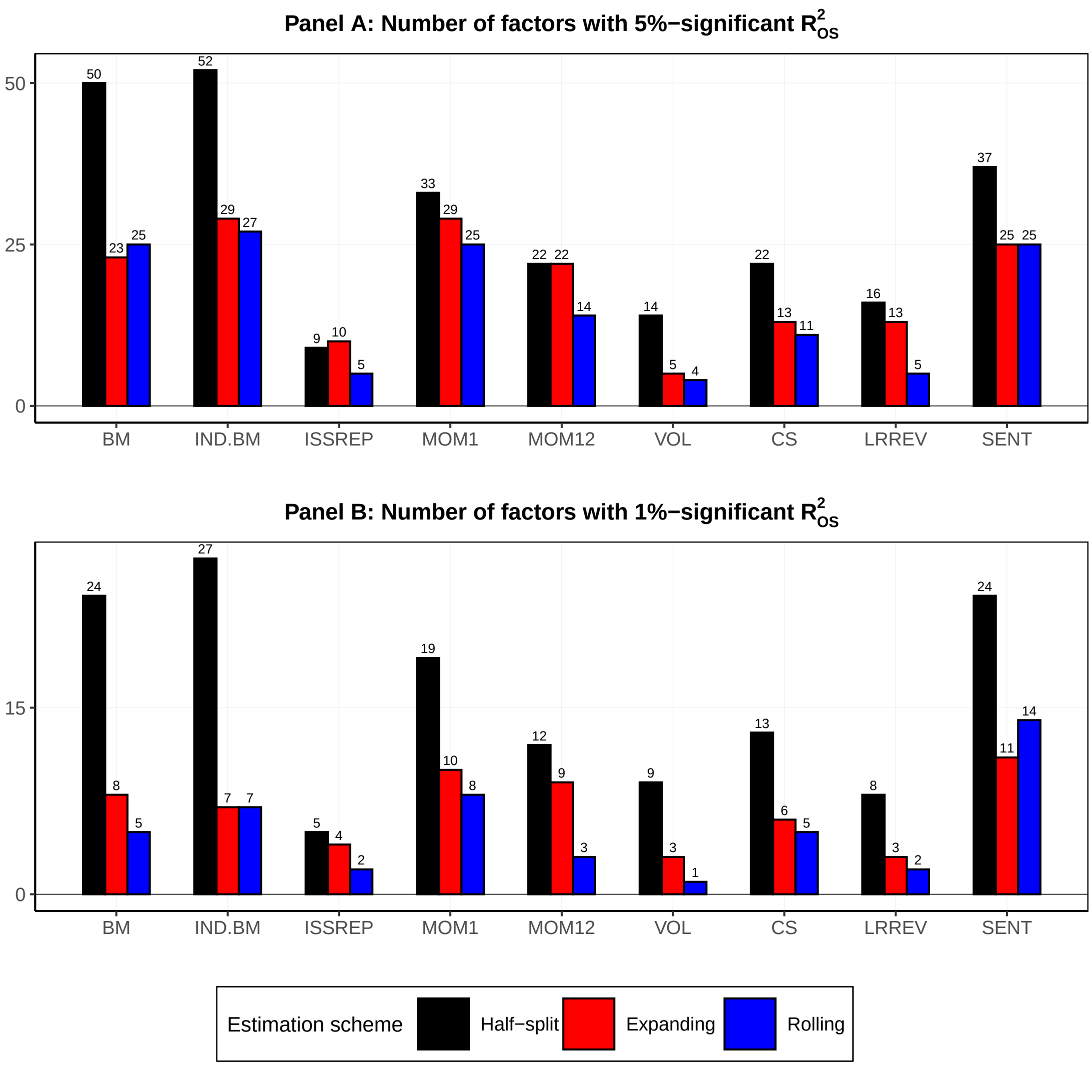

Figure 3. Factor prediction with principal component portfolio approach: Optimal anomaly timing portfolio
This figure shows performance of the optimal anomaly timing portfolio under the principal component (PC) portfolio approach relative to the optimal factor investing portfolio across three estimation designs. The sample period is from 1970:01 to 2021:12. The number of factors is 92. Additional detail on the factors is in Table A1 in the Appendix. The nine predictors are *Book-to-market ratio (BM)*, *Industry-adjusted book-to-market ratio (IND.BM)*, *Issuer-repurchaser spread (ISSREP)*, *One-month momentum (MOM1)*, *12-month momentum (MOM12)*, *Volatility (VOL)*, *Characteristic spread (CS)*, *Long-run reversal (LRREV)*, and *Sentiment index (SENT)*. For *SENT*, I reconstruct the series for each period from its underlying five components using the same recursive design of the out-of-sample tests. Predictor construction detail is described in Table A2 in the Appendix. I obtain the optimal portfolio based on the first five PC portfolios with anomaly timing and the market factor without timing. The optimal weight is estimated as $(\gamma\hat{\boldsymbol{\Sigma}}_t)^{-1}\hat{\boldsymbol{R}}_t$, where $\hat{\boldsymbol{\Sigma}}_t$ is the estimated covariance matrix of forecast errors, $\hat{\boldsymbol{R}}_t$ is the forecast returns, and γ is a risk aversion parameter equal to one. Panels A and B show the relative change (in percentage) in annualized Sharpe ratio and Certainty Equivalent Return (CER), respectively. Panel C shows the relative change (in percentage) in the annualized variance of the stochastic discount factor (SDF). The SDF is estimated as $\boldsymbol{\omega}_t'\hat{\boldsymbol{\Sigma}}_t\boldsymbol{\omega}_t$, where $\boldsymbol{\omega}$ is the optimal weight. The forecasting designs are split-sample (black bar), expanding-window (red bar), and rolling-window (blue bar). For all designs, I use an initial training window equal to the first half of the sample (1970:01 - 1995:12). For the split-sample design, I use the parameter estimates to construct forecasts for the second half. For expanding- (rolling-)window design, I expand (roll) the training window each month to update parameter estimates and construct forecasts for the rest of the sample. The rolling window is equal to the initial training window. The out-of-sample evaluation period is from 1996:01 to 2021:12.

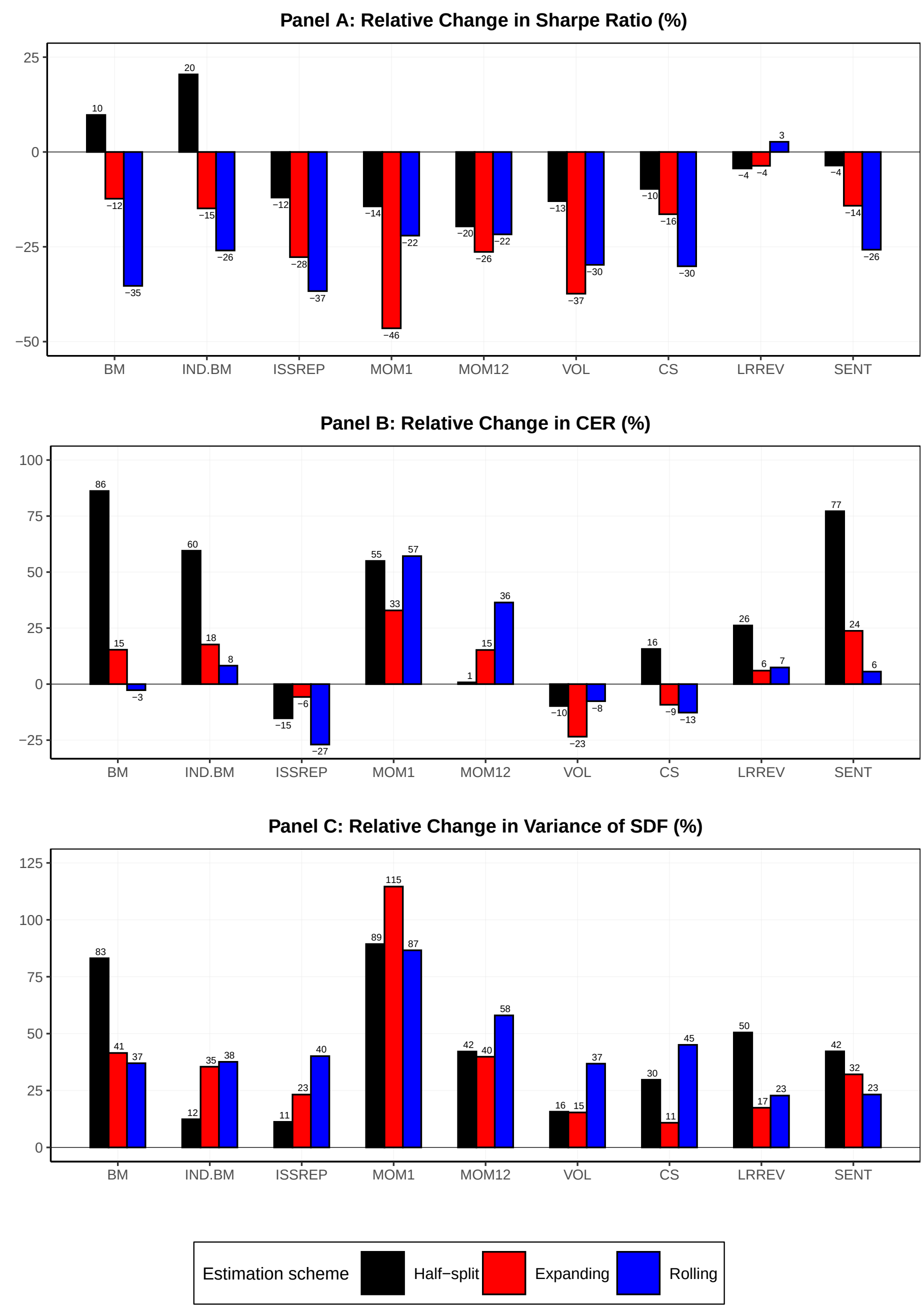

Figure 4. Factor prediction with shrinkage methods: Out-of-sample $\mathbf{R}^2$
This figure shows the out-of-sample performance of factor prediction under the shrinkage methods. Each plot shows the R^2_{OS} in percent for factors in one of six categories: *Momentum*, *Value*, *Investment*, *Profitability*, *Intangibles*, and *Trading frictions*. I use the classification scheme of Hou et al. (2020) to classify the factors. The sample period is from 1970:01 to 2021:12. The number of factors is 92. Additional detail on the factors is in Table A1 in the Appendix. The predictors are are *Book-to-market ratio*, *Industry-adjusted book-to-market ratio*, *Issuer-repurchaser spread*, *One-month momentum*, *12-month momentum*, *Volatility*, *Characteristic spread*, *Long-run reversal*, and five components of *Sentiment index*. Predictor construction detail is described in Table A2 in the Appendix. The shrinkage methods include forecast combination (*FC*), discount mean square of prediction errors (*DMSPE*), predictor average (*Average*), principal component regression (*PCR*), and partial least squares (*PLS*). *FC* is the equal-weighted average of univariate predictive regression forecasts from all predictors. *DMSPE* is the weighted average of univariate predictive regression forecasts, in which forecasts that have lower prediction errors over the holdout period have greater weight. *Average* is the univariate predictive regression forecast based on the cross-sectional average of all predictors. *PCR* (*PLS*) is a univariate predictive regression forecast based on the first principal component (return-relevant component) of all predictors. I adopt an expanding-window design for the out-of-sample tests. For *FC*, *Average*, *PCR*, and *PLS*, I use an initial training window equal to the first half of the sample (1970:01 - 1995:12) to estimate forecasts for 1996:01. For *DMSPE* that requires holdout periods, I use an initial training window from 1970:01 to 1985:12, and the subsequent 120 months (1986:01 to 1995:12) as the holdout period to estimate forecasts for 1996:01. The holdout window length is the same for subsequent forecasts. I expand the training window each month to estimate forecasts for the rest of the sample. The out-of-sample evaluation period is from 1996:01 to 2021:12. I calculate Campbell and Thompson's (2008) R^2_{OS} using Equation 8. I use Clark and West's (2007) procedure for the R^2_{OS} tests.

Table 1. Predictor variables in the literature

The table lists the variables that have been shown to be good predictors of factor returns. Columns (1) and (2) show the variables and studies that use them, respectively. Columns (3) and (4) report the sample period and number of test factors used in each study, respectively. Column (5) reports the return horizon in each study. Column (6) shows whether or not a study uses out-of-sample tests. Panel A shows single time-series variables (i.e., common predictors for all factors), and Panel B shows factor-specific variables (i.e., one predictor for each factor).

No.	Variable	Study	Sample period	Number of test portfolios	Horizon	Out of sample (Yes/No)
	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: Single time-series variables			
1	Sentiment index	Stambaugh et al. (2012)	1965-2008	11	Monthly	N
2	Aggregate mutual fund flows	Akbas et al. (2015)	1994-2012	11	Monthly	N
			Panel B: Factor-specific variables			
1	Book-to-market ratio	Cohen et al. (2003)	1938-1997	1	Annual	N
		Haddad et al. (2020)	1974-2017	50	Monthly	Y
		Baba-Yara et al. (2021)	1972-2017	6	Monthly, Annual	Y
2	Industry adjusted book-to-market ratio	Baba-Yara et al. (2021)	1972-2017	1	Monthly, Annual	Y
3	Issuer-repurchaser spread	Greenwood and Hanson (2012)	1962-2006	11	Monthly	N
4	Time-series momentum	Moskowitz et al. (2012)	1965-2009	58	Monthly	N
5	Volatility	Moreira and Muir (2017)	1926-2015	9	Monthly	N
6	Factor momentum	Gupta and Kelly (2019)	1965-2017	65	Monthly	N
		Ehsani and Linnainmaa (2022)	1963-2019	22	Monthly	N
7	Characteristic spread	Kelly et al. (2023)	1963-2019	138	Monthly	Y
		Kagkadis et al. (2023)	1970-2019	72	Monthly	Y

Table 2. Predictive regressions: Out-of-sample results

The table summarizes the out-of-sample performance of factor return prediction using the conventional predictive regression approach. The sample period is from 1970:01 to 2021:12. The number of factors is 92. Additional detail on the factors is in Table A1 in the Appendix. The nine predictors are *Book-to-market ratio (BM)*, *Industry-adjusted book-to-market ratio (IND.BM)*, *Issuer-repurchaser spread (ISSREP)*, *One-month momentum (MOM1)*, *12-month momentum (MOM12)*, *Volatility (VOL)*, *Characteristic spread (CS)*, *Long-run reversal (LRREV)*, and *Sentiment index (SENT)*. For *SENT*, I reconstruct the series for each period from its underlying five components using the same recursive design of the out-of-sample tests. Predictor construction detail is described in Table A2 in the Appendix. I adopt an expanding-window estimation design for the out-of-sample tests. I use an initial training window equal to the first half of the sample (1970:01 - 1995:12) to estimate forecasts for 1996:01, and expand the training window each month to estimate forecasts for the rest of the sample. The out-of-sample evaluation period is from 1996:01 to 2021:12. Panel A reports the out-of-sample statistics across the factors. Columns (1) to (4) report the mean (standard deviation), the 25th percentile, the median, and the 75th percentile R^2_{OS}. Column (5) reports the total R^2_{OS} across all factors. Column (6) reports the number of factors that have positive R^2_{OS} and the numbers in brackets indicate those with statistical significance (at least at the 10% level). Panel B reports the predictive performance across factor categories. I use the classification scheme of Hou et al. (2020) to classify each factor into one of six categories: *Momentum*, *Value*, *Investment*, *Profitability*, *Intangibles*, and *Trading frictions*. The number in bracket next to each category heading is the total number of factors in that category. I report the total R^2_{OS} (in percent) across the factors in each category, and the numbers in brackets indicate positive and statistically significant (at least at the 10% level) R^2_{OS}. I calculate Campbell and Thompson's (2008) R^2_{OS} using Equation 8. The total R^2_{OS} is calculated using Equation 9. I use Clark and West's (2007) procedure for the R^2_{OS} tests.

Panel A: Out-of-sample performance across individual factors

Predictor	R^2_{OS} (%)				Total R^2_{OS} (%)	> 0% [Sig.]
	Mean (SD)	P25	Median	P75		
	(1)	(2)	(3)	(4)	(5)	(6)
BM	-0.10 (0.96)	-0.72	-0.32	0.45	0.17	33 [23]
IND.BM	-0.12 (1.09)	-0.76	-0.30	0.43	0.18	34 [21]
ISSREP	-0.13 (0.74)	-0.46	-0.25	0.11	-0.17	30 [17]
MOM1	0.17 (0.99)	-0.43	-0.09	0.65	0.18	42 [25]
MOM12	0.03 (0.79)	-0.34	-0.15	0.32	-0.01	36 [17]
VOL	-0.60 (0.86)	-1.12	-0.66	-0.25	-0.55	16 [5]
CS	-0.06 (0.51)	-0.31	-0.09	0.21	0.00	35 [11]
LRREV	-0.16 (0.63)	-0.40	-0.14	0.00	-0.15	22 [8]
SENT	0.52 (1.15)	-0.17	0.09	1.06	0.68	56 [26]

Panel B: Out-of-sample performance across factor categories

Predictor	Total R^2_{OS} (%) [$R^2_{OS} > 0\%$ & Sig.]					
	Momentum [6]	Value [20]	Investment [18]	Profitability [16]	Intangibles [13]	Trading frictions [19]
BM	0.34 [3]	0.16 [7]	-0.22 [3]	-0.20 [2]	-0.63 [1]	0.65 [7]
IND.BM	0.36 [2]	0.21 [7]	-0.14 [3]	-0.25 [3]	-0.75 [2]	0.68 [4]
ISSREP	-0.32 [0]	-0.41 [3]	-0.21 [3]	0.04 [4]	-0.11 [3]	-0.07 [4]
MOM1	-0.48 [0]	0.24 [7]	-0.27 [0]	1.26 [12]	0.06 [2]	0.16 [4]
MOM12	-0.21 [0]	0.45 [5]	-0.12 [1]	-0.18 [2]	0.26 [5]	-0.20 [4]
VOL	0.40 [1]	-0.42 [2]	-0.52 [0]	-1.31 [0]	-0.74 [0]	-0.63 [2]
CS	-0.08 [0]	0.12 [4]	-0.05 [3]	-0.08 [1]	-0.42 [0]	0.14 [3]
LRREV	-0.06 [1]	-0.35 [2]	-0.22 [0]	-0.25 [1]	-0.29 [1]	0.04 [3]
SENT	-0.05 [0]	0.25 [4]	0.25 [3]	1.89 [10]	0.54 [2]	0.91 [7]

Table 3. Principal component portfolio approach: Out-of-sample results with split-sample specification
The table summarizes the out-of-sample performance of factor return prediction using the principal component (PC) portfolio approach. The sample period is from 1970:01 to 2021:12. The number of factors is 92. Additional detail on the factors is in Table A1 in the Appendix. The predictor is *Book-to-market ratio* (BM). I adopt a split-sample design for the out-of-sample tests. I estimate parameters using the first half of the sample (1970:01 - 1995:12), and use the estimates to construct forecasts for the second half. The out-of-sample evaluation period is from 1996:01 to 2021:12. Panel A reports the results on the predictability for the first five PC portfolios from the univariate predictive regression

$$PC^k_{t+1} = \lambda^k_0 + \lambda^k_1 X^k_t + \epsilon^k_{t+1},$$

where PC^k_{t+1} is the excess return of PC portfolio k ($k = \overline{1,5}$) in month $t+1$, and X_t is the book-to-market ratio of PC portfolio k in month t. I obtain the PC portfolio returns and their predictor using the eigenvectors of the covariance matrix of factor returns. Rows (1) and (2) report the predictive coefficient estimate $\hat{\lambda}_1$ and t-statistic (in brackets), respectively. Rows (3) and (4) report the full-sample and out-of-sample monthly R^2, respectively. Panel B reports the results on the predictability for 92 factors under the conventional predictive regressions and the PC portfolio approach. I obtain a PC-based return forecast for factor i as

$$\hat{R}_{i,t+1} = \sum_{k=1}^{5} \hat{\omega}^k_{i,t+1} \widehat{PC}^k_{t+1},$$

where $\hat{R}_{i,t+1}$ and $\widehat{PC}^k_{t+1}$ are the excess return forecasts of factor portfolio i and PC portfolio k in month $t+1$, respectively. $\hat{\omega}^k_{i,k,t+1}$ is the loading of factor portfolio i on PC portfolio k from the PC estimation. Columns (1) and (2) report the mean (standard deviation), and the median R^2_{OS}, respectively. Column (3) reports the total R^2_{OS}. Column (4) reports the number of R^2_{OS}s that are non-negative, and statistically significant at the 5% level in brackets. I calculate Campbell and Thompson's (2008) R^2_{OS} using Equation 8. The total R^2_{OS} is calculated using Equation 9. I use Clark and West's (2007) procedure for the R^2_{OS} tests.

	Panel A: Prediction of five largest equity components				
	PC1	PC2	PC3	PC4	PC5
	(1)	(2)	(3)	(4)	(5)
Own BM	1.25	2.63	0.44	-0.35	-0.48
t-statistic	[1.06]	[2.74]	[0.89]	[-0.61]	[-0.76]
R^2 (%)	1.00	2.33	0.16	0.08	0.18
R^2_{OS} (%)	1.47	2.46	-0.43	-0.24	-0.62

	Panel B: Out-of-sample prediction across individual factors			
Method	R^2_{OS} (%)		Total R^2_{OS} (%)	R^2_{OS} > 0% [5%-*Sig.*]
	Mean (SD)	Median		
	(1)	(2)	(3)	(4)
	(1)	(2)	(3)	(4)
Predictive regression	-0.10 (1.59)	0.03	0.23	47 [21]
PC portfolio	0.64 (2.24)	0.99	1.15	67 [50]

Table 4. Factor prediction with shrinkage methods: Out-of-sample performance
The table summarizes the out-of-sample performance of factor return prediction using the shrinkage methods. The sample period is from 1970:01 to 2021:12. The number of factors is 92. Additional detail on the factors is in Table A1 in the Appendix. The predictors are are *Book-to-market ratio*, *Industry-adjusted book-to-market ratio*, *Issuer-repurchaser spread*, *One-month momentum*, *12-month momentum*, *Volatility*, *Characteristic spread*, *Long-run reversal*, and five components of *Sentiment index*. Predictor construction detail is described in Table A2 in the Appendix. The shrinkage methods include forecast combination (*FC*), discount mean square of prediction errors (*DMSPE*), predictor average (*Average*), principal component regression (*PCR*), and partial least squares (*PLS*). *FC* is the equal-weighted average of univariate predictive regression forecasts from all predictors. *DMSPE* is the weighted average of univariate predictive regression forecasts, in which forecasts that have lower prediction errors over the holdout period have greater weight. *Average* is the univariate predictive regression forecast based on the cross-sectional average of all predictors. *PCR* (*PLS*) is a univariate predictive regression forecast based on the first principal component (return-relevant component) of all predictors. I adopt an expanding-window design for the out-of-sample tests. For *FC*, *Average*, *PCR*, and *PLS*, I use an initial training window equal to the first half of the sample (1970:01 - 1995:12) to estimate forecasts for 1996:01. For *DMSPE* that requires holdout periods, I use an initial training window from 1970:01 to 1985:12, and the subsequent 120 months (1986:01 to 1995:12) as the holdout period to estimate forecasts for 1996:01. The holdout window length is the same for subsequent forecasts. I expand the training window each month to estimate forecasts for the rest of the sample. The out-of-sample evaluation period is from 1996:01 to 2021:12. Panel A reports the results for the shrinkage methods. Columns (1) to (4) report the mean (standard deviation), the 25th percentile, the median, and the 75th percentile of R^2_{OS}. Column (5) report the total R^2_{OS} across all 92 factors. Column (6) reports the number of R^2_{OS}s that are non-negative, and statistically significant at the 10% (5%) level in brackets (parentheses). For comparison, Panel B reports the lowest and the highest values of the column statistics across all predictors for the conventional predictive regression and the PC portfolio approach using the same estimation design. I calculate Campbell and Thompson's (2008) R^2_{OS} using Equation 8. The total R^2_{OS} is calculated using Equation 9. I use Clark and West's (2007) procedure for the R^2_{OS} tests.

Method	R^2_{OS} (%)				Total R^2_{OS} (%)	$R^2_{OS} \geq 0$
	Mean (SD)	P25	Median	P75		[10%−Sig.] (5%−Sig.)
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Summary of predictive performance for the shrinkage methods						
FC	0.52 (0.50)	0.15	0.45	0.85	0.60	80 [52] (38)
DMSPE	0.53 (0.51)	0.15	0.46	0.86	0.61	81 [51] (38)
Average	0.52 (1.06)	-0.34	0.27	1.24	0.58	55 [43] (30)
PCR	1.08 (1.49)	0.07	0.89	1.95	1.30	70 [58] (39)
PLS	0.52 (2.23)	-0.87	0.43	1.95	0.91	52 [48] (42)
Panel B: Summary of predictive performance for the conventional and PC portfolio methods						
Predictive regression						
Lowest	-0.60 (0.51)	-1.12	-0.66	-0.25	-0.55	16 [5] (1)
Highest	0.52 (1.15)	-0.17	0.09	1.06	0.68	56 [26] (22)
PC						
Lowest	-0.60 (0.78)	-1.11	-0.58	0.08	-0.43	26 [7] (5)
Highest	0.60 (1.78)	-0.29	0.56	1.29	0.63	61 [42] (29)

Table 5. Direct comparison

The table report the number of factors that have higher out-of-sample R^2 under the shrinkage methods than the conventional and principal component (PC) portfolio approach. The shrinkage methods include forecast combination (*FC*), discount mean square of prediction errors (*DMSPE*), predictor average (*Average*), principal component regression (*PCR*), and partial least squares (*PLS*). The sample period is from 1970:01 to 2021:12. The number of factors is 92. Additional detail on the factors is in Table A1 in the Appendix. The predictors are *Book-to-market ratio (BM)*, *Industry-adjusted book-to-market ratio (IND.BM)*, *Issuer-repurchaser spread (ISSREP)*, *One-month momentum (MOM1)*, *12-month momentum (MOM12)*, *Volatility (VOL)*, *Characteristic spread (CS)*, *Long-run reversal (LRREV)*, and *Sentiment index (SENT)*. For *SENT*, I use its underlying five components for the shrinkage methods. For the conventional and principal-component approach, I reconstruct the series for each period from the five components using the same recursive design of the out-of-sample tests. Predictor construction detail is described in Table A2 in the Appendix. I adopt an expanding-window design for the out-of-sample tests. For *FC*, *Average*, *PCR*, and *PLS*, I use an initial training window equal to the first half of the sample (1970:01 - 1995:12) to estimate forecasts for 1996:01. For *DMSPE* that requires holdout periods, I use an initial training window from 1970:01 to 1985:12, and the subsequent 120 months (1986:01 to 1995:12) as the holdout period to estimate forecasts for 1996:01. The holdout window length is the same for subsequent forecasts. I expand the training window each month to estimate forecasts for the rest of the sample. The out-of-sample evaluation period is from 1996:01 to 2021:12. Panels A and B report the comparison results for the conventional and PC approach, respectively. I calculate Campbell and Thompson's (2008) R^2_{OS} using Equation 8. The numbers in brackets are the p-value from a one-sided binomial test with the null hypothesis that the proportion of reported factors is less than half.

Predictor	FC	DMSPE	Average	PCR	PLS
	(1)	(2)	(3)	(4)	(5)
Panel A: Predictive regression					
BM	75 [0.00]	75 [0.00]	75 [0.00]	79 [0.00]	64 [0.00]
IND.BM	72 [0.00]	72 [0.00]	67 [0.00]	78 [0.00]	62 [0.00]
ISSREP	78 [0.00]	78 [0.00]	64 [0.00]	70 [0.00]	50 [0.23]
MOM1	65 [0.00]	68 [0.00]	54 [0.06]	65 [0.00]	47 [0.46]
MOM12	71 [0.00]	71 [0.00]	62 [0.00]	66 [0.00]	51 [0.17]
VOL	85 [0.00]	85 [0.00]	75 [0.00]	78 [0.00]	63 [0.00]
CS	78 [0.00]	77 [0.00]	60 [0.00]	73 [0.00]	59 [0.00]
LRREV	78 [0.00]	78 [0.00]	62 [0.00]	71 [0.00]	56 [0.02]
SENT	56 [0.02]	57 [0.01]	45 [0.62]	63 [0.00]	44 [0.70]
Panel B: Principal-component approach					
BM	57 [0.01]	57 [0.01]	57 [0.01]	61 [0.00]	49 [0.30]
IND.BM	57 [0.01]	54 [0.06]	54 [0.06]	59 [0.00]	49 [0.30]
ISSREP	73 [0.00]	73 [0.00]	64 [0.00]	64 [0.00]	56 [0.02]
MOM1	51 [0.17]	53 [0.09]	56 [0.02]	65 [0.00]	52 [0.13]
MOM12	64 [0.00]	65 [0.00]	59 [0.00]	66 [0.00]	55 [0.04]
VOL	80 [0.00]	80 [0.00]	69 [0.00]	72 [0.00]	64 [0.00]
CS	75 [0.00]	74 [0.00]	66 [0.00]	68 [0.00]	59 [0.00]
LRREV	70 [0.00]	72 [0.00]	59 [0.00]	67 [0.00]	54 [0.06]
SENT	49 [0.30]	48 [0.38]	45 [0.62]	58 [0.01]	43 [0.77]

Table 6. Distribution of Sharpe ratios across factor timing strategies
The table reports the distribution of annualized Sharpe ratio from timed factor strategies that use return forecasts from the shrinkage methods, the conventional predictive regressions, and the principal-component (PC) approach. The shrinkage methods include forecast combination (*FC*), discount mean square of prediction errors (*DMSPE*), predictor average (*Average*), principal component regression (*PCR*), and partial least squares (*PLS*). The sample period is from 1970:01 to 2021:12. The number of factors is 92. Additional detail on the factors is in Table A1 in the Appendix. The predictors are *Book-to-market ratio (BM)*, *Industry-adjusted book-to-market ratio (IND.BM)*, *Issuer-repurchaser spread (ISSREP)*, *One-month momentum (MOM1)*, *12-month momentum (MOM12)*, *Volatility (VOL)*, *Characteristic spread (CS)*, *Long-run reversal (LRREV)*, and *Sentiment index (SENT)*. For *SENT*, I use its underlying five components for the shrinkage methods. For the conventional and principal-component approach, I reconstruct the series for each period from the five components using the same recursive design of the out-of-sample tests. Predictor construction detail is described in Table A2 in the Appendix. I adopt an expanding-window design to obtain return forecasts. For *FC*, *Average*, *PCR*, and *PLS*, I use an initial training window equal to the first half of the sample (1970:01 - 1995:12) to estimate forecasts for 1996:01. For *DMSPE* that requires holdout periods, I use an initial training window from 1970:01 to 1985:12, and the subsequent 120 months (1986:01 to 1995:12) as the holdout period to estimate forecasts for 1996:01. The holdout window length is the same for subsequent forecasts. I expand the training window each month to estimate forecasts for the rest of the sample. The out-of-sample evaluation period is from 1996:01 to 2021:12. The optimal weight for factor timing strategy i in month t is estimated as $\hat{R}^m_{i,t+1}/(\gamma\hat{\sigma}^2_{i,t+1})$, where $\hat{R}^m_{i,t+1}$ is the return forecast using forecasting method m, $\hat{\sigma}^2_{i,t+1}$ is the sample variance using all data up to month t, and γ is a risk aversion parameter. Panel A reports the distribution that use the shrinkage methods to obtain return forecasts. Row (1) reports the results from factor investing, in which the optimal weight uses the historical mean as return forecast. Columns (1) to (4) report the mean, the standard deviation (SD), the 25th percentile, the median, and the 75th percentile Sharpe ratio, respectively. Panel B reports the mean and median Sharpe ratio that use the conventional predictive regressions (Columns (1) and (2)), and the PC approach (Columns (3) and (4)) to obtain return forecasts. I use a risk aversion parameter of one, and impose a leverage constraint that the absolute weight on the factor portfolio is less than or equal to two.

Panel A: Shrinkage methods

Method	Mean	SD	P25	Median	P75
	(1)	(2)	(3)	(4)	(5)
Factor investing	0.11	0.25	-0.02	0.09	0.25
FC	0.23	0.20	0.09	0.23	0.37
DMSPE	0.23	0.20	0.10	0.23	0.36
Average	0.22	0.20	0.08	0.22	0.34
PCR	0.28	0.23	0.11	0.28	0.43
PLS	0.27	0.24	0.09	0.27	0.48

Panel B: Predictive regression and PC approach

Predictor	Predictive regression		PC	
	Mean	Median	Mean	Median
	(1)	(2)	(3)	(4)
BM	0.16	0.18	0.15	0.17
IND.BM	0.16	0.16	0.13	0.16
ISSREP	0.16	0.13	0.06	0.05
MOM1	0.22	0.21	0.24	0.21
MOM12	0.18	0.16	0.13	0.12
VOL	0.13	0.10	0.01	0.01
CS	0.15	0.14	0.03	0.01
LRREV	0.13	0.11	0.07	0.05
SENT	0.22	0.17	0.18	0.18

Table 7. Distribution of Certainty Equivalent Return across factor timing strategies
The table reports the distribution of annualized (in percent) Certainty Equivalent Return (CER) from timed factor strategies that use return forecasts from the shrinkage methods, the conventional predictive regressions, and the principal-component (PC) approach. The shrinkage methods include forecast combination (FC), discount mean square of prediction errors ($DMSPE$), predictor average (*Average*), principal component regression (PCR), and partial least squares (PLS). The sample period is from 1970:01 to 2021:12. The number of factors is 92. Additional detail on the factors is in Table A1 in the Appendix. The predictors are *Book-to-market ratio (BM)*, *Industry-adjusted book-to-market ratio (IND.BM)*, *Issuer-repurchaser spread (ISSREP)*, *One-month momentum (MOM1)*, *12-month momentum (MOM12)*, *Volatility (VOL)*, *Characteristic spread (CS)*, *Long-run reversal (LRREV)*, and *Sentiment index (SENT)*. For *SENT*, I use its underlying five components for the shrinkage methods. For the conventional and principal-component approach, I reconstruct the series for each period from the five components using the same recursive design of the out-of-sample tests. Predictor construction detail is described in Table A2 in the Appendix. I adopt an expanding-window design to obtain return forecasts. For *FC*, *Average*, *PCR*, and *PLS*, I use an initial training window equal to the first half of the sample (1970:01 - 1995:12) to estimate forecasts for 1996:01. For *DMSPE* that requires holdout periods, I use an initial training window from 1970:01 to 1985:12, and the subsequent 120 months (1986:01 to 1995:12) as the holdout period to estimate forecasts for 1996:01. The holdout window length is the same for subsequent forecasts. I expand the training window each month to estimate forecasts for the rest of the sample. The out-of-sample evaluation period is from 1996:01 to 2021:12. The optimal weight for factor timing strategy i in month t is estimated as $\hat{R}^m_{i,t+1}/(\gamma\hat{\sigma}^2_{i,t+1})$, where $\hat{R}^m_{i,t+1}$ is the return forecast using forecasting method m, $\hat{\sigma}^2_{i,t+1}$ is the sample variance using all data up to month t, and γ is a risk aversion parameter. The CER for factor strategy i is estimated as $\bar{R}_i - (\gamma/2)\hat{\sigma}^2_i$, where $\bar{R}_i$ and $\hat{\sigma}^2_i$ are the average return and sample variance of the timed strategy during the evaluation period, respectively. Panel A reports the distribution that use the shrinkage methods to obtain return forecasts. Row (1) reports the results from factor investing, in which the optimal weight uses the historical mean as return forecast. Columns (1) to (4) report the mean, the standard deviation (SD), the 25th percentile, the median, and the 75th percentile CER, respectively. Panel B reports the mean and median CER that use the conventional predictive regressions (Columns (1) and (2)), and the PC approach (Columns (3) and (4)) to obtain return forecasts. I use a risk aversion parameter of one, and impose a leverage constraint that the absolute weight on the factor portfolio is less than or equal to two.

Panel A: Shrinkage methods

Method	Mean	SD	P25	Median	P75
	(1)	(2)	(3)	(4)	(5)
Factor investing	0.40	5.40	-3.04	-0.67	2.98
FC	2.39	5.13	-1.16	1.85	5.89
DMSPE	2.40	5.14	-1.26	1.96	5.36
Average	2.00	5.64	-1.68	2.17	5.71
PCR	4.00	6.61	-0.86	3.11	8.57
PLS	3.64	7.29	-2.11	3.43	9.41

Panel B: Predictive regression and PC approach

Predictor	Predictive regression		PC	
	Mean	Median	Mean	Median
	(1)	(2)	(3)	(4)
BM	0.52	0.34	0.28	-0.16
IND.BM	0.60	0.52	0.04	0.56
ISSREP	0.35	-0.12	-1.85	-2.17
MOM1	1.71	1.77	2.56	1.63
MOM12	0.72	0.57	-0.61	-0.59
VOL	-0.62	-1.01	-2.75	-3.81
CS	0.39	-0.08	-1.85	-2.51
LRREV	0.22	-0.63	-0.66	-0.82
SENT	1.97	0.93	1.49	0.97

Table 8. Performance of anomaly timing portfolios

The table compares performance of the optimal anomaly timing portfolio under the principal component (PC) portfolio approach across nine predictors and the shrinkage methods. The sample period is from 1970:01 to 2021:12. The number of factors is 92. Additional detail on the factors is in Table A1 in the Appendix. The nine predictors are *Book-to-market ratio (BM)*, *Industry-adjusted book-to-market ratio (IND.BM)*, *Issuer-repurchaser spread (ISSREP)*, *One-month momentum (MOM1)*, *12-month momentum (MOM12)*, *Volatility (VOL)*, *Characteristic spread (CS)*, *Long-run reversal (LRREV)*, and *Sentiment index (SENT)*. For *SENT*, I reconstruct the series for each period from its underlying five components using the same recursive design of the out-of-sample tests. Predictor construction detail is described in Table A2 in the Appendix. I obtain the optimal portfolio based on the first five PC portfolios with anomaly timing and the market factor without timing. The optimal weight is estimated as $(\gamma\hat{\boldsymbol{\Sigma}})^{-1}\hat{\boldsymbol{R}}_t$, where $\hat{\boldsymbol{\Sigma}}$ is the estimated covariance matrix of forecast errors, $\hat{\boldsymbol{R}}_t$ is the forecast returns, and γ is a risk aversion parameter equal to one. The first row reports the average performance across all nine predictors that use the PC approach. For the shrinkage methods, expected PC returns are estimated as the loading-weighted forecast returns of individual factor portfolios. The shrinkage methods include forecast combination (FC), discount mean square of prediction errors ($DMSPE$), predictor average ($Average$), principal component regression (PCR), and partial least squares (PLS). I adopt an expanding-window design to obtain return forecasts. For FC, $Average$, PCR, and PLS, I use an initial training window equal to the first half of the sample (1970:01 - 1995:12) to estimate forecasts for 1996:01. For $DMSPE$ that requires holdout periods, I use an initial training window from 1970:01 to 1985:12, and the subsequent 120 months (1986:01 to 1995:12) as the holdout period to estimate forecasts for 1996:01. The holdout window length is the same for subsequent forecasts. I expand the training window each month to estimate forecasts for the rest of the sample. The out-of-sample evaluation period is from 1996:01 to 2021:12. Columns (1), (2), (3), and (4) report the annualized Sharpe ratio, Information ratio, annualized Certainty Equivalent Return (in percentage), and annualized average of stochastic discount factor. Information ratio is obtained versus the untimed optimal portfolio , and the Fama-French five-factor model.

Method	Sharpe ratio	Information ratio	CER	Variance of SDF
	(1)	(2)	(3)	(4)
PC	0.72	-0.01	1.22	1.70
FC	0.94	0.31	1.41	1.29
DMSPE	0.93	0.28	1.43	1.31
Average	0.78	0.15	1.58	1.44
PCR	0.84	0.24	2.21	2.00
PLS	0.74	0.21	2.58	2.89

Appendix

Figure A1. Predicting factor returns with predictive regression approach: In-sample t-statistic
This figure shows the t-statistic from the conventional predictive regressions of one-month ahead factor returns on their predictors. Each plot shows the results for factors in one of six categories: *Momentum*, *Value*, *Investment*, *Profitability*, *Intangibles*, and *Trading frictions*. I use the classification scheme of Hou et al. (2020) to classify the factors. The sample period is from 1970:01 to 2021:12. The number of factors is 92. Additional detail on the factors is in Table A1 in the Appendix. The nine predictors are *Book-to-market ratio (BM)*, *Industry-adjusted book-to-market ratio (IND.BM)*, *Issuer-repurchaser spread (ISSREP)*, *One-month momentum (MOM1)*, *12-month momentum (MOM12)*, *Volatility (VOL)*, *Characteristic spread (CS)*, *Long-run reversal (LRREV)*, and *Sentiment index (SENT)*. For *SENT*, I use Baker and Wurgler's (2006) full-sample orthogonalized series. Predictor construction detail is described in Table A2 in the Appendix. For each factor and predictor, I estimate Equation 2 using the full sample of data, and record the coefficient estimate, the t-statistic and the p-value from the t-test. The colors indicate whether the coefficient estimate is consistent in sign with original studies, and whether the estimate is statistically significant at either of three levels (10%, 5%, or 1%). I use the Newey and West's (1987) $t-$statistic with a 2-year window for the kernel.

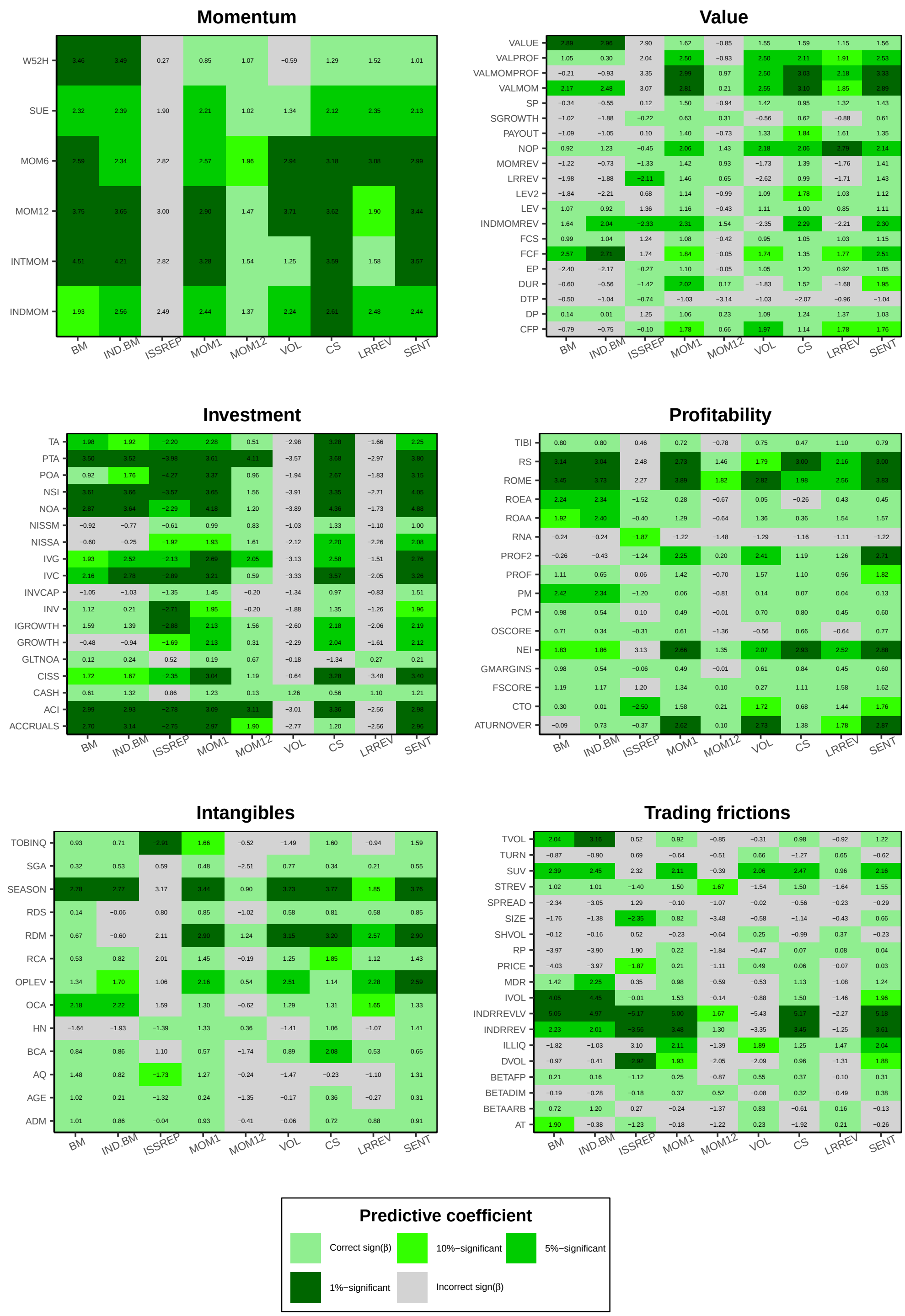

Figure A2. Predicting factor returns with predictive regression approach: $\mathbf{R^2_{OS}}$

This figure shows the R^2_{OS} from the conventional predictive regressions of one-month ahead factor returns on their predictors. Each plot shows the results for factors in one of six categories: *Momentum*, *Value*, *Investment*, *Profitability*, *Intangibles*, and *Trading frictions*. I use the classification scheme of Hou et al. (2020) to classify the factors. The sample period is from 1970:01 to 2021:12. The number of factors is 92. Additional detail on the factors is in Table A1 in the Appendix. The nine predictors are *Book-to-market ratio (BM)*, *Industry-adjusted book-to-market ratio (IND.BM)*, *Issuer-repurchaser spread (ISSREP)*, *One-month momentum (MOM1)*, *12-month momentum (MOM12)*, *Volatility (VOL)*, *Characteristic spread (CS)*, *Long-run reversal (LRREV)*, and *Sentiment index (SENT)*. For *SENT*, I reconstruct the series for each period from its underlying five components using the same recursive design of the out-of-sample tests. Predictor construction detail is described in Table A2 in the Appendix. I adopt an expanding-window estimation design for the out-of-sample tests. I use an initial training window equal to the first half of the sample (1970:01 - 1995:12) to estimate forecasts for 1996:01, and expand the training window each month to estimate forecasts for the rest of the sample. The out-of-sample evaluation period is from 1996:01 to 2021:12. I calculate Campbell and Thompson's (2008) R^2_{OS} using Equation 8.

Figure A3. Factor prediction with principal component approach: Cumulative total $\mathbf{R^2_{OS}}$
This figure shows the cumulative total R^2_{OS} using two predictors under the principal component (PC) portfolio approach. The sample period is from 1970:01 to 2021:12. The number of factors is 92. Additional detail on the factors is in Table A1 in the Appendix. The two predictors are *Book-to-market ratio (BM)*, *One-month momentum (MOM1)*, and *Volatility (VOL)*. Predictor construction detail is described in Table A2 in the Appendix. Each line plots $(\Sigma_i SST_i^t - \Sigma_i SSE_i^t)/\Sigma_i SST_i^T$, where $SSE_i^t = \Sigma_{\tau=1}^t (R_{i,\tau} - \hat{R}_{i,\tau})^2, SST_i^t = \Sigma_{\tau=1}^t (R_{i,\tau} - \bar{R}_{i,\tau})^2$, and $SST_i^T = \Sigma_{\tau=1}^T (R_{i,\tau} - \bar{R}_{i,\tau})^2$, with $R_{i,\tau}, \hat{R}_{i,\tau}$ and $\bar{R}_{i,\tau}$ the factor i's return, its one-month ahead return forecast, and its prevailing mean forecast, respectively. For all designs, I use an initial training window equal to the first half of the sample (1970:01 - 1995:12). For the split-sample design, I use the parameter estimates to construct forecasts for the second half. For expanding- (rolling-)window design, I expand (roll) the training window each month to update parameter estimates and construct forecasts for the rest of the sample. The rolling window is equal to the initial training window. The out-of-sample evaluation period is from 1996:01 to 2021:12. The solid, dashed, and dotted line indicates the split-sample, recursive, and rolling window estimation design, respectively. Shaded areas denote NBER-dated recessions.

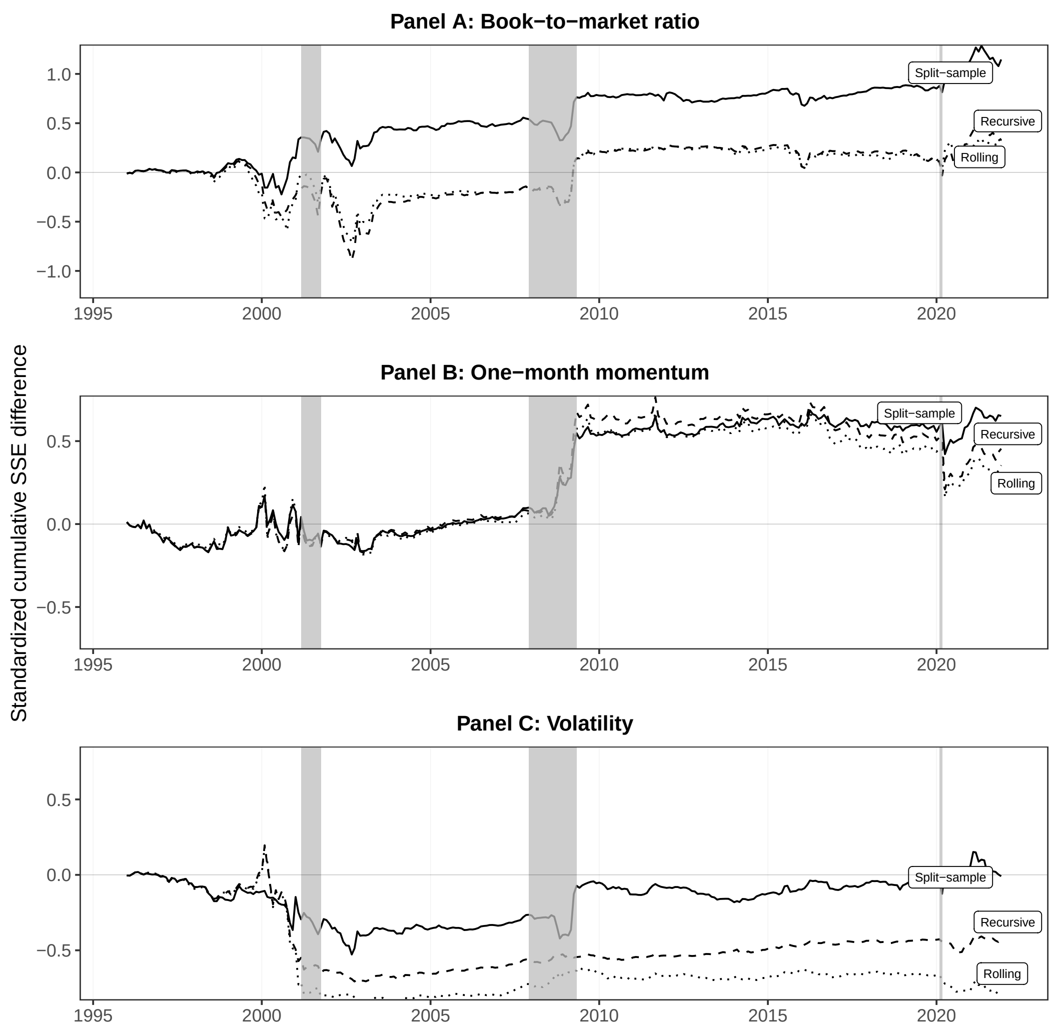

Figure A4. Factor timing strategies with shrinkage methods: Sharpe ratios

This figure shows the annualized Sharpe ratio for factor timing strategies that use return forecasts from the shrinkage methods. The sample period is from 1970:01 to 2021:12. The number of factors is 92. Additional detail on the factors is in Table A1 in the Appendix. The predictors are are *Book-to-market ratio*, *Industry-adjusted book-to-market ratio*, *Issuer-repurchaser spread*, *One-month momentum*, *12-month momentum*, *Volatility*, *Characteristic spread*, *Long-run reversal*, and five components of *Sentiment index*. Predictor construction detail is described in Table A2 in the Appendix. The shrinkage methods include forecast combination (*FC*), discount mean square of prediction errors (*DMSPE*), predictor average (*Average*), principal component regression (*PCR*), and partial least squares (*PLS*). *FC* is the equal-weighted average of univariate predictive regression forecasts from all predictors. *DMSPE* is the weighted average of univariate predictive regression forecasts, in which forecasts that have lower prediction errors over the holdout period have greater weight. *Average* is the univariate predictive regression forecast based on the cross-sectional average of all predictors. *PCR* (*PLS*) is a univariate predictive regression forecast based on the first principal component (return-relevant component) of all predictors. I adopt an expanding-window design to obtain return forecasts. For *FC*, *Average*, *PCR*, and *PLS*, I use an initial training window equal to the first half of the sample (1970:01 - 1995:12) to estimate forecasts for 1996:01. For *DMSPE* that requires holdout periods, I use an initial training window from 1970:01 to 1985:12, and the subsequent 120 months (1986:01 to 1995:12) as the holdout period to estimate forecasts for 1996:01. The holdout window length is the same for subsequent forecasts. I expand the training window each month to estimate forecasts for the rest of the sample. The out-of-sample evaluation period is from 1996:01 to 2021:12. The optimal weight for factor timing strategy i in month t is estimated as $\hat{R}^m_{i,t+1}/(\gamma\hat{\sigma}^2_{i,t+1})$, where $\hat{R}^m_{i,t+1}$ is the return forecast using forecasting method m, $\hat{\sigma}^2_{i,t+1}$ is the sample variance using all data up to month t, and γ is a risk aversion parameter. The colors indicate the deciles of the Sharpe ratio distribution across all factors within each method. I use a risk aversion parameter of one, and impose a leverage constraint that the absolute weight on the factor portfolio is less than or equal to two.

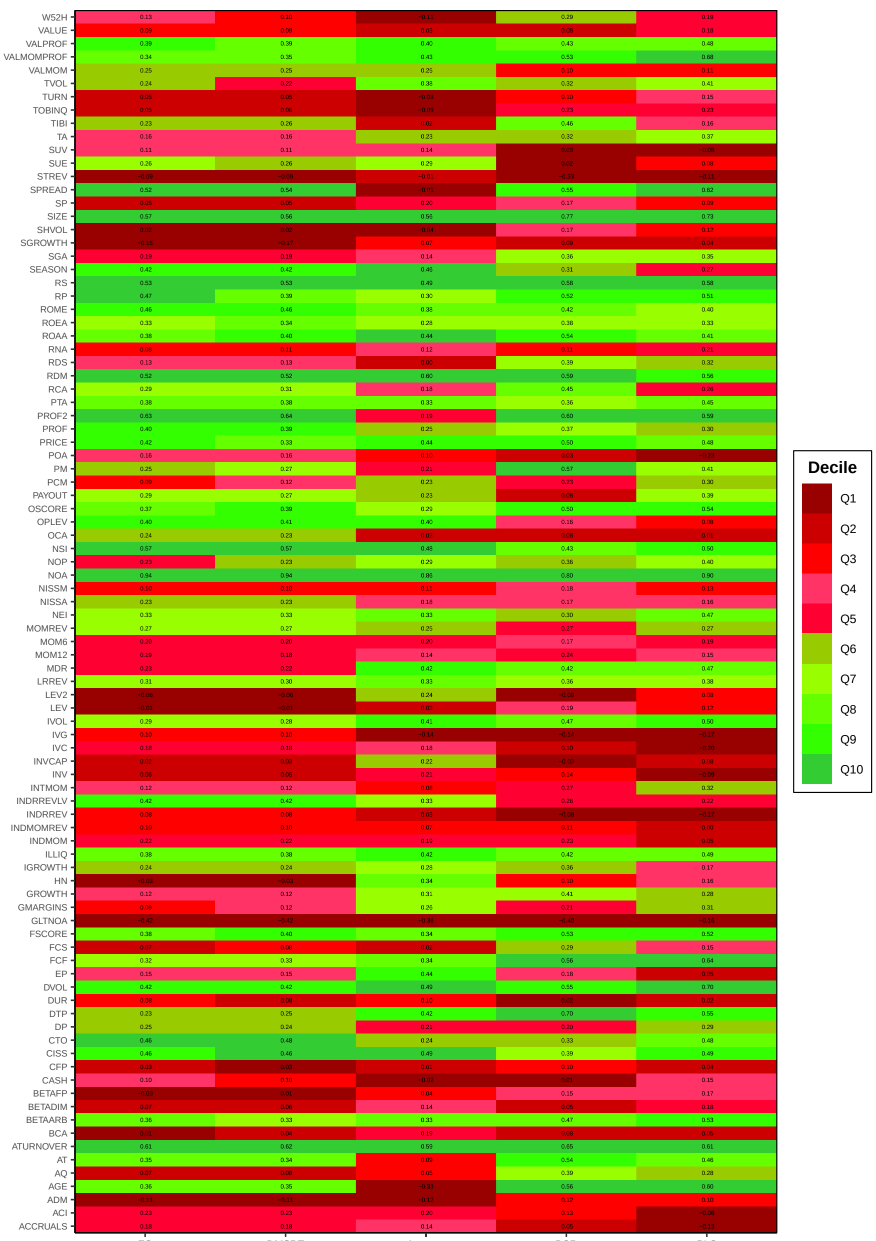

Figure A5. Factor timing strategies with shrinkage methods: Certainty Equivalent Returns

This figure shows the annualized Certainty Equivalent Returns (CER) for factor timing strategies that use return forecasts from the shrinkage methods. The sample period is from 1970:01 to 2021:12. The number of factors is 92. Additional detail on the factors is in Table A1 in the Appendix. The predictors are are *Book-to-market ratio*, *Industry-adjusted book-to-market ratio*, *Issuer-repurchaser spread*, *One-month momentum*, *12-month momentum*, *Volatility*, *Characteristic spread*, *Long-run reversal*, and five components of *Sentiment index*. Predictor construction detail is described in Table A2 in the Appendix. The shrinkage methods include forecast combination (*FC*), discount mean square of prediction errors (*DMSPE*), predictor average (*Average*), principal component regression (*PCR*), and partial least squares (*PLS*). *FC* is the equal-weighted average of univariate predictive regression forecasts from all predictors. *DMSPE* is the weighted average of univariate predictive regression forecasts, in which forecasts that have lower prediction errors over the holdout period have greater weight. *Average* is the univariate predictive regression forecast based on the cross-sectional average of all predictors. *PCR* (*PLS*) is a univariate predictive regression forecast based on the first principal component (return-relevant component) of all predictors. I adopt an expanding-window design to obtain return forecasts. For *FC*, *Average*, *PCR*, and *PLS*, I use an initial training window equal to the first half of the sample (1970:01 - 1995:12) to estimate forecasts for 1996:01. For *DMSPE* that requires holdout periods, I use an initial training window from 1970:01 to 1985:12, and the subsequent 120 months (1986:01 to 1995:12) as the holdout period to estimate forecasts for 1996:01. The holdout window length is the same for subsequent forecasts. I expand the training window each month to estimate forecasts for the rest of the sample. The out-of-sample evaluation period is from 1996:01 to 2021:12. The optimal weight for factor timing strategy i in month t is estimated as $\hat{R}^m_{i,t+1}/(\gamma\hat{\sigma}^2_{i,t+1})$, where $\hat{R}^m_{i,t+1}$ is the return forecast using forecasting method m, $\hat{\sigma}^2_{i,t+1}$ is the sample variance using all data up to month t, and γ is a risk aversion parameter. The CER for factor strategy i is estimated as $\bar{R}_i - (\gamma/2)\hat{\sigma}^2_i$, where $\bar{R}_i$ and $\hat{\sigma}^2_i$ are the average return and sample variance of the timed strategy during the evaluation period, respectively. The colors indicate the deciles of the CER distribution across all factors within each method. I use a risk aversion parameter of one, and impose a leverage constraint that the absolute weight on the factor portfolio is less than or equal to two.

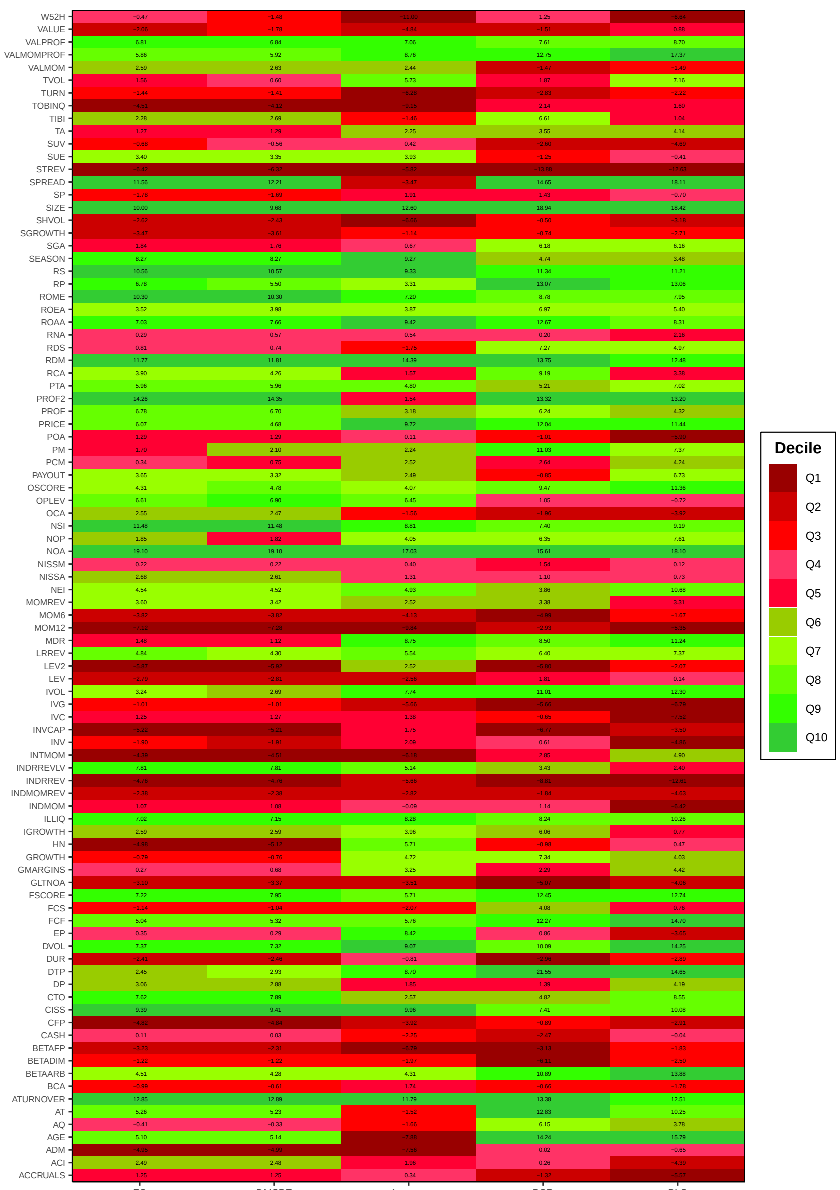

Table A1. Firm characteristics and factor portfolios

The table summarizes the firm characteristics used to construct the long-short factor decile portfolios in the paper. The panels indicate factor categories following the classification scheme of Hou et al. (2020). For factors that Hou et al. (2020) do not cover, ∗ indicates a classification based on the closeness of the characteristics to other characteristics in the category. I follow the descriptions in Hou et al. (2015) (HXZ), Kelly et al. (2019) (KPS), and Haddad et al. (2020) (HKS) in constructing the characteristics. Columns (1) and (2) report the characteristic and their acronym used in the paper. Column (3) reports the original study. Columns (4) and (5) report the average return and Fama and French's (1993) 3-factor alpha (annualized in percentage) for 92 equity factors from 1970:01 to 2021:12. The numbers in brackets report Newey and West's (1987) t-statistic with a 2-year window for the kernel. Columns (6)-(8) indicate which studies (HKS, KPS, HXZ) uses the characteristic.

Anomaly (1)	Acronym (2)	Original study (3)	Average return (4)	FF 3-factor alpha (5)	HKS (6)	KPS (7)	HXZ (8)
			Panel A: Momentum				
Industry momentum	INDMOM	Moskowitz and Grinblatt (1999)	5.92 [2.43]	8.42 [3.22]	✓		✓
Intermediate momentum	INTMOM	Novy-Marx (2012)	10.78 [3.40]	13.69 [4.59]		✓	
Price momentum (12-month prior returns)	MOM12	Jegadeesh and Titman (1993)	11.80 [3.43]	17.28 [5.86]	✓	✓	✓
Price momentum (6-month prior returns)	MOM6	Jegadeesh and Titman (1993)	8.29 [2.84]	12.49 [4.45]	✓		✓
Earnings surprise	SUE	Chan, Jegadeesh, and Lakonishok (1996)	3.88 [2.13]	4.34 [2.26]	✓		✓
Closeness to 52-week high	W52H	George and Hwang (2004)	3.41 [0.95]	11.74 [4.10]		✓	
			Panel B: Value				
Cash flow-to-price	CFP	Lakonishok, Shleifer, and Vishny (1994)	4.60 [1.78]	2.18 [1.22]	✓		✓
Dividend yield	DP	Litzenberger and Ramaswamy (1979)	2.89 [1.01]	2.07 [0.96]			✓
Debt-to-price	DTP	Litzenberger and Ramaswamy (1979)	-4.45 [-1.03]	-10.67 [-2.50]		✓	
Equity duration	DUR	Dechow, Sloan, and Soliman (2004)	4.86 [1.96]	0.89 [0.58]	✓		✓
Earnings-to-price	EP	Basu (1977)	3.14 [1.03]	-0.75 [-0.35]	✓	✓	✓
Free cash flow∗	FCF	Hou, Karolyi, and Kho (2011)	4.68 [2.04]	8.18 [4.11]		✓	
Change in gross margins minus change in sales∗	FCS	Abarbanell and Bushee (1998)	2.24 [1.12]	2.31 [1.03]		✓	
Industry momentum-reversal∗	INDMOMREV	Moskowitz and Grinblatt (1999)	4.64 [2.30]	1.83 [0.91]	✓		
Assets-to-market equity	LEV	Bhandari (1988)	2.83 [1.11]	-2.93 [-1.59]	✓	✓	✓
Leverage∗	LEV2	Lewellen (2015)	2.99 [1.12]	-3.68 [-2.29]		✓	
Long-term reversal	LRREV	De Bondt and Thaler (1985)	4.53 [1.41]	0.01 [0.00]	✓	✓	✓
Momentum-reversal∗	MOMREV	Jegadeesh and Titman (1993)	4.04 [1.40]	0.69 [0.28]	✓		
Net payout yield	NOP	Boudoukh, Michaely, Richardson, and Roberts (2007)	5.31 [2.04]	4.87 [2.46]			✓
Payout yield	PAYOUT	Boudoukh et al. (2007)	3.15 [1.34]	0.53 [0.24]			✓
Sales growth	SGROWTH	Lakonishok et al. (1994)	1.14 [0.62]	-0.86 [-0.52]	✓		✓
Sales-to-price∗	SP	Barbee Jr, Mukherji, and Raines (1996)	3.65 [1.43]	-1.46 [-0.85]	✓	✓	
Value-momentum∗	VALMOM	Novy-Marx (2013)	7.07 [2.91]	4.97 [2.83]	✓		
Value-momentum-profitability∗	VALMOMPROF	Novy-Marx (2013)	7.94 [3.10]	7.63 [3.79]	✓		
Value-profitability∗	VALPROF	Novy-Marx (2013)	6.47 [2.42]	4.22 [2.39]	✓		
Book-to-market equity (annual)	VALUE	Fama and French (1993)	4.80 [1.55]	-1.42 [-0.70]	✓	✓	✓

(Continued on next page)

Table A1 (*continued*)

Anomaly	Acronym	Original study	Average return	FF 3-factor alpha	HKS	KPS	HXZ
		Panel C: Investment					
Operating accruals	ACCRUALS	Sloan (1996)	4.29 [2.96]	4.46 [2.94]	✓	✓	✓
Abnormal corporate investment	ACI	Titman, Wei, and Xie (2004)	4.02 [3.02]	3.84 [2.96]			✓
Cash-to-assets∗	CASH	Palazzo (2012)	2.71 [1.23]	4.30 [2.33]		✓	
Composite issuance	CISS	Daniel and Titman (2006)	6.83 [3.32]	9.12 [5.20]	✓		✓
Growth in LTNOA	GLTNOA	Fairfield, Whisenant, and Yohn (2003)	0.25 [0.19]	0.77 [0.61]	✓		
Investment-to-assets	GROWTH	Cooper, Gulen, and Schill (2008)	4.98 [2.12]	3.01 [1.90]	✓	✓	✓
Investment growth	IGROWTH	Xing (2008)	4.13 [2.13]	2.85 [1.67]	✓		✓
Changes in property, plant, and equipment	INV	Lyandres, Sun, and Zhang (2008)	3.59 [1.96]	2.90 [1.78]	✓	✓	✓
Investment-to-capital∗	INVCAP	Xing (2008)	4.13 [1.49]	1.85 [0.95]	✓		
Inventory changes	IVC	Thomas and Zhang (2002)	4.83 [3.30]	4.61 [3.21]			✓
Inventory growth	IVG	Belo and Lin (2012)	5.00 [2.75]	4.38 [2.59]			✓
Share issuance (annual)∗	NISSA	Pontiff and Woodgate (2008)	3.98 [1.91]	3.19 [1.76]	✓		
Share issuance (monthly)∗	NISSM	Pontiff and Woodgate (2008)	1.67 [0.97]	0.83 [0.45]	✓		
Net operating assets	NOA	Hirshleifer, Hou, Teoh, and Zhang (2004)	8.41 [4.08]	9.31 [4.39]	✓	✓	✓
Net stock issues	NSI	Fama and French (2008)	7.35 [3.68]	7.91 [4.31]			✓
Percent operating accruals	POA	Hafzalla, Lundholm, and Matthew Van Winkle (2011)	4.51 [3.15]	4.47 [2.88]			✓
Percent total accruals	PTA	Hafzalla et al. (2011)	5.16 [3.67]	5.32 [3.71]			✓
Total accruals	TA	Richardson, Sloan, Soliman, and Tuna (2005)	3.22 [2.25]	1.86 [1.49]			✓
		Panel D: Profitability					
Asset turnover	ATURNOVER	Soliman (2008)	5.31 [2.70]	6.46 [3.29]	✓	✓	✓
Capital turnover	CTO	Haugen and Baker (1996)	3.08 [1.64]	3.97 [2.05]		✓	✓
F-score	FSCORE	Piotroski (2000)	6.85 [1.44]	10.88 [2.50]	✓		✓
Gross margins	GMARGINS	Novy-Marx (2013)	0.97 [0.54]	4.49 [2.84]	✓	✓	
Consecutive quarters with earnings increases	NEI	Barth, Elliott, and Finn (1999)	9.68 [2.71]	13.78 [3.62]			✓
O-score	OSCORE	Dichev (1998)	1.73 [0.67]	5.91 [3.24]			✓
Price-to-cost margin∗	PCM	Gorodnichenko and Weber (2016)	0.97 [0.54]	4.49 [2.84]		✓	
Profit margin	PM	Soliman (2008)	0.18 [0.08]	4.74 [2.24]			✓
Gross profits-to-assets	PROF	Novy-Marx (2013)	3.63 [1.46]	7.31 [2.97]	✓		✓
Gross profits-to-book value of equity∗	PROF2	Ball, Gerakos, Linnainmaa, and Nikolaev (2016)	5.45 [2.23]	7.40 [2.63]		✓	
Return on net operating assets	RNA	Soliman (2008)	-2.41 [-1.17]	0.49 [0.27]		✓	✓
Return on assets (annual)	ROAA	Chen, Novy-Marx, and Zhang (2011)	3.73 [1.35]	8.42 [3.48]	✓	✓	✓
Return on equity (annual)	ROEA	Chen et al. (2011)	0.87 [0.35]	4.51 [1.88]	✓	✓	✓
Return on market equity∗	ROME	Chen et al. (2011)	11.80 [3.79]	12.85 [4.52]	✓		
Revenue surprise	RS	Jegadeesh and Livnat (2006)	5.80 [2.95]	8.39 [5.10]			✓
Taxable income-to-book income	TIBI	Green, Hand, and Zhang (2017)	1.28 [0.79]	2.76 [1.70]			✓

(Continued on next page)

Table A1 (*continued*)

Anomaly	Acronym	Original study	Average return	FF 3-factor alpha	HKS	KPS	HXZ
		Panel E: Intangibles					
Advertisement expense-to-market	ADM	Chan, Lakonishok, and Sougiannis (2001)	2.46 [0.91]	-1.91 [-0.84]			✓
Firm age	AGE	Barry and Brown (1985)	1.23 [0.33]	-0.18 [-0.07]	✓		
Accrual quality	AQ	Francis, LaFond, Olsson, and Schipper (2005)	2.27 [1.28]	0.41 [0.40]			✓
Brand capital-to-assets	BCA	Belo, Lin, and Vitorino (2014)	1.32 [0.57]	3.20 [1.40]			✓
Hiring rate	HN	Belo, Lin, and Bazdresch (2014)	2.55 [1.34]	1.02 [0.81]			✓
Organizational capital-to-assets	OCA	Eisfeldt and Papanikolaou (2013)	2.64 [1.34]	3.09 [2.06]			✓
Operating leverage	OPLEV	Novy-Marx (2011)	4.16 [2.19]	4.40 [2.25]		✓	✓
R&D capital-to-assets	RCA	Li (2011)	3.76 [1.47]	5.28 [2.36]			✓
R&D-to-market	RDM	Chan et al. (2001)	7.70 [2.95]	3.56 [1.44]			✓
R&D-to-sales	RDS	Chan et al. (2001)	2.58 [0.88]	4.73 [1.83]			✓
Seasonality	SEASON	Heston and Sadka (2010)	10.85 [3.74]	11.29 [4.11]	✓		
SG&A-to-sales∗	SGA	Freyberger, Neuhierl, and Weber (2020)	1.61 [0.57]	4.85 [1.92]		✓	
Tobin's Q∗	TOBINQ	Freyberger et al. (2020)	5.04 [1.59]	-1.23 [-0.72]		✓	
		Panel F: Trading frictions					
Total assets∗	AT	Gandhi and Lustig (2015)	-0.48 [-0.19]	-2.78 [-2.07]		✓	
Beta arbitrage	BETAARB	Cooper et al. (2008)	-0.64 [-0.16]	7.00 [2.65]	✓		
Dimson's beta	BETADIM	Dimson (1979)	0.83 [0.35]	4.62 [2.02]			✓
Market beta	BETAFP	Frazzini and Pedersen (2014)	1.13 [0.27]	8.50 [2.59]		✓	✓
Dollar trading volume	DVOL	Brennan, Chordia, and Subrahmanyam (1998)	6.13 [2.04]	2.22 [1.96]			✓
Illiquidity	ILLIQ	Amihud (2002)	6.13 [2.04]	2.22 [1.96]			✓
Industry relative reversals∗	INDRREV	Da, Liu, and Schaumburg (2014)	9.08 [3.55]	5.13 [1.89]	✓		
Industry relative reversals (low volatility)∗	INDRREVLV	Da et al. (2014)	12.74 [5.18]	9.83 [4.34]	✓		
Idiosyncratic volatility	IVOL	Ang, Hodrick, Xing, and Zhang (2006)	6.02 [1.71]	11.10 [4.67]	✓	✓	✓
Maximum daily return	MDR	Bali, Cakici, and Whitelaw (2011)	3.31 [1.03]	7.37 [3.00]			✓
Price∗	PRICE	Blume and Husic (1973)	0.24 [0.07]	-7.07 [-3.26]	✓		
1/share price	RP	Miller and Scholes (1982)	0.26 [0.07]	-7.05 [-3.22]			✓
Share volume∗	SHVOL	Datar, Naik, and Radcliffe (1998)	-0.89 [-0.25]	2.41 [0.96]	✓		
The market equity	SIZE	Fama and French (1992)	2.19 [0.66]	-1.88 [-1.46]	✓	✓	✓
Bid-ask spread∗	SPREAD	Amihud (2002)	-0.75 [-0.20]	-6.35 [-2.22]		✓	
Short-term reversal	STREV	Jegadeesh (1990)	3.62 [1.54]	-0.35 [-0.13]	✓	✓	✓
Standard unexplained volume∗	SUV	Garfinkel (2009)	4.67 [2.17]	4.20 [2.38]		✓	
Share turnover	TURN	Datar et al. (1998)	-1.99 [-0.65]	0.71 [0.32]		✓	✓
Total volatility	TVOL	Ang et al. (2006)	3.84 [1.03]	9.79 [3.80]			✓

Table A2. Predictors of factor returns
The table describes the construction of predictors of factor returns. Panel A describes five factor-specific predictors: *Book-to-market ratio (BM)*, *Industry-adjusted book-to-market ratio (IND.BM)*, *Issuer-repurchaser spread (ISSREP)*, *One-month momentum (MOM1)*, *12-month momentum (MOM12)*, *Volatility (VOL)*, *Characteristic spread (CS)*, and *Long-run reversal (LRREV)*. Panel B describes the common time-series predictor for all factors: *Sentiment index (SENT)*.

No.	Predictor	Construction detail
		Panel A: Factor-specific predictors
1	Book-to-market ratio (*BM*)	The difference in natural logarithm of the book-to-market ratio between the long and the short portfolio of a characteristic. I calculate a firm's annual book value as in Fama and French (1992) and its monthly market value as in Asness and Frazzini (2013). Following Haddad et al. (2020), I calculate a portfolio's book-to-market ratio as the sum of the book value relative to the total market value of all firms in that portfolio.
2	Industry adjusted book-to-market ratio (*IND.BM*)	The difference in natural logarithm of the industry-adjusted book-to-market ratio between the long and the short portfolio of a characteristic. I calculate a firm's annual book value as in Fama and French (1992) and its monthly market value as in Asness and Frazzini (2013). Following Baba-Yara et al. (2021), I subtract from the firm's book-to-market ratio the value-weighted average book-to-market ratio of the industry the firm belongs to. I use Fama & French 48-industry classification and firms' 4-digit SIC code for industry assignment. A portfolio's industry-adjusted book-to-market ratio is defined as the value-weighted average of the industry-adjusted book-to-market ratio of all firms in that portfolio.
3	Issuer-repurchaser spread (*ISSREP*)	I follow Greenwood and Hanson (2012) to construct the characteristic-based issuer-repurchaser spread for a factor portfolio as $$ISSREP^X_{t-1} = \frac{\sum_{j \in Issuers} X_{j,t-1}}{N^{Issuers}_{t-1}} - \frac{\sum_{j \in Repurchasers} X_{j,t-1}}{N^{Repurchasers}_{t-1}}, \quad (15)$$ where $X_{j,t-1}$ denote firm j's cross-sectional decile for characteristic X in year $t-1$, and N_{t-1} is the total number of issuers or purchasers. Greenwood and Hanson (2012) define issuers (purchasers) as firms whose net stock issuance is greater than 10% (less than -0.5%) at December of year $t-1$. I follow Fama and French (2008) and define net stock issuance as the change in log split-adjusted shares outstanding from Compustat annual file.
4	One-month momentum (*MOM1*)	Following Gupta and Kelly (2019), I calculate a factor portfolio's one-month momentum as its return from the prior month, scaled by the prior 3-year variance of factor returns.
5	12-month momentum (*MOM12*)	Following Ehsani and Linnainmaa (2022), I define a factor portfolio's 12-month momentum as an indicator variable that equals one if the factor portfolio's average monthly returns over the past 12 months is positive and zero otherwise.

(Continued on next page)

Table A2 (*continued*)

No.	Predictor	Construction detail
6	Volatility (VOL)	Following Moreira and Muir (2017), I estimate factor volatility as the realized variance of daily factor returns in the prior month, scaled by the average of monthly variances up to the prior month. I use the natural logarithm of volatility.
7	Characteristic spread (CS)	Following Kelly et al. (2023), I transform each characteristic into a [-0.5,+0.5] interval and calculate a factor's characteristic spread as the difference in the value-weighted characteristic between the long and the short leg of the factor.
8	Long-run reversal ($LRREV$)	Following Moskowitz et al. (2012), I calculate a factor's long-run reversal signal as cumulative returns over the past 5 years, scaled by the realized variance of the factor's returns over the same period.
		Panel B: Single time-series predictors
9	Sentiment index ($SENT$)	For in-sample tests and out-of-sample tests that use a split-sample estimation design, I use Baker and Wurgler (2006) full-sample orthogonalized sentiment series. For out-of-sample tests that use either a recursive or a rolling window estimation design, I follow Huang et al. (2015) to create look-ahead bias-free series. First, at each out-of-sample month $t+1$ I use the data on five individual sentiment series (the close-end fund discount rate, the number of IPOs, the 12-month lagged first-day returns of IPOs, the 12-month lagged dividend premium, and the equity share in new issues) from 1965:07 to only month t, and standardize them to have mean of 0 and standard deviation of 1. Next, I regress each series on six macroeconomic variables (the growth of industrial production, the growth of durable consumption, the growth of nondurable consumption, the growth of service consumption, the growth of employment, and a dummy variable for NBER-dated recessions) to obtain five orthogonalized series. I smooth five orthogonalized series with six-month average values to mitigate outliers in the individual series. For the conventional predictive regressions and Haddad et al. (2020) principal-component portfolio approach, I use the first principal component extracted from the five individual sentiment series to predict factor returns for month $t+1$. For all shrinkage methods, I use all five individual sentiment measures as predictors. Data on the original variables is available on Jeffrey Wurgler's website at `https://pages.stern.nyu.edu/~jwurgler/`.

Table A3. Principal component portfolio approach: Split-sample specification with reverse samples
The table summarizes the out-of-sample performance of factor return prediction using the principal component (PC) portfolio approach. The sample period is from 1970:01 to 2021:12. The number of factors is 92. Additional detail on the factors is in Table A1 in the Appendix. The predictor is *Book-to-market ratio* (BM). I adopt a split-sample design for the out-of-sample tests. I estimate parameters using half of the sample, and use the estimates to construct forecasts for the other half. *Forward* (*Reverse*) design refers to the use of the first half (second half) of the sample for parameter estimation. Panel A reports the results on the predictability for the first five PC portfolios from the univariate predictive regression

$$PC_{t+1}^k = \lambda_0^k + \lambda_1^k X_t^k + \epsilon_{t+1}^k,$$

where PC_{t+1}^k is the excess return of PC portfolio k ($k = \overline{1,5}$) in month $t+1$, and X_t^k is the book-to-market ratio of PC portfolio k in month t. I obtain the PC portfolio returns and their predictor using the eigenvectors of the covariance matrix of factor returns. Rows (1) and (2) report the predictive coefficient estimate $\hat{\beta}_1$ and t-statistic (in brackets), respectively. Rows (3) reports the out-of-sample monthly R^2, respectively. Panel B reports the results on the predictability for 92 factors under the PC portfolio approach for two split-sample designs. I obtain a PC-based return forecast for factor i as

$$\hat{R}_{i,t+1} = \sum_{k=1}^{5} \hat{\omega}_{i,t+1}^k \widehat{PC}_{t+1}^k,$$

where $\hat{R}_{i,t+1}$ and $\widehat{PC}_{t+1}^k$ are the excess return forecasts of factor portfolio i and PC portfolio k in month $t+1$, respectively. $\hat{\omega}_{i,k,t+1}^k$ is the loading of factor portfolio i on PC portfolio k from the PC estimation. Columns (1) and (2) report the mean (standard deviation), and the median R_{OS}^2, respectively. Column (3) reports the total R_{OS}^2. Column (4) reports the number of R_{OS}^2s that are non-negative, and statistically significant at the 5% level in brackets. I calculate Campbell and Thompson's (2008) R_{OS}^2 using Equation 8. The total R_{OS}^2 is calculated using Equation 9. I use Clark and West's (2007) procedure for the R_{OS}^2 tests.

Panel A: Prediction of five largest equity components					
	PC1	PC2	PC3	PC4	PC5
	(1)	(2)	(3)	(4)	(5)
Forward					
Own BM	1.25	2.63	0.44	-0.35	-0.48
t-statistic	[1.06]	[2.74]	[0.89]	[-0.61]	[-0.76]
R_{OS}^2 (%)	1.47	2.46	-0.43	-0.24	-0.62
Reverse					
Own BM	4.20	0.45	0.68	0.71	-0.09
t-statistic	[3.41]	[0.62]	[0.60]	[0.64]	[-0.10]
R_{OS}^2 (%)	3.95	0.23	-0.54	0.41	-0.08

Panel B: Out-of-sample prediction across individual factors				
Method	R_{OS}^2 (%)		Total R_{OS}^2 (%)	$R_{OS}^2 \geq 0\%$ [5%−Sig.]
	Mean (SD)	Median		
	(1)	(2)	(3)	(4)
PC portfolio				
Forward	0.64 (2.24)	0.99	1.15	67 [50]
Reverse	-0.62 (3.16)	-0.11	-0.15	44 [31]

Table A4. Predictive regressions: In-sample results

The table summarizes the in-sample performance of factor return prediction using the conventional predictive regression approach. The sample period is from 1970:01 to 2021:12. The number of factors is 92. Additional detail on the factors is in Table A1 in the Appendix. The nine predictors are *Book-to-market ratio (BM)*, *Industry-adjusted book-to-market ratio (IND.BM)*, *Issuer-repurchaser spread (ISSREP)*, *One-month momentum (MOM1)*, *12-month momentum (MOM12)*, *Volatility (VOL)*, *Characteristic spread (CS)*, *Long-run reversal (LRREV)*, and *Sentiment index (SENT)*. For *SENT*, I use Baker and Wurgler's (2006) full-sample orthogonalized series. Predictor construction detail is described in Table A2 in the Appendix. For each factor and predictor, I run the univariate predictive regression

$$R_{i,t+1} = \beta_{i0} + \beta_{i1} X_t + \varepsilon_{i,t+1},$$

where $R_{i,t+1}$ is the excess return of factor portfolio i in month $t+1$, and X_t is the predictor variable in month t. I record the coefficient estimate $\hat{\beta}_1$, the p-value from the t-test, and the regression R^2. Columns (1) to (3) report the the total R^2 across all factors for three sample periods: full, first half, and second half, respectively. Total R^2 is $1 - (\Sigma_i SSE_i / \Sigma_i SST_i)$, in which $SSE_i = \Sigma_{t=1}^{T} (R_{i,t} - \hat{R}_{i,t})^2$, and $SST_i = \Sigma_{t=1}^{T} (R_{i,t} - \bar{R}_i)^2$, with $R_{i,t}$, $\hat{R}_{i,t}$, and $\bar{R}_f$ being the excess return of factor portfolio i in month t, its return forecast in month t, and the sample mean return, respectively. T is the number of months in the sample. Columns (4) to (6) reports the number of $\hat{\beta}_1$s that have the same sign as in original studies for three sample periods, and Column (7) reports the number of estimates that have the same correct sign in both first and second half of the full sample. The numbers in brackets indicate the number of statistically significant estimates (p-value in the t-test less than 10%). I use the Newey and West's (1987) $t-$statistic with a 2-year window for the kernel.

Predictor	Total R^2 (%)			Correct [Sig.]			
	Full sample	First-half	Second-half	Full sample	First-half	Second-half	First to second
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
BM	0.71	0.68	0.75	64 [32]	69 [24]	59 [15]	54 [7]
IND.BM	0.73	0.72	0.78	62 [35]	66 [30]	57 [18]	52 [11]
ISSREP	0.28	0.53	0.36	49 [25]	54 [21]	45 [11]	34 [5]
MOM1	0.67	0.98	0.41	85 [41]	79 [44]	76 [16]	66 [10]
MOM12	0.31	0.53	0.09	47 [8]	50 [10]	51 [10]	32 [2]
VOL	0.37	0.54	0.31	42 [20]	52 [25]	36 [5]	29 [4]
CS	0.19	0.13	0.52	81 [34]	72 [42]	66 [17]	53 [12]
LRREV	0.06	0.12	0.03	40 [16]	52 [21]	38 [6]	31 [5]
SENT	0.96	0.74	1.94	85 [43]	80 [46]	75 [19]	65 [11]

Table A5. Structural break tests for PC predictive regressions and factor loadings

The table reports results from structural break tests in predictive regressions of PC portfolio returns on their book-to-market ratio (Panel A), and in the factor loadings (Panel B). The sample period is from 1965:07 to 2021:12. The number of factors is 92. Additional detail on the factors is in Table A1 in the Appendix. For tests of structural changes in predictive regressions, the tests follow Bai and Perron (1998) and Bai and Perron (2003), and allow for an unknown number of structural breaks. In particular, panel A reports the results from the univariate predictive regression

$$PC^k_{t+1} = \beta^k_0 + \beta^k_1 BM^k_t + \epsilon^k_{t+1},$$

where PC^k_{t+1} is the excess return of PC portfolio k ($k = \overline{1,5}$) in month $t+1$, and BM^k_t is the book-to-market ratio of PC portfolio k in month t. For tests of structural changes in the factor loadings, the tests follow Chen et al. (2014) and allow for one unknown structural break in the factor loadings. In particular, panel B reports the results from the regression of the first PC on the other four PCs for two subperiods before and after the break. Column (2) reports the $UDmax$ statistic and *** denotes significance at the 1% level in a test of the null hypothesis of zero breaks. Column (3) reports the $WDmax(10\%)$ for tests in panel A or $supWald$ statistic for tests in panel B. The $WDmax(10\%)$ statistic indicates the 10% significance level, and * denotes significance at the 10% level. Columns (4)-(6) show the estimated structural break dates. Columns (7)-(10) report the coefficient estimates β for each subperiod. The numbers in parentheses are t-statistics based on Andrews (1991) standard errors.

	Bai and Perron (1998) statistics		Bai and Perron (1998) break dates			Subperiod predictive coefficient			
PC	$UDmax$	$WDmax$ (10%) / $supWald$	1st break	2nd break	3rd break	$\beta(1)$	$\beta(2)$	$\beta(3)$	$\beta(4)$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Predictive regressions									
1	6.64	8.80*	1983:06	2000:02	—	0.15	5.28	2.82	—
					—	[0.11]	[4.53]	[2.77]	—
2	43.93***	54.70*	1988:11	2000:08	2011:02	0.72	-0.97	7.59	0.75
						[0.94]	[-1.01]	[7.44]	[0.74]
3	—	—	—	—	—	—	—	—	—
4	15.84***	15.80*	1987:08	2001:05	—	0.51	4.25	-0.13	—
					—	[0.48]	[3.58]	[-0.14]	—
5	—	—	—	—	—	—	—	—	—
Panel B: Time-varying factor loadings									
2		197.72***	2000:05			-0.27	0.24		
						[-2.11]	[2.76]		
3		197.72***	2000:05			0.02	0.30		
						[0.05]	[1.45]		
4		197.72***	2000:05			-0.83	0.53		
						[-4.94]	[2.70]		
5		197.72***	2000:05			-1.01	1.17		
						[-3.18]	[6.30]		

Table A6. Factor prediction with shrinkage methods: Robustness across estimation designs
The table summarizes results on the out-of-sample factor predictability under the shrinkage methods across different estimation designs. The sample period is from 1970:01 to 2021:12. The number of factors is 92. Additional detail on the factors is in Table A1 in the Appendix. The predictors are are *Book-to-market ratio*, *Industry-adjusted book-to-market ratio*, *Issuer-repurchaser spread*, *One-month momentum*, *12-month momentum*, *Volatility*, *Characteristic spread*, *Long-run reversal*, and five components of *Sentiment index*. Predictor construction detail is described in Table A2 in the Appendix. The shrinkage methods include forecast combination (*FC*), discount mean square of prediction errors (*DMSPE*), predictor average (*Average*), principal component regression (*PCR*), and partial least squares (*PLS*). *FC* is the equal-weighted average of univariate predictive regression forecasts from all predictors. *DMSPE* is the weighted average of univariate predictive regression forecasts, in which forecasts that have lower prediction errors over the holdout period have greater weight. *Average* is the univariate predictive regression forecast based on the cross-sectional average of all predictors. *PCR* (*PLS*) is a univariate predictive regression forecast based on the first principal component (return-relevant component) of all predictors. Panel A (B) reports the results using an expanding-window design with the initial training period of 240 (360) months. Panel C reports the results using a rolling-window design with the initial training period equal to the first half of the sample (1970:01-1995:12). The rolling window equals to the initial training period. Columns (1)-(3) report the mean, the standard deviation (SD), and the median of R^2_{OS}. Column (3) reports the total R^2_{OS}. Column (4) reports the number of R^2_{OS}s that are non-negative, and statistically significant with p-value less than 0.1 in brackets. I calculate Campbell and Thompson's (2008) R^2_{OS} using Equation 8. The total R^2_{OS} is calculated using Equation 9. I use Clark and West's (2007) procedure for the R^2_{OS} tests.

Method	Individual R^2_{OS} (%)			Total R^2_{OS} (%)	$R^2_{OS} \geq 0[Sig.]$
	Mean	SD	Median		
	(1)	(2)	(3)	(4)	(5)
Panel A: Expanding-window estimation design (240 months)					
	(1)	(2)	(3)	(4)	(5)
FC	0.57	0.49	0.50	0.65	81 [62]
DMSPE	0.59	0.50	0.47	0.65	82 [64]
Average	0.65	1.12	0.42	0.71	57 [45]
PCR	1.15	1.47	1.00	1.35	70 [61]
PLS	0.72	2.21	0.91	1.08	62 [60]
Panel B: Expanding-window estimation design (360 months)					
FC	0.49	0.54	0.42	0.56	72 [45]
DMSPE	0.50	0.54	0.40	0.56	74 [45]
Average	0.54	1.21	0.44	0.61	52 [42]
PCR	1.01	1.70	0.88	1.25	63 [53]
PLS	0.31	2.40	0.33	0.67	50 [42]
Panel C: Rolling-window estimation design					
FC	0.60	0.72	0.39	0.75	73 [39]
DMSPE	0.45	0.96	0.37	0.59	58 [29]
Average	0.31	1.23	0.05	0.48	47 [38]
PCR	0.35	1.82	-0.04	0.82	45 [39]
PLS	-0.21	2.46	-0.70	0.34	38 [36]

Table A7. Factor prediction with shrinkage methods: Predictor exclusion
The table summarizes results on the out-of-sample factor predictability under the shrinkage methods when select predictor from the rows are excluded for prediction. The sample period is from 1970:01 to 2021:12. The number of factors is 92. Additional detail on the factors is in Table A1 in the Appendix. The predictors are *Book-to-market ratio (BM)*, *Industry-adjusted book-to-market ratio (IND.BM)*, *Issuer-repurchaser spread (ISSREP)*, *One-month momentum (MOM1)*, *12-month momentum (MOM12)*, *Volatility (VOL)*, *Characteristic spread (CS)*, *Long-run reversal (LRREV)*, and five components of *Sentiment index (SENT)*. For *SENT*, I exclude all five components when making predictions. Predictor construction detail is described in Section 2 in the main text. The shrinkage methods include forecast combination (*FC*), discount mean square of prediction errors (*DMSPE*), predictor average (*Average*), principal component regression (*PCR*), and partial least squares (*PLS*). *FC* is the equal-weighted average of univariate predictive regression forecasts from all predictors. *DMSPE* is the weighted average of univariate predictive regression forecasts, in which forecasts that have lower prediction errors over the holdout period have greater weight. *Average* is the univariate predictive regression forecast based on the cross-sectional average of all predictors. *PCR* (*PLS*) is a univariate predictive regression forecast based on the first principal component (return-relevant component) of all predictors. I adopt an expanding-window design for the out-of-sample tests. For *FC*, *Average*, *PCR*, and *PLS*, I use an initial training window equal to the first half of the sample (1970:01 - 1995:12) to estimate forecasts for 1996:01. For *DMSPE* that requires holdout periods, I use an initial training window from 1970:01 to 1985:12, and the subsequent 120 months (1986:01 to 1995:12) as the holdout period to estimate forecasts for 1996:01. The holdout window length is the same for subsequent forecasts. I expand the training window each month to estimate forecasts for the rest of the sample. The out-of-sample evaluation period is from 1996:01 to 2021:12. Panel A reports the mean and the total R^2_{OS} (in brackets). Panel B reports the annualized mean and median Sharpe ratio (in parentheses) for factor timing strategies. Panel C reports the annualized mean and median Certainty Equivalent Returns (in parentheses) for factor timing strategies. The first row in each panel reports the base case results as in Tables 4, 6, and 7. The optimal weight for factor timing strategy i in month t is estimated as $\hat{R}^m_{i,t+1}/(\gamma\hat{\sigma}^2_{i,t+1})$, where $\hat{R}^m_{i,t+1}$ is the return forecast using forecasting method m, $\hat{\sigma}^2_{i,t+1}$ is the sample variance using all data up to month t, and γ is a risk aversion parameter. I calculate Campbell and Thompson's (2008) R^2_{OS} using Equation 8. The total R^2_{OS} is calculated using Equation 9. I use a risk aversion parameter of one, and impose a leverage constraint that the absolute weight on the factor portfolio is less than or equal to two.

Excluded predictor	FC	DMSPE	Average	PCR	PLS
	(1)	(2)	(3)	(4)	(5)
Panel A: Mean R^2_{OS} [total R^2_{OS}] (%)					
Base case	0.52 [0.60]	0.53 [0.61]	0.52 [0.58]	1.08 [1.30]	0.52 [0.91]
BM	0.51 [0.57]	0.52 [0.57]	0.47 [0.46]	1.03 [1.17]	0.67 [0.99]
IND.BM	0.51 [0.56]	0.52 [0.56]	0.45 [0.45]	0.99 [1.07]	0.66 [0.96]
ISSREP	0.51 [0.61]	0.52 [0.61]	0.57 [0.58]	1.10 [1.40]	0.55 [0.99]
MOM1	0.48 [0.57]	0.49 [0.57]	0.37 [0.42]	0.93 [1.16]	0.45 [0.83]
MOM12	0.51 [0.61]	0.52 [0.62]	0.40 [0.46]	0.93 [1.12]	0.50 [0.91]
VOL	0.56 [0.64]	0.57 [0.64]	0.57 [0.72]	1.07 [1.33]	0.72 [1.07]
CS	0.54 [0.63]	0.55 [0.63]	0.57 [0.64]	1.04 [1.33]	0.52 [0.91]
LRREV	0.54 [0.64]	0.56 [0.64]	0.50 [0.59]	0.99 [1.27]	0.47 [0.88]
SENT	0.53 [0.66]	0.54 [0.66]	0.28 [0.60]	0.15 [0.40]	-0.04 [0.32]
Panel B: Annualized mean (median) Sharpe ratio					
Base case	0.23 [0.23]	0.23 [0.23]	0.22 [0.22]	0.28 [0.28]	0.27 [0.27]
BM	0.23 (0.22)	0.23 (0.22)	0.21 (0.21)	0.29 (0.26)	0.28 (0.29)
IND.BM	0.23 (0.21)	0.23 (0.22)	0.21 (0.22)	0.28 (0.27)	0.27 (0.28)
ISSREP	0.23 (0.24)	0.23 (0.23)	0.22 (0.22)	0.29 (0.29)	0.27 (0.27)
MOM1	0.22 (0.22)	0.22 (0.21)	0.18 (0.18)	0.25 (0.25)	0.24 (0.25)
MOM12	0.23 (0.23)	0.23 (0.23)	0.19 (0.18)	0.26 (0.26)	0.25 (0.24)
VOL	0.24 (0.24)	0.24 (0.23)	0.23 (0.21)	0.28 (0.27)	0.28 (0.28)
CS	0.24 (0.23)	0.23 (0.22)	0.22 (0.20)	0.28 (0.30)	0.27 (0.27)
LRREV	0.23 (0.22)	0.23 (0.22)	0.21 (0.21)	0.29 (0.29)	0.27 (0.26)
SENT	0.23 (0.21)	0.24 (0.22)	0.23 (0.25)	0.25 (0.24)	0.24 (0.24)
Panel C: Annualized mean (median) CER (%)					
Base case	2.39 [1.85]	2.40 [1.96]	2.00 [2.17]	4.00 [3.11]	3.64 [3.43]
BM	2.32 (1.58)	2.33 (1.65)	1.87 (1.64)	4.08 (3.03)	3.91 (3.80)
IND.BM	2.29 (1.36)	2.30 (1.60)	1.82 (1.90)	3.62 (2.81)	3.61 (3.39)
ISSREP	2.43 (2.38)	2.42 (2.18)	2.01 (2.30)	4.18 (3.71)	3.66 (3.69)
MOM1	2.18 (1.30)	2.21 (1.61)	1.20 (1.28)	3.19 (2.34)	2.84 (2.92)
MOM12	2.32 (1.90)	2.34 (2.06)	1.26 (1.31)	3.19 (2.41)	3.17 (2.64)
VOL	2.52 (2.49)	2.52 (2.28)	2.59 (1.97)	3.93 (3.22)	4.00 (3.54)
CS	2.54 (2.16)	2.52 (2.17)	2.05 (2.00)	4.04 (3.30)	3.80 (3.30)
LRREV	2.50 (2.03)	2.51 (1.77)	1.79 (1.61)	4.06 (3.33)	3.59 (3.25)
SENT	2.64 (1.82)	2.71 (2.05)	2.56 (2.33)	2.85 (2.18)	2.63 (2.55)

Table A8. Factor prediction with shrinkage methods: Predictor inclusion

The table summarizes results on the out-of-sample factor predictability under the shrinkage methods when select variable from the rows are included in the set of base predictors for prediction. The sample period is from 1970:01 to 2021:12. The number of factors is 92. Additional detail on the factors is in Table A1 in the Appendix. The base predictors are *Book-to-market ratio (BM)*, *Industry-adjusted book-to-market ratio (IND.BM)*, *Issuer-repurchaser spread (ISSREP)*, *One-month momentum (MOM1)*, *12-month momentum (MOM12)*, *Volatility (VOL)*, *Characteristic spread (CS)*, *Long-run reversal (LRREV)*, and five components of *Sentiment index (SENT)*. Predictor construction detail is described in Section 2 in the main text. The included predictors include 14 economic variables in Goyal and Welch (2008). The last row reports the results when all 14 variables are included. The shrinkage methods include forecast combination (*FC*), discount mean square of prediction errors (*DMSPE*), predictor average (*Average*), principal component regression (*PCR*), and partial least squares (*PLS*). *FC* is the equal-weighted average of univariate predictive regression forecasts from all predictors. *DMSPE* is the weighted average of univariate predictive regression forecasts, in which forecasts that have lower prediction errors over the holdout period have greater weight. *Average* is the univariate predictive regression forecast based on the cross-sectional average of all predictors. *PCR* (*PLS*) is a univariate predictive regression forecast based on the first principal component (return-relevant component) of all predictors. I adopt an expanding-window design for the out-of-sample tests. For *FC*, *Average*, *PCR*, and *PLS*, I use an initial training window equal to the first half of the sample (1970:01 - 1995:12) to estimate forecasts for 1996:01. For *DMSPE* that requires holdout periods, I use an initial training window from 1970:01 to 1985:12, and the subsequent 120 months (1986:01 to 1995:12) as the holdout period to estimate forecasts for 1996:01. The holdout window length is the same for subsequent forecasts. I expand the training window each month to estimate forecasts for the rest of the sample. The out-of-sample evaluation period is from 1996:01 to 2021:12. I report the mean and the total R^2_{OS} (in brackets). The first row reports the base case results as in Table 4. I calculate Campbell and Thompson's (2008) R^2_{OS} using Equation 8. The total R^2_{OS} is calculated using Equation 9.

Included predictor	Mean R^2_{OS} [total R^2_{OS}] (%)				
	FC	DMSPE	Average	PCR	PLS
	(1)	(2)	(3)	(4)	(5)
Base case	0.52 [0.60]	0.53 [0.61]	0.52 [0.58]	1.08 [1.30]	0.52 [0.91]
Dividend-price ratio	0.53 [0.60]	0.55 [0.61]	0.36 [0.40]	1.02 [1.22]	0.57 [0.95]
Dividend yield	0.53 [0.60]	0.54 [0.61]	0.36 [0.40]	1.00 [1.20]	0.57 [0.95]
Earnings-price ratio	0.50 [0.57]	0.52 [0.58]	0.39 [0.49]	1.06 [1.26]	0.49 [0.85]
Dividend-earnings ratio	0.53 [0.64]	0.54 [0.64]	0.36 [0.36]	0.85 [1.14]	0.48 [0.98]
Stock variance	0.45 [0.54]	0.47 [0.55]	0.30 [0.33]	0.93 [1.22]	0.27 [0.73]
Book-to-market ratio	0.50 [0.57]	0.52 [0.57]	0.54 [0.63]	0.97 [1.20]	0.47 [0.84]
Net equity expansion	0.48 [0.55]	0.50 [0.56]	0.43 [0.47]	0.96 [1.16]	0.39 [0.76]
Treasury bill yield	0.50 [0.59]	0.52 [0.59]	0.52 [0.60]	0.97 [1.26]	0.52 [0.92]
Long-term Treasury bond yield	0.49 [0.58]	0.51 [0.58]	0.55 [0.64]	0.98 [1.26]	0.47 [0.87]
Long-term Treasury bond return	0.49 [0.57]	0.50 [0.58]	0.48 [0.50]	1.08 [1.29]	0.47 [0.93]
Term spread	0.48 [0.56]	0.50 [0.57]	0.57 [0.62]	0.84 [1.05]	0.47 [0.87]
Default yield spread	0.52 [0.61]	0.54 [0.62]	0.53 [0.58]	0.92 [1.13]	0.59 [1.08]
Default return spread	0.45 [0.52]	0.47 [0.53]	0.44 [0.51]	1.06 [1.27]	0.17 [0.54]
Inflation	0.49 [0.58]	0.51 [0.58]	0.58 [0.68]	0.95 [1.22]	0.47 [0.93]
All 14 predictors	0.37 [0.42]	0.39 [0.44]	0.23 [0.31]	0.37 [0.52]	-0.23 [0.34]

Appendix references

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

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets*, *5*(1), 31–56.

Andrews, D. W. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 817–858.

Ang, A., Hodrick, R. J., Xing, Y., and Zhang, X. (2006). The cross-section of volatility and expected returns. *Journal of Finance*, *61*(1), 259–299.

Asness, C. S., and Frazzini, A. (2013). The devil in HML's details. *Journal of Portfolio Management*, *39*(4), 49–68.

Baba-Yara, F., Boons, M., and Tamoni, A. (2021). Value return predictability across asset classes and commonalities in risk premia. *Review of Finance*, *25*(2), 449–484.

Baker, M., and Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *Journal of Finance*, *61*(4), 1645–1680.

Bali, T. G., Cakici, N., and Whitelaw, R. F. (2011). Maxing out: Stocks as lotteries and the cross-section of expected returns. *Journal of Financial Economics*, *99*(2), 427–446.

Ball, R., Gerakos, J., Linnainmaa, J. T., and Nikolaev, V. (2016). Accruals, cash flows, and operating profitability in the cross section of stock returns. *Journal of Financial Economics*, *121*(1), 28–45.

Barbee Jr, W. C., Mukherji, S., and Raines, G. A. (1996). Do sales–price and debt–equity explain stock returns better than book–market and firm size? *Financial Analysts Journal*, *52*(2), 56–60.

Barry, C. B., and Brown, S. J. (1985). Differential information and security market equilibrium. *Journal of Financial and Quantitative Analysis*, *20*(4), 407–422.

Barth, M. E., Elliott, J. A., and Finn, M. W. (1999). Market rewards associated with patterns of increasing earnings. *Journal of Accounting Research*, *37*(2), 387–413.

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*(3), 663–682.

Belo, F., and Lin, X. (2012). The inventory growth spread. *Review of Financial Studies*, *25*(1), 278–313.

Belo, F., Lin, X., and Bazdresch, S. (2014). Labor hiring, investment, and stock return predictability in the cross section. *Journal of Political Economy*, *122*(1), 129–177.

Belo, F., Lin, X., and Vitorino, M. A. (2014). Brand capital and firm value. *Review of Economic Dynamics*, *17*(1), 150–169.

Bhandari, L. C. (1988). Debt/equity ratio and expected common stock returns: Empirical evidence. *Journal of Finance*, *43*(2), 507–528.

Blume, M. E., and Husic, F. (1973). Price, beta, and exchange listing. *Journal of Finance*, *28*(2), 283–299.

Boudoukh, J., Michaely, R., Richardson, M., and Roberts, M. R. (2007). On the importance of measuring payout yield: Implications for empirical asset pricing. *Journal of Finance*, *62*(2), 877–915.

Brennan, M. J., Chordia, T., and Subrahmanyam, A. (1998). Alternative factor specifications, security characteristics, and the cross-section of expected stock returns. *Journal of Financial Economics*, *49*(3), 345–373.

Chan, L. K., Jegadeesh, N., and Lakonishok, J. (1996). Momentum strategies. *Journal of Finance*, *51*(5), 1681–1713.

Chan, L. K., Lakonishok, J., and Sougiannis, T. (2001). The stock market valuation of research and development expenditures. *Journal of Finance*, *56*(6), 2431–2456.

Chen, L., Dolado, J. J., and Gonzalo, J. (2014). Detecting big structural breaks in large factor models. *Journal of Econometrics*, *180*(1), 30–48.

Chen, L., Novy-Marx, R., and Zhang, L. (2011). An alternative three-factor model. *Working Paper*.

Cooper, M. J., Gulen, H., and Schill, M. J. (2008). Asset growth and the cross-section of stock returns. *the Journal of Finance*, *63*(4), 1609–1651.

Da, Z., Liu, Q., and Schaumburg, E. (2014). A closer look at the short-term return reversal. *Management Science*, *60*(3), 658–674.

Daniel, K., and Titman, S. (2006). Market reactions to tangible and intangible information. *Journal of Finance*, *61*(4), 1605–1643.

Datar, V. T., Naik, N. Y., and Radcliffe, R. (1998). Liquidity and stock returns: An alternative test. *Journal of Financial Markets*, *1*(2), 203–219.

De Bondt, W. F., and Thaler, R. (1985). Does the stock market overreact? *Journal of Finance*, *40*(3), 793–805.

Dechow, P. M., Sloan, R. G., and Soliman, M. T. (2004). Implied equity duration: A new measure of equity risk. *Review of Accounting Studies*, *9*(2), 197–228.

Dichev, I. D. (1998). Is the risk of bankruptcy a systematic risk? *Journal of Finance*, *53*(3), 1131–1147.

Dimson, E. (1979). Risk measurement when shares are subject to infrequent trading. *Journal of Financial Economics*, *7*(2), 197–226.

Ehsani, S., and Linnainmaa, J. T. (2022). Factor momentum and the momentum factor. *Journal of Finance*, *77*(3), 1877–1919.

Eisfeldt, A. L., and Papanikolaou, D. (2013). Organization capital and the cross-section of expected returns. *Journal of Finance*, *68*(4), 1365–1406.

Fairfield, P. M., Whisenant, J. S., and Yohn, T. L. (2003). Accrued earnings and growth: Implications for future profitability and market mispricing. *The Accounting Review*, *78*(1), 353–371.

Fama, E. F., and French, K. R. (1992). The cross-section of expected stock returns. *Journal of Finance*, *47*(2), 427–465.

Fama, E. F., and French, K. R. (2008). Dissecting anomalies. *Journal of Finance*, *63*(4), 1653–1678.

Francis, J., LaFond, R., Olsson, P., and Schipper, K. (2005). The market pricing of accruals quality. *Journal of Accounting and Economics*, *39*(2), 295–327.

Frazzini, A., and Pedersen, L. H. (2014). Betting against beta. *Journal of Financial Economics*, *111*(1), 1–25.

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

Gandhi, P., and Lustig, H. (2015). Size anomalies in us bank stock returns. *Journal of Finance*, *70*(2), 733–768.

Garfinkel, J. A. (2009). Measuring investors' opinion divergence. *Journal of Accounting Research*, *47*(5), 1317–1348.

George, T. J., and Hwang, C.-Y. (2004). The 52-week high and momentum investing. *Journal of Finance*, *59*(5), 2145–2176.

Gorodnichenko, Y., and Weber, M. (2016). Are sticky prices costly? evidence from the stock market. *American Economic Review*, *106*(1), 165–99.

Green, J., Hand, J. R., and Zhang, X. F. (2017). The characteristics that provide independent information about average us monthly stock returns. *Review of Financial Studies*, *30*(12), 4389–4436.

Greenwood, R., and Hanson, S. G. (2012). Share issuance and factor timing. *Journal of Finance*, *67*(2), 761–798.

Gupta, T., and Kelly, B. T. (2019). Factor momentum everywhere. *Journal of Portfolio Management*, *45*(3), 13–36.

Haddad, V., Kozak, S., and Santosh, S. (2020). Factor timing. *Review of Financial Studies*, *33*(5), 1980–2018.

Hafzalla, N., Lundholm, R., and Matthew Van Winkle, E. (2011). Percent accruals. *The Accounting Review*, *86*(1), 209–236.

Haugen, R. A., and Baker, N. L. (1996). Commonality in the determinants of expected stock returns. *Journal of Financial Economics*, *41*(3), 401–439.

Heston, S. L., and Sadka, R. (2010). Seasonality in the cross section of stock returns: the international evidence. *Journal of*

Financial and Quantitative Analysis, *45*(5), 1133–1160.

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

Hou, K., Karolyi, G. A., and Kho, B.-C. (2011). What factors drive global stock returns? *Review of Financial Studies*, *24*(8), 2527–2574.

Huang, D., Jiang, F., Tu, J., and Zhou, G. (2015). Investor sentiment aligned: A powerful predictor of stock returns. *Review of Financial Studies*, *28*(3), 791–837.

Jegadeesh, N. (1990). Evidence of predictable behavior of security returns. *Journal of Finance*, *45*(3), 881–898.

Jegadeesh, N., and Livnat, J. (2006). Revenue surprises and stock returns. *Journal of Accounting and Economics*, *41*(1-2), 147–171.

Jegadeesh, N., and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance*, *48*(1), 65–91.

Lakonishok, J., Shleifer, A., and Vishny, R. W. (1994). Contrarian investment, extrapolation, and risk. *The journal of finance*, *49*(5), 1541–1578.

Lewellen, J. (2015). The cross section of expected stock returns. *Critical Finance Review*, *49*.

Li, D. (2011). Financial constraints, r&d investment, and stock returns. *Review of Financial Studies*, *24*(9), 2974–3007.

Litzenberger, R. H., and Ramaswamy, K. (1979). The effect of personal taxes and dividends on capital asset prices: Theory and empirical evidence. *Journal of Financial Economics*, *7*(2), 163–195.

Lyandres, E., Sun, L., and Zhang, L. (2008). The new issues puzzle: Testing the investment-based explanation. *Review of Financial Studies*, *21*(6), 2825–2855.

Miller, M. H., and Scholes, M. S. (1982). Dividends and taxes: Some empirical evidence. *Journal of Political Economy*, *90*(6), 1118–1141.

Moskowitz, T. J., and Grinblatt, M. (1999). Do industries explain momentum? *Journal of Finance*, *54*(4), 1249–1290.

Novy-Marx, R. (2011). Operating leverage. *Review of Finance*, *15*(1), 103–134.

Novy-Marx, R. (2012). Is momentum really momentum? *Journal of Financial Economics*, *103*(3), 429–453.

Novy-Marx, R. (2013). The other side of value: The gross profitability premium. *Journal of Financial Economics*, *108*(1), 1–28.

Palazzo, B. (2012). Cash holdings, risk, and expected returns. *Journal of Financial Economics*, *104*(1), 162–185.

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

Pontiff, J., and Woodgate, A. (2008). Share issuance and cross-sectional returns. *Journal of Finance*, *63*(2), 921–945.

Richardson, S. A., Sloan, R. G., Soliman, M. T., and Tuna, I. (2005). Accrual reliability, earnings persistence and stock prices. *Journal of accounting and economics*, *39*(3), 437–485.

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

Soliman, M. T. (2008). The use of dupont analysis by market participants. *The accounting review*, *83*(3), 823–853.

Thomas, J. K., and Zhang, H. (2002). Inventory changes and future returns. *Review of Accounting Studies*, *7*(2), 163–187.

Titman, S., Wei, K. J., and Xie, F. (2004). Capital investments and stock returns. *Journal of Financial and Quantitative Analysis*, *39*(4), 677–700.

Xing, Y. (2008). Interpreting the value effect through the q-theory: An empirical investigation. *Review of Financial Studies*, *21*(4), 1767–1795.