Firm Characteristics and Expected Stock Returns

Yufeng Han
University of North Carolina at Charlotte
yufeng.han@uncc.edu

Ai He
University of South Carolina
ai.he@moore.sc.edu

David E. Rapach
Washington University in St. Louis and Saint Louis University
david.rapach@wustl.edu

Guofu Zhou*
Washington University in St. Louis and China Academy of Financial Research
zhou@wustl.edu

September 10, 2020

*Corresponding author. Send correspondence to Guofu Zhou, Olin School of Business, Washington University in St. Louis, St. Louis, MO 63130; e-mail: zhou@wustl.edu; phone: 314-935-6384. We thank conference and seminar participants at the 2018 Conference on Financial Predictability and Big Data, 2018 International Symposium on Forecasting, 2019 American Finance Association Annual Meeting, 2019 China Finance Review International Conference, Baruch CUNY, Beijing University, Emory University, Georgia State University, Gothenburg University, Indiana University, London Business School, Lund University, Renmin University of China, Rutgers University, Saint Louis University, Shanghai University of Finance and Economics, Sichuan University, Southwestern University of Finance and Economics, State University of New York at Buffalo, Stockholm University, University of Cincinnati, University of Illinois at Urbana-Champaign, University of North Carolina at Charlotte, Washington University in St. Louis, and Zhejiang University, as well as Svetlana Bryzgalova, Victor DeMiguel, Jeremiah Green, John Hand, Jonathan Lewellen (AFA discussant), Sicong (Allen) Li, Jean Pauphilet, Minchul Shin, and Michael Weber, for valuable comments. We are also grateful to Andrew Chen, Jeremiah Green, and especially Fuwei Jiang for help in constructing the data.

Firm Characteristics and Expected Stock Returns

Abstract

We analyze the joint out-of-sample predictive ability of a comprehensive set of 299 firm characteristics for cross-sectional stock returns. We develop a cross-sectional out-of-sample R^2 statistic that provides an informative measure of the accuracy of cross-sectional return forecasts in terms of mean squared forecast error. To improve cross-sectional return forecasts based on a large number of firm characteristics, we propose an E-LASSO approach that implements shrinkage in a flexible manner. Our new approach produces significant cross-sectional out-of-sample R^2 gains on a consistent basis over time and provides the most accurate out-of-sample estimates of cross-sectional expected returns to date. The E-LASSO approach also generates substantial economic value in the context of long-short portfolios. Finally, we present evidence that more characteristics work better than fewer with respect to forecasting cross-sectional stock returns.

JEL classifications: C53, C55, C58, G12, G14, G17

Key words: Cross-sectional expected stock returns, Characteristic premia, Shrinkage, LASSO, Forecast combination, Forecast encompassing

1 Introduction

Firm characteristics play a fundamental role in finance, as they contain a wealth of potentially relevant information about a firm. Well-known factor models, such as the Fama and French (1993) three-factor model, extend the capital asset pricing model by including factors based on a few firm characteristics. Haugen and Baker (1996) appear to be the first to jointly analyze the out-of-sample predictive ability of a large number of firm characteristics (around 40) for cross-sectional stock returns. They construct long-short portfolios by sorting on the forecasted cross-sectional returns and find that the performance of the portfolios cannot be explained by conventional risk-based models. Lewellen (2015) investigates the joint predictive ability of 15 firm characteristics by constructing long-short portfolios based on the sorted forecasted returns, as well as by directly evaluating the return forecasts via a cross-sectional version of a Mincer and Zarnowitz (1969) time-series regression. Lewellen (2015) finds that the collective information in the characteristics is useful for predicting cross-sectional returns on an out-of-sample basis. Green et al. (2017) construct cross-sectional out-of-sample return forecasts using around 100 firm characteristics and find that long-short portfolios formed from the sorted forecasted returns perform well before 2003 but that value-weighted long-short portfolios fail to provide significant gains thereafter.

Using machine learning methods, Kozak et al. (2020), Freyberger et al. (2020), and Gu et al. (2020) examine the joint out-of-sample predictive ability of 50 to 94 firm characteristics. These studies also find that characteristics contain useful information for predicting returns, primarily in the context of long-short portfolios. Instead of analyzing joint predictive power, McLean and Pontiff (2016) construct long-short portfolios by sorting directly on nearly 100 individual firm characteristics to examine whether the relevance of individual characteristics is affected by the publication of academic research. Similarly, Harvey et al. (2016) study anomalies based on approximately 300 individual characteristics, while Hou et al. (2020)

study 452 anomalies based on nearly 240 characteristics.[1] In summary, there are now a number of studies that investigate the predictive ability of large sets of firm characteristics, predominantly via the construction of long-short portfolios.

In this paper, we analyze the joint out-of-sample predictive power of 299 firm characteristics, which presently constitutes the most comprehensive set of characteristics considered in studies of joint predictive ability. To simultaneously incorporate the information in all of the characteristics, we generate one-month-ahead return forecasts by estimating cross-sectional multiple regressions relating returns to the lagged characteristics. Conventional ordinary or weighted least squares (OLS or WLS, respectively) estimation of such a high-dimensional regression is susceptible to overfitting. We consider a variety of strategies to mitigate overfitting. As in Haugen and Baker (1996), Lewellen (2015), and Green et al. (2017), we smooth the cross-sectional OLS or WLS coefficient estimates over time when computing the forecasts (smoothed OLS or WLS forecast). In addition, we use the popular *least absolute shrinkage and selection operator* (LASSO, Tibshirani 1996) from the machine learning literature to estimate the high-dimensional cross-sectional regressions used to compute the forecasts (direct LASSO forecast). The LASSO is a penalized regression technique that helps to guard against overfitting by directly shrinking the coefficients. The LASSO's ℓ_1 penalty permits shrinkage to zero, so that it performs variable selection.

We also propose a *forecast combination* method for constructing cross-sectional return forecasts.[2] Instead of estimating a high-dimensional cross-sectional multiple regression, we pool individual forecasts generated by estimating cross-sectional univariate regressions based on each characteristic in turn. Because it exerts a strong shrinkage effect, forecast combination allows us to incorporate information from the entire set of characteristics in a manner that avoids overfitting. Incorporating insights from Diebold and Shin (2019) in the time-

[1]Hou et al. (2020) define 237 characteristics in their appendix. They arrive at 452 anomalies by constructing long-short portfolios with holding periods of one, six, and twelve months based on a number of the individual characteristics.

[2]See Timmermann (2006) for a survey of forecast combination in the time-series domain and Rapach et al. (2010) for a finance application.

series domain, we refine the cross-sectional combination forecast by using the LASSO to select the "best" univariate forecasts to include in the pooled forecast (C-LASSO forecast). Finally, we use the notion of *forecast encompassing* to blend the smoothed OLS or WLS and C-LASSO forecasts (E-LASSO forecast). Intuitively, the smoothed OLS or WLS forecast is likely to perform relatively well when characteristic premia are stable, while the C-LASSO forecast is better able to accommodate time-varying characteristic premia in a manner that avoids overfitting. Blending the smoothed OLS or WLS and C-LASSO forecasts represents a flexible shrinkage procedure that makes the forecast more robust to periods of stable and unstable characteristic premia. We use a data-driven method to optimally select the parameter used to blend the two forecasts. The C-LASSO and E-LASSO strategies constitute new approaches for generating cross-sectional out-of-sample return forecasts based on the information in a large number of firm characteristics.

We also develop new methods for evaluating cross-sectional out-of-sample return forecasts. The most popular approach for measuring forecast accuracy is the mean squared forecast error (MSFE) criterion. While the Campbell and Thompson (2008) out-of-sample R^2 statistic is based on MSFE and is used extensively in the literature on *time-series* return predictability, it is not adequate for directly assessing *cross-sectional* return forecasts. To fill this gap in the literature, we develop a cross-sectional version of the Campbell and Thompson (2008) out-of-sample R^2 statistic. Specifically, for each month, we compute an adjusted cross-sectional measure of MSFE for a forecast that incorporates the information in the firm characteristics, as well as the cross-sectional MSFE for a naïve benchmark forecast that ignores the information in the characteristics; the cross-sectional out-of-sample R^2 (R^2_{CSOS}) statistic for a given month is the proportional reduction in MSFE for the competing vis-á-vis the benchmark forecast. We then use the Fama and MacBeth (1973) technique and take the time-series average of the monthly R^2_{CSOS} statistics.[3]

[3]Gu et al. (2020) use the Campbell and Thompson (2008) out-of-sample R^2 statistics to assess time-series forecasts of individual stock returns based on firm characteristics generated in a panel predictive regression. In contrast, we analyze cross-sectional return forecasts based on firm characteristics in terms of cross-sectional MSFE via our new R^2_{CSOS} statistic.

In addition to the R^2_{CSOS} statistic, we develop a cross-sectional version of the Harvey et al. (1998) time-series encompassing test to compare the information content in competing cross-sectional out-of-sample return forecasts. For each month, we estimate a cross-sectional regression to estimate the parameter for an optimal convex combination of the competing forecasts. Again using the Fama and MacBeth (1973) technique, we then compute the time-series average of the monthly parameter estimates. By testing whether the time-series average of the parameter estimates is significantly different from zero or one, we can test whether each competing forecast contains relevant information for predicting cross-sectional returns relative to that already contained in the other forecast.

In our empirical analysis, we employ an out-of-sample period that covers more than four decades (1975:01 to 2018:12) and a variety of economic conditions. To examine the influence of small-cap stocks, we consider both value and equal weighting. In line with overfitting, we find that the conventional WLS or OLS forecast significantly underperforms the naïve benchmark forecast in terms of cross-sectional MSFE. Although the smoothed WLS or OLS forecast performs better than the conventional forecast in terms of MSFE, the smoothed forecast still significantly underperforms the naïve benchmark for value weighting, so that overfitting remains a concern. While the direct LASSO forecast performs better than the conventional forecast, it significantly underperforms the naïve benchmark for value and equal weighting and is thus inadequate for alleviating overfitting in our context.

The C-LASSO forecast appears helpful for mitigating overfitting, as it outperforms the naïve benchmark with respect to MSFE for value and equal weighting, although only significantly so for the latter case. By blending the smoothed WLS or OLS and C-LASSO forecasts, the E-LASSO forecast performs the best overall, delivering significant reductions in MSFE vis-á-vis the naïve benchmark for value and equal weighting. In general, we detect a decline in cross-sectional return predictability from the first to the second half of the out-of-sample period. Nevertheless, our new E-LASSO forecast provides significant improvements in forecast accuracy consistently over time and represents the most reliable out-of-sample method developed to date for estimating relative cross-sectional expected returns.

The E-LASSO forecast also provides impressive out-of-sample gains in terms of economic value. For each month, we sort stocks into quintiles according to their forecasted returns and construct a zero net-investment portfolio that goes long (short) the quintile with the highest (lowest) forecasted returns. The long-short portfolio based on the E-LASSO forecast generates an economically and statistically significant annualized Sharpe ratio of 1.61 (3.98) for value (equal) weighting. Similarly to the MSFE results, the gains generally decline from the first to the second half of the out-of-sample period, but they remain economically and statistically significant. The long-short portfolio based on the E-LASSO forecast also delivers substantial alpha in the context of leading factor models, including those of Carhart (1997), Fama and French (2015), Hou et al. (2015), and Stambaugh and Yuan (2017).

Finally, we explore the number and types of firm characteristics that are relevant for determining cross-sectional expected returns. In the spirit of Harvey et al. (2016), McLean and Pontiff (2016), Freyberger et al. (2020), and Hou et al. (2020), we categorize the 299 characteristics into six subgroups based on economic concepts (momentum, value vs. growth, investment, profitability, intangibles, and trading frictions). We find that forecasts based on the entire set of characteristics perform substantially better than those based on each of the subgroups, so that a variety of different types of characteristics appear relevant for cross-sectional returns. Next, we form collections of characteristics from across the categories, ranging in size from 25 to 299 characteristics. We find that more is preferred to less when it comes to using firm characteristics to forecast cross-sectional returns. In addition, to identify the relevant characteristics over time, we use the LASSO to estimate monthly cross-sectional versions of a Granger and Ramanathan (1984) multiple regression that relates actual returns to return forecasts based on the individual characteristics. The LASSO estimation results indicate that nearly all of the characteristics matter a substantial portion of the time and that approximately 50 to 70 characteristics matter on average each month.

The remainder of the paper is organized as follows. Section 2 describes the construction of the cross-sectional out-of-sample return forecasts. Section 3 discusses the evaluation of cross-sectional return forecasts. Section 4 reports our main empirical results, while Section 5 presents evidence on the number and types of relevant firm characteristics. Section 6 concludes.

2 Constructing Cross-Sectional Return Forecasts

This section discusses methodologies for generating cross-sectional out-of-sample return forecasts on the basis of $j = 1, \ldots, J$ firm characteristics. The first is the standard OLS forecast. The second is the smoothed OLS forecast, which improves the standard forecast by smoothing the OLS coefficient estimates over time. The third is a straightforward application of the LASSO. Motivated by Diebold and Shin (2019) and forecast encompassing, we provide two new methods, the C-LASSO and E-LASSO forecasts, respectively. Section A1 of the Internet Appendix provides detailed step-by-step instructions for implementing the methods. All of the forecasts are based on data available at the time of forecast formation, so that they do not entail "look-ahead" bias.

2.1 OLS Forecast

The natural starting point for generating month-$(t+1)$ cross-sectional return forecasts based on information available through month t is a cross-sectional multiple regression that includes all J characteristics as explanatory variables:

$$r_{i,t} = a_t + \sum_{j=1}^{J} b_{j,t} z_{i,j,t-1} + \varepsilon_{i,t} \quad \text{for } i = 1, \ldots, n_t, \tag{2.1}$$

where $r_{i,t}$ is the month-t return on stock i, n_t is the number of available firm observations for month t, and $z_{i,j,t}$ is the month-t value for the jth characteristic for stock i. A month-$(t+1)$

cross-sectional return forecast based on Equation (2.1) is given by

$$\hat{r}^{\text{OLS}}_{i,t+1|t} = \hat{a}_t + \sum_{j=1}^{J} \hat{b}_{j,t} z_{i,j,t} \quad \text{for } i = 1, \ldots, n_{t+1}, \tag{2.2}$$

where $\hat{a}_t$ and $\hat{b}_{j,t}$ are the OLS estimates of a_t and $b_{j,t}$, respectively, for $j = 1, \ldots, J$ in Equation (2.1). Note that, based on data availability, J can change over time in our applications; for notational simplicity, we omit a time subscript for J. When generating cross-sectional return forecasts, we use all of the characteristics available in a given month.[4]

The main drawback to the OLS forecast in Equation (2.2) is that it is susceptible to overfitting. This concern is especially relevant in our context, as we consider a large number of predictors, and returns inherently contain a large unpredictable component (i.e., the data are noisy). Next, we consider strategies for mitigating overfitting when forecasting cross-sectional stock returns.

2.2 Smoothed OLS Forecast

The smoothed OLS (SOLS) forecast modifies Equation (2.2) by smoothing the OLS coefficient estimates over time (e.g., Haugen and Baker 1996; Lewellen 2015; Green et al. 2017):

$$\hat{r}^{\text{SOLS}}_{i,t+1|t} = \tilde{a}_t + \sum_{j=1}^{J} \tilde{b}_{j,t} z_{i,j,t} \quad \text{for } i = 1, \ldots, n_{t+1}, \tag{2.3}$$

where

$$\tilde{a}_t = \frac{1}{M} \sum_{m=0}^{M-1} \hat{a}_{t-m}, \tag{2.4}$$

[4] Freyberger et al. (2020) and Gu et al. (2020) forecast individual stock returns via panel predictive regressions. A panel specification requires that observations are available for all of the characteristics for each month, which substantially limits the number of characteristics that can be considered; in contrast, our forecasting methods use all available characteristics in a given month, so that we can incorporate information from a larger number of characteristics over time.

$$\tilde{b}_{j,t} = \frac{1}{M} \sum_{m=0}^{M-1} \hat{b}_{j,t-m} \quad \text{for } j = 1, \ldots, J, \tag{2.5}$$

and M is the length of the smoothing window. Equation (2.3) reduces to Equation (2.2) for $M = 1$, while using the largest value of M available at each point in time constitutes an expanding window for smoothing the coefficient estimates over time. In our applications, an expanding window generally performs the best.

Smoothing the coefficient estimates over time works to stabilize the coefficient estimates and thus the cross-sectional return forecasts, thereby helping to guard against overfitting. The SOLS forecast is likely to be effective when the slope coefficients in Equation (2.1) are relatively stable over time, which occurs when characteristic premia are stable. By smoothing, we reduce the influence of noise in the data to more precisely estimate the characteristic premia.

2.3 Direct LASSO Forecast

The LASSO (Tibshirani 1996) is a popular machine learning tool for performing both shrinkage and variable selection in regressions with a large number of explanatory variables. To implement LASSO estimation in our context, instead of estimating Equation (2.1) via OLS, we use the following objective function:

$$\underset{a_t, b_{1,t}, \ldots, b_{J,t} \in \mathbb{R}}{\arg\min} \left\{ \frac{1}{2n_t} \sum_{i=1}^{n_t} \left[r_{i,t} - \left(a_t + \sum_{j=1}^{J} b_{j,t} z_{i,j,t-1} \right) \right]^2 + \lambda_t \sum_{j=1}^{J} |b_{j,t}| \right\}, \tag{2.6}$$

where $\lambda_t \geq 0$ is a regularization parameter that governs the degree of shrinkage. When $\lambda_t = 0$, Equation (2.6) reduces to the familiar OLS objective function. The presence of λ_t in Equation (2.6) shrinks the slope estimates, thereby helping to guard against overfitting. The ℓ_1 penalty term in Equation (2.6) allows for shrinkage to zero (for sufficiently large λ_t), so that it performs variable selection.

We need to choose the regularization parameter, λ_t, in Equation (2.6). Although K-fold cross validation is a popular method for tuning the regularization parameter, specifying the number of folds and their definitions is largely arbitrary, and the results can be sensitive to these choices. Instead, based on results in Flynn et al. (2013) and Taddy (2017), we select λ_t using the Hurvich and Tsai (1989) corrected version of the Akaike information criterion (AIC, Akaike 1973).[5]

The LASSO forecast is given by

$$\hat{r}_{i,t+1|t}^{\text{LASSO}} = \hat{a}_t^{\text{LASSO}} + \sum_{j=1}^{J} \hat{b}_{j,t}^{\text{LASSO}} z_{i,j,t}, \quad \text{for } i = 1, \ldots, n_{t+1}, \tag{2.7}$$

where $\hat{a}_t^{\text{LASSO}}$ and $\hat{b}_{j,t}^{\text{LASSO}}$ are the LASSO estimates of a_t and $b_{j,t}$, respectively, for $j = 1, \ldots, J$ in Equation (2.1). As reported in Tables A2 and A3 of the Internet Appendix, we obtain similar results for forecasts based on well-known variants of the LASSO, including the elastic net (Zou and Hastie 2005), adaptive LASSO (Zou 2006), and adaptive elastic net (Zou and Zhang 2009; Ghosh 2011).

2.4 Combination LASSO Forecast

Instead of generating forecasts based on estimation of the cross-sectional multiple regression in Equation (2.1), the forecast combination approach begins by estimating a series of cross-sectional univariate regressions, each of which relates returns to an individual characteristic:

$$r_{i,t} = c_{j,t} + d_{j,t} z_{i,j,t-1} + \varepsilon_{i,t} \quad \text{for } 1, \ldots, n_t;\ j = 1, \ldots, J. \tag{2.8}$$

For each characteristic, we compute a month-$(t+1)$ cross-sectional univariate regression forecast:

$$\hat{r}_{i,t+1|t}^{(j)} = \hat{c}_{j,t} + \hat{d}_{j,t} z_{i,j,t} \quad \text{for } 1, \ldots, n_{t+1};\ j = 1, \ldots, J, \tag{2.9}$$

[5]Selecting λ_t via the corrected AIC outperforms conventional five-fold cross validation for our applications in Section 4. The complete cross-validation results are reported in Table A1 of the Internet Appendix.

where $\hat{c}_{j,t}$ and $\hat{d}_{j,t}$ are the OLS estimates of $c_{j,t}$ and $d_{j,t}$, respectively, in Equation (2.8). We then pool the return forecasts based on the individual characteristics to form a combination forecast of $r_{i,t+1}$. A simple combination forecast takes the arithmetic mean of the individual forecasts in Equation (2.9):

$$\hat{r}^{\text{C-Mean}}_{i,t+1|t} = \frac{1}{J}\sum_{j=1}^{J}\hat{r}^{(j)}_{i,t+1|t} \quad \text{for } i = 1, \ldots, n_{t+1}. \tag{2.10}$$

As shown by Rapach et al. (2010) in a time-series context, the simple combination mean forecast in Equation (2.10) exerts a strong shrinkage effect. To see the shrinkage property of the forecast, first note that

$$\hat{a}_t = \bar{r}_t - \sum_{j=1}^{J}\hat{b}_{j,t}\bar{z}_{j,t-1}, \tag{2.11}$$

$$\hat{c}_{j,t} = \bar{r}_t - \hat{d}_{j,t}\bar{z}_{j,t-1} \quad \text{for } j = 1, \ldots, J, \tag{2.12}$$

in Equations (2.2) and (2.9), respectively. Using Equations (2.11) and (2.12), we can rewrite the forecasts in Equations (2.2) and (2.9) as

$$\hat{r}^{\text{OLS}}_{i,t|t+1} = \bar{r}_t + \sum_{j=1}^{J}\hat{b}_{j,t}(z_{i,j,t} - \bar{z}_{j,t-1}) \quad \text{for } i = 1, \ldots, n_{t+1}, \tag{2.13}$$

$$\hat{r}^{(j)}_{i,t|t+1} = \bar{r}_t + \hat{d}_{j,t}(z_{i,j,t} - \bar{z}_{j,t-1}) \quad \text{for } i = 1, \ldots, n_{t+1};\ j = 1, \ldots, J, \tag{2.14}$$

respectively, where $\bar{z}_{j,t}$ is the cross-sectional mean for $z_{i,j,t}$ for $j = 1, \ldots, J$. Finally, we use Equation (2.14) to rewrite the combination mean forecast in Equation (2.10) as

$$\hat{r}^{\text{C-Mean}}_{i,t+1|t} = \bar{r}_t + \sum_{j=1}^{J}\frac{1}{J}\hat{d}_{j,t}(z_{i,j,t} - \bar{z}_{j,t-1}) \quad \text{for } i = 1, \ldots, n_{t+1}. \tag{2.15}$$

Comparing Equations (2.13) and (2.15), relative to the forecast based on OLS estimation of the high-dimensional multiple regression in Equation (2.1), the combination mean forecast makes two adjustments: (i) it replaces the OLS multiple regression slope coefficient estimates with their univariate counterparts, thereby improving estimation efficiency (a manifestation of the bias-efficiency tradeoff); (ii) it shrinks the magnitude of each slope coefficient by the factor $1/J$, which has the effect of shrinking the cross-sectional forecast toward the cross-sectional mean.

Incorporating time-series insights from Diebold and Shin (2019), we use the LASSO to refine the cross-sectional combination forecast in Equation (2.10). Consider the following cross-sectional version of a Granger and Ramanathan (1984) multiple regression for month t involving the univariate regression forecasts in Equation (2.9):

$$r_{i,t} = \xi_t + \sum_{j=1}^{J} \phi_{j,t} \hat{r}_{i,t|t-1}^{(j)} + \varepsilon_{i,t} \quad \text{for } i = 1, \ldots, n_t. \tag{2.16}$$

We estimate Equation (2.16) using the LASSO objective function:

$$\underset{\xi_t \in \mathbb{R};\phi_{1,t},\ldots,\phi_{J,t} \in \mathbb{R}_{\geq 0}}{\arg\min} \left\{ \frac{1}{2n_t} \sum_{i=1}^{n_t} \left[r_{i,t} - \left(\xi_t + \sum_{j=1}^{J} \phi_{j,t} \hat{r}_{i,t|t-1}^{(j)} \right) \right]^2 + \lambda_t \sum_{j=1}^{J} |\phi_{j,t}| \right\}, \tag{2.17}$$

where we again select λ_t in Equation (2.17) using the corrected AIC. Let $\hat{\mathcal{J}}_t \subseteq \{1, \ldots, J\}$ denote the index set of cross-sectional univariate regression forecasts selected by the LASSO in Equation (2.17). The combination LASSO (C-LASSO) forecast is given by

$$\hat{r}_{i,t+1|t}^{\text{C-LASSO}} = \frac{1}{\left|\hat{\mathcal{J}}_t\right|} \sum_{j \in \hat{\mathcal{J}}_t} \hat{r}_{i,t+1|t}^{(j)} = \bar{r}_t + \sum_{j \in \hat{\mathcal{J}}_t} \frac{1}{\left|\hat{\mathcal{J}}_t\right|} \hat{d}_{j,t}(z_{i,j} - \bar{z}_{j,t-1}) \quad \text{for } i = 1, \ldots, n_{t+1}, \tag{2.18}$$

where $|\hat{\mathcal{J}}_t|$ is the cardinality of $\hat{\mathcal{J}}_t$. The C-LASSO forecast retains the shrinkage properties of the combination mean forecast in Equation (2.10) and refines the forecast by using the LASSO to select the most relevant univariate forecasts to include in the pooled forecast.[6] The

[6] Analogously to Equation (2.3), we can smooth the coefficient estimates over time when computing the cross-sectional univariate regression forecasts in Equation (2.9) used to form the combination forecast in Equation (2.18). However, because the combination approach typically employs substantial shrinkage when J is large—as evinced by Equations (2.15) and (2.18)—smoothing the univariate coefficient estimates over

C-LASSO forecast should prove relatively beneficial when characteristic premia are changing substantively over time. Because the slope coefficients for the fitted forecasting models in Equation (2.9) are not smoothed over time, they readily accommodate time-varying risk premia. The strong shrinkage property of the forecast combination approach accommodates time-varying characteristic premia, while still guarding against overfitting.

2.5 Encompassing LASSO Forecast

Different forecasting strategies have relative strengths and weaknesses, so that it can be beneficial to blend forecasting strategies. We use the notion of forecast encompassing, as discussed in greater detail below in Section 3.3, to optimally blend forecasting strategies.

Specifically, suppose that we want to form a month-$(t+1)$ cross-sectional return forecast that takes the form of a convex combination of the SOLS and C-LASSO forecasts in Equations (2.3) and (2.18), respectively, using data available through month t. As described in Section 3.3, we can conveniently estimate the parameter for optimally blending the forecasts via the cross-sectional univariate regression in Equation (3.15). The encompassing LASSO (E-LASSO) forecast is given by

$$\hat{r}_{i,t+1|t}^{\text{E-LASSO}} = \left(1 - \tilde{\theta}_t\right) \hat{r}_{i,t+1|t}^{\text{SOLS}} + \tilde{\theta}_t \hat{r}_{i,t+1|t}^{\text{C-LASSO}} \quad \text{for } i = 1, \ldots, n_{t+1}. \tag{2.19}$$

where

$$\tilde{\theta}_t = \frac{1}{M} \sum_{m=0}^{M-1} \hat{\theta}_{t-m} \tag{2.20}$$

and $\hat{\theta}_t$ is the OLS estimate of θ_t in Equation (3.15). We set $M = 36$ in our applications; results are similar for larger values of M, as well as an expanding smoothing window.

Intuitively, by blending the SOLS and C-LASSO forecasts, we diversify the forecast to make it more robust to alternating periods of stable and unstable characteristic premia. As

time when forming the combination forecast tends to make the cross-sectional return forecast too conservative (i.e., it has a substantive downward bias).

discussed in Section 2.2 (Section 2.4), the SOLS (C-LASSO) forecast is likely to perform relatively well when characteristic premia are stable (unstable). The E-LASSO forecast provides a data-driven strategy for optimally blending the two forecasts over time; as such, it can be interpreted as a flexible shrinkage forecast.

2.6 Value-Weighted Forecasts

For value weighting, we construct forecasts by using weighted least squares (WLS) or the weighted LASSO to estimate the regressions in Equations (2.1), (2.8), (2.16) and (3.15). Section A1 of the Internet Appendix provides details for constructing the value-weighted forecasts.

3 Evaluating Cross-Sectional Return Forecasts

In this section, we develop a cross-sectional out-of-sample R^2 statistic, R^2_{CSOS}, which is analogous to the Campbell and Thompson (2008) time-series out-of-sample R^2 statistic. The R^2_{CSOS} statistic provides an informative metric for comparing competing and benchmark forecasts in terms of cross-sectional MSFE. We also discuss the Lewellen (2015) cross-sectional version of a Mincer and Zarnowitz (1969) times-series regression, which is convenient for measuring the bias in the cross-sectional forecasts. Finally, we propose an encompassing test for comparing competing cross-sectional return forecasts.

3.1 Cross-Sectional MSFE

We use $\hat{r}_{i,t|t-1}$ to generically denote a forecast of $r_{i,t}$ based on information available through month $t-1$. We define the month-t cross-sectional MSFE in terms of deviations from cross-sectional means:

$$\text{MSFE}_{\hat{r},t} = \frac{1}{n_t} \sum_{i=1}^{n_t} \left[(r_{i,t} - \bar{r}_t) - \left(\hat{r}_{i,t|t-1} - \bar{\hat{r}}_{t|t-1} \right) \right]^2 \quad \text{for } t = 1, \ldots, T, \tag{3.1}$$

where T is the number of observations for the out-of-sample period and $\bar{r}_t$ and $\bar{\hat{r}}_{t|t-1}$ are the cross-sectional means for $r_{i,t}$ and $\hat{r}_{i,t|t-1}$, respectively.

To see why the MSFE in Equation (3.1) is a relevant metric for assessing cross-sectional return forecasts, consider the traditional MSFE metric (i.e., not specified in terms of deviations):

$$\text{MSFE}^{\dagger}_{\hat{r},t} = \frac{1}{n_t}\sum_{i=1}^{n_t}\left(r_{i,t} - \hat{r}_{i,t|t-1}\right)^2 \quad \text{for } t = 1, \ldots, T. \tag{3.2}$$

If the forecast is perfect (i.e., $r_{i,t} = \hat{r}_{i,t|t-1}$ for $i = 1, \ldots, n_t$), then it is clear that both MSFE measures in Equations (3.1) and (3.2) equal zero. However, if $\hat{r}_{i,t|t-1} = r_{i,t} + x$ for $i = 1, \ldots, n_t$, then $\text{MSFE}^{\dagger}_{\hat{r},t} = x^2$ in Equation (3.2); in contrast, according to our new measure, $\text{MSFE}_{\hat{r},t} = 0$ in Equation (3.1). The latter case is the relevant one for assessing cross-sectional return forecasts, as such forecasts are concerned with *relative* expected returns across stocks (or the ranking of stocks based on relative expected returns); in other words, we are interested in predicting cross-sectional differences in expected returns. Hence, both the realized and forecasted returns should be demeaned when computing the MSFE, which is what Equation (3.1) does, so that it constitutes an appropriate measure of forecast accuracy in our cross-sectional context.

To motivate the R^2_{CSOS} statistic, consider a naïve benchmark forecast. In a time-series context, the naïve benchmark is the historical average (i.e., the time-series mean of the available return observations), so that it ignores the information in the predictor variables (Goyal and Welch 2008). Analogously, in our cross-sectional context, the naïve benchmark forecast is the cross-sectional mean:

$$\hat{r}^{\text{Naïve}}_{i,t|t-1} = \bar{r}_{t-1} \quad \text{for } i = 1, \ldots, n_t, \tag{3.3}$$

which ignores the information in firm characteristics. Because the naïve benchmark predicts no cross-sectional variation in returns ($\hat{r}^{\text{Naïve}}_{i,t|t-1} - \bar{\hat{r}}^{\text{Naïve}}_{t|t-1} = 0$ for $i = 1, \ldots, n_t$), its cross-sectional

MSFE in Equation (3.2) is simply the cross-sectional return variance:

$$\hat{\sigma}_{r,t}^2 = \frac{1}{n_t} \sum_{i=1}^{n_t} (r_{i,t} - \bar{r}_t)^2 \quad \text{for } t = 1, \ldots, T. \tag{3.4}$$

The popular Campbell and Thompson (2008) R_{OS}^2 statistic measures the proportional reduction in the time-series MSFE for a competing forecast relative to the naïve historical average benchmark. In our cross-sectional setting, we are interested in the proportional reduction in the cross-sectional MSFE for a competing forecast based on firm characteristics relative to the naïve benchmark in Equation (3.3) that ignores the information in the firm characteristics. Following the intuition of the R_{OS}^2 statistic, we use Equations (3.1) and (3.4) to define the month-t R_{CSOS}^2 statistic:

$$R_{\mathrm{CSOS},t}^2 = 1 - \frac{\sum_{i=1}^{n_t} \left[(r_{i,t} - \bar{r}_t) - \left(\hat{r}_{i,t|t-1} - \bar{\hat{r}}_{t|t-1} \right) \right]^2}{\sum_{i=1}^{n_t} (r_{i,t} - \bar{r}_t)^2} \quad \text{for } t = 1, \ldots, T. \tag{3.5}$$

We then use the Fama and MacBeth (1973) procedure and take the time-series average of Equation (3.5):

$$R_{\mathrm{CSOS}}^2 = \frac{1}{T} \sum_{t=1}^{T} R_{\mathrm{CSOS},t}^2, \tag{3.6}$$

where we compute the standard error for R_{CSOS}^2 using $\{R_{\mathrm{CSOS},t}^2\}_{t=1}^{T}$. Finally, we calculate a t-statistic for testing H_0: $R_{\mathrm{CSOS}}^2 = 0$ against H_A: $R_{\mathrm{CSOS}}^2 \neq 0$.[7]

3.2 Cross-Sectional Bias

Lewellen (2015) relates actual to forecasted cross-sectional returns in the following regression:

$$r_{i,t} = \gamma_t + \delta_t \hat{r}_{i,t|t-1} + \varepsilon_{i,t} \quad \text{for } 1, \ldots, n_t;\ t = 1, \ldots, T, \tag{3.7}$$

[7]We base the t-statistic on a heteroskedasticity and autocorrelation robust standard error (Newey and West 1987).

which can be interpreted as a cross-sectional version of a Mincer and Zarnowitz (1969) time-series regression. The ordinary least squares (OLS) estimate of δ_t in Equation (3.7) is given by

$$\hat{\delta}_t = \frac{\sum_{i=1}^{n_t}(r_{i,t} - \bar{r}_t)\big(\hat{r}_{i,t|t-1} - \bar{\hat{r}}_{t|t-1}\big)}{\sum_{i=1}^{n_t}\big(\hat{r}_{i,t|t-1} - \bar{\hat{r}}_{t|t-1}\big)^2} \quad \text{for } t = 1, \ldots, T. \tag{3.8}$$

Again using the Fama and MacBeth (1973) procedure, we take the time-series average of the $\hat{\delta}_t$ estimates in Equation (3.8):

$$\hat{\delta} = \frac{1}{T}\sum_{t=1}^{T}\hat{\delta}_t, \tag{3.9}$$

where the standard error for $\hat{\delta}$ is computed using $\{\hat{\delta}_t\}_{t=1}^{T}$. If $\hat{\delta} > 0$, then the cross-sectional return forecasts are positively related to actual returns, a reasonable necessary condition for the cross-sectional forecasts to be informative.

Moreover, as pointed out by Lewellen (2015), the estimate of the slope coefficient in Equation (3.7) measures the bias in the month-t cross-sectional forecasts: if the slope coefficient equals one, then the forecasts are unbiased, since a unit increase in the forecasted return corresponds to a unit increase in the realized return on average; if the slope coefficient is less (greater) than one, then the forecasts overstate (understate) the cross-sectional difference in expected returns, as a unit increase in the forecasted return is associated with a less-than-unit (greater-than-unit) increase in the actual return on average. Because an intercept term is included in Equation (3.7), the Frisch-Waugh-Lovell theorem implies that the forecast bias is measured with respect to deviations from cross-sectional means, in line with our interest in relative expected returns in Equation (3.1). To test for cross-sectional forecast bias, we compute a t-statistic for testing H_0: $\delta = 1$ against H_A: $\delta \neq 1$.

We also compute the time-series average of the monthly R^2 statistics for Equation (3.7):

$$R^2_{\text{CSMZ}} = \frac{1}{T}\sum_{t=1}^{T} R^2_{\text{CSMZ},t}, \tag{3.10}$$

where $R^2_{\text{CSMZ},t}$ is the R^2 statistic for Equation (3.7) and we compute the standard error for R^2_{CSMZ} using $\{R^2_{\text{CSMZ},t}\}_{t=1}^{T}$.

Both $\hat{\delta}_t$ and the $R^2_{\text{CSMZ},t}$ statistic in Equation (3.7) play roles in determining the cross-sectional MSFE in Equation (3.1), as indicated by the following relation:[8]

$$\text{MSFE}_{\hat{r},t} = \left(\hat{\delta}_t - 1\right)^2 \hat{\sigma}^2_{\hat{r},t} + \left(1 - R^2_{\text{CSMZ},t}\right)\hat{\sigma}^2_{r,t} \quad \text{for } t = 1, \ldots, T, \tag{3.11}$$

where $\hat{\delta} - 1$ is the forecast bias and $\hat{\sigma}^2_{\hat{r},t}$ is the cross-sectional forecast variance:

$$\hat{\sigma}^2_{\hat{r},t} = \frac{1}{n_t} \sum_{i=1}^{n_t} \left(\hat{r}_{i,t} - \bar{\hat{r}}_{t|t-1}\right)^2 \quad \text{for } t = 1, \ldots, T. \tag{3.12}$$

It is instructive to consider the following two special cases for Equation (3.11):

- *Naïve benchmark forecast.* The naïve benchmark forecast ignores the information in firm characteristics and predicts no cross-sectional variation in returns. In this case, $\hat{\sigma}^2_{\hat{r},t} = R^2_{\text{CSMZ},t} = 0$, so that $\text{MSFE}_{\hat{r},t} = \hat{\sigma}^2_{r,t}$ in Equation (3.11), corresponding to Equation (3.4).

- *Perfect forecast.* A perfect cross-sectional return forecast produces $\hat{\delta}_t = R^2_{\text{CSMZ},t} = 1$ in Equation (3.7), so that $\text{MSFE}_{\hat{r},t} = 0$ in Equation (3.11).

3.3 Forecast Encompassing

Incorporating insights from Harvey et al. (1998) in the time-series domain, we also develop a forecast encompassing test for comparing the information content of two competing cross-sectional return forecasts. Suppose that we have two competing cross-sectional return forecasts for month t, $\hat{r}^{\text{A}}_{i,t|t-1}$ and $\hat{r}^{\text{B}}_{i,t|t-1}$, and consider the following composite forecast:

$$\hat{r}^{*}_{i,t|t-1} = (1 - \zeta_t)\hat{r}^{\text{A}}_{i,t|t-1} + \zeta_t \hat{r}^{\text{B}}_{i,t|t-1} \quad \text{for } i = 1, \ldots, n_t;\ t = 1, \ldots, T, \tag{3.13}$$

[8] Section A2 of the Internet Appendix provides the derivation of Equation (3.11).

where $0 \leq \zeta_t \leq 1$. Its cross-sectional MSFE is given by

$$\mathrm{MSFE}_t^* = \frac{1}{n_t} \sum_{i=1}^{n_t} \left[(r_{i,t} - \bar{r}_t) - \left(\hat{r}_{i,t|t-1}^* - \bar{\hat{r}}_{t|t-1}^* \right) \right]^2 \text{ for } t = 1, \ldots, T, \tag{3.14}$$

where $\bar{\hat{r}}_{t|t-1}^*$ is the cross-sectional mean of $\hat{r}_{i,t|t-1}^*$.

As shown in Section A2 of the Internet Appendix, the value of ζ_t in Equation (3.13) that minimizes the month-t cross-sectional MSFE in Equation (3.14) is identical to the OLS estimate of θ_t, $\hat{\theta}_t$, in the following cross-sectional univariate regression:

$$\hat{e}_{i,t|t-1}^{\mathrm{A}} = \eta_t + \theta_t \left(\hat{e}_{i,t|t-1}^{\mathrm{A}} - \hat{e}_{i,t|t-1}^{\mathrm{B}} \right) + \varepsilon_{i,t} \text{ for } i = 1, \ldots, n_t;\ t = 1, \ldots, T, \tag{3.15}$$

where

$$\hat{e}_{i,t|t-1}^{k} = r_{i,t} - \hat{r}_{i,t|t-1}^{k} \text{ for } k = \mathrm{A}, \mathrm{B}. \tag{3.16}$$

Once again using the Fama and MacBeth (1973) procedure, we take the time-series average of the monthly slope coefficient estimates in Equation (3.15):

$$\hat{\theta} = \frac{1}{T} \sum_{t=1}^{T} \hat{\theta}_t, \tag{3.17}$$

where we compute the standard error for $\hat{\theta}$ using $\{\hat{\theta}_t\}_{t=1}^T$.

Equation (3.17) allows us to compare the information content of two competing cross-sectional return forecasts. When $\theta = 0$, A encompasses B on average over the forecast evaluation period, meaning that B does not contain useful information for forecasting cross-sectional returns (in terms of cross-sectional MSFE) beyond the information already contained in A; alternatively, when $\theta > 0$, A does not encompass B, so that B does provide useful information beyond that already contained in A. Analogously, if $\theta = 1$, then B encompasses A on average; if $\theta < 1$, then B does not encompass A.

3.4 Economic Value

To measure the economic value of the cross-sectional out-of-sample return forecasts, we construct long-short portfolios by sorting stocks according to their forecasted returns (e.g., Haugen and Baker 1996; Lewellen 2015; Green et al. 2017; Freyberger et al. 2020). Specifically, at the end of each month, we sort stocks into equal-weighted quintiles based on their forecasted returns for the subsequent month. We then construct a zero-investment portfolio that goes long (short) the last (first) quintile. As in Green et al. (2017), the quintile breakpoints are computed using NYSE stocks (so that the long and short legs do not necessarily contain the same number of stocks).

3.5 Value-Weighted Statistics

We also analyze cross-sectional forecasts based on value weighting. To this end, we define the following weighted version of the cross-sectional MSFE in Equation (3.1):

$$\text{wMSFE}_{\hat{r},t} = \frac{1}{n_t} \sum_{i=1}^{n_t} w_{i,t} \left[(r_{i,t} - \bar{r}_t) - \left(\hat{r}_{i,t|t-1} - \bar{\hat{r}}_{t|t-1} \right) \right]^2 \quad \text{for } t = 1, \ldots, T, \tag{3.18}$$

where $w_{i,t} \geq 0$ for $i = 1, \ldots, n_t$; $\sum_{i=1}^{n_t} w_{i,t} = n_t$; and $\bar{r}_t$ and $\bar{\hat{r}}_{t|t-1}$ are now the weighted cross-sectional means of $r_{i,t}$ and $\hat{r}_{i,t|t-1}$, respectively. The month-t weights are proportional to market capitalization at the end of month $t-1$. Section A2 of the Internet Appendix provides versions of the statistics in Sections 3.1 to 3.3 for the value-weighted case. For value weighting, the quintiles used to form the long-short portfolio in Section 3.4 are value weighted.

4 Empirical Results

4.1 Data

We consider common stocks on the NYSE, AMEX, and NASDAQ with a market value on CRSP at the end of the previous month and a non-missing value for common equity in the firm's annual financial statement. Based on data available from CRSP, Compustat, I/B/E/S, and researchers' websites, we analyze the out-of-sample predictive ability of 299 firm characteristics. The firm characteristics are culled from characteristics appearing in Haugen and Baker (1996), Lewellen (2015), Harvey et al. (2016), McLean and Pontiff (2016), Green et al. (2017), Freyberger et al. (2020), and Hou et al. (2020), among others. Presently, we analyze the largest number of characteristics in a study of joint predictive ability.[9] Similarly to Harvey et al. (2016), McLean and Pontiff (2016), Freyberger et al. (2020), and Hou et al. (2020), we arrange the characteristics into subgroups based on economic concepts: momentum, value vs. growth, investment, profitability, intangibles, and trading frictions. The characteristics and subgroups are listed in Table 1. Section A3 of the Internet Appendix provides detailed definitions and relevant studies for the characteristics.

Our sample period covers 1965:01 to 2018:12. We use the first ten years as the initial in-sample period for estimating the forecasting models (including smoothing the slope coefficient estimates), so that we compute and evaluate cross-sectional out-of-sample return forecasts for 1975:01 to 2018:12. The out-of-sample period spans more than four decades. According to business-cycle turning points dated by National Bureau of Economic Research, the out-of-sample period covers six recessions, including the Great Recession (2007:12 to 2009:06). We have a total of 2,343,904 firm-level return observations over the out-of-sample period.

[9]For example, Haugen and Baker (1996), Lewellen (2015), Green et al. (2017), and Freyberger et al. (2020) consider 41, 15, 94, and 62 characteristics, respectively. Note that Harvey et al. (2016), McLean and Pontiff (2016), and Hou et al. (2020) use characteristics to form long-short anomaly portfolios, so that they do not directly analyze the predictive ability of firm characteristics. As discussed in Section 1, the latter studies also examine characteristics individually, while we analyze the collective information in the 299 characteristics that we consider.

For each month, we winsorize each characteristic at the first and 99th cross-sectional percentiles. Some characteristics have missing values for some months. For a characteristic to be included when computing the cross-sectional forecast for a given month, we require at least 20% of the characteristic observations for that month to be non-missing. When a characteristic is included, we fill in any missing values for the month with the cross-sectional mean (median for indicator variables) of the available observations. Without loss of generality, before estimating the forecasting models, we standardize each characteristic to have a monthly cross-sectional mean (standard deviation) of zero (one). The number of characteristics included in a month ranges from 192 to 298; all of the 299 characteristics are used in multiple months.

4.2 Forecast Accuracy

To assess the accuracy of the different forecasts, Panel A of Table 2 reports the time-series average of the monthly weighted R^2_{CSOS} statistics, as well as the averages of the monthly weighted slope coefficients and R^2 statistics for the cross-sectional regression in Equation (3.7). The table also reports the time-series average of the monthly weighted cross-sectional forecast standard deviations.

Panel A provides results for value weighting. As expected, the conventional WLS forecast appears plagued by severe overfitting. The R^2_{CSOS} statistic in the third column indicates that the MSFE for the WLS forecast is nearly 46% higher than that for the naïve benchmark forecast. The $\hat{\delta}$ estimate of 0.05 in the fourth column is significantly greater than zero; however, its small magnitude represents a substantial upward bias in the forecast, and there is strong evidence against the null hypothesis that $\delta = 1$ in the sixth column.

The volatility for the WLS forecast is 5.24% in the second column, which compares to an actual return volatility of 7.75%. Because cross-sectional returns inherently contain a large unpredictable component, the relatively high volatility of the WLS forecast vis-á-vis the actual return volatility indicates that the forecast misinterprets noise in the data for a predictive signal, a manifestation of overfitting. In summary, the WLS forecast does not provide a reliable out-of-sample measure of relative cross-sectional expected returns.

The results for the SWLS forecast in Panel A of Table 2 indicate that smoothing the coefficient estimates over time helps to mitigate overfitting. The forecast volatility for the SWLS forecast falls to 1.02%. The R^2_{CSOS} statistic for the forecast is -0.40%, so that it performs much better than the conventional WLS forecast in terms of cross-sectional MSFE. The $\hat{\delta}$ estimate for the SWLS forecast is 0.47, which again represents an improvement over the WLS forecast in terms of the bias. Nevertheless, the R^2_{CSOS} statistic for the SWLS forecast is still negative and significant, so that the forecast significantly underperforms the naïve benchmark with respect to cross-sectional MSFE. According to the t-statistic in the sixth column, the SWLS forecast also has a significant upward bias, in line with overfitting.

The LASSO forecast seeks to alleviate overfitting by directly shrinking the coefficients for the fitted forecasting model. The results for the LASSO forecast in Panel A show that the LASSO does relatively little to mitigate overfitting in our context, as the results are similar to those for the conventional WLS forecast. Compared to the WLS forecast, the LASSO forecast does produce a higher R^2_{CSMZ} statistic (2.75% and 2.58% for the LASSO and WLS forecasts, respectively). Based on Equation (3.11), this works to reduce the cross-sectional MSFE; the effect is moderate, however, so that the R^2_{CSOS} statistic is still sizably negative (-41.04%, compared to -45.73% for the WLS forecast).

The strong shrinkage effect of forecast combination is evident in the results for the C-LASSO forecast in Panel A of Table 2. The volatility of the C-LASSO forecast is only 0.21%, which is considerably below that of the WLS forecast. The $\hat{\delta}$ estimate for the C-LASSO forecast is significantly positive; at the same time, the estimate of 2.35 indicates a downward bias in the forecast, and the estimate is significantly greater than one. The R^2_{CSMZ} statistic increases to 4.20% for the C-LASSO forecast. Together with the reduction in forecast volatility, according to Equation (3.11), this contributes to a reduction in the cross-sectional MSFE. Indeed, unlike the WLS, SWLS, and LASSO forecasts, the R^2_{CSOS}

statistic for the C-LASSO forecast is positive, so that it outperforms the naïve benchmark in terms of cross-sectional MSFE. However, the R^2_{CSOS} statistic is close to zero (0.06%) and insignificant at conventional levels.

Although both the SWLS and C-LASSO forecasts perform markedly better than the WLS forecast, the former still appears to be characterized by overfitting, while the latter appears too conservative; for example, the SWLS and C-LASSO forecasts have upward and downward biases, respectively. This suggests that there are benefits to blending the two forecasts, which is what the E-LASSO approach does. Indeed, the results for the E-LASSO forecast in Panel A of Table 2 show that blending the two forecasts produces something close to a "Goldilocks" forecast. The R^2_{CSOS} statistic for the E-LASSO forecast is 0.43%, which is significant at the 1% level, so that the E-LASSO approach delivers a significant improvement in cross-sectional MSFE vis-á-vis the naïve benchmark. The $\hat{\delta}$ estimate for the E-LASSO forecast is 1.17, which is significantly greater than zero; it is also significantly greater than one, but only at the 10% level. Overall, the E-LASSO forecast is close to being unbiased and is significantly more accurate than the naïve benchmark forecast that ignores the information in the firm characteristics.

Intuitively, as discussed in Section 2, the SWLS (C-LASSO) forecast is likely to perform well when characteristic premia are relatively stable (unstable). By diversifying across the SWLS and C-LASSO forecasts, the E-LASSO forecast is robust to periods of stable and unstable characteristic premia. Panel A of Figure 1 shows the time series of the blending parameter, $\tilde{\theta}_t$ in Equation (2.19), used to compute the E-LASSO forecast. This is the weight attached to the C-LASSO forecast in a convex combination of the SWLS and C-LASSO forecasts. The blending parameter uses information available at the time of forecast formation to optimally blend the SWLS and C-LASSO forecasts. The figure reveals interesting fluctuations in the blending parameter around its average value of approximately 0.5 (corresponding to a simple average of the SWLS and C-LASSO forecasts). Specifically, there is a tendency for the weight attached to the C-LASSO forecast to increase in the run-up to a business-cycle recession, especially the recent Great Recession, suggesting that recessions are linked to time-varying characteristic premia.

Panel B of Table 2 provides results for equal weighting, which places much more emphasis on small-cap stocks. The conventional OLS forecast produces an R^2_{CSOS} statistic of -15.51% (which is significantly negative), so that it underperforms the naïve benchmark by a sizable margin, although the statistic is much smaller in magnitude than that for the WLS forecast in Panel A. The $\hat{\delta}$ estimate for the OLS forecast is 0.12, so that the forecast exhibits a considerable upward bias. The estimate is significantly greater than zero, but also significantly less than one. The OLS forecast volatility is 6.71%, compared to an actual return volatility of 16.24%. Overall, like the WLS forecast in Panel A, the conventional OLS forecast in Panel B is plagued by substantial overfitting.

As demonstrated by the results for the SOLS forecast in Panel B of Table 2, smoothing the coefficient estimates over time significantly improves forecasting performance. The forecast volatility falls to 1.94%, indicating that smoothing reduces the degree of overfitting. Moreover, the R^2_{CSOS} statistic is 0.38%, which is significant at the 1% level, so that the SOLS forecast is significantly more accurate than the naïve benchmark. The $\hat{\delta}$ estimate of 0.75 represents a moderate upward bias in the SOLS forecast; the $\hat{\delta}$ estimate is significantly greater than zero and significantly less than one.

The LASSO forecast in Panel B helps to alleviate overfitting, but to a quite limited extent. Compared to the OLS forecast volatility of 6.71%, the LASSO forecast volatility declines to 4.86%, which is still relatively high. The R^2_{CSOS} statistic of -7.40% is significantly negative, so that the LASSO forecast significantly underperforms the naïve benchmark in terms of cross-sectional MSFE. The $\hat{\delta}$ estimate of 0.20 indicates a sizable upward bias in the LASSO forecast, which is significant according to the t-statistic in the sixth column.

Similarly to Panel A, the C-LASSO forecast exerts a strong shrinkage effect in Panel B of Table 2, with a forecast volatility of 0.28%. The shrinkage effect enables the C-LASSO forecast to outperform the naïve benchmark, as evinced by the R^2_{CSOS} statistic of 0.11%, which is significant at the 1% level. Again demonstrating the shrinkage induced by the C-LASSO approach, the forecast exhibits a substantial and statistically significant downward bias ($\hat{\delta} = 3.95$).

Reminiscent of the results in Panel A, by blending the SOLS and C-LASSO forecasts, the E-LASSO forecast in Panel B of Table 2 generates the best performance in terms of cross-sectional MSFE. The R^2_{CSOS} statistic is 0.76% (significant at the 1% level). Although the E-LASSO forecast has a significant downward bias, it is moderate in magnitude ($\hat{\delta} = 1.29$). Panel B of Figure 1 displays the time series for the parameter used to blend the SOLS and C-LASSO forecasts. Similarly to Panel A, the weight attached to the C-LASSO forecast tends to increase in the run-up to a recession in Panel B. There are also distinct lower-frequency trends in the blending parameter. Through the mid 1990s, the weight for the C-LASSO forecast is always less than 0.5. It then exhibits an upward trend and remains above 0.5 near the mid 2000s through the end of the out-of-sample period. The characteristic premia appear more subject to time variation in the later part of the out-of-sample period, and the data-driven E-LASSO forecast accounts for this by placing greater weight on the C-LASSO forecast.

In summary, comparing the results in Panels A and B of Table 2, we see that the conventional forecast approach is marred by substantive overfitting for value and equal weighting. For both cases, the direct LASSO approach alleviates overfitting only moderately, so that the LASSO forecast significantly underperforms the naïve benchmark with respect to cross-sectional MSFE. Smoothing the coefficient estimates over time substantially improves forecasting performance, but the SWLS forecast still significantly underperforms the naïve benchmark for value weighting; for equal weighting, the SOLS forecast significantly outperforms the naïve benchmark. The C-LASSO approach exerts a strong shrinkage effect, which helps the C-LASSO forecast to outperform the naïve benchmark for value and equal weighting (significantly so for the latter case). The E-LASSO method performs the best in terms of cross-sectional MSFE for value and equal weighting, delivering significantly more accurate forecasts than the naïve benchmark for both cases. Finally, the ability of firm characteristics to improve cross-sectional out-of-sample return forecasts appears stronger for equal relative to value weighting. Specifically, the R^2_{CSOS} statistic for the E-LASSO forecast for equal weighting in Panel B is 77% larger than the corresponding statistic for value weighting in Panel A.

The forecast encompassing test in Section 3.3 helps to explain the value of blending the SWLS or SOLS and C-LASSO forecasts. For value weighting, the $\hat{\theta}$ estimate in Equation (3.17) is 0.54, and we easily reject the null hypothesis that the SWLS forecast encompasses the C-LASSO forecast (t-statistic of 30.12). We also easily reject the null hypothesis that the C-LASSO forecast encompasses the SWLS forecast (t-statistic of -25.57). Because each forecast fails to encompass the other, we have statistical evidence that each forecast contains unique useful information. A similar set of results holds for equal weighting: the $\hat{\theta}$ estimate is 0.41, and we reject the null that the SOLS forecast encompasses the C-LASSO forecast (t-statistic of 19.96), as well as the null that the C-LASSO forecast encompasses the SOLS forecast (t-statistic of -28.51).

Table 3 reports results for the first and second halves of the out-of-sample period (1975:01 to 1996:12 and 1997:01 to 2018:12, respectively). This provides a sense of the predictive ability of the firm characteristics over time. For value weighting in Panel A, the results for both subsamples follow the same pattern as those in Panel A of Table 2; most notably, the E-LASSO approach delivers the best performance in terms of cross-sectional MSFE. Panel A of Table 3 also shows that the degree of out-of-sample predictability diminishes from the first to the second subsample: the R^2_{CSOS} statistic for the E-LASSO forecast falls from 0.58% in the first subsample to 0.29% in the second. Nevertheless, the R^2_{CSOS} statistic is significant at the 1% level for each subsample, so that cross-sectional out-of-sample return predictability remains significant over time (despite the reduced number of observations available for each subsample). Similar results are evident for equal weighting in Panel B of Table 3, where the

R^2_{CSOS} statistic for the E-LASSO forecast goes from 1.08% in the first subsample to 0.45% in the second, with both statistics statistically significant at the 1% level.

Figure 2 provides additional perspective on the degree of out-of-sample cross-sectional return predictability over time. The figure depicts the 120-month rolling average of the monthly $R^2_{\text{CSOS},t}$ statistics, along with 95% confidence bands. The horizontal axis gives the end of the 120-month period. For value weighting in Panel A, cross-sectional out-of-sample return predictability is relatively stable through the mid 2000s. The degree of out-of-sample predictability then falls, although it remains significant most of the time (despite the relatively small number of observations corresponding to the rolling window), including at the end of the sample. The results for equal weighting in Panel B follow a similar pattern. However, out-of-sample predictability declines more sharply for equal weighting after the mid 2000s. Before that time, the degree of out-of-sample predictability is stronger for equal vis-á-vis value weighting; by the end of the sample, the degree of predictability is quite similar for value and equal weighting.

The time variation in predictive ability in Table 3 and Figure 2 are consistent with McLean and Pontiff (2016). In the context of long-short portfolios, they find that the predictive ability of individual firm characteristics tends to diminish over time as investors learn about relevant characteristics from academic studies; although diminished, the value of the information in characteristics often remains significant after it is "discovered."[10] We show that the collective information in a large number of firm characteristics continues to be useful for forecasting cross-sectional returns out of sample. While the degree of out-of-sample predictive ability declines over time, the collective information in the characteristics remains statistically significant. In Section 4.3, we show that it also remains economically valuable.

[10] Figure 2 also generally accords with the in-sample results in Green et al. (2017), who find that the predictability of many characteristics declines after 2003.

4.3 Economic Value

Results for forecast evaluation based on long-short portfolio performance are provided in Table 4. The table reports annualized means, volatilities, and Sharpe ratios for long-short portfolios that go long (short) stocks with the highest (lowest) forecasted returns (as described in Section 3.4). As a benchmark, Panel A (B) of Table 4 also reports performance measures for the excess return on the CRSP value-weighted (equal-weighted) market portfolio. The results largely align with those based on cross-sectional MSFE in Section 4.2.

For value weighting in Panel A of Table 4 and the full 1975:01 to 2018:12 out-of-sample period (see the second through fourth columns), the market portfolio generates an annualized mean and volatility of 7.86% and 15.29%, respectively, which translate to an annualized Sharpe ratio of 0.51. The long-short portfolio based on the conventional WLS forecast performs similarly to the market portfolio, with a Sharpe ratio of 0.57. The SWLS forecast performs substantially better. The long-short portfolio for the SWLS forecast has an average return over twice that of the market portfolio and lower volatility, so that its Sharpe ratio of 1.32 is more than twice that of the market. The portfolio based on the LASSO forecast has a Sharpe ratio that is slightly higher than that for the portfolio formed from the WLS forecast. The C-LASSO forecast improves upon the WLS forecast more substantially, with the long-short portfolio providing a Sharpe ratio of 0.74. As with the results based on cross-sectional MSFE in Table 2, the E-LASSO approach produces the best results in Panel A of Table 4: the long-short portfolio based on the E-LASSO forecast generates the highest average return (20.67%), lowest volatility (12.86%), and highest Sharpe ratio (1.61); the latter is over three times that of the market portfolio.

The fifth through seventh and eighth through tenth columns of Panel A report performance measures for the 1975:01 to 1996:12 and 1997:01 to 2018:12 subsamples, respectively. As in Table 3, the relative performance of the long-short portfolios in each subsample is similar to that for the full out-of-sample period, with the portfolio based on the E-LASSO forecast performing the best in each subsample. Again like Table 3, the collective information in the forecasts diminishes in value from the first to the second subsample, but it still remains valuable for the E-LASSO forecast in Panel A of Table 4: the Sharpe ratio goes from 2.43 in the first subsample to 1.21 in the second; the latter is sizable and nearly three times that of the market portfolio (0.43).

Panel B of Table 4 shows results for equal weighting, which follow the same pattern as those for value weighting in Panel A. Focusing on the long-short portfolio based on the E-LASSO forecast in Panel B, for the full out-of-sample period, its average return is 41.17%, which is nearly four times larger than that of the market portfolio (11.08%), while its volatility (10.35%) is well below that of the market (18.89%). These lead to a Sharpe ratio of 3.98 for the long-short portfolio formed from the E-LASSO forecasts. Again reminiscent of the cross-sectional MSFE results in Section 4.2, the collective information in the characteristics appears more valuable for equal weighting, which places considerably more weight on small-cap stocks. For the long-short portfolio based on the E-LASSO forecast in Panel B of Table 4, the Sharpe ratio is 5.74 and 2.91 in the first and second subsamples, respectively, both of which are economically sizable.

Further evidence on the time variation in long-short portfolio performance is provided in Figure 3. The figure portrays 120-month rolling window estimates of the Sharpe ratio, where the horizontal axis again gives the end of the 120-month window. For value weighting in Panel A, the Sharpe ratio is fairly stable at a value of approximately 2.5 through the late 1990s; it then declines to around 1.2 in the early 2000s and remains close to that level through the end of the sample. The rolling Sharpe ratios in Panel A are always significant at the 1% level. A similar pattern emerges for equal weighting in Panel B, although the rolling Sharpe ratios are larger than those in Panel A. There is also a more persistent decline in the Sharpe ratios throughout the 2000s for equal weighting in Panel B. As in Panel A, the rolling Sharpe ratios are all significant at the 1% level.

Figure 4 depicts log cumulative returns for the long-short portfolios based on the E-LASSO forecast and the market benchmark. For value weighting in Panel A, the E-LASSO portfolio appears to substantially outperform the market benchmark, in line with the performance measures and rolling Sharpe ratios reported in Panel A of Table 4 and Figure 3. Panel A of Figure 4 highlights interesting differences in portfolio performance during recessions. Specifically, the market portfolio tends to suffer large drawdowns during recessions; in contrast, the long-short portfolio based on the E-LASSO forecast typically generates large gains during recessions, especially the recent Great Recession. Panel B of Figure 4 paints a similar picture for equal weighting.[11]

4.4 Alpha

Next, we examine whether leading factor models from the literature can account for the substantial average excess return earned by the long-short portfolio based on the E-LASSO forecast. We consider four multifactor models: the Carhart (1997) four-factor, Fama and French (2015) five-factor, Hou et al. (2015) q-factor, and Stambaugh and Yuan (2017) mispricing-factor models. Table 5 reports estimates of the annualized alpha and factor exposures for each of the multifactor models for the full out-of-sample period, as well as the 1975:01 to 1996:12 and 1997:01 to 2018:12 subsamples.

Panel A of Table 5 reports results for the Carhart (1997) model, which includes the Fama and French (1993) market, size, and value factors, along with a momentum factor. According to the second, fourth, and sixth columns, the long-short portfolio based on the E-LASSO forecast generates a sizable risk-adjusted return consistently over time. The annualized alpha is 17.90% for the full sample and 18.79% and 16.89% for the first and second subsamples, respectively, all of which are significant at the 1% level. The long-short portfolio evinces essentially zero exposure to the market factor, significant positive exposures to the size and

[11] Table A4 of the Internet Appendix reports breakeven transaction costs for the long-short portfolios, with the market portfolio serving as the benchmark. Focusing on the long-short portfolio based on the E-LASSO forecast, unit transaction costs would need to be quite high to eliminate the gains; for example, the breakeven transaction cost with respect to the Sharpe ratio is 95 (255) basis points for value (equal) weighting.

momentum factors, and significant negative exposure to the value factor. Taken together, the factors can explain around 47% of the fluctuations in portfolio return; nevertheless, the factors do not account for the portfolio's performance. For equal weighting (see the third, fifth, and seventh columns), the factors again cannot explain the average excess return for the long-short portfolio: the annualized alphas are 39.74%, 44.11%, and 33.98% for the full sample and first and second subsamples, respectively (all of which are significant at the 1% level). The factors exposures are generally smaller in magnitude compared to those for value weighting.

The Fama and French (2015) model in Panel B of Table 5 augments market, size, and value factors with profitability and investment factors. The annualized alphas continue to be sizable; for example, they are 21.87% and 39.83% for the full sample for value and equal weighting, respectively. All of the alphas in Panel B are significant at the 1% level. For value and equal weighting, the long-short portfolio exhibits limited exposures to the profitability and investment factors (only the exposure to the profitability factor during the first subsample for equal weighting is significant). The results are similar in Panel C for the Hou et al. (2015) q-factor model, which includes investment and return-on-equity factors in addition to market and size factors. Although there are significant exposures to the investment and return-on-equity factors in some instances, the annualized alphas remain sizable and all are significant at the 1% level.

Finally, Panel D of Table 5 reports results for the Stambaugh and Yuan (2017) mispricing-factor model, which consists of market and size factors, together with management and performance mispricing factors derived from eleven prominent anomalies (Stambaugh et al. 2012, 2015). For both value and equal weighting, we continue to see significant positive exposure to the size factor. For value weighting, there is significant positive (negative) exposure to the performance (management) factor. Again, the factor exposures cannot explain the substantive average excess return for the long-short portfolio, as the annualized alphas are close in magnitude to those in Panels A through C and remain significant at the 1% level.

In summary, leading multifactor models cannot readily explain the performance of the long-short portfolio based on the E-LASSO forecast. The portfolio generates economically and statistically significant alpha consistently over time. The results in this section complement those based on statistical accuracy in Section 4.2 and demonstrate the efficacy of the E-LASSO strategy for incorporating the collection information in the 299 firm characteristics.

5 How Many Characteristics Matter?

In this section, we provide insight into how many and what types of firm characteristics are relevant for cross-sectional returns. We begin by constructing out-of-sample forecasts based on the six subgroups of characteristics formed according to economic concepts in Table 1. Table 6 reports results for the E-LASSO forecast, which is the most accurate.

Panel A of Table 6 reports results for the 19 firm characteristics in the momentum category. For value weighting, although the E-LASSO forecast appears unbiased, the R^2_{CSOS} statistic is quite small and statistically insignificant. The E-LASSO forecast based on the entire set characteristics in Table 2 thus performs markedly better in terms of cross-sectional MSFE than the forecast limited to the set of momentum-related characteristics. For equal weighting, the R^2_{CSOS} statistic is 0.29%, which is significant at the 1% level, so that the E-LASSO forecast significantly outperforms the naïve benchmark. However, it is still less than half of the value (0.76%) of the R^2_{CSOS} statistic for the E-LASSO forecast based on all of the characteristics in Table 2.

The E-LASSO forecast based on the 44 characteristics in the value vs. growth category significantly outperforms the naïve benchmark for value and equal weighting in Panel B. While significant, the R^2_{CSOS} statistics for value and equal weighting (0.15% and 0.28%, respectively) are less than half of the corresponding statistics (0.43% and 0.76%, respectively) for the entire set of characteristics in Table 2. Similar results obtain for E-LASSO forecasts based on the 35, 52, and 92 characteristics in the investment, profitability, and intangibles categories, respectively, in Panels C through E of Table 6; specifically, the R^2_{CSOS} statistics are significant but considerably smaller in magnitude than those for the complete set of characteristics. For value weighting, the E-LASSO forecast based on the 57 characteristics in the trading frictions category fails to outperform the naïve benchmark in Panel F. Although it significantly outperforms the naïve benchmark for equal weighting, the R^2_{CSOS} statistic is less than a fourth of that based on the entire set of characteristics.

Comparing the results in Tables 2 and 6, the information in the complete set of 299 firm characteristics consistently produces more accurate out-of-sample forecasts of cross-sectional returns than the information in subgroups of characteristics formed according to various economic concepts. It thus appears that a broad array of economic influences affect cross-sectional expected returns and that relevant information is lost by focusing on a subset of variables in a particular category.

Next, to get a further sense of whether more characteristics work better than fewer characteristics, we proceed as follows. We first form a subgroup of 25 characteristics by selecting the 11th, 23rd,..., 287th, 299th characteristics from the list in Table 1. We then form a subgroup of 50 characteristics by selecting the 5th, 11th,..., 293rd, 299th characteristics from the list. Continuing in the manner, we form subgroups of 25, 50, 75, 100, 150, and 299 characteristics, where the last subgroup corresponds to the entire set of characteristics in Table 1. This allows us to select "random" subgroups of various sizes with representatives from each economic category.

Table 7 provides results for the different subgroup sizes for the E-LASSO forecast. A clear pattern emerges. For value weighting, the R^2_{CSOS} statistic increases monotonically with the number of characteristics. The same holds for equal weighting, with the exception of the transition from 75 to 100 characteristics, where the R^2_{CSOS} statistic is slightly lower in Panel D compared to Panel C (0.34 and 0.35, respectively). For value weighting, the R^2_{CSOS} statistic is 0.10% for 25 characteristics, which is significant at the 5% level. It is significant at the 1% level for 50 or more characteristics; for equal weighting, the R^2_{CSOS} statistic is significant at the 1% level for each subgroup size. Overall, Table 7 indicates that more characteristics are preferred to less for forecasting cross-sectional returns.

To glean additional insight into the relevance of firm characteristics over time, we examine the individual characteristics selected by the LASSO in the cross-sectional Granger and Ramanathan (1984) multiple regression in Equation (2.16). We analyze the selected characteristics from two perspectives.

First, Figure 5 shows the number of characteristics selected by the LASSO each month. For value weighting in Panel A, a substantial number of characteristics appear relevant each month. The time-series average for the number of selected characteristics is 71, and between 40 and 100 characteristics are selected nearly every month. We find similar results for equal weighting in Panel B, where the time-series average for the number of selected characteristics is 53 and between 30 and 80 characteristics are selected for the majority of months.

Second, Figure 6 provides a heatmap for the selection frequencies for individual characteristics for the full 1975:01 to 2018:12 out-of-sample period, as well as the 1975:01 to 1996:12 and 1997:01 to 2018:12 subsamples. To conserve space, Figure 6 reports frequencies for the first 60 characteristics listed in Table 1; heatmaps for the remaining characteristics are provided in Figure A1 of the Internet Appendix. The figures indicate that many characteristics are selected between 20% to 50% of the time. The selection frequencies appear to be relatively stable over the subsamples for most characteristics. R1 (1-month momentum), EFP (analyst earnings forecast to price), FQ (quarterly Piotroski fundamental score), OB (order backlog), RA1 (year 1 lagged return, annual), RA25 (years 2-5 lagged returns, annual), SV (systematic volatility risk), and SUV (standardized unexplained volume) are among the characteristics with the highest selection frequencies for both value and equal weighting. It would be difficult to identify these characteristics a priori.

Taken together, the results in Tables 6 and 7 and Figures 5 and 6 suggest that the collective information in 299 firm characteristics from a variety of economic categories is relevant for tracking cross-sectional expected returns. However, not all of the characteristics matter every month. On average, about 50 to 70 characteristics matter in a given month, with substantive "churn" in the relevant characteristics over time, consistent with time-varying characteristic premia. Overall, our results indicate that firm characteristics affect cross-sectional expected returns in a quite complicated manner.

6 Conclusion

We ask the question: What is the maximum out-of-sample predictive ability of firm characteristics for cross-sectional stock returns? To address this question, we need to consider an extensive set of firm characteristics. To this end, we analyze the joint predictive power of 299 characteristics, the largest number considered to date in studies of joint predictive ability. Due to the high-dimensional nature of the problem, existing cross-sectional approaches are insufficient for analyzing the collective information in the 299 firm characteristics. To incorporate the information in the 299 characteristics in a manner that avoids overfitting, we develop an E-LASSO approach that relies on a flexible shrinkage strategy to improve cross-sectional return forecasts. We also develop an R^2_{CSOS} statistic to provide an informative means for assessing cross-sectional return forecasts in terms of MSFE.

Empirically, the collective information in the 299 firm characteristics, as captured by the E-LASSO forecast, evinces significant predictive ability for cross-sectional returns over the 1975:01 to 2018:12 out-of-sample period. Although weaker, cross-sectional out-of-sample return predictability remains significant after the early 2000s, which appears to be a structural break identified in the literature. The E-LASSO forecast also provides substantial economic value in the context of long-short portfolios both pre- and post-2003. Overall, the E-LASSO approach generates out-of-sample gains consistently over time and presently provides the most accurate out-of-sample estimates of cross-sectional expected returns.

We further find that subsets of firm characteristics grouped by economic features (e.g., profitability) or selected quasi-randomly have weaker predictive power than the entire set of 299 characteristics. The characteristics that matter appear to change over time; the relevance of individual characteristics comes and goes, and then comes back again. Around 50 to 70 characteristic appear relevant on average in a given month, with the relevant characteristics rotating over time. Hence, relying on a small set of firm characteristics selected ex ante to predict cross-sectional stock returns is likely to neglect a considerable amount of relevant information.

References

Akaike, H. (1973). Information Theory and an Extension of the Maximum Likelihood Principle. In *Proceedings of the 2nd International Symposium on Information Theory.* Ed. by B. N. Petrov and F. Csaki. Budapest: Akadémiai Kiadó, pp. 267–281.

Bao, Y. (2009). Estimation Risk-Adjusted Sharpe Ratio and Fund Performance Ranking under a General Return Distribution. *Journal of Financial Econometrics* 7:2, 152–173.

Campbell, J. Y. and S. B. Thompson (2008). Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average? *Review of Financial Studies* 21:4, 1509–1531.

Carhart, M. M. (1997). On Persistence in Mutual Fund Performance. *Journal of Finance* 52:1, 57–82.

Diebold, F. X. and M. Shin (2019). Machine Learning for Regularized Survey Forecast Combination: Partially-Egalitarian Lasso and its Derivatives. *International Journal of Forecasting* 35:4, 1679–1691.

Fama, E. F. and K. R. French (1993). Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics* 33:1, 3–56.

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

Fama, E. F. and J. D. MacBeth (1973). Risk, Return, and Equilibrium: Empirical Tests. *Journal of Political Economy* 81:3, 607–636.

Flynn, C. J., C. M. Hurvich, and J. S. Simonoff (2013). Efficiency for Regularization Parameter Selection in Penalized Likelihood Estimation of Misspecified Models. *Journal of the American Statistical Association* 108:503, 1031–1043.

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

Ghosh, S. (2011). On the Grouped Selection and Model Complexity of the Adaptive Elastic Net. *Statistics and Computing* 21:3, 451–462.

Goyal, A. and I. Welch (2008). A Comprehensive Look at the Empirical Performance of Equity Premium Prediction. *Review of Financial Studies* 21:4, 1455–1508.

Granger, C. W. J. and R. Ramanathan (1984). Improved Methods of Combining Forecasts. *Journal of Forecasting* 3:2, 197–204.

Green, J., J. R. M. Hand, and X. F. Zhang (2017). The Characteristics that Provide Independent Information about Average U.S. Monthly Stock Returns. *Review of Financial Studies* 30:12, 4389–4436.

Gu, S., B. Kelly, and D. Xiu (2020). Empirical Asset Pricing via Machine Learning. *Review of Financial Studies* 33:5, 2223–2273.

Harvey, C. R., Y. Liu, and H. Zhu (2016). ...and the Cross-Section of Expected Returns. *Review of Financial Studies* 29:1, 5–68.

Harvey, D. I., S. J. Leybourne, and P. Newbold (1998). Tests for Forecast Encompassing. *Journal of Business and Economic Statistics* 16:2, 254–259.

Haugen, R. A. and N. L. Baker (1996). Commonality in the Determinants of Expected Stock Returns. *Journal of Financial Economics* 41:3, 401–439.

Hou, K., C. Xue, and L. Zhang (2015). Digesting Anomalies: An Investment Approach. *Review of Financial Studies* 28:3, 650–703.

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

Hurvich, C. M. and C.-L. Tsai (1989). Regression and Time Series Model Selection in Small Samples. *Biometrika* 76:2, 297–307.

Kozak, S., S. Nagel, and S. Santosh (2020). Shrinking the Cross Section. *Journal of Financial Economics* 135:2, 271–292.

Lewellen, J. (2015). The Cross-Section of Expected Stock Returns. *Critical Finance Review* 4:1, 1–44.

McLean, R. D. and J. Pontiff (2016). Does Academic Research Destroy Return Predictability? *Journal of Finance* 71:1, 5–32.

Mincer, J. A. and V. Zarnowitz (1969). The Evaluation of Economic Forecasts. In *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*. Ed. by J. A. Mincer. Cambridge, MA: National Bureau of Economic Research, pp. 3–46.

Newey, W. K. and K. D. West (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica* 55:3, 703–708.

Rapach, D. E., J. K. Strauss, and G. Zhou (2010). Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy. *Review of Financial Studies* 23:2, 821–862.

Stambaugh, R. F., J. Yu, and Y. Yuan (2012). The Short of It: Investor Sentiment and Anomalies. *Journal of Financial Economics* 102:4, 288–302.

Stambaugh, R. F., J. Yu, and Y. Yuan (2015). Arbitrage Asymmetry and the Idiosyncratic Volatility Puzzle. *Journal of Finance* 70:5, 1903–1948.

Stambaugh, R. F. and Y. Yuan (2017). Mispricing Factors. *Review of Financial Studies* 30:4, 1270–1315.

Taddy, M. (2017). One-Step Estimator Paths for Concave Regularization. *Journal of Computational and Graphical Statistics* 26:3, 525–536.

Tibshirani, R. (1996). Regression Shrinkage and Selection via the LASSO. *Journal of the Royal Statistical Society. Series B (Methodological)* 58:1, 267–288.

Timmermann, A. (2006). Forecast Combinations. In *Handbook of Economic Forecasting*. Ed. by G. Elliott, C. W. J. Granger, and A. Timmermann. Vol. 1. Amsterdam: Elsevier, pp. 135–196.

Zou, H. (2006). The Adaptive Lasso and Its Oracle Properties. *Journal of the American Statistical Association* 101:476, 1418–1429.

Zou, H. and T. Hastie (2005). Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 67:2, 301–320.

Zou, H. and H. H. Zhang (2009). On the Adaptive Elastic-Net with a Diverging Number of Parameters. *Annals of Statistics* 37:4, 1733–1751.

Table 1: Firm characteristics

(1) Abbrev.	(2) Characteristic	(3) Abbrev.	(4) Characteristic
Panel A: Momentum (19)			
R1	1-month momentum	IM	Industry momentum
R6	6-month momentum	SUE	Standardized unexpected earnings
DR6	Change in 6-month momentum	EAR	Earnings announcement returns
R11	12-month momentum	ABR	Earnings announcement abnormal returns
RINT	Intermediate momentum	RE	Revisions in analyst earnings forecasts
R18	Momentum reversal	DEF	Change in forecasted earnings per share
R36	Long-term reversal, 36 months	NEI	Number of consecutive earnings increases
R60	Long-term reversal, 60 months	RS	Revenue surprise
RESID6	Residual momentum, prior 6-month returns	TES	Tax expense surprise
RESID11	Residual momentum, prior 11-month returns		
Panel B: Value vs. growth (44)			
BM	Book to market equity	OCP	Operating cash flow to price
BMIA	Industry-adjusted book to market equity	OCPQ	Quarterly operating cash flow to price
BMQ	Quarterly book to market equity	CD	Cash flow to debt
BMJ	Book to June-end market equity	FCB	Free cash flow to book equity
DP	Dividend yield	SP	Sales to price
DPQ	Quarterly dividend yield	SPQ	Quarterly sales to price
OP	Payout yield	EM	Enterprise multiple
OPQ	Quarterly payout yield	EMQ	Quarterly enterprise multiple
NOP	Net payout yield	EBP	Enterprise book to price
NOPQ	Quarterly net payout yield	EBPQ	Quarterly enterprise book to price
EP	Earnings to price	VHP	Intrinsic value to market equity
EPQ	Quarterly earnings to price	Q	Tobin's q
EFP	Analyst earnings forecast to price	SGR	5-year sales growth rank
DM	Debt to market equity	SG	Annual sales growth
DMQ	Quarterly debt to market equity	SGQ	Quarterly sales growth
NDP	Net debt to price	LTG	Analyst long-term growth forecasts
NDPQ	Quarterly net debt to price	G5Y	Forecasted growth in 5-year earnings per share
AM	Assets to market equity	IRB	Intangible return, book to market
AMQ	Quarterly assets to market equity	IRC	Intangible return, cash flow to price
CP	Cash flow to price	IRE	Intangible return, earnings to price
CPIA	Industry-adjusted cash flow to price	IRS	Intangible return, sales to price
CPQ	Quarterly cash flow to price	DUR	Equity duration

The table lists and provides abbreviations for the 299 firm characteristics used in this study. The characteristics are grouped according to six economic categories.

Table 1 (continued)

(1) Abbrev.	(2) Characteristic	(3) Abbrev.	(4) Characteristic
Panel C: Investment (35)			
IG	Investment growth	NXF	Net external finance
IG2	2-year investment growth	NEF	Net equity finance
IG3	3-year investment growth	NDF	Net debt finance
IGIA	Industry-adjusted percent change in investment	NOA	Net operating assets
IVG	Inventory growth	DNOA	Change in net operating assets
IVC	Inventory changes	DLNO	Change in long-term net operating assets
CI	Corporate investment	DWC	Change in net noncash working capital
ACI	Abnormal corporate investment	DCOA	Change in current operating assets
DEP	Depreciation to PP&E	DCOL	Change in current operating liabilities
DDEP	Percent change in depreciation	DNCO	Change in net noncurrent operating assets
DA	Asset growth	DNCA	Change in noncurrent operating assets
IA	Investment to assets	DNCL	Change in noncurrent operating liabilities
IAQ	Quarterly investment to assets	DFIN	Change in financial assets
DSO	Change in shares outstanding	DSTI	Change in short-term investments
IPO	New equity issues	DLTI	Change in long-term investments
NSI	Net stock issues	DFNL	Change in financial liabilities
CEI	Composite equity issuance	DBE	Change in common equity
CDI	Composite debt issuance		
Panel D: Profitability (52)			
ROE	Return on equity	GLAQ	Quarterly gross profits to lagged assets
ROEQ	Quarterly return on equity	OPE	Operating profits to book equity
DROEQ	4-quarter change in return on equity	OLE	Operating profits to lagged book equity
ROA	Return on assets	OLEQ	Quarterly operating profits to lagged book equity
ROAQ	Quarterly return on assets	OPA	Operating profits to book assets
DROAQ	4-quarter change in return on assets	OLA	Operating profits to lagged book assets
RNA	Return on net operating assets	OLAQ	Quarterly operating profits to lagged book assets
RNAQ	Quarterly return on net operating assets	COP	Cash-based operating profits to book assets
ROIC	Return on invested capital	CLA	Cash-based operating profits to lagged book assets
EPS	Earnings per share	CLAQ	Quarterly cash-based operating profits to lagged book assets
PM	Profit margin	PROF	Gross profitability to book equity
PMIA	Industry-adjusted profit margin	ROC	Cash productivity

Table 1 (continued)

(1) Abbrev.	(2) Characteristics	(3) Abbrev.	(4) Characteristics
PMQ	Quarterly profit margin	TBI	Taxable income to book income
DPM	Change in profit margin	TBIQ	Quarterly taxable income to book income
DPMIA	Industry-adjusted change in profit margin	IPM	Pre-tax income to sales
PCM	Sales minus costs of goods sold to sales	BL	Book leverage
SAT	Sales to total assets	BLQ	Quarterly book leverage
SATIA	Industry-adjusted sales to total assets	G	Mohanram growth score
ATO	Asset turnover	F	Piotroski fundamental score
ATOQ	Quarterly asset turnover	FQ	Quarterly Piotroski fundamental score
DATO	Change in asset turnover	O	Ohlson O-score
DATOIA	Industry-adjusted change in asset turnover	OQ	Quarterly Ohlson O-score
CTO	Capital turnover	Z	Altman Z-score
CTOQ	Quarterly capital turnover	ZQ	Quarterly Altman Z-score
GPA	Gross profits to assets	CR	Credit rating
GLA	Gross profits to lagged assets	FP	Failure profitability
Panel E: Intangibles (92)			
OA	Operating accruals	HA	Industry concentration in total assets
TA	Total accruals	HE	Industry concentration in book equity
DAC	Discretionary accruals	ALA	Liquidity of book assets
POA	Percent operating accruals	ALAQ	Quarterly liquidity of book assets
PTA	Percent total accruals	ALM	Liquidity of market assets
PDA	Percent discretionary accruals	ALMQ	Quarterly liquidity of market assets
AOA	Absolute value of operating accruals	VCF	Cash flow volatility
VOA	Accrual volatility	OB	Order backlog
ACQ	Accrual quality	ETR	Effective tax rate
OCA	Organizational capital to book assets	HN	Hiring rate
OCAIA	Industry-adjusted organizational capital to book assets	HNIA	Industry-adjusted hiring rate
ADM	Advertising expense to market equity	LFE	Labor force efficiency
GAD	Growth in advertising expense	SIN	Sin stocks
RDM	R&D expense to market equity	AGE	Firm age
RDIND	R&D increase	ANA	Analyst coverage
RDMQ	Quarterly R&D expense to market equity	DANA	Change in analyst coverage
RDS	R&D expense to sales	AOP	Analyst optimism
RDSQ	Quarterly R&D expense to sales	DLS	Disparity between long- and short-term earnings growth forecasts
RCA	R&D capital to book asset	DIS	Dispersion in analyst earnings forecasts
BCA	Brand capital to book assets	DLG	Dispersion in analyst long-term growth forecasts
LBP	Leverage component of book to price	PAFE	Predicted analyst forecast error

Table 1 (continued)

(1) Abbrev.	(2) Characteristics	(3) Abbrev.	(4) Characteristics
CTA	Cash to assets	EPER	Earnings persistence
CAL	Current ratio	EPRD	Earnings predictability
DCAL	Percent change in current ratio	ESM	Earnings smoothness
QUCIK	Quick ratio	EVR	Value relevance of earnings
DQUICK	Percent change in quick ratio	ETL	Earnings timeliness
OL	Operating leverage	ECS	Earnings conservatism
OLQ	Quarterly operating leverage	VROA	Earnings volatility
TAN	Tangibility of assets	DIVI	Dividend initiation
TANQ	Quarterly tangibility of assets	DIVO	Dividend omission
RER	Real estate ratio	GIND	Corporate governance
FRM	Pension plan funding to market equity	SA	SA index of financing constraints
FRA	Pension plan funding to book assets	KZ	Kaplan-Zingales index of financing constraints
SDD	Secured debt to total debt	KZQ	Quarterly Kaplan-Zingales index of financing constraints
SDIND	Secured debt indicator	WW	Whited-Wu index of financing constraints
CDIND	Convertible debt indicator	WWQ	Quarterly Whited-Wu index of financing constraints
DLD	Growth in long-term debt	RA1	Year 1 lagged return, annual
SC	Sales to cash	RN1	Year 1 lagged return, nonannual
SIV	Sales to inventories	RA25	Years 2-5 lagged returns, annual
DSIV	Percent change in sales to inventories	RN25	Years 2-5 lagged returns, nonannual
SR	Sales to receivables	RA610	Years 6-10 lagged returns, annual
DSI	Percent change in sales minus percent change in inventories	RN610	Years 6-10 lagged returns, nonannual
DSA	Percent change in sales minus percent change in accounts receivable	RA1115	Years 11-15 lagged returns, annual
DSS	Percent change in sales minus percent change in SG&A	RN1115	Years 11-15 lagged returns, nonannual
DGS	Percent change in gross margin minus percent change in sales	RA1620	Years 16-20 lagged returns, annual
HS	Industry concentration in sales	RN1620	Years 16-20 lagged returns, nonannual
Panel F: Trading frictions (57)			
ME	Market equity	IVQ	Idiosyncratic volatility from the q-factor model
MEIA	Industry-adjusted market equity	TS	Total skewness
PPS	Price per share	ISC	Idiosyncratic skewness from the CAPM
AT	Total assets	ISFF	Idiosyncratic skewness from the Fama-French 3-factor model

Table 1 (continued)

(1) Abbrev.	(2) Characteristics	(3) Abbrev.	(4) Characteristics
ATQ	Quarterly total assets	ISQ	Idiosyncratic skewness from the q-factor model
HIGH52	52-week high price	CS1	Coskewness, 1 month
MDR	Maximum daily return	CS60	Coskewness, 60 months
BETAEW	CAPM beta using daily returns and equal-weighted market excess return	TAIL	Tail risk
BETAEWSQ	CAPM beta squared	SBA	Bid-ask spread
BETADAILY	CAPM beta using daily returns	SHL	High-low bid-ask spread
BETAC	CAPM beta	TUR	Share turnover
BETAFP	Frazzini-Pedersen beta	DTO	Detrended turnover minus market turnover
BETAD	Dimson beta	VTUR	Turnover volatility
BETAFF	Fama-French 3-factor beta	CVT	Coefficient of variation of share turnover
BETAHS	Hong-Sraer beta	DTV	Dollar trading volume
BETALSY	Liu-Stambaugh-Yuan beta	VDTV	Volatility of dollar trading volume
BETADOWN	Downside beta	AMI	Absolute return to volume
BETARET	Acharya-Pedersen liquidity beta, return-return	SUV	Standardized unexplained volume
BETALCC	Acharya-Pedersen liquidity beta, illiquidity-illiquidity	CVD	Coefficient of variation of dollar trading volume
BETALRC	Acharya-Pedersen liquidity beta, return-illiquidity	VT	Volume trend
BETALCR	Acharya-Pedersen liquidity beta, illiquidity-return	LM1	Turnover-adjusted number of zero daily trading volume
BETANET	Acharya-Pedersen net liquidity beta	LM7	Prior 6-month turnover-adjusted number of zero daily trading volume
BETAPS	Pástor-Stambaugh liquidity beta	LM12	Prior 12-month turnover-adjusted number of zero daily trading volume
BETALEV	Financial intermediary leverage beta	VEA	Abnormal earnings announcement volume
TV	Total volatility	D1	Price delay based on R^2
SV	Systematic volatility risk	D2	Price delay based on slopes
IVEW	Idiosyncratic volatility using equal-weighted market excess return	D3	Price delay based on adjusted slopes
IVCA	Idiosyncratic volatility from the CAPM	PIN	Probability of information-based trading
IVFF	Idiosyncratic volatility from the Fama-French 3-factor model		

Table 2: R^2_{CSOS} statistics

(1)	(2)	(3)	(4)	(5)	(6)	(7)
Forecast	Forecast volatility (%)	R^2_{CSOS} (%)	$\hat{\delta}$	t-statistic, $\delta = 0$	t-statistic, $\delta = 1$	R^2_{CSMZ} (%)
Panel A: Value weighting						
WLS	5.24	-45.73^{***}	0.05	5.54^{***}	-95.79^{***}	2.58^{***}
SWLS	1.02	-0.40^{***}	0.47	9.56^{***}	-10.94^{***}	1.66^{***}
LASSO	4.98	-41.04^{***}	0.06	5.73^{***}	-86.03^{***}	2.75^{***}
C-LASSO	0.21	0.06	2.35	5.93^{***}	3.41^{***}	4.20^{***}
E-LASSO	0.47	0.43^{***}	1.17	11.89^{***}	1.69^{*}	1.72^{***}
Panel B: Equal weighting						
OLS	6.71	-15.51^{***}	0.12	11.44^{***}	-85.35^{***}	1.17^{***}
SOLS	1.94	0.38^{***}	0.75	19.78^{***}	-6.44^{***}	1.24^{***}
LASSO	4.86	-7.40^{***}	0.20	11.91^{***}	-46.53^{***}	1.61^{***}
C-LASSO	0.28	0.11^{***}	3.95	11.20^{***}	8.37^{***}	1.81^{***}
E-LASSO	1.17	0.76^{***}	1.29	20.13^{***}	4.48^{***}	1.21^{***}

The table reports statistics for out-of-sample forecasts of cross-sectional stock returns based on the 299 firm characteristics listed in Table 1. The out-of-sample period is 1975:01 to 2018:12. The forecasts are described in Section 2. Forecast volatility is the time-series average of the monthly forecast cross-sectional standard deviations; the time-series average of the monthly return cross-sectional standard deviations is 7.75% (16.24%) in Panel A (B). R^2_{CSOS} is the time-series average of the monthly cross-sectional out-of-sample R^2 statistics; $\hat{\delta}$ and R^2_{CSMZ} are the time-series averages of the monthly slope coefficient estimates and R^2 statistics, respectively, for cross-sectional Mincer and Zarnowitz (1969) regressions that relate actual to forecasted returns. For the third and fifth through seventh columns, *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively, based on heteroskedasticity and autocorrelation robust t-statistics. The t-statistic in the fifth (sixth) column is for testing H_0: $\delta = 0$ against H_A: $\delta \neq 0$ (H_0: $\delta = 1$ against H_A: $\delta \neq 1$). The statistics in Panel A are based on the weighted cross-sectional standard deviation, weighted cross-sectional mean squared forecast error, and weighted least squares estimation of the cross-sectional Mincer and Zarnowitz (1969) regressions, where the weights are based on market capitalization.

Table 3: Subsample results

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	1975:01 to 1996:12 subsample					1997:01 to 2018:12 subsample				
Forecast	R^2_{CSOS} (%)	$\hat{\delta}$	t-stat., $\delta = 0$	t-stat., $\delta = 1$	R^2_{CSMZ} (%)	R^2_{CSOS} (%)	$\hat{\delta}$	t-stat., $\delta = 0$	t-stat., $\delta = 1$	R^2_{CSMZ} (%)
Panel A: Value weighting										
WLS	−36.28***	0.09	6.44***	−66.04***	2.36***	−55.19***	0.02	1.57	−74.65***	2.80***
SWLS	−0.50**	0.44	9.40***	−11.88***	1.57***	−0.31*	0.49	5.74***	−5.96***	1.75***
LASSO	−31.60***	0.10	6.82***	−59.64***	2.56***	−50.48***	0.02	1.54	−67.48***	2.95***
C-LASSO	0.19**	2.86	5.18***	3.37***	3.77***	−0.06	1.85	3.28***	1.51	4.63***
E-LASSO	0.58***	1.22	10.97***	1.95**	1.62***	0.29***	1.12	6.89***	0.72	1.82***
Panel B: Equal weighting										
OLS	−12.21***	0.15	12.80***	−71.03***	0.87***	−18.81***	0.08	5.26***	−57.42***	1.47***
SOLS	0.92***	0.90	22.55***	−2.58***	1.47***	−0.16	0.61	10.04***	−6.37***	1.01***
LASSO	−4.70***	0.26	12.70***	−35.25***	1.24***	−10.10***	0.14	5.68***	−34.17***	1.98***
C-LASSO	0.15***	4.92	10.89***	8.68***	1.36***	0.07	2.97	5.75***	3.82***	2.26***
E-LASSO	1.08***	1.30	23.57***	5.44***	1.50***	0.45***	1.27	10.99***	2.35**	0.92***

The table reports statistics for out-of-sample forecasts of cross-sectional stock returns based on the 299 firm characteristics listed in Table 1. The forecasts are described in Section 2. R^2_{CSOS} is the time-series average of the monthly cross-sectional out-of-sample R^2 statistics; $\hat{\delta}$ and R^2_{CSMZ} are the time-series averages of the monthly slope coefficient estimates and R^2 statistics, respectively, for cross-sectional Mincer and Zarnowitz (1969) regressions that relate actual to forecasted returns. For the second, fourth through seventh, and ninth through eleventh columns, *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively, based on heteroskedasticity and autocorrelation robust t-statistics. The t-statistic in the fourth and ninth (fifth and tenth) columns is for testing H_0: $\delta = 0$ against H_A: $\delta \neq 0$ (H_0: $\delta = 1$ against H_A: $\delta \neq 1$). The statistics in Panel A are based on the weighted cross-sectional standard deviation, weighted cross-sectional mean squared forecast error, and weighted least squares estimation of the cross-sectional Mincer and Zarnowitz (1969) regressions, where the weights are based on market capitalization.

Table 4: Long-short portfolio performance

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	1975:01 to 2018:12 sample			1975:01 to 1996:12 subsample			1997:01 to 2018:12 subsample		
Forecast	Ann. mean (%)	Ann. volatility (%)	Ann. Sharpe ratio	Ann. mean (%)	Ann. volatility (%)	Ann. Sharpe ratio	Ann. mean (%)	Ann. volatility (%)	Ann. Sharpe ratio
Panel A: Value weighting									
Market	7.86	15.29	0.51***	9.09	14.99	0.61***	6.64	15.61	0.43**
WLS	8.81	15.45	0.57***	13.10	12.73	1.03***	4.52	17.69	0.26
SWLS	17.12	12.95	1.32***	17.74	9.64	1.84***	16.50	15.59	1.06***
LASSO	9.56	15.45	0.62***	14.93	13.03	1.15***	4.20	17.43	0.24
C-LASSO	13.47	18.33	0.74***	16.57	14.91	1.11***	10.38	21.20	0.49**
E-LASSO	20.67	12.86	1.61***	22.23	9.14	2.43***	19.10	15.73	1.21***
Panel B: Equal weighting									
Market	11.08	18.89	0.59***	13.06	18.58	0.70***	9.11	19.22	0.47**
OLS	23.45	15.67	1.50***	27.69	9.74	2.84***	19.22	19.84	0.97***
SOLS	39.38	11.36	3.47***	47.72	8.95	5.33***	31.04	12.91	2.40***
LASSO	26.40	18.20	1.45***	30.57	12.50	2.45***	22.23	22.47	0.99***
C-LASSO	21.91	18.98	1.15***	25.86	14.00	1.85***	17.96	22.88	0.79***
E-LASSO	41.17	10.35	3.98***	48.82	8.51	5.74***	33.52	11.50	2.91***

The table reports annualized summary statistics for long-short portfolios constructed from out-of-sample forecasts of cross-sectional stock returns based on the 299 firm characteristics listed in Table 1. The forecasts are described in Section 2. At the end of each month, we sort all available stocks into quintiles according to their forecasted returns for the next month. The breakpoints for the quintile portfolios are computed using NYSE stocks. The long-short portfolio goes long (short) the fifth (first) quintile. The quintiles for the long-short portfolios in Panel A (B) are value (equal) weighted according to market capitalization. Market in Panel A (B) is the CRSP value-weighted (equal-weighted) market portfolio return minus the risk-free return. For the fourth, seventh, and tenth columns, *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively, according to t-statistics based on Bao (2009) standard errors.

Table 5: Alphas and factor exposures

(1)	(2)	(3)	(4)	(5)	(6)	(7)
	1975:01 to 2018:12 sample		1975:01 to 1996:12 subsample		1997:01 to 2018:12 subsample	
Coefficient	Value weighting	Equal weighting	Value weighting	Equal weighting	Value weighting	Equal weighting
Panel A: Carhart (1997) four-factor model						
Ann. α (%)	17.90***	39.74***	18.79***	44.11***	16.89***	33.98***
MKT	0.00	0.02	0.03	0.12***	0.02	−0.08
SMB	0.52***	0.30***	0.54***	0.46***	0.47***	0.18**
HML	−0.29***	0.06	−0.14**	0.25***	−0.39***	−0.10
WML	0.32***	0.03	0.21***	0.08*	0.37***	−0.02
R^2 (%)	46.72	9.49	42.28	33.53	50.68	5.04
Panel B: Fama and French (2015) five-factor model						
Ann. α (%)	21.87***	39.83***	20.09***	43.42***	20.42***	33.89***
MKT	−0.07	0.02	0.03	0.13***	−0.14*	−0.07
SMB	0.50***	0.31***	0.52***	0.49***	0.46***	0.15*
HML	−0.49***	−0.03	−0.16**	0.20***	−0.63***	−0.15
RMW	−0.13	0.03	0.04	0.22**	−0.15	−0.05
CMA	0.02	0.08	−0.12	0.13	0.07	0.08
R^2 (%)	32.81	9.16	37.77	35.61	34.63	4.80

The table reports multifactor model estimation results for long-short portfolios constructed from out-of-sample forecasts of cross-sectional stock returns based on the 299 firm characteristics listed in Table 1 and the E-LASSO approach described in Section 2.5. At the end of each month, we sort all available stocks into quintiles according to their forecasted returns for the next month. The breakpoints for the quintile portfolios are computed using NYSE stocks. The long-short portfolio goes long (short) the fifth (first) quintile. The quintiles for the long-short portfolios in the second, fourth, and sixth (third, fifth, and seventh) columns are value (equal) weighted according to market capitalization. 0.00 indicates less than 0.005 in absolute value; *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively, based on heteroskedasticity and autocorrelation robust t-statistics. The factors are as follows: MKT = market excess return, SMB = "small minus big" size factor, HML = "high minus low" value factor, WML = "winner minus loser" momentum factor, RMW = "robust minus weak" profitability factor, CMA = "conservative minus aggressive" investment factor, ME = market equity factor, IA = investment factor, ROE = return on equity factor, MGMT = management factor, PERF = performance factor. The available sample in Panel D ends in 2016:12.

Table 5 (continued)

(1)	(2)	(3)	(4)	(5)	(6)	(7)
	1975:01 to 2018:12 sample		1975:01 to 1996:12 subsample		1997:01 to 2018:12 subsample	
Coefficient	Value weighting	Equal weighting	Value weighting	Equal weighting	Value weighting	Equal weighting
Panel C: Hou et al. (2015) q-factor model						
Ann. α (%)	19.83***	40.14***	17.14***	45.91***	19.66***	34.49***
MKT	−0.05	0.02	0.04	0.12***	−0.14	−0.10
ME	0.55***	0.27***	0.57***	0.43***	0.54***	0.14**
IA	−0.54***	0.03	−0.09	0.22**	−0.74***	−0.11
ROE	0.19**	−0.03	0.31***	−0.14**	0.10	−0.12
R^2 (%)	25.29	8.32	38.70	31.77	24.00	4.05
Panel D: Stambaugh and Yuan (2017) mispricing-factor model						
Ann. α (%)	17.54***	40.50***	20.15***	43.82***	15.14***	36.16***
MKT	0.00	0.03	0.01	0.12***	0.07	−0.07
SMB	0.55***	0.34***	0.48***	0.47***	0.60***	0.23***
MGMT	−0.33***	−0.01	−0.28***	0.12*	−0.43***	−0.11
PERF	0.34***	0.04	0.14**	−0.01	0.47***	0.04
R^2 (%)	37.82	10.66	42.09	28.32	40.23	5.44

Table 6: R^2_{CSOS} statistics for economic subgroups

(1) Weighting	(2) R^2_{CSOS} (%)	(3) $\hat{\delta}$	(4) t-stat., $\delta = 0$	(5) t-stat., $\delta = 1$	(6) R^2_{CSMZ} (%)	(7) R^2_{CSOS} (%)	(8) $\hat{\delta}$	(9) t-stat., $\delta = 0$	(10) t-stat., $\delta = 1$	(11) R^2_{CSMZ} (%)
	Panel A: Momentum (19)					*Panel B: Value vs. growth (44)*				
Value	0.01	0.87	5.20***	−0.75	2.27***	0.15**	1.12	8.36***	0.89	1.94***
Equal	0.29***	1.10	15.44***	1.42	0.74***	0.28***	1.39	17.74***	5.01***	0.62***
	Panel C: Investment (35)					*Panel D: Profitability (52)*				
Value	0.07**	0.90	5.10***	−0.58	1.24***	0.08*	0.98	5.67***	−0.10	1.51***
Equal	0.08***	1.45	10.75***	3.32***	0.38***	0.15***	1.59	11.24***	4.16***	0.68***
	Panel E: Intangibles (92)					*Panel F: Trading Frictions (57)*				
Value	0.18***	1.39	8.65***	2.41**	1.69***	−0.09	0.75	4.08***	−1.34	2.98***
Equal	0.13***	1.69	12.51***	5.12***	0.58***	0.18***	1.26	11.84***	2.45**	1.14***

The table reports statistics for out-of-sample forecasts of cross-sectional stock returns based on the firm characteristics for the subgroup in the panel heading and the E-LASSO approach described in Section 2.5; the characteristics in each subgroup are listed in Table 1. The out-of-sample period is 1975:01 to 2018:12. R^2_{CSOS} is the time-series average of the monthly cross-sectional out-of-sample R^2 statistics; $\hat{\delta}$ and R^2_{CSMZ} are the time-series averages of the monthly slope coefficient estimates and R^2 statistics, respectively, for cross-sectional Mincer and Zarnowitz (1969) regressions that relate actual to forecasted returns. For the second, fourth through seventh, and ninth through eleventh columns, *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively, based on heteroskedasticity and autocorrelation robust t-statistics. The t-statistic in the fourth and ninth (fifth and tenth) columns is for testing H_0: $\delta = 0$ against H_A: $\delta \neq 0$ (H_0: $\delta = 1$ against H_A: $\delta \neq 1$). For value weighting, the statistics are based on the weighted cross-sectional mean squared forecast error and weighted least squares estimation of the cross-sectional Mincer and Zarnowitz (1969) regressions, where the weights are based on market capitalization.

Table 7: R^2_{CSOS} statistics for quasi-random subgroups

(1) Weighting	(2) R^2_{CSOS} (%)	(3) $\hat{\delta}$	(4) t-stat., $\delta = 0$	(5) t-stat., $\delta = 1$	(6) R^2_{CSMZ} (%)	(7) R^2_{CSOS} (%)	(8) $\hat{\delta}$	(9) t-stat., $\delta = 0$	(10) t-stat., $\delta = 1$	(11) R^2_{CSMZ} (%)
	Panel A: 25 characteristics					*Panel B: 50 characteristics*				
Value	0.10**	1.42	6.30***	1.87*	2.51***	0.22***	1.40	6.61***	1.87*	2.74***
Equal	0.18***	1.48	11.58***	3.76***	0.84***	0.24***	1.44	10.92***	3.35***	1.02***
	Panel C: 75 characteristics					*Panel D: 100 characteristics*				
Value	0.23***	1.34	8.07***	2.06**	2.17***	0.33***	1.36	10.57***	2.88***	2.10***
Equal	0.35***	1.40	16.20***	4.62***	0.82***	0.34***	1.31	13.94***	3.26***	0.95***
	Panel E: 150 characteristics					*Panel F: 299 characteristics*				
Value	0.36***	1.25	10.01***	2.05**	1.66***	0.43***	1.17	11.89***	1.69*	1.72***
Equal	0.57***	1.28	18.03***	3.99***	0.99***	0.76***	1.29	20.13***	4.48***	1.21***

The table reports statistics for out-of-sample forecasts of cross-sectional stock returns based on the number of firm characteristics in the panel heading and the E-LASSO approach described in Section 2.5. We select the characteristics listed in Table 1 in increments of 300 divided by the number of characteristics in the panel heading; for example, in Panel A, we use the 11th, 23rd,..., 287th, 299th characteristics listed in Table 1; Panel F uses all of the 299 characteristics in Table 1. The out-of-sample period is 1975:01 to 2018:12. R^2_{CSOS} is the time-series average of the monthly cross-sectional out-of-sample R^2 statistics; $\hat{\delta}$ and R^2_{CSMZ} are the time-series averages of the monthly slope coefficient estimates and R^2 statistics, respectively, for cross-sectional Mincer and Zarnowitz (1969) regressions that relate actual to forecasted returns. For the second, fourth through seventh, and ninth through eleventh columns, *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively, based on heteroskedasticity and autocorrelation robust t-statistics. The t-statistic in the fourth and ninth (fifth and tenth) columns is for testing H_0: $\delta = 0$ against H_A: $\delta \neq 0$ (H_0: $\delta = 1$ against H_A: $\delta \neq 1$). For value weighting, the statistics are based on the weighted cross-sectional mean squared forecast error and weighted least squares estimation of the cross-sectional Mincer and Zarnowitz (1969) regressions, where the weights are based on market capitalization.

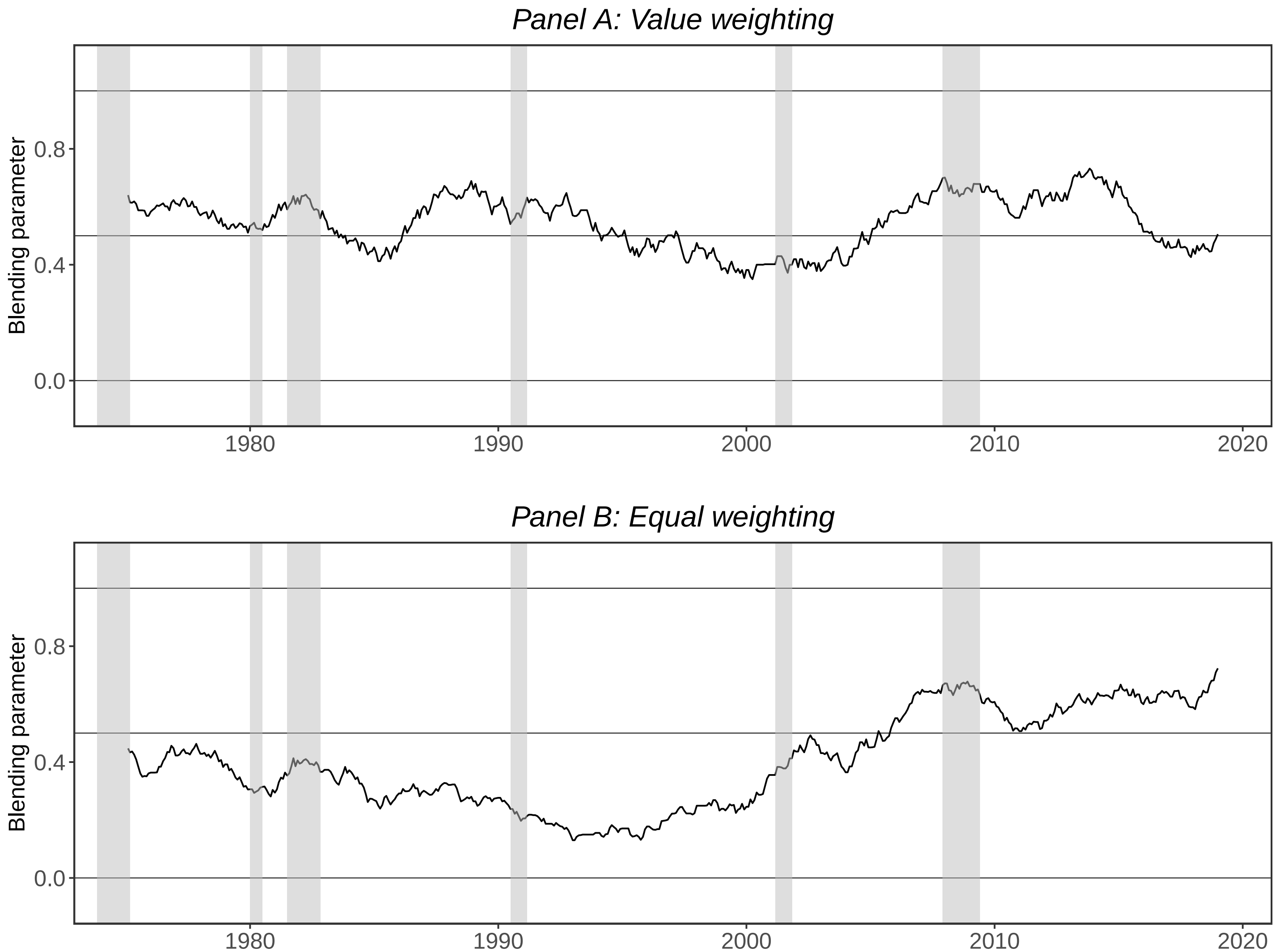

Figure 1: Blending parameter for the E-LASSO forecast

The figure depicts the parameter for blending the SOLS (Panel A) or SWLS (Panel B) and C-LASSO forecasts to form the E-LASSO forecast. Vertical bars delineate business-cycle recessions as dated by the National Bureau of Economic Research.

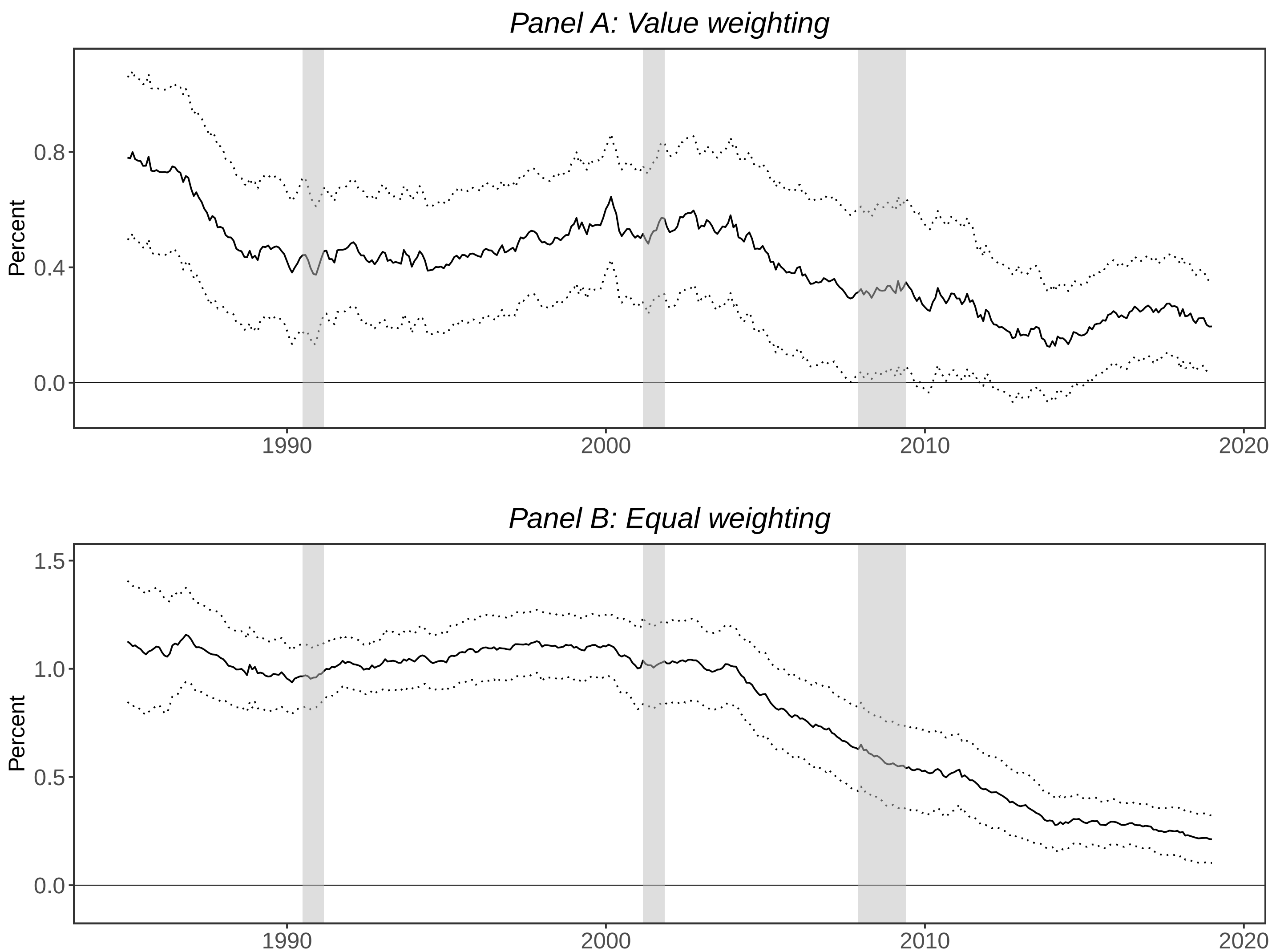

Figure 2: Rolling $R^2_{\mathrm{CSOS},t}$ means for the E-LASSO forecast

The solid line in each panel depicts 120-month rolling time-series means for monthly $R^2_{\mathrm{CSOS},t}$ statistics for the E-LASSO forecast. The dotted lines delineate 95% confidence bands. Vertical bars delineate business-cycle recessions as dated by the National Bureau of Economic Research.

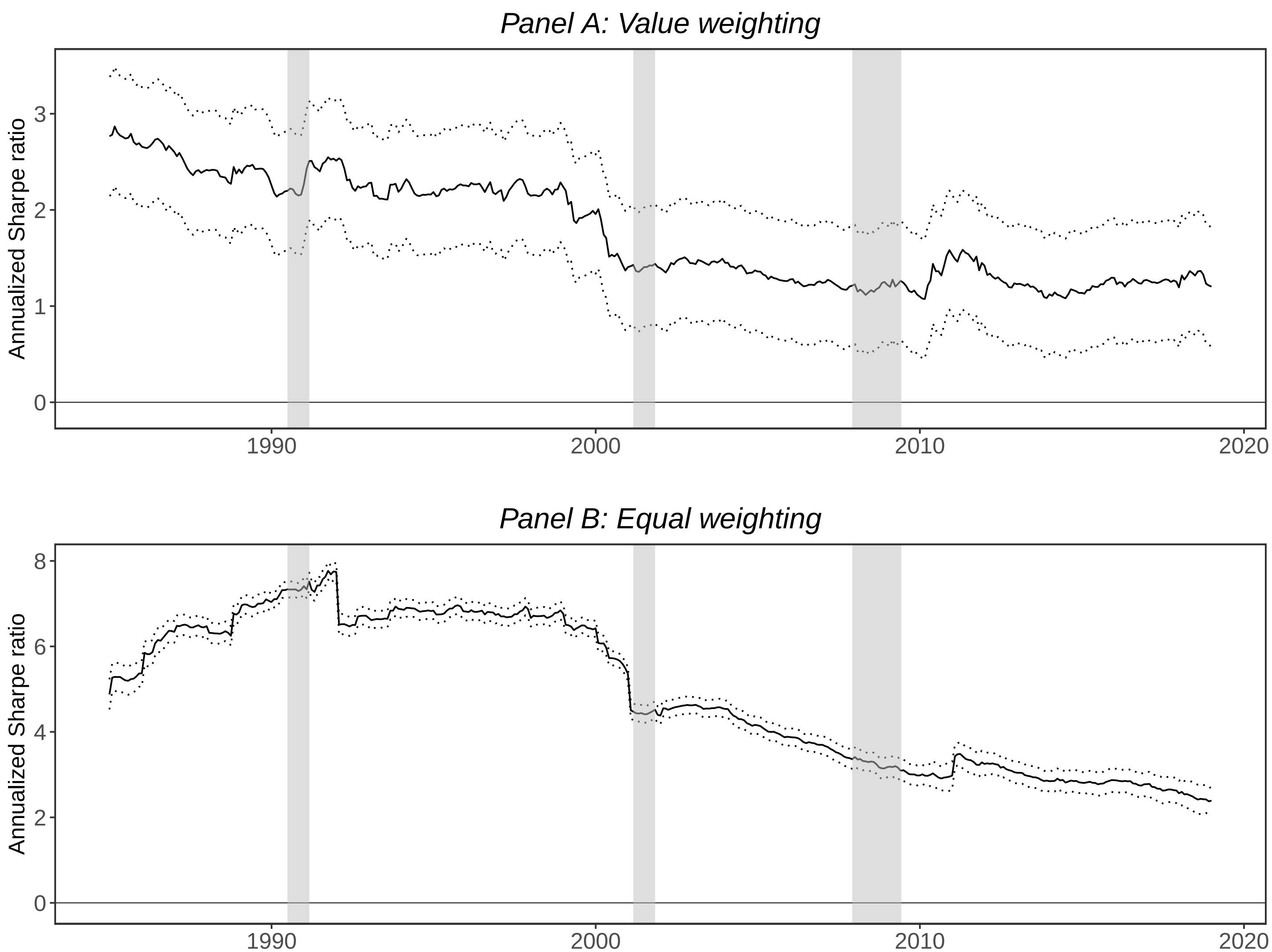

Figure 3: Rolling Sharpe ratios for long-short portfolios based on the E-LASSO forecast

The solid line in each panel depicts 120-month rolling Sharpe ratios for the long-short portfolio based on the E-LASSO forecast. The dotted lines delineate 95% confidence bands. Vertical bars delineate business-cycle recessions as dated by the National Bureau of Economic Research.

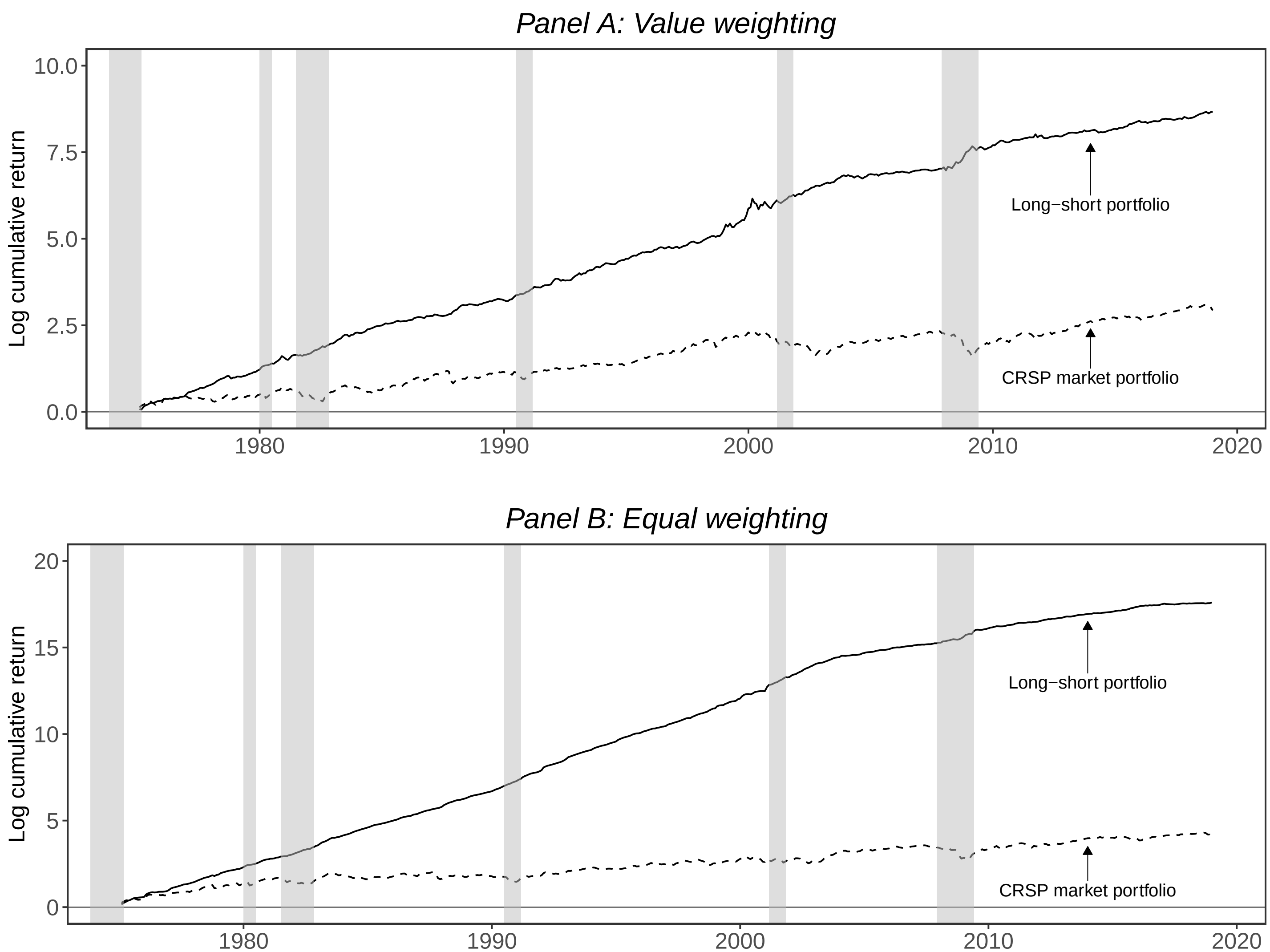

Figure 4: Log cumulative return for long-short portfolios based on the E-LASSO forecast

The solid (dashed) line in each panel depicts the log cumulative return for the long-short portfolio based on the E-LASSO forecast (CRSP market portfolio). Vertical bars delineate business-cycle recessions as dated by the National Bureau of Economic Research.

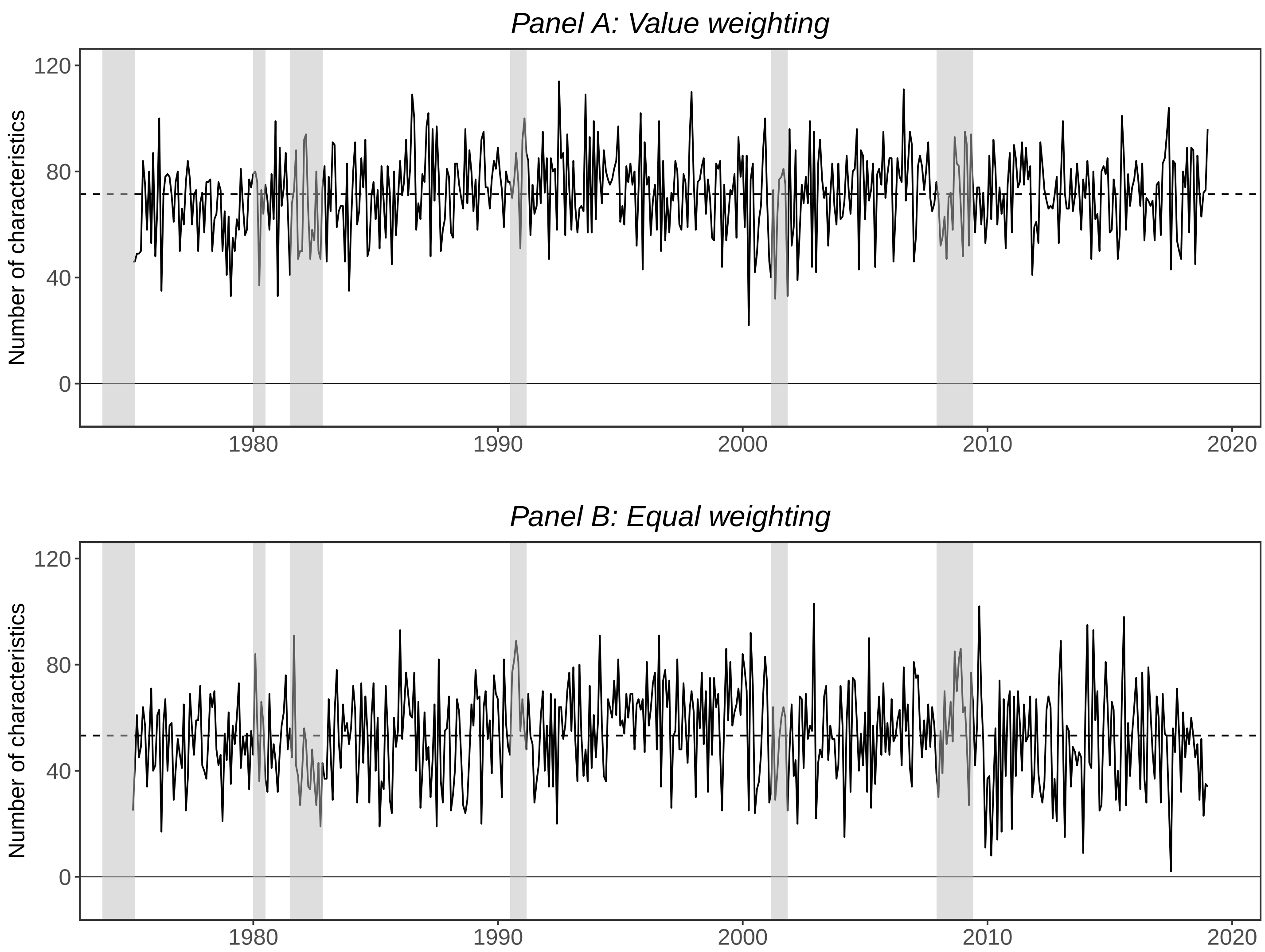

Figure 5: Number of firm characteristics selected each month

The solid line depicts the number of firm characteristics selected by the LASSO each month in cross-sectional Granger and Ramanathan (1984) multiple regressions. The horizontal dashed line gives the time-series average of the monthly values (71 and 53 in Panels A and B, respectively).

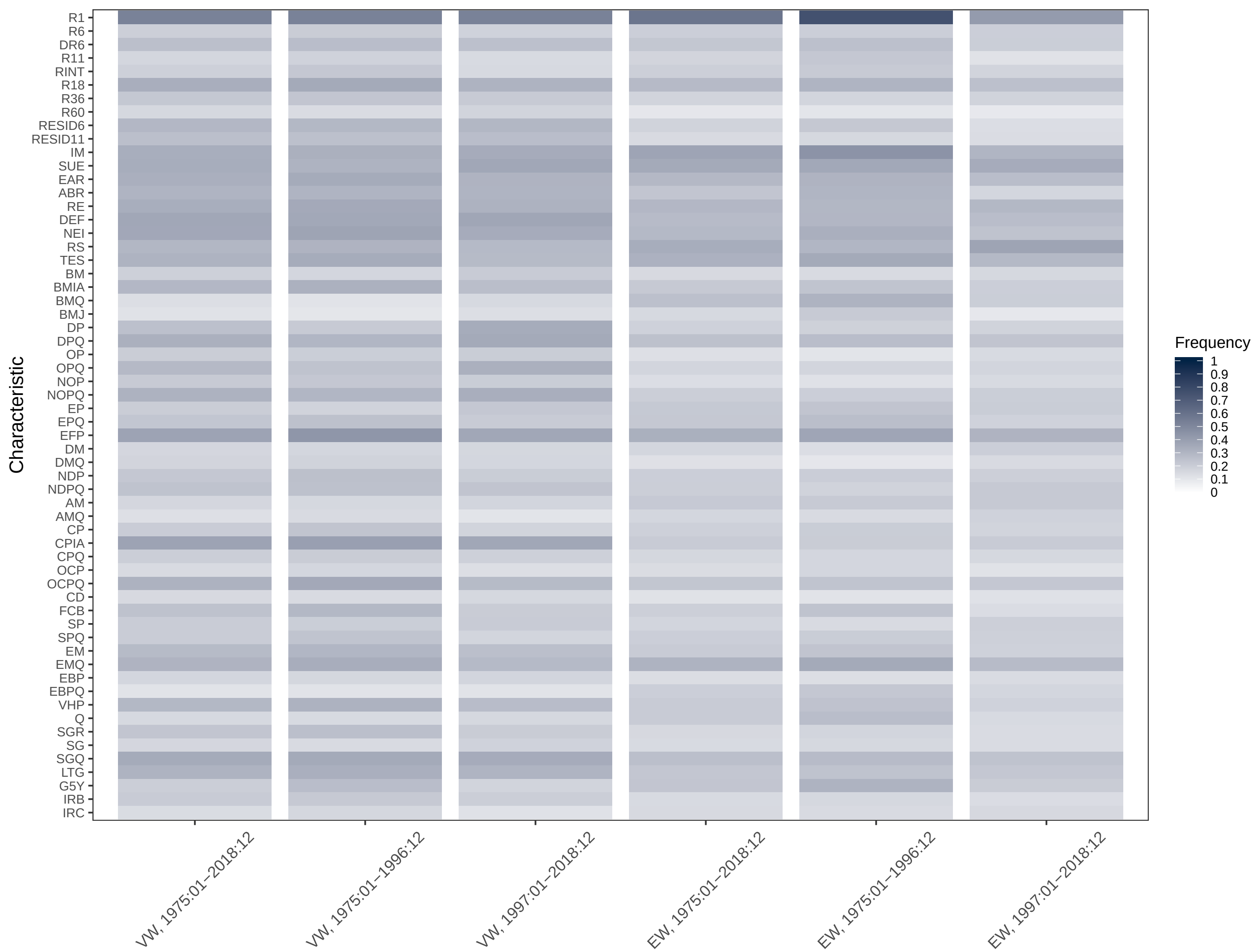

Figure 6: Selection frequencies for firm characteristics

The figure provides a heatmap for the selection frequencies for the first 60 of the 299 firm characteristics listed in Table 1 selected by the LASSO in monthly cross-sectional Granger and Ramanathan (1984) multiple regressions. VW (EW) denotes value (equal) weighting. Heatmaps for the selection frequencies for the remaining characteristics are presented in Table A1 of the Internet Appendix.