What Firm Characteristics Drive US Stock Returns?

Yufeng Han
University of North Carolina at Charlotte
yhan15@uncc.edu

Ai He
Emory University
ai.he@emory.edu

David E. Rapach
Saint Louis University
david.rapach@slu.edu

Guofu Zhou*
Washington University in St. Louis and CAFR
zhou@wustl.edu

October 12, 2018

*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, Baruch CUNY, Beijing University, Georgia State University, Indiana University, Renmin University of China, Rutgers University, Saint Louis University, Shanghai University of Finance and Economics, and Washington University in St. Louis, as well as Minchul Shin and Michael Weber, for insightful comments. We are also extremely grateful to John Hand, Jeremiah Green, and Frank Zhang for providing extensive and insightful comments on a previous draft, which helped to clarify various issues and improve the exposition of the paper.

What Firm Characteristics Drive US Stock Returns?

Abstract

We employ a forecast combination approach to analyze the ability of 94 firm characteristics from Green, Hand, and Zhang (2017) to predict US stock returns. Using machine learning tools to pool forecasts, we find that most of the firm characteristics matter over time—and approximately 30 matter on average at each point in time—for forecasting value-weighted cross-sectional returns before and after 2003, a year when Green, Hand, and Zhang (2017) detect a major structural break. By processing the information in a plethora of predictors in a manner that alleviates overfitting, our combination approach provides economically significant out-of-sample forecasts of cross-sectional returns consistently over time.

JEL classifications: G11, G14

Key words: Firm characteristics, Cross-sectional expected stock returns, Forecast combination, Machine learning, Fama-MacBeth regression, Forecast encompassing

1. Introduction

Green, Hand, and Zhang (2017, GHZ) construct a comprehensive set of 94 firm characteristics and analyze their ability to explain cross-sectional US stock returns, addressing a challenge posed by Cochrane (2011). They detect a sharp deterioration in the ability of firm characteristics to predict cross-sectional value-weighted returns after 2003. They further conclude that twelve firm characteristics affect cross-sectional value-weighted expected returns before 2003, while only two characteristics influence expected returns after 2003. GHZ's thought-provoking findings point to a substantially diminished role for firm characteristics in determining cross-sectional expected returns since 2003 and suggest that relatively few characteristics matter for expected returns.

A singular strength of GHZ is the analysis of a large number of firm characteristics from the literature. In the present paper, we reexamine the ability of the 94 firm characteristics compiled by GHZ to predict cross-sectional returns. Our analysis focuses on the construction of out-of-sample cross-sectional return forecasts, and we use a forecast combination approach to incorporate the information from a plethora of predictor variables in a manner that guards against *overfitting*.[1] GHZ employ conventional ordinary or weighted least squares (OLS or WLS, respectively) to estimate cross-sectional multiple regressions that simultaneously include all 94 firm characteristics as predictor variables. Conventional estimation is asymptotically optimal in environments with a stable data-generating process (DGP). However, because conventional OLS or WLS maximizes the fit of the model over a finite estimation (or training) sample, it potentially leads to unreliable out-of-sample performance, especially in the presence of a large noise component and structural instability in the DGP. These are relevant concerns for cross-sectional stock returns, as returns inherently

[1] Our notion of overfitting accords with the definition from the Oxford English Dictionary: "The production of an analysis which corresponds too closely or exactly to a particular set of data, and may therefore fail to fit additional data or predict future observations reliably" (`https://en.oxforddictionaries.com/definition/overfitting`). See Timmermann (2006) for an extensive survey of forecast combination.

contain a large unpredictable component, and GHZ's important analysis points to structural instability in the relationship between firm characteristics and cross-sectional returns.

Specifically, instead of generating stock return forecasts by fitting a cross-sectional multiple regression that includes all 94 characteristics as predictor variables, our forecast combination approach produces return forecasts by first fitting a series of cross-sectional univariate regressions, each of which includes an individual firm characteristic as a predictor variable. We then pool the cross-sectional return forecasts corresponding to the individual characteristics by taking the arithmetic mean, thereby providing us with combination forecasts of cross-sectional returns. In a time-series multiple regression context, Rapach, Strauss, and Zhou (2010) show that the mean combination forecast alleviates multicollinearity and acts as a *shrinkage* device. In this way, forecast combination stabilizes out-of-sample forecasts and mitigates overfitting.[2]

The mean combination forecast is a simple average of individual forecasts based on all of the firm characteristics. As recently proposed by Diebold and Shin (forthcoming), we also employ machine learning tools in an effort to refine the combination forecast. Specifically, we use the Tibshirani (1996) least absolute shrinkage and selection operator (LASSO) or Zou and Hastie (2005) elastic net to select the individual forecasts to include in the combination forecast. Machine learning tools are becoming increasingly popular in finance; for example, Feng, Giglio, and Xiu (2017), Freyberger, Neuhierl, and Weber (2018), Gu, Kelly, and Xiu (2018), Chinco, Clark-Joseph, and Ye (forthcoming), and Kozak, Nagel, and Santosh (forthcoming) utilize machine learning to analyze cross-sectional returns. In contrast to these studies, we employ machine learning in the construction of combination forecasts of cross-sectional returns. As a byproduct of the construction of the combination forecasts, the individual forecasts selected by the LASSO or elastic net provide insight into the relevance of individual firm characteristics over time.

[2]Rapach, Strauss, and Zhou (2010) find that forecast combination substantially improves time-series forecasts of the US aggregate market excess return. Applying forecast combination to the cross section for the first time, the present paper shows that it also significantly improves cross-sectional return forecasts.

To assess the cross-sectional return forecasts, we follow Lewellen (2015) and analyze the slope coefficients in Fama and MacBeth (1973, FM) regressions of realized on forecasted cross-sectional returns. We extend analytical results in Lewellen (2015) from equal to general weighting of cross-sectional return observations. The analytical results facilitate interpretation of the FM slope coefficient estimates. We also develop a new procedure based on forecast encompassing (Chong and Hendry 1986; Fair and Shiller 1990; Harvey, Leybourne, and Newbold 1998) for comparing the information content of two competing cross-sectional return forecasts in an FM framework.

As a benchmark, we first compute monthly cross-sectional return forecasts based on conventional estimation of cross-sectional multiple regressions. For the 1990 to 2017 forecast evaluation period and WLS estimation of the cross-sectional regressions, the FM slope coefficient estimate is well below unity, implying that the conventional return forecasts substantially overstate the cross-sectional dispersion in expected returns (Lewellen 2015), a direct indication of overfitting. The relatively small value for the FM slope estimate also implies that the conventional forecasts are less accurate on average in terms of cross-sectional value-weighted mean squared forecast error (MSFE) than a "naive" forecast that ignores the information in firm characteristics. The FM slope estimate becomes statistically indistinguishable from zero after 2003, so that overfitting is magnified and essentially renders the conventional forecasts unrelated to realized returns after 2003.

In contrast to the conventional forecasts, the FM slope coefficient estimates for the various combination forecasts are all above (but not statistically different from) unity, demonstrating that the combination forecasts effectively guard against overfitting and provide a more accurate measure of the cross-sectional dispersion in expected returns. The FM slope estimates are quite stable before and after 2003, so that—unlike the conventional forecasts—the combination forecasts remain informative for tracking cross-sectional value-weighted returns after 2003. The combination forecasts also deliver a lower cross-sectional value-weighted MSFE on average than the naive forecast that ignores the information in firm characteristics.

Before 2003, forecast encompassing tests indicate that the conventional and combination forecasts contain unique information relative to each other for predicting cross-sectional value-weighted returns. After 2003, however, the conventional forecasts no longer provide useful information for predicting cross-sectional returns beyond the information already contained in the combination forecasts. Indeed, for the post-2003 period, an optimal composite forecast comprised of the conventional and combination forecasts attaches a coefficient very near zero (unity) to the conventional (combination) forecast, indicating that the conventional forecast provides little information gain after 2003. Because the conventional and combination approaches share the same information set—namely, data for 94 firm characteristics—the differences in outcomes relate to *how* the information in firm characteristics is processed. Forecast combination helps to improves out-of-sample performance by processing large information sets in a manner that guards against overfitting.

The LASSO and elastic net both suggest that a large number of firm characteristics matter for forecasting cross-sectional value-weighted returns. For the 1990 to 2017 period, both methods identify approximately 30 characteristics on average as relevant cross-sectional return predictors. The number of characteristics selected each month by the LASSO or elastic net is quite stable over time, and more than ten characteristics are selected each month. Along this line, approximately 30 characteristics are selected on average before and after 2003.

Although approximately 30 characteristics are selected on average by the LASSO and elastic net, this figure tends to understate the number of relevant firm characteristics, as there is significant "churn" in the individual characteristics selected over time. The selection frequencies for the vast majority of individual characteristics exceed 30%, with the largest frequencies just below 50%. Overall, our results indicate that most of the 94 characteristics matter over time, and around 30 characteristics matter on average at each point in time. This is consistent with the view that the importance of individual economic risks changes over time and that a considerable number of individual risks—rather than only a few—matter at a point in time. To the extent that the characteristics reflect behavioral biases or related factors that generate mispricing, our findings similarly imply that the relevance of individual behavioral factors varies significantly over time and that a sizable number of such factors are pertinent at each point in time.

As in Lewellen (2015) and Green, Hand, and Zhang (2017), we also assess the economic value of cross-sectional return forecasts by forming monthly spread portfolios that go long (short) the top (bottom) decile of firms with the highest (lowest) forecasted returns. Spread portfolios based on the combination forecasts deliver sizable average monthly returns, as well as substantial risk-adjusted average returns in the context of the Carhart (1997) four-factor, Fama and French (2015) five-factor, and Hou, Xue, and Zhang (2015) q-factor models. Confirming results in GHZ, a value-weighted spread portfolio based on the conventional multiple regression approach exhibits a significant decline in performance after 2003; in contrast, spread portfolios formed from the combination forecasts generate sizable gains both before and after 2003.

The rest of the paper is organized as follows. Section 2 describes the forecast combination approach, as well as the econometric procedures we use to analyze cross-sectional return forecasts. Section 3 reports results for the 1990 to 2017 out-of-sample period and pre- and post-2003 subperiods. Section 4 concludes.

2. Data and Methodology

2.1. Data

We use the same 94 firm characteristics as GHZ, updated through December of 2017.[3] The data begin in January of 1980, as most characteristics only become widely available starting in 1980. We retain common stocks on the NYSE, AMEX, and NASDAQ that have a market value on CRSP at the end of the previous month and nonmissing value for common equity in

[3]We thank Jeremiah Green for providing SAS code to extract the data from CRSP, Compustat, and I/B/E/S on his webpage (`https://sites.google.com/site/jeremiahrgreenacctg/home`).

the firm's annual financial statement. As in GHZ, we relate a firm's month-t return to firm characteristics available at the end of month $t-1$. We assume that annual firm characteristics are available in month $t-1$ if the firm's fiscal year ended at least six months before month $t-1$. Similarly, we assume that quarterly accounting data are available in month $t-1$ if the fiscal quarter ended at least four months before month $t-1$. I/B/E/S and CRSP data are aligned in calendar time using the I/B/E/S statistical period date and CRSP monthly end date.

Following GHZ, we winsorize the characteristics at the 1st and 99th percentiles of their monthly observations, and we use the cross-sectional mean and standard deviation to standardize the observations for each characteristic for each month (so that they have zero mean and unit standard deviation). To avoid excluding a stock from a cross-sectional multiple regression when it is missing an observation for only one or a few characteristics, we again follow GHZ and replace the missing values with the characteristic's cross-sectional standardized mean value of zero.[4]

For convenience, Table 1 provides acronyms and definitions for the 94 firm characteristics, as given in Table 1 of GHZ. The Appendix in GHZ provides detailed information for the characteristics.

2.2. Forecast Combination

We generate a combination forecast of the month-$(t+1)$ return for stock i based on information available in month t as follows. For month t, we first estimate a series of cross-sectional univariate regressions, each of which relates returns to an individual characteristic:

$$r_{i,t} = a_{j,t} + b_{j,t} z_{i,j,t-1} + \varepsilon_{i,t} \quad \text{for } i = 1, \ldots, I_t;\ j = 1, \ldots, J_{t-1}, \tag{2.1}$$

[4]Note that our mean combination forecast does not have any problem with missing values and hence does not require us to fill in missing values. Nevertheless, the mean combination approach yields virtually the same results whether we ignore missing values or use missing-value-filled data à la GHZ.

where $r_{i,t}$ is the month-t return for stock i, $z_{i,j,t-1}$ is the jth firm characteristic for stock i in month $t-1$, I_t is the number of stocks available in month t, and J_{t-1} is the number of characteristics available at the end of month $t-1$. Due to missing values for some characteristics (especially in earlier periods), J_t can be smaller than 94. We estimate Equation (2.1) via OLS or WLS. For the latter, the observation weight for stock i in month t corresponds to the market value of stock i at the end of month $t-1$, which we denote by $w_{i,t}$.[5]

In the next step, we use the fitted cross-sectional regression in Equation (2.1) to construct month-$(t+1)$ return forecasts for each stock based on each characteristic:

$$\hat{r}^{(j)}_{i,t+1|t} = \hat{a}_{j,t} + \hat{b}_{j,t} z_{i,j,t} \quad \text{for } i = 1, \ldots, I_{t+1};\ j = 1, \ldots, J_t, \tag{2.2}$$

where $\hat{a}_{j,t}$ and $\hat{b}_{j,t}$ are the OLS or WLS estimates of $a_{j,t}$ and $b_{j,t}$, respectively, in Equation (2.1). The final step pools the return forecasts based on the individual characteristics to form a combination forecast of $r_{i,t+1}$. We compute a simple combination forecast of $r_{i,t+1}$ by taking the arithmetic mean of the individual forecasts:[6]

$$\hat{r}^{\text{Mean}}_{i,t+1|t} = \frac{1}{J_t} \sum_{j=1}^{J_t} \hat{r}^{(j)}_{i,t+1|t} \quad \text{for } i = 1, \ldots, I_{t+1}. \tag{2.3}$$

In addition to the mean combination forecast, we compute a trimmed mean combination forecast, which omits the 5% smallest and 5% largest values of $\hat{r}^{(j)}_{i,t+1|t}$ for $j = 1, \ldots, J_t$ before taking the average in Equation (2.3).

GHZ compute cross-sectional return forecasts for month $t+1$ using a conventional two-step procedure.[7] The first step entails OLS or WLS estimation of the following cross-sectional

[5] We scale the weights such that $\sum_{i=1}^{I_t} w_{i,t} = I_t$.

[6] It is typically the case that $J_{t-1} = J_t$. When a characteristic is unavailable in month $t-1$ but becomes available in month t ($J_{t-1} < J_t$), we obviously cannot estimate Equation (2.1) for the missing characteristic to obtain $\hat{a}_{j,t}$ and $\hat{b}_{j,t}$ in Equation (2.2); in this case, we simply do not include the forecasts based on the missing characteristic when computing the month-$(t+1)$ combination forecast in Equation (2.3). Similarly, when a characteristic is available in month $t-1$ but not month t ($J_{t-1} > J_t$), we obviously do not have the necessary characteristic observations for Equation (2.2), and we again exclude them when computing Equation (2.3).

[7] Also see Haugen and Baker (1996), Hanna and Ready (2005), and Lewellen (2015).

multiple regression for month t:

$$r_{i,t} = a_t^{\mathrm{MR}} + \sum_{j=1}^{J_{t-1}} b_{j,t}^{\mathrm{MR}} z_{i,j,t-1} + \varepsilon_{i,t} \quad \text{for } i = 1, \ldots, I_t. \tag{2.4}$$

Denoting the OLS or WLS estimates of a_t^{MR} and $b_{j,t}^{\mathrm{MR}}$ for $j = 1, \ldots, J_{t-1}$ in Equation (2.4) by $\hat{a}_t^{\mathrm{MR}}$ and $\hat{b}_{j,t}^{\mathrm{MR}}$, respectively, the second step computes month-$(t+1)$ return forecasts for each stock based on rolling 120-month averages of the estimated coefficients in Equation (2.4):

$$\hat{r}_{i,t+1|t}^{\mathrm{MR}} = \bar{a}_t^{\mathrm{MR}} + \sum_{j=1}^{J_t} \bar{b}_{j,t}^{\mathrm{MR}} z_{i,j,t} \quad \text{for } i = 1, \ldots, I_{t+1}, \tag{2.5}$$

where

$$\bar{a}_t^{\mathrm{MR}} = \frac{1}{120} \sum_{s=0}^{119} \hat{a}_{t-s}^{\mathrm{MR}}, \tag{2.6}$$

$$\bar{b}_{j,t}^{\mathrm{MR}} = \frac{1}{120} \sum_{s=0}^{119} \hat{b}_{j,t-s}^{\mathrm{MR}} \quad \text{for } j = 1, \ldots, J_t. \tag{2.7}$$

Observe that both the forecast combination and conventional approaches only utilize data available at the time of forecast formation. The approaches thus generate *out-of-sample* cross-sectional return forecasts, mimicking the situation of an investor in real time.

The first step of the conventional approach entails OLS or WLS estimation of a high-dimensional regression model. This approach provides a natural baseline. Especially when a regression includes a large number of predictor variables, however, conventional estimation runs the risk of overfitting the model to the estimation sample. Intuitively, by maximizing the fit of a model over the estimation sample, OLS or WLS estimation potentially reads too much into the estimation sample, which can detract from the fitted model's out-of-sample predictive ability. Overfitting risks are magnified when the DGP is characterized by a large noise component and structural instability.

How does forecast combination guard against overfitting? As discussed in Rapach, Strauss, and Zhou (2010), the mean combination forecast can be interpreted as a shrinkage forecast. Relative to a forecast based on OLS or WLS estimation of the high-dimensional regression in Equation (2.4), the mean combination forecast makes two adjustments: (i) it replaces the multiple regression slope coefficient estimates with their univariate counterparts; (ii) it shrinks the return forecasts to the cross-sectional mean. Using the Frisch-Waugh-Lovell (FWL) theorem, these adjustments become evident by substituting Equation (2.2) into Equation (2.3):

$$\hat{r}_{i,t+1|t}^{\text{Mean}} = \bar{r}_t + \frac{1}{J_t} \sum_{j=1}^{J_t} \hat{b}_{j,t}(z_{i,j,t} - \bar{z}_{j,t}) \text{ for } i = 1, \ldots, I_{t+1}, \tag{2.8}$$

where

$$\bar{r}_t = \frac{1}{I_t} \sum_{i=1}^{I_t} w_{i,t} r_{i,t}, \tag{2.9}$$

$$\bar{z}_{j,t} = \frac{1}{I_t} \sum_{i=1}^{I_t} w_{i,t} z_{i,j,t-1}, \tag{2.10}$$

and $w_{i,t} = 1$ for $i = 1, \ldots, I_t$ for OLS estimation of Equation (2.1). By replacing the multiple regression slope coefficient estimates with their univariate counterparts, the combination approach eliminates the role of multicollinearity in generating imprecise coefficient estimates. Although the univariate estimates are potentially biased, the gain in estimation precision can be worthwhile in light of the bias-efficiency tradeoff. Shrinking the return forecasts to the cross-sectional mean further guards against overfitting. Intuitively, shrinking the slope coefficients towards zero in Equation (2.8) helps to prevent a researcher from reading too much into the estimation sample when generating the out-of-sample forecast.[8]

Incorporating insights from Diebold and Shin (forthcoming), we also construct combination forecasts that use the LASSO and elastic net in an effort to refine the forecasts.

[8]As we discuss in Section 2.3, we are concerned with forecasting cross-sectional returns relative to the cross-sectional mean return and not with forecasting the cross-sectional mean return itself.

The LASSO and elastic net are powerful machine learning tools for variable selection in high-dimensional settings (e.g., Zhang and Huang 2008; Bickel, Ritov, and Tsybakov 2009; Meinshausen and Yu 2009). We harness these tools to select the most relevant individual forecasts to include in the pooled forecast. As an added benefit, by tracking the individual forecasts selected by the LASSO and elastic net each month, we can analyze the relevance of individual firm characteristics for predicting cross-sectional returns over time.

We compute a combination forecast based on the LASSO as follows. Along the lines of Granger and Ramanathan (1984), consider the following multiple regression that relates realized cross-sectional returns to the return forecasts based on the individual characteristics in Equation (2.2):

$$r_{i,t} = a_t^{\mathrm{GR}} + \sum_{j=1}^{J_{t-1}} b_{j,t}^{\mathrm{GR}} \hat{r}_{i,t|t-1}^{(j)} + \varepsilon_{i,t} \quad \text{for } i = 1, \ldots, I_t. \tag{2.11}$$

We estimate Equation (2.11) via the weighted LASSO:

$$\underset{\tilde{a}_t^{\mathrm{GR}} \in \mathbb{R},\, \tilde{\mathbf{b}}_t^{\mathrm{GR}} \in \mathbb{R}_{\geq 0}^{J_{t-1}}}{\arg\min} \left(\frac{1}{2I_t} \sum_{i=1}^{I_t} w_{i,t} \tilde{\varepsilon}_{i,t}^2 + \lambda_t \left\| \tilde{\mathbf{b}}_t^{\mathrm{GR}} \right\|_1 \right), \tag{2.12}$$

where

$$\tilde{\mathbf{b}}_t^{\mathrm{GR}} = \begin{bmatrix} \tilde{b}_{1,t}^{\mathrm{GR}} & \ldots & \tilde{b}_{J_{t-1},t}^{\mathrm{GR}} \end{bmatrix}', \tag{2.13}$$

$$\tilde{\varepsilon}_{i,t} = r_{i,t} - \left(\tilde{a}_t^{\mathrm{GR}} + \sum_{j=1}^{J_{t-1}} \tilde{b}_{j,t}^{\mathrm{GR}} \hat{r}_{i,t|t-1}^{(j)} \right), \tag{2.14}$$

$$\left\| \tilde{\mathbf{b}}_t^{\mathrm{GR}} \right\|_1 = \sum_{j=1}^{J_{t-1}} \left| \tilde{b}_{j,t}^{\mathrm{GR}} \right|, \tag{2.15}$$

and $\lambda_t \geq 0$ is a regularization parameter.[9] Setting $w_{i,t} = 1$ for $i = 1, \ldots, I_t$ yields the ordinary LASSO. When $\lambda_t = 0$, Equation (2.12) reduces to the familiar WLS objective

[9] Note that Equation (2.12) imposes the restriction that $\tilde{b}_{j,t}^{\mathrm{GR}} \geq 0$ for $j = 1, \ldots, J_{t-1}$, so that we preclude the forecasted returns from being negatively related to the realized returns in the fitted model.

function (OLS objective function when $w_{i,t} = 1$ for $i = 1, \ldots, I_t$). The presence of λ_t in Equation (2.12) shrinks the slope estimates. The LASSO's ℓ_1 penalty term allows for shrinkage to zero (for a sufficiently large λ_t), so that it performs variable selection.[10] The LASSO combination forecast is simply the average of the individual forecasts selected by the LASSO in Equation (2.12). Because we use Equation (2.12), which is based on information available in month t, to select the individual forecasts to include in the LASSO combination forecast for month $t+1$, the LASSO combination forecast does not entail look-ahead bias. The LASSO combination forecast retains the shrinkage properties of the mean combination forecast; at the same time, it fine-tunes the combination forecast by focusing on the individual forecasts that are the most relevant for explaining cross-sectional returns in the prior month.

When a group of variables is strongly correlated, the LASSO has a tendency to select one of the variables from the group and not the others. To help ensure that we do not neglect relevant characteristics, we also estimate Equation (2.11) via the weighted elastic net:

$$\underset{\tilde{a}_t^{\mathrm{GR}} \in \mathbb{R},\, \tilde{\mathbf{b}}_t^{\mathrm{GR}} \in \mathbb{R}_{\geq 0}^{J_{t-1}}}{\arg\min} \left\{ \frac{1}{2I_t} \sum_{i=1}^{I_t} w_{i,t} \tilde{\varepsilon}_{i,t}^2 + \lambda_t \left[0.5(1-\gamma) \left\| \tilde{\mathbf{b}}_t^{\mathrm{GR}} \right\|_2^2 + \gamma \left\| \tilde{\mathbf{b}}_t^{\mathrm{GR}} \right\|_1 \right] \right\}, \tag{2.16}$$

where

$$\left\| \tilde{\mathbf{b}}_t^{\mathrm{GR}} \right\|_2 = \left[\sum_{j=1}^{J_{t-1}} \left(\tilde{b}_{j,t}^{\mathrm{GR}} \right)^2 \right]^{0.5} \tag{2.17}$$

and $0 \leq \gamma \leq 1$ is a blending parameter between ridge regression ($\gamma = 0$) and the LASSO ($\gamma = 1$). When $\gamma = 0.5$, the elastic net tends to select strongly correlated predictors as a group. Using $\gamma = 0.5$, we compute an elastic net combination forecast in the same manner

[10]Based on simulation evidence in Flynn, Hurvich, and Simonoff (2013) and Taddy (2017), we use the Hurvich and Tsai (1989) corrected version of the Akaike information criterion (Akaike 1973, AIC) to choose λ_t in Equation (2.12).

as the LASSO combination forecast, except that we use the elastic net in Equation (2.16) in lieu of the LASSO to select the individual forecasts to include in the pooled forecast.[11]

With respect to the treatment of small-cap stocks, we proceed analogously to GHZ and consider three different cases. The first is "Value Weighted," where we estimate Equations (2.1) and (2.4) via WLS and use the weighted LASSO or elastic net with observations weighted by previous month-end market values. The second case is "Equal Weighted excl. Microcap," where we estimate Equations (2.1) and (2.4) via OLS and use the ordinary LASSO or elastic net after excluding stocks with market values below the NYSE 20th percentile. The final case is "Equal Weighted," which relies on OLS estimation of Equations (2.1) and (2.4) and uses the ordinary LASSO or elastic net based on data for all available stocks.

2.3. Predictive Slope

Following Lewellen (2015), we analyze the cross-sectional return forecasts by estimating FM regressions that relate realized to forecasted returns.[12] Specifically, for each month t, we estimate the following cross-sectional univariate regression:

$$r_{i,t} = a_t^{\text{FM}} + b_t^{\text{FM}} \hat{r}_{i,t|t-1} + \varepsilon_{i,t} \quad \text{for } i = 1, \ldots, I_t; \; t = 1, \ldots, T, \tag{2.18}$$

where $\hat{r}_{i,t|t-1}$ generically denotes a forecast of $r_{i,t}$ and T is the total number of months for which we compute cross-sectional return forecasts. We then compute the time-series average of the monthly slope coefficient estimates in Equation (2.18):

$$\hat{b}^{\text{FM}} = \frac{1}{T} \sum_{t=1}^{T} \hat{b}_t^{\text{FM}}, \tag{2.19}$$

[11] We again use the corrected AIC to choose λ_t in Equation (2.16). The results in Section 3 are similar for combination forecasts based on adaptive versions of the LASSO and elastic net (Zou 2006; Zou and Zhang 2009).

[12] GHZ do not directly analyze their conventional out-of-sample cross-sectional return forecasts using the FM regression approaches in this and the next subsections.

where $\hat{b}_t^{\text{FM}}$ is the OLS or WLS estimate of b_t^{FM} in Equation (2.18). To account for potential autocorrelation in the $\hat{b}_t^{\text{FM}}$ estimates, the t-statistic for $\hat{b}^{\text{FM}}$ is based on a Newey and West (1987) standard error computed using a lag of twelve months.[13] For the Value Weighted case, we estimate Equation (2.18) via WLS with previous month-end market values serving as the observation weights; for the Equal Weighted excl. Microcap and Equal Weighted cases, we estimate Equation (2.18) via OLS using the relevant population of stocks. We use a one-sided, upper-tailed test to assess the significance of $\hat{b}^{\text{FM}}$, as we treat cross-sectional return forecasts as economically reasonable when they are positively related to realized returns.

The size of $\hat{b}^{\text{FM}}$ reveals how well the cross-sectional return forecasts capture the cross-sectional dispersion in expected returns (Lewellen 2015) and provides a direct measure of overfitting. Similarly to a Mincer and Zarnowitz (1969) regression, $\hat{b}^{\text{FM}} = 1$ indicates that the cross-sectional return forecasts are unbiased: a unit increase in the forecasted return corresponds to a unit increase in the realized return on average.[14] If $\hat{b}^{\text{FM}} < 1$, then the cross-sectional forecasts overstate the cross-sectional dispersion in expected returns: a unit increase in the forecasted return corresponds to a less-than-unit increase in the realized return on average, which immediately indicates overfitting in the construction of the forecasts. When $\hat{b}^{\text{FM}} > 1$, the cross-sectional return forecasts are generally conservative, in the sense that a unit increase in the forecasted return is associated with a greater-than-unit increase in the realized return on average. Because of the inherently large unpredictable component in monthly stock returns, a forecast characterized by $\hat{b}^{\text{FM}} > 1$ is likely to perform more reliably than a forecast characterized by $\hat{b}^{\text{FM}} < 1$.

Lewellen (2015, footnote 3) points out that the slope coefficient in Equation (2.18) is related to the difference in cross-sectional MSFEs between a naive forecast that ignores any information in firm characteristics and a competing forecast. We extend his result from

[13]The results are robust to the number of lags used to compute the Newey and West (1987) standard error.

[14]Due to the presence of the intercept term in Equation (2.18), the FWL theorem implies that the bias is measured in terms of deviations from the cross-sectional mean return, so that we are not concerned with forecasting the cross-sectional mean return itself.

equal to general weighting of cross-sectional return observations. Denote the competing return forecast by $\hat{r}_{i,t|t-1}$, with its forecast error given by

$$\hat{e}_{i,t|t-1} = r_{i,t} - \hat{r}_{i,t|t-1} \text{ for } i = 1, \ldots, I_t. \tag{2.20}$$

Because we are concerned with forecasting cross-sectional returns relative to the cross-sectional mean return (and not with forecasting the cross-sectional mean return per se), we demean the forecast error in Equation (2.20):

$$\hat{u}_{i,t|t-1} = \underbrace{\left(r_{i,t} - \hat{r}_{i,t|t-1}\right)}_{\hat{e}_{i,t|t-1}} - \underbrace{\left(\bar{r}_t - \bar{\hat{r}}_{t|t-1}\right)}_{\bar{\hat{e}}_{t|t-1}} = (r_{i,t} - \bar{r}_t) - \left(\hat{r}_{i,t|t-1} - \bar{\hat{r}}_{t|t-1}\right) \tag{2.21}$$

for $i = 1, \ldots, I_t$, where a bar again indicates a variable's cross-sectional value-weighted average (i.e., $\bar{x}_t = (1/I_t)\sum_{i=1}^{I_t} w_{i,t}x_{i,t}$).

The naive forecast, which ignores any information in the characteristics, is simply the cross-sectional value-weighted average return for the previous month:

$$\hat{r}_{i,t|t-1}^{\text{Naive}} = \bar{r}_{t-1} \text{ for } i = 1, \ldots, I_t. \tag{2.22}$$

Its demeaned forecast error is given by

$$\hat{u}_{i,t|t-1}^{\text{Naive}} = \underbrace{(r_{i,t} - \bar{r}_{t-1})}_{\hat{e}_{i,t|t-1}^{\text{Naive}}} - \underbrace{(\bar{r}_t - \bar{r}_{t-1})}_{\bar{\hat{e}}_{t|t-1}^{\text{Naive}}} = r_{i,t} - \bar{r}_t \text{ for } i = 1, \ldots, I_t. \tag{2.23}$$

As shown in Appendix A, the month-t difference in cross-sectional value-weighted MSFEs between the naive and competing forecasts can be expressed as

$$\frac{1}{I_t}\sum_{i=1}^{I_t} w_{i,t}\left[\left(\hat{u}_{i,t|t-1}^{\text{Naive}}\right)^2 - \hat{u}_{i,t|t-1}^2\right] = \left(2\hat{\delta}_t - 1\right)\hat{\sigma}_{\hat{r},t}^2, \tag{2.24}$$

where

$$\hat{\delta}_t = \frac{\sum_{i=1}^{I_t} w_{i,t}(r_{i,t} - \bar{r}_t)\big(\hat{r}_{i,t|t-1} - \bar{\hat{r}}_{t|t-1}\big)}{\sum_{i=1}^{I_t} w_{i,t}\big(\hat{r}_{i,t|t-1} - \bar{\hat{r}}_{t|t-1}\big)^2} \tag{2.25}$$

and

$$\hat{\sigma}^2_{\hat{r},t} = \frac{1}{I_t}\sum_{i=1}^{I_t} w_{i,t}\big(\hat{r}_{i,t|t-1} - \bar{\hat{r}}_{t|t-1}\big)^2 \tag{2.26}$$

is the cross-sectional value-weighted variance of $\hat{r}_{i,t|t-1}$ for $i = 1, \ldots, I_t$. By the FWL theorem, $\hat{\delta}_t$ in Equation (2.25) is identical to the WLS estimate of b_t^{FM} in Equation (2.18); for equal observation weights ($w_{i,t} = 1$ for $i = 1, \ldots, I_t$), $\hat{\delta}_t$ is identical to the OLS estimate of b_t^{FM} in Equation (2.18). When $\hat{b}_t^{\mathrm{FM}} > 0.5$, Equation (2.24) implies that the competing forecast is more accurate than the naive forecast in terms of the month-t cross-sectional value-weighted MSFE. Accordingly, $\hat{b}^{\mathrm{FM}} > 0.5$ in Equation (2.19) means that the cross-sectional value-weighted MSFE for the competing forecast is lower on average than that for the naive forecast over the evaluation period.

To understand the importance of the goodness-of-fit measure in Equation (2.18), as shown in Appendix A, we can express the month-t cross-sectional value-weighted MSFE for a generic forecast $\hat{r}_{i,t|t-1}$ as

$$\underbrace{\frac{1}{I_t}\sum_{i=1}^{I_t} w_{i,t}\hat{u}^2_{i,t|t-1}}_{\mathrm{MSFE}_t} = \mathrm{bias}_t^2\hat{\sigma}^2_{\hat{r},t} + \big(1 - R_t^2\big)\hat{\sigma}^2_{r,t}, \tag{2.27}$$

where $\mathrm{bias}_t = \hat{b}_t^{\mathrm{FM}} - 1$ is the forecast bias, R_t^2 is the R^2 statistic for Equation (2.18), and

$$\hat{\sigma}^2_{r,t} = \frac{1}{I_t}\sum_{i=1}^{I_t} w_{i,t}(r_{i,t} - \bar{r}_t)^2 \tag{2.28}$$

is the cross-sectional value-weighted variance of $r_{i,t}$ for $i = 1, \ldots, I_t$. Equation (2.27) shows that a decrease (increase) in $|\mathrm{bias}_t|$ (R_t^2) leads to a decrease in MSFE_t. As discussed above, overfitting is associated with $\hat{b}_t^{\mathrm{FM}} < 1$, so that it creates a bias ($\mathrm{bias}_t < 0$) that increases MSFE_t in Equation (2.27); at the same time, overfitting adversely affects the ability of the forecasts to track realized returns in Equation (2.18), corresponding to a small R_t^2 that further inflates MSFE_t in Equation (2.27). Although a conservative forecast also produces a bias ($\hat{b}_t^{\text{FM}} > 1$ and $\text{bias}_t > 0$), by guarding against overfitting, it is better able to track realized returns; the corresponding increase in R_t^2 works to offset the effects of the bias by decreasing MSFE_t in Equation (2.27).

2.4. Forecast Encompassing

We can also interpret Equation (2.18) from the perspective of forecast encompassing. Consider a composite forecast comprised of the naive and competing forecasts:

$$\hat{r}^*_{i,t|t-1} = (1-\theta_t)\bar{r}_{t-1} + \theta_t \hat{r}_{i,t|t-1} = \bar{r}_{t-1} + \theta_t\big(\hat{r}_{i,t|t-1} - \bar{r}_{t-1}\big) \quad \text{for } i = 1, \ldots, I_t. \tag{2.29}$$

The demeaned forecast error for the composite forecast is given by

$$\hat{u}^*_{i,t|t-1} = \underbrace{\big(r_{i,t} - \hat{r}^*_{i,t|t-1}\big)}_{\hat{e}^*_{i,t|t-1}} - \underbrace{\big(\bar{r}_t - \bar{\hat{r}}^*_{t|t-1}\big)}_{\bar{\hat{e}}^*_{t|t-1}} = (r_{i,t} - \bar{r}_t) - \big(\hat{r}^*_{i,t|t-1} - \bar{\hat{r}}^*_{t|t-1}\big) \tag{2.30}$$

for $i = 1, \ldots, I_t$, while its cross-sectional value-weighted MSFE is given by

$$\frac{1}{I_t}\sum_{i=1}^{I_t} w_{i,t}\big(\hat{u}^*_{i,t|t-1}\big)^2. \tag{2.31}$$

Again generalizing Lewellen (2015, footnote 3), Appendix A shows that the value of θ_t that minimizes Equation (2.31) is given by $\hat{\delta}_t$ in Equation (2.25). By the FWL theorem, the WLS estimate of b_t^{FM} in Equation (2.18) thus coincides with the value of θ_t that minimizes the month-t cross-sectional value-weighted MSFE for the composite forecast; for equal observation weights ($w_{i,t} = 1$ for $i = 1, \ldots, I_t$), the OLS estimate of b_t^{FM} in Equation (2.18) represents the value of θ_t that minimizes Equation (2.31). Furthermore, the $\hat{b}^{\text{FM}}$ estimate in Equation (2.19) can be viewed as the average optimal value for θ_t over the forecast evaluation period.

When $\theta_t \leq 0$ in Equation (2.29), the naive forecast is said to encompass the competing forecast: the competing forecast does not contain information useful for predicting returns beyond the information already contained in the naive forecast; when $\theta_t > 0$, the naive forecast does not encompass the competing forecast: the competing forecast does provide useful information beyond that already found in the naive forecast. Testing the null hypothesis that $b^{\mathrm{FM}} \leq 0$ against the alternative that $b^{\mathrm{FM}} > 0$ in Equation (2.19) thus represents a test of whether the naive forecast encompasses the competing forecast on average over the forecast evaluation period.

Next, consider two competing forecasts, $\hat{r}^{\mathrm{A}}_{i,t|t-1}$ and $\hat{r}^{\mathrm{B}}_{i,t|t-1}$, each of which potentially incorporates information from the firm characteristics. The demeaned errors for the competing forecasts are given by

$$\hat{u}^{\mathrm{A}}_{i,t|t-1} = \underbrace{\left(r_{i,t} - \hat{r}^{\mathrm{A}}_{i,t|t-1}\right)}_{\hat{e}^{\mathrm{A}}_{i,t|t-1}} - \underbrace{\left(\bar{r}_t - \bar{\hat{r}}^{\mathrm{A}}_{t|t-1}\right)}_{\bar{\hat{e}}^{\mathrm{A}}_{t|t-1}} = (r_{i,t} - \bar{r}_t) - \left(\hat{r}^{\mathrm{A}}_{i,t|t-1} - \bar{\hat{r}}^{\mathrm{A}}_{t|t-1}\right), \tag{2.32}$$

$$\hat{u}^{\mathrm{B}}_{i,t|t-1} = \underbrace{\left(r_{i,t} - \hat{r}^{\mathrm{B}}_{i,t|t-1}\right)}_{\hat{e}^{\mathrm{B}}_{i,t|t-1}} - \underbrace{\left(\bar{r}_t - \bar{\hat{r}}^{\mathrm{B}}_{t|t-1}\right)}_{\bar{\hat{e}}^{\mathrm{B}}_{t|t-1}} = (r_{i,t} - \bar{r}_t) - \left(\hat{r}^{\mathrm{B}}_{i,t|t-1} - \bar{\hat{r}}^{\mathrm{B}}_{t|t-1}\right) \tag{2.33}$$

for $i = 1, \dots, I_t$. Analogously to Equation (2.29), we define a composite forecast comprised of the two competing forecasts:

$$\hat{r}^{\dagger}_{i,t|t-1} = (1 - \eta_t)\hat{r}^{\mathrm{A}}_{i,t|t-1} + \eta_t \hat{r}^{\mathrm{B}}_{i,t|t-1} = \hat{r}^{\mathrm{A}}_{i,t|t-1} + \eta_t\left(\hat{r}^{\mathrm{B}}_{i,t|t-1} - \hat{r}^{\mathrm{A}}_{i,t|t-1}\right) \text{ for } i = 1, \dots, I_t. \tag{2.34}$$

Its corresponding cross-sectional value-weighted MSFE is given by

$$\frac{1}{I_t} \sum_{i=1}^{I_t} w_{i,t} \left(\hat{u}^{\dagger}_{i,t|t-1}\right)^2, \tag{2.35}$$

where

$$\hat{u}^{\dagger}_{i,t|t-1} = \underbrace{\left(r_{i,t} - \hat{r}^{\dagger}_{i,t|t-1}\right)}_{\hat{e}^{\dagger}_{i,t|t-1}} - \underbrace{\left(\bar{r}_t - \bar{\hat{r}}^{\dagger}_{t|t-1}\right)}_{\bar{\hat{e}}^{\dagger}_{t|t-1}} = (r_{i,t} - \bar{r}_t) - \left(\hat{r}^{\dagger}_{i,t|t-1} - \bar{\hat{r}}^{\dagger}_{t|t-1}\right) \tag{2.36}$$

for $i = 1, \ldots, I_t$. Appendix A establishes that the value of η_t that minimizes the cross-sectional value-weighted MSFE in Equation (2.35) is identical to the WLS estimate of $b^{\dagger}_t$ in the following cross-sectional univariate regression:

$$\hat{e}^{\mathrm{A}}_{i,t|t-1} = a^{\dagger}_t + b^{\dagger}_t\left(\hat{e}^{\mathrm{A}}_{i,t|t-1} - \hat{e}^{\mathrm{B}}_{i,t|t-1}\right) + \varepsilon_{i,t} \text{ for } i = 1, \ldots, I_t;\ t = 1, \ldots, T. \tag{2.37}$$

For equal observation weights ($w_{i,t} = 1$ for $i = 1, \ldots, I_t$), the optimal value of η_t matches the OLS estimate of $b^{\dagger}_t$ in Equation (2.37). As in an FM regression, we take the time-series average of the monthly slope coefficient estimates in Equation (2.37):

$$\hat{b}^{\dagger} = \frac{1}{T}\sum_{t=1}^{T} \hat{b}^{\dagger}_t, \tag{2.38}$$

where $\hat{b}^{\dagger}_t$ is the OLS or WLS estimate of $b^{\dagger}_t$ in Equation (2.37).

Equation (2.38) allows us to conveniently compare the information content of two competing cross-sectional return forecasts. When $b^{\dagger} \leq 0$, A encompasses B on average over the forecast evaluation period, as B does not contribute useful information to the composite forecast beyond that already found in A. Alternatively, when $b^{\dagger} > 0$, A does not encompass B, so that B does provide useful information beyond that contained in A. Conversely, if $1 - b^{\dagger} \leq 0$, then B encompasses A on average; if $1 - b^{\dagger} > 0$, then B does not encompass A. If A does not encompass B and B encompasses A, then B "dominates" A from the standpoint of forecast encompassing.[15] In Section 3.2, we use Equation (2.38) to compare the information content of the GHZ and combination cross-sectional return forecasts.

[15] Similarly, if A encompasses B and B does not encompass A, then A dominates B.

2.5. Spread Portfolios

Following Lewellen (2015) and Green, Hand, and Zhang (2017), we also use cross-sectional return forecasts to construct spread portfolios. To construct a spread portfolio based on a particular forecasting strategy, at the beginning of each month, we first sort stocks into deciles according to the their forecasted returns for the month. We then form a zero-investment spread portfolio that goes long (short) the highest (lowest) decile portfolio. For the Value Weighted case, deciles are defined using return forecasts for NYSE stocks, and the decile portfolios are value weighted. For the Equal Weighted excl. Microcap case, deciles are formed using return forecasts for stocks with market values above the NYSE 20% percentile, and the decile portfolios are equal weighted. Finally, for the Equal Weighted case, deciles are defined using return forecasts for NYSE stocks, and the decile portfolios are equal weighted.

3. Empirical Results

3.1. Predictive Slope

As a starting point, Table 2 reports FM regression results for the cross-sectional return forecasts in Equation (2.2) based on the individual characteristics for the 1990:01 to 2017:12 out-of-sample period.[16] Beginning with the Value Weighted case in the second through fourth columns, of the 94 characteristics, only eight and five are significant at the 10% and 5% levels, respectively. Furthermore, a substantial number (42) of the FM slope coefficient estimates in the second column have the "wrong" (i.e., negative) sign. For both the Equal Weighted excl. Microcap and Equal Weighted cases in the fifth through seventh and eighth through tenth columns, respectively, 17 and eight of the characteristics are significant at the 10% and 5% levels, respectively. However, there is little overlap in the significant characteristics across the two cases, and many of the FM slope coefficients are again negative in the fifth and eighth columns. Overall, Table 2 indicates that firm characteristics taken individually have

[16]The 1990:01 starting date for the out-of-sample period accommodates the 120-month moving averages used to construct the conventional forecast in Equation (2.5).

limited out-of-sample predictive ability with respect to cross-sectional returns, especially for the Value Weighed case.

Table 3 reports FM regression results for the conventional and combination forecasts. Panel A presents results for the 1990:01 to 2017:12 period. Starting with the Value Weighted case in the second through fourth columns, the FM slope coefficient estimate for the conventional forecast is 0.31, which is significant at the 1% level. Although it is greater than zero in terms of statistical significance, the FM slope estimate is well below unity in magnitude—in fact, it is significantly below unity based on its (unreported) standard error. As discussed in Section 2.3, the relatively small magnitude of the FM slope estimate means that the conventional forecast substantially overstates the cross-sectional dispersion in expected returns, a direct indication of overfitting. In addition, because the FM slope estimate is below 0.5, the conventional forecast fails to outperform the naive forecast in terms of average cross-sectional value-weighted MSFE for the 1990:01 to 2017:12 period.

The FM slope coefficients for all of the combination forecasts are also significant at the 1% level in Panel A of Table 3. Unlike the conventional forecast, however, the coefficient estimates are all above unity, but not significantly so based on their (unreported) standard errors. The combination forecasts thus better measure the cross-sectional dispersion in expected returns, highlighting the efficacy of forecast combination for guarding against overfitting. Because the FM slope coefficient estimates for the combination forecasts are all above 0.5, the combination forecasts deliver a lower cross-sectional value-weighted MSFE on average than the naive forecast for the 1990:01 to 2017:12 period. The R^2 statistics for the combination forecasts are more than twice as large as that for the conventional forecast, providing additional support for the effectiveness of forecast combination.

As shown in Panels B and C of Table 3, the conventional forecast exhibits a notable decline in predictive ability in the context of the FM regressions for the cross-sectional return forecasts. In particular, the FM slope coefficient estimate for the conventional forecast falls from 0.63 (significant at the 1% level) for the 1990:01 to 2002:12 period in Panel B to only 0.05 (insignificant at conventional levels) for the 2004:01 to 2017:12 period in Panel C. Overfitting thus becomes more acute after 2003, essentially rendering the conventional return forecast unrelated to the realized return after 2003.[17]

In contrast to the conventional forecast, the predictive ability of the combination forecasts does not disappear after 2003 for the Value Weighted case. The estimated FM slope coefficient for the mean (trimmed mean) combination forecast goes from 1.47 (1.97) before 2003 to 1.94 (2.92) after 2003, where the former (latter) is significant at the 10% (1%) level. The FM slope coefficient estimates for the LASSO and elastic net combination forecasts are quite stable before and after 2003: the FM slope estimates for the LASSO (elastic net) combination forecast are 1.73 and 1.55 (1.76 and 1.58) for the pre- and post-2003 periods, respectively, both of which are significant at the 1% level.[18] The FM slope estimates for the combination forecasts are all well above 0.5 during both subperiods, so that the combination forecasts produce a lower cross-sectional value-weighted MSFE on average than the naive forecast both before and after 2003.[19]

We further investigate the predictive ability of the conventional and elastic net combination forecasts over time in Figure 1.[20] The figure depicts ten-year rolling averages of the $\hat{b}_t^{\mathrm{FM}}$ estimates in Equation (2.18) for the Value Weighted case. The figure includes two-sided 90% confidence intervals, as well as lower bounds for one-sided, upper-tailed 90% confidence intervals. For the conventional forecast in Panel A, apart from ten-year periods ending between approximately 2002 to 2004, the FM slope estimate is always significantly less than unity according to the two-sided confidence intervals. Moreover, the slope estimate exhibits a consistent and marked decline for ten-year periods ending in 2004 through 2010, where

[17]Because the slope coefficient estimate for the conventional forecast is below 0.5 for the post-2003 period, the forecast becomes less accurate in terms of average cross-sectional value-weighted MSFE than the naive forecast after 2003. Furthermore, we cannot cannot reject the null hypothesis that the naive forecast encompasses the conventional forecast for the post-2003 period, so that the conventional forecast does not offer significant information gains vis-à-vis the naive forecast after 2003.

[18]Based on their (unreported) standard errors, the FM slope estimates for the LASSO and elastic net combination forecasts are insignificantly different from unity for the pre- and post-2003 periods.

[19]In addition, the naive forecast does not encompass any of the combination forecasts before or after 2003.

[20]To conserve space, we focus on the elastic net combination forecast in Figure 1; the results are qualitatively similar for the other combination forecasts.

it drops from around 0.6 to close to zero; thereafter, the slope estimate remains near zero, and it is almost always insignificantly different from zero based on the lower bounds for the one-sided, upper-tailed confidence intervals. Furthermore, the slope estimate falls below 0.5 for ten-year periods ending in 2007 through 2017, so that the conventional forecast fails to outperform the naive forecast in terms of cross-sectional value-weighted MSFE for these ten-year periods.

The FM slope coefficient estimate for the elastic net combination forecast is substantially more stable in Panel B—typically lying between one and two—and it is rarely significantly different from unity according to the two-sided confidence intervals. Based on the lower bounds for the one-sided, upper-tailed confidence intervals, the slope estimate is significant for nearly every ten-year period. In addition, because the slope estimate is always greater than 0.5, the elastic net combination forecast delivers a lower cross-sectional value-weighted MSFE than the naive forecast for every ten-year period in Panel B of Figure 1.

The results for the Equal Weighted excl. Microcap case in the fifth through seventh columns of Table 3 are similar to those for the Value Weighted case. The FM slope coefficient estimate for the conventional forecast is 0.35 for the 1990:01 to 2017:12 period in Panel A. Although the FM slope estimate is significant at the 1% level, its relatively small value again indicates that the conventional forecast overstates the cross-sectional dispersion in expected returns. Because the FM slope estimate is less than 0.5, the conventional forecast is also less accurate than the naive forecast in terms of average cross-sectional MSFE for the 1990:01 to 2017:12 period. Turning to the results for the combination forecasts in Panel A, the FM slope coefficient estimates are all significant at the 1% level and greater than unity.[21] Unlike the conventional forecast, the combination forecasts thus deliver a lower cross-sectional MSFE on average than the naive forecast. The R^2 statistics for the combination forecasts in Panel A are approximately three to four times larger than that for the conventional forecast.

[21] The FM slope estimates for the LASSO and elastic net combination forecasts are insignificantly different from unity based on their (unreported) standard errors.

For the Equal Weighted excl. Microcap case, the conventional forecast again experiences a breakdown in predictive ability as we move from Panel B to Panel C in Table 3. The FM slope coefficient estimate for the conventional forecast drops from 0.72 (significant at the 1% level) for the pre-2003 period to 0.04 (insignificant at conventional levels) for the post-2003 period. Based on the FM slope estimates, the conventional forecast outperforms the naive forecast on average in terms of cross-sectional MSFE for the pre-2003 period, but fails to do so for the post-2003 period. In contrast to the conventional forecast, the FM slope estimates for the combination forecasts are all significant and above unity for both the pre- and post-2003 periods. Because the FM slope estimates are greater than 0.5 for both subperiods, the combination forecasts are more accurate than the naive forecast in terms of average cross-sectional MSFE before and after 2003.

The last three columns of Table 3 reports results for the Equal Weighted case, which places much greater weight on small-cap stocks. The FM slope coefficient estimate for the conventional forecast in Panel A is 0.67, which is significant at the 1% level. As the FM slope estimate is above 0.5, the conventional forecast produces a lower cross-sectional MSFE on average than the naive forecast for the 1990:01 to 2017:12 period.[22] All of the FM slope estimates for the combination forecasts are above unity and significant at the 1% level in Panel A, continuing the pattern for the Value Weighted and Equal Weighted excl. Microcap cases. Again continuing the pattern, the R^2 statistics for the combination forecasts are larger than that for the conventional forecast by factors of approximately two to three.

Unlike the Value Weighted and Equal Weighted excl. Microcap cases, the FM slope estimates for the conventional forecast are significant for the Equal Weighted case for both the pre- and post-2003 periods in Panels B and C, respectively, of Table 3. However the FM slope estimate falls from 0.90 before 2003 to 0.48 after 2003, so that the conventional forecast does not provide a lower cross-sectional MSFE on average than the naive forecast after 2003. The FM slope estimates for the combination forecasts are all significant at the 1% level for

[22]Based on its (unreported) standard error, the FM slope estimate is significantly below unity, so that the conventional forecast continues to manifest overfitting.

both subperiods, and they are always above unity. The R^2 statistics for the combination forecasts are approximately two to three times larger than that for the conventional forecast before and after 2003.

3.2. Forecast Encompassing

In this subsection, we use forecast encompassing tests to compare the information content of the conventional and combination forecasts. In terms of the notation in Section 2.4, A (B) represents the conventional (combination) forecast. Table 4 reports estimates of $b^{\dagger}$ and $1 - b^{\dagger}$ in Equation (2.38), which correspond to the coefficients attached to the combination and conventional forecasts, respectively, in an optimal composite forecast comprised of the two competing forecasts. Under the null hypothesis that $b^{\dagger} \leq 0$ $(1 - b^{\dagger} \leq 0)$, the conventional forecast encompasses the combination forecast (combination forecast encompasses the conventional forecast).

Value Weighted results are reported in the second through fifth columns of Table 4. For the 1990:01 to 2017:12 period in Panel A, the $\hat{b}^{\dagger}$ estimates are all significant at the 1% level, so that the conventional forecast does not encompass any of the combination forecasts. At the same time, the $1 - \hat{b}^{\dagger}$ estimates are also all significant at the 1% level, meaning that the combination forecasts do not encompass the conventional forecast. The conventional and combination forecasts thus contain unique information vis-à-vis one another for predicting cross-sectional value-weighted returns. The estimated coefficients for the combination and conventional forecasts in the optimal composite forecast are equal to or near 0.70 and 0.30, respectively.

The results for the Value Weighted case in Panels B and C of Table 4 reveal conspicuous differences in the relative information content of the conventional and combination forecasts before and after 2003. For the pre-2003 period in Panel B, the results are similar to those in Panel A: the conventional forecast does not encompass the combination forecasts, and the combination forecasts do not encompass the conventional forecast. The situation changes markedly for the post-2003 period in Panel C, where the combination forecasts now dominate the conventional forecast with respect to information content. Specifically, the conventional forecast no longer encompasses the combination forecasts, while the combination forecasts encompass the conventional forecast. Indeed, the estimated coefficients attached to the combination and conventional forecasts in the optimal composite forecast are very close to unity and zero, respectively, so that the conventional forecast delivers essentially no information gain after 2003.

Figure 2 provides additional perspective on the time variation in the encompassing test results, focusing on the comparison of the conventional forecast with the elastic net combination forecast. Panels A and B report ten-year rolling averages of the $1 - \hat{b}^{\dagger}$ and $\hat{b}^{\dagger}$ estimates, respectively, in Equation (2.37), along with two-sided 90% confidence intervals and lower bounds for one-sided, upper-tailed 90% confidence intervals.[23] Similarly to Panel A of Figure 1, the coefficient estimate for the conventional forecast evinces a steady and marked decline for ten-year periods ending in 2004 through 2010 in Panel A of Figure 2, and the coefficient estimate is close to zero for ten-year periods ending in 2010 through 2017. According to the lower bounds for the one-sided, upper-tailed confidence intervals, the coefficient estimate is significant for ten-year periods ending in 2000 through 2009, but it becomes insignificant for nearly every ten-year period ending in 2010 through 2017. With the exceptions of a few ten-year periods ending between 2002 through 2004, the coefficient estimate for the elastic net combination forecast is always significant in Panel B, and it becomes very close to unity for ten-year periods ending in 2011 through 2017.

Turning to the Equal Weighted excl. Microcap case in the sixth through ninth columns of Table 4, the results are similar to those for the Value Weighted case. For the 1990:01 to 2017:12 and 1990:01 to 2002:12 periods in Panels A and B, respectively, the conventional and combination forecasts do not encompass each other, so that both forecasts contribute significantly to the optimal composite forecast. For the 2004:01 to 2017:12 period in Panel

[23]By construction, the graphs in Panels A and B of Figure 2 are mirror images of each other.

C, the conventional forecast does not encompass the combination forecasts, while the combination forecasts encompass the conventional forecast. The combination forecasts thus again dominate the combination forecast in terms of information content for the post-2003 period.

Similarly to the results for the Equal Weighted case in Table 3, the conventional forecast displays more consistent predictive ability in the last four columns of Table 4. The conventional and combination forecasts do not encompass each other for the 1990:01 to 2017:12 period in Panel A, and the same finding holds before and after 2003 in Panels B and C, respectively.

In sum, the results in Tables 3 and 4 indicate the following. In contrast to the combination forecasts, conventional forecasts of cross-sectional returns are prone to overfitting, as evinced by the below-unity slope coefficient estimates for the latter in Table 3. The shrinkage entailed by the combination approach enables the combination forecasts to better track realized returns, as demonstrated by the higher R^2 statistics for the combination vis-á-vis the conventional forecasts in Table 3. For all three cases in Table 3, combination return forecasts are significantly related to actual returns before and after 2003. For the Value Weighted and Equal Weighted excl. Microcap cases, conventional return forecasts are significantly related to actual returns before—but not after—2003, while conventional returns forecasts are significantly related to actual returns before and after 2003 for the Equal Weighted case. The forecast encompassing results in Table 4 indicate that conventional and combination forecasts contain useful information vis-á-vis each other for forecasting cross-sectional returns for all three cases before 2003. After 2003, the conventional and combination forecasts continue to contain useful information relative to each other for the Equal Weighted case, while the combination forecasts dominate the conventional forecasts in terms of information content for the Value Weighted and Equal Weighted excl. Microcap cases. The patterns in the pre- and post-2003 results are similar across Tables 3 and 4.

3.3. The Number and Nature of Relevant Characteristics

By examining the individual forecasts selected by the LASSO or elastic net in Equation (2.11), we can get a sense of how the number and nature of relevant firm characteristics evolve over time. Figure 3 presents the number of characteristics selected each month by the LASSO (Panel A) or elastic net (Panel B) in Equation (2.11) for the Value Weighted case.[24] For the 1990:01 to 2017:12 period, the LASSO and elastic net select 30.71 and 31.06 characteristics, respectively, on average. As expected, the elastic net selects a higher number of characteristics on average, but the difference is small, so that the results for the LASSO and elastic net are quite similar. The number of characteristics selected by each procedure in Figure 3 is fairly stable over time, typically lying between 20 and 40. Along this line, more than ten characteristics are selected each month by both procedures, and the average number of selected characteristics is close to 30 for both the pre- and post-2003 periods. Overall, Figure 3 indicates that a relatively large number of firm characteristics are consistently relevant for forecasting cross-sectional value-weighted returns. Because Equation (2.11) relates realized returns to out-of-sample return forecasts, Figure 3 provides *out-of-sample* evidence that approximately 30 characteristics matter on average for cross-sectional value-weighted expected returns.[25]

To glean insight into the nature of the relevant firm characteristics, Table 5 reports selection frequencies by the elastic net for the individual characteristics.[26] An interesting feature of Table 5 is the churn implied by the selection frequencies. For the 1990:01 to 2017:12 period, the vast majority of selection frequencies are above 30%—only two are below 20%—and all are less than 50%. The selection frequencies are also quite stable before and after 2003. The results in Table 5 point to important time variation in the DGP for

[24] Results for the Equal Weighted excl. Microcap and Equal Weighted cases are similar.

[25] GHZ analyze the number of relevant firm characteristics via in-sample tests based on Fama-MacBeth two-step estimation of cross-sectional multiple regressions for 1980:01 to 2014:12, as well as pre- and post-2003 subperiods. If GHZ estimated cross-sectional regressions each month via OLS or WLS, then it would not be surprising if they found a larger number of statistically significant characteristics, whose identities could vary considerably over time.

[26] To conserve space, we do not report results for the LASSO, which are similar.

cross-sectional value-weighted returns, which helps to explain the usefulness of the forecast combination approach.

Taken together, Figure 3 and Table 5 paint a nuanced picture of the number of relevant firm characteristics for cross-sectional expected returns: Table 5 suggests that nearly all of the firm characteristics matter over time, in the sense that they affect cross-sectional expected returns at least 20% of the time, while Figure 3 indicates that approximately 30 of the firm characteristics matter on average for cross-sectional expected returns at each point in time. From an economic standpoint, to the extent that the characteristics relate to risk factors, Table 5 supports the notion that the importance of individual risk factors varies considerably over time, and Figure 3 indicates that a sizable number of risk factors are relevant at each point in time. Similarly, to the degree that the firm characteristics instead reflect behavioral biases or related factors, our findings imply that the salience of individual behavioral factors changes over time, while a substantial number of such factors affect cross-sectional expected returns at each point in time.[27]

In terms of the most important characteristics in Table 5, there are nine characteristics with selection frequencies of 40% or more for both the pre- and post-2003 periods. The nine characteristics (with their definitions and acronyms from Table 1), their selection frequencies for the 1990:01 to 2017:12 period, and the studies proposing them are as follows:

- One-month momentum, `mom1m`: 49% (Jegadeesh and Titman 1993)
- Tax income to book income, `tb`: 45% (Lev and Nissim 2004)
- Sin stocks, `sin`:[28] 44% (Hong and Kacperczyk 2009)
- Number of earnings increases, `nincr`:[29] 43% (Barth, Elliott, and Finn 1999)

[27]The selection frequencies in Table 5 help to explain the limited ability of individual characteristics to predict cross-sectional value-weighted returns in Table 2: each of the individual characteristics is only relevant for predicting cross-sectional returns approximately 20% to 50% of the time.

[28]This characteristic indicates whether a firm's primary industry classification is in the smoke or tobacco, beer or alcohol, or gaming industries.

[29]This is measured as the number of consecutive quarters (up to eight) with an increase in earnings over the same quarter in the prior year.

- Sales to receivables, `salerec`: 43% (Ou and Penman 1989)
- Change in number of analysts, `chnanalyst`: 43% (Scherbina 2008)
- Corporate investment, `cinvest`: 42% (Titman, Wei, and Xie 2004)
- Industry-adjusted change in profit margin, `chpmia`: 42% (Soliman 2008)
- New equity issue, `IPO`: 41% (Loughran and Ritter 1995)

GHZ use the classifications in McLean and Pontiff (2016)—event, market, valuation, and fundamental—to organize the relevant characteristics they identify. When we similarly apply the McLean and Pontiff (2016) classifications to the nine characteristics above, five of the characteristics are in the fundamental category (`tb`, `nincr`, `salerec`, `cinvest`, `chpmia`), two can be placed in the event category (`chnanalyst`, `IPO`), and one is in the market category (`mom1m`). The other characteristic (`sin`) falls outside of the four categories in McLean and Pontiff (2016).

Considering a set of 36 firm characteristics, Freyberger, Neuhierl, and Weber (2018) use nonparametric methods and the group LASSO (Huang, Horowitz, and Wei 2010) to estimate a generalized version of Equation (2.4) that allows for nonlinearities in the conditional mean. Although the differences in approaches make it difficult to directly compare the results in Freyberger, Neuhierl, and Weber (2018) and the present paper, an interesting parallel exists. Specifically, Freyberger, Neuhierl, and Weber (2018) find substantial time variation in the predictive ability of individual firm characteristics for cross-sectional returns, which is consistent with the churn in the relevant characteristics implied by Table 5.

GHZ find evidence of a structural break in 2003, and the results for the conventional forecasts in Tables 3 and 4 are consistent with their finding. As potential explanations for the break, GHZ point to important changes in the US equity market from July 2002 to June 2003, such as the passage of the Sarbanes-Oxley Act, acceleration of 10-Q and 10-K filing requirements by the SEC, and introduction of autoquoting by the NYSE. In conjunction with increasingly cheap computing power at that time, such changes likely lowered the costs of exploiting characteristics-based mispricing, thereby diminishing the ability of at least some characteristics to predict cross-sectional returns. In addition to a sharp break in 2003, the results in Figure 3 and Table 5 suggest that the DGP relating firm characteristics to cross-sectional returns changes over time in significant ways on a more regular basis.[30] A constantly evolving DGP occasionally punctuated by sharp breaks presents keen out-of-sample challenges and renders conventional forecasts highly susceptible to overfitting. Nevertheless, by guarding against overfitting, our forecast combination approach produces informative cross-sectional return forecasts on a consistent basis over time.

3.4. Spread Portfolios

Next, we evaluate the cross-sectional return forecasts by constructing spread portfolios, as described in Section 2.5. Performance metrics for spread portfolios formed from the cross-sectional forecasts indirectly assess the forecasts by measuring their value as inputs in asset allocation decisions. Table 6 reports means and volatilities for spread portfolio returns based on the conventional and combination forecasts. We again report results for the 1990:01 to 2017:12 period (Panel A), as well as pre- and post-2003 subperiods (Panels B and C, respectively). In addition, Panel D reports the pre-versus-post-2003 change in average return. Note that the rankings of forecasts can differ for conventional statistical criteria and portfolio performance metrics (e.g., Leitch and Tanner 1991; Cenesizoglu and Timmermann 2012). This means that forecasts characterized by overfitting—as indicated by below-unity slope coefficient estimates in Table 3—can still have value as inputs for constructing the spread portfolios.

The second through fourth columns of Table 6 report results for the Value Weighted case. The portfolios based on the combination forecasts generate substantial average returns for the 1990:01 to 2017:12 period, ranging from 0.95% to 1.16%. The portfolio based on

[30] This is reminiscent of the literature on aggregate stock market predictability, where individual economic variables evince episodic predictive ability (e.g., Rapach, Strauss, and Zhou 2010; Henkel, Martin, and Nadari 2011; Rapach and Zhou 2013).

the conventional forecast also yields a sizable average return of 0.98%, which shows that the forecast is useful as an input for constructing the spread portfolio, despite its below-unity slope coefficient estimate in Panel A of Table 3. However, the average return for the portfolio based on the conventional forecast falls from near 2% for the pre-2003 period to close to zero for the post-2003 period in Table 6, and the decline in average return is statistically significant (at the 1% level). It thus appears that the high degree of overfitting indicated by the near-zero slope coefficient estimate in the second column of Panel C of Table 3 substantively reduces the value of the conventional forecast as an input for asset allocation decisions after 2003. Although the portfolios based on the combination forecasts also experience decreases in average return after 2003, the declines are considerably smaller—around 50 basis points—and statistically insignificant. The results for the Value Weighted case in Table 6 follow the same pattern as those in Tables 3 and 4.

Figure 4 plots log cumulative returns for the spread portfolios based on the conventional and elastic net combination forecasts for the Value Weighted case. Panel A shows that the spread portfolio based on the conventional forecast generally produces strong gains from the mid 1990s through the early 2000s and subsequently experiences a drop-off in performance that lasts through the end of the sample. The spread portfolio based on the elastic net combination forecast in Panel B delivers gains on a more consistent basis. Interestingly, the spread portfolio based on the elastic net combination forecast performs well during business-cycle recessions, especially the recent Great Recession. In line with the discussion in Section 3.3, the sizable gains realized by the spread portfolio during cyclical downturns suggest that the economic risks and/or behavioral influences captured by the characteristics become more important during times of macroeconomic stress.

The results for the Equal Weighted excl. Microcap case in the fifth through seventh columns of Table 6 tell a similar story. For the 1990:01 to 2017:12 period, the spread portfolios based on the combination forecasts provide average returns ranging from 1.36% to 1.60%, while the spread portfolio based on the conventional forecast produces an average return of

1.24%. The average return for the spread portfolio based on the conventional forecast again falls markedly from 2.59% before 2003 to 0.11% after 2003, and the decline is statistically significant (at the 1% level). Similarly to the Value Weighted case, the substantive overfitting indicated by the near-zero slope coefficient estimate in the fifth column of Panel C of Table 3 apparently reduces the value of the conventional forecast as an input for constructing the spread portfolio after 2003. The spread portfolios based on the combination forecasts display more limited declines in average return, and the decline is only significant for the LASSO combination forecast (at the 10% level).

The last three columns of Table 6 report results for the Equal Weighted case. The average return for the spread portfolio based on the conventional forecast is 2.95% for the full period. Looking across the subperiods, the average return experiences a decline (significant at the 1% level) from 4.45% before 2003 to 1.68% after 2003. The spread portfolios based on the combination forecasts generate average returns ranging from 1.71% to 1.93%. Apart from the trimmed mean (at the 10% level), the spread portfolios based on the combination forecasts do not exhibit significant decreases in average return before and after 2003. Observe that, in contrast to the Value Weighted and Equal Weighted excl. Microcap cases, the average return for the conventional forecast in Panel C of Table 6 is significant (at the 1% level) and greater than the average returns for the combination forecasts. The overfitting in the conventional forecast detected in the eighth column of Panel C of Table 3 thus does not prevent the forecast from proving useful as an input for asset allocation decisions after 2003.[31]

In Tables 7 and 8, we measure risk-adjusted average returns for the spread portfolios using three leading asset pricing models from the literature: the Carhart (1997) four-factor, Fama and French (2015) five-factor, and Hou, Xue, and Zhang (2015) q-factor models.[32]

[31] The results for the spread portfolios based on the conventional forecasts in Panels B through D of Table 6 are similar to those in Table 8 of GHZ, where the full and post-2003 periods end in 2014:12. Sharpe ratios based on the summary statistics in Table 6 follow the same pattern as the average returns. For example, the spread portfolio based on the conventional forecast produces an annualized Sharpe ratio of 0.69 for the Value Weighted case for the 1990:01 to 2017:12 period, but the annualized Sharpe ratio is only 0.004 for the 2004:01 to 2017:12 period; the annualized Sharpe ratios for the spread portfolio based on the elastic net combination forecast are 0.56 and 0.50 for the 1990:01 to 2017:12 and 2004:01 to 2017:12 periods, respectively.

[32] GHZ do not measure risk-adjusted returns for their spread portfolios.

Table 7 reports estimates of the alphas and factor exposures for the Value Weighted case for the 1990:01 to 2017:12 period.[33] For the spread portfolio based on the conventional forecast, adjusting for risk using the Carhart (1997) four-factor model in Panel A lowers the average return from 0.98% in Table 6 to 0.62% in Table 7. The spread portfolio's significant exposure to the momentum factor primarily accounts for the decline in average return after adjusting for risk. The risk-adjusted average returns for the spread portfolios based on the combination forecasts are all higher than the corresponding unadjusted average returns in Table 6, due in large measure to the portfolios' substantial negative exposures to the market factor in Table 7.

Panel B of Table 7 reports risk-adjusted average returns estimated in the context of the Fama and French (2015) five-factor model. The risk adjustment has little effect on the average return for the spread portfolio based on the conventional forecast, with the portfolio's sizable negative exposure to the value factor offsetting its positive exposures to the other factors. Similarly to the results in Panel A, the risk-adjusted average returns for the spread portfolios based on the combination forecasts are all larger than their unadjusted counterparts in Table 6. The higher risk-adjusted average returns in Table 7 primarily reflect substantial negative exposures to the market and value factors.

As in Panel A, the average return for the spread portfolio based on the conventional forecast falls by approximately a third when we adjust for risk using the Hou, Xue, and Zhang (2015) q-factor model in Panel C. The decline in average return is primarily due to the spread portfolio's substantial exposures to the size and return on equity factors. The risk adjustment increases the average returns for the spread portfolios based on the combination forecasts in the context of the q-factor model, reflecting the portfolios' significant negative exposures to the market factor and insignificant exposures to the remaining factors.[34]

[33] To conserve space, we only report results for the Value Weighted case. The results for the Equal Weighted excl. Microcap and Equal Weighted cases are qualitatively similar.

[34] The significant alphas corresponding to the conventional forecasts in Table 7 indicate that the overfitting signaled by the below-unity slope coefficient estimate in the second column of Panel A of Table 3 does not prevent the conventional forecast from providing useful information as an input for constructing the spread portfolio on average over the 1990:01 to 2017:12 period.

Finally, Table 8 reports estimates of the post-2003 change in alpha. The change in alpha corresponds to the coefficient for a dummy variable that takes a value of zero (one) for 1990:01 to 2003:12 (2004:01 to 2017:12) in the estimated multifactor model. The results in Table 8 repeat the pattern evident throughout the present paper. For the spread portfolios based on the conventional forecasts, the risk-adjusted average return always decreases significantly (at the 1% level) after 2003. For the Value Weighted and Equal Weighted excl. Microcap cases, the risk-adjusted average return essentially becomes zero after 2003, regardless of the model used to make the risk adjustment. In contrast, the risk-adjusted average returns for the spread portfolios based on the combination forecasts do not experience significant declines after 2003, with the exceptions of the mean and trimmed mean combination forecasts for the Equal Weighted case in the context of the Carhart (1997) four-factor model (at the 10% level). Table 8 also continues to show that overfitting in the conventional forecasts does not necessarily preclude the forecasts from serving as useful inputs for constructing the spread portfolios, especially before 2003.

The results in Tables 7 and 8 indicate that leading multifactor models from the literature fail to account for the average returns generated by the spread portfolios based on the combination forecasts. In light of the results in Section 3.3, this finding is not surprising. In Section 3.3, we show that nearly all of the firm characteristics matter for cross-sectional expected returns. It will thus be difficult for an asset pricing model based on a relatively small set of factors to account for the myriad of risk and/or behavioral factors that apparently determine cross-sectional expected returns.

4. Conclusion

In this paper, we apply robust forecast combination methods to the cross section of stock returns. We find that the information in 94 firm characteristics from GHZ is useful for forecasting cross-sectional returns on a consistent basis over time. The key to our finding is how we process the information in the 94 firm characteristics. Although cross-sectional return forecasts constructed using a conventional multiple regression approach incorporate information from the entire set of 94 firm characteristics, they do so in a manner that is susceptible to overfitting. Indeed, the conventional forecasts overstate the cross-sectional dispersion in expected returns—a direct indication of overfitting—and suffer a substantive decline in predictive ability after 2003. While our forecast combination approach also utilizes information from the entire set of 94 firm characteristics, it does so in a manner that guards against overfitting. When we process the information in the 94 firm characteristics via our approach, the combination forecasts indicate that the characteristics are collectively valuable for forecasting cross-sectional returns consistently over time. Combination forecasts that utilize machine learning tools suggest that most of the firm characteristics matter over time—and approximately 30 matter on average at each point in time—for cross-sectional value-weighted expected returns.

Because a relatively large number of firm characteristics appear relevant for forecasting cross-sectional returns, it is unlikely that a factor model of low dimensionality—say, five or less—will be able to account for the cross section of expected returns. Indeed, although the Fama and French (1993) three-factor, Carhart (1997) four-factor, Fama and French (2015) five-factor, Hou, Xue, and Zhang (2015) q-factor, and Stambaugh and Yuan (2017) mispricing-factor models explain various portfolio returns substantially better than the CAPM, they provide little improvement at the individual stock level (e.g., He, Huang, and Zhou 2018). Our results suggest that asset pricing models with significantly more factors are needed to adequately explain the cross section of individual firm returns. An important area for future research is to identify additional factors for explaining cross-sectional returns beyond the well-established factors from the literature. Furthermore, in light of Freyberger, Neuhierl, and Weber (2018), the development of nonlinear factor models may be needed. As we also find that the relevance of individual characteristics varies over time, a keen challenge is to incorporate additional factors into asset pricing models in a manner that avoids overfitting the data and facilitates economic interpretation.

Appendix A. Proofs

A.1. Derivation of $\hat{\delta}_t$

To derive Equation (2.24) in the main text, we use Equations (2.21) and (2.23) to write the difference in cross-sectional value-weighted MSFEs between the naive and competing forecasts as

$$
\begin{aligned}
&\frac{1}{I_t}\sum_{i=1}^{I_t} w_{i,t}\Big[\big(\hat{u}_{i,t|t-1}^{\text{Naive}}\big)^2 - \hat{u}_{i,t|t-1}^2\Big] \\
&\quad = \frac{1}{I_t}\sum_{i=1}^{I_t} w_{i,t}\Big\{(r_{i,t} - \bar{r}_t)^2 - \big[(r_{i,t} - \bar{r}_t) - \big(\hat{r}_{i,t|t-1} - \bar{\hat{r}}_{t|t-1}\big)\big]^2\Big\} \\
&\quad = \frac{1}{I_t}\sum_{i=1}^{I_t} w_{i,t}\Big[2(r_{i,t} - \bar{r}_t)\big(\hat{r}_{i,t|t-1} - \bar{\hat{r}}_{t|t-1}\big) - \big(\hat{r}_{i,t|t-1} - \bar{\hat{r}}_{t|t-1}\big)^2\Big] \\
&\quad = 2\frac{1}{I_t}\sum_{i=1}^{I_t} w_{i,t}(r_{i,t} - \bar{r}_t)\big(\hat{r}_{i,t|t-1} - \bar{\hat{r}}_{t|t-1}\big) - \frac{1}{I_t}\sum_{i=1}^{I_t} w_{i,t}\big(\hat{r}_{i,t|t-1} - \bar{\hat{r}}_{t|t-1}\big)^2 \\
&\quad = \left[2\frac{\sum_{i=1}^{I_t} w_{i,t}(r_{i,t} - \bar{r}_t)\big(\hat{r}_{i,t|t-1} - \bar{\hat{r}}_{t|t-1}\big)}{\sum_{i=1}^{I_t} w_{i,t}\big(\hat{r}_{i,t|t-1} - \bar{\hat{r}}_{t|t-1}\big)^2} - 1\right]\frac{1}{I_t}\sum_{i=1}^{I_t} w_{i,t}\big(\hat{r}_{i,t|t-1} - \bar{\hat{r}}_{t|t-1}\big)^2.
\end{aligned}
$$

A.2. Derivation of Equation (2.27)

According to the FWL theorem, we can write the fitted residual for Equation (2.18) as

$$
\hat{\varepsilon}_{i,t} = (r_{i,t} - \bar{r}_t) - \hat{b}_t^{\text{FM}}\big(\hat{r}_{i,t|t-1} - \bar{\hat{r}}_{t|t-1}\big),
$$

so that the month-t value-weighted residual sum of squares is given by

$$
\begin{aligned}
\text{RSS}_t &= \sum_{i=1}^{I_t} w_{i,t}\hat{\varepsilon}_{i,t}^2 \\
&= \sum_{i=1}^{I_t} w_{i,t}\Big[(r_{i,t} - \bar{r}_t) - \hat{b}_t^{\text{FM}}\big(\hat{r}_{i,t|t-1} - \bar{\hat{r}}_{t|t-1}\big)\Big]^2
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{i=1}^{I_t} w_{i,t}(r_{i,t} - \bar{r}_t)^2 - 2\hat{b}_t^{\text{FM}} \sum_{i=1}^{I_t} w_{i,t}(r_{i,t} - \bar{r}_t)\big(\hat{r}_{i,t|t-1} - \bar{\hat{r}}_{t|t-1}\big) \\
&\quad + \left(\hat{b}_t^{\text{FM}}\right)^2 \sum_{i=1}^{I_t} w_{i,t}\big(\hat{r}_{i,t|t-1} - \bar{\hat{r}}_{t|t-1}\big)^2.
\end{aligned}
$$

Defining the month-t value-weighted total sum of squares as

$$
\text{TSS}_t = \sum_{i=1}^{I_t} w_{i,t}(r_{i,t} - \bar{r}_t)^2
$$

and using

$$
\hat{b}_t^{\text{FM}} = \frac{\sum_{i=1}^{I_t} w_{i,t}(r_{i,t} - \bar{r}_t)\big(\hat{r}_{i,t|t-1} - \bar{\hat{r}}_{t|t-1}\big)}{\sum_{i=1}^{I_t} w_{i,t}\big(\hat{r}_{i,t|t-1} - \bar{\hat{r}}_{t|t-1}\big)^2},
$$

we can express the R^2 statistic for Equation (2.18) as

$$
\begin{aligned}
R_t^2 &= \frac{\text{TSS}_t - \text{RSS}_t}{\text{TSS}_t} \\
&= \frac{2\hat{b}_t^{\text{FM}} \sum_{i=1}^{I_t} w_{i,t}(r_{i,t} - \bar{r}_t)\big(\hat{r}_{i,t|t-1} - \bar{\hat{r}}_{t|t-1}\big) - \left(\hat{b}_t^{\text{FM}}\right)^2 \sum_{i=1}^{I_t} w_{i,t}\big(\hat{r}_{i,t|t-1} - \bar{\hat{r}}_{t|t-1}\big)^2}{\sum_{i=1}^{I_t} w_{i,t}(r_{i,t} - \bar{r}_t)^2} \\
&= \frac{\left(\hat{b}_t^{\text{FM}}\right)^2 (1/I_t) \sum_{i=1}^{I_t} w_{i,t}\big(\hat{r}_{i,t|t-1} - \bar{\hat{r}}_{t|t-1}\big)^2}{(1/I_t) \sum_{i=1}^{I_t} w_{i,t}(r_{i,t} - \bar{r}_t)^2} \\
&= \left(\hat{b}_t^{\text{FM}}\right)^2 \frac{\hat{\sigma}_{\hat{r},t}^2}{\hat{\sigma}_{r,t}^2},
\end{aligned}
$$

so that

$$
\left(\hat{b}_t^{\text{FM}}\right)^2 = R_t^2 \frac{\hat{\sigma}_{r,t}^2}{\hat{\sigma}_{\hat{r},t}^2}.
$$

We can write the squared forecast bias as

$$
\underbrace{\left(\hat{b}_t^{\text{FM}} - 1\right)^2}_{\text{bias}_t} = \left(\hat{b}_t^{\text{FM}}\right)^2 - 2\hat{b}_t^{\text{FM}} + 1,
$$

so that

$$
\begin{aligned}
2\hat{b}_t^{\mathrm{FM}} - 1 &= \left(\hat{b}_t^{\mathrm{FM}}\right)^2 - \mathrm{bias}_t^2 \\
&= R_t^2 \frac{\hat{\sigma}_{r,t}^2}{\hat{\sigma}_{\hat{r},t}^2} - \mathrm{bias}_t^2.
\end{aligned}
$$

Finally, we can express the month-t cross-sectional value-weighted MSFE as

$$
\begin{aligned}
\mathrm{MSFE}_t &= \frac{1}{I_t}\sum_{i=1}^{I_t} w_{i,t}\left[(r_{i,t} - \bar{r}_t) - \left(\hat{r}_{i,t|t-1} - \bar{\hat{r}}_{t|t-1}\right)\right]^2 \\
&= \frac{1}{I_t}\sum_{i=1}^{I_t} w_{i,t}(r_{i,t} - \bar{r}_t)^2 - 2\underbrace{\frac{1}{I_t}\sum_{i=1}^{I_t} w_{i,t}(r_{i,t} - \bar{r}_t)\left(\hat{r}_{i,t|t-1} - \bar{\hat{r}}_{t|t-1}\right)}_{\hat{b}_t^{\mathrm{FM}} \frac{1}{I_t}\sum_{i=1}^{I_t} w_{i,t}\left(\hat{r}_{i,t|t-1} - \bar{\hat{r}}_{t|t-1}\right)^2} \\
&\quad + \frac{1}{I_t}\sum_{i=1}^{I_t} w_{i,t}\left(\hat{r}_{i,t|t-1} - \bar{\hat{r}}_{t|t-1}\right)^2 \\
&= \underbrace{\frac{1}{I_t}\sum_{i=1}^{I_t} w_{i,t}(r_{i,t} - \bar{r}_t)^2}_{\hat{\sigma}_{r,t}^2} - \underbrace{\left(2\hat{b}_t^{\mathrm{FM}} - 1\right)}_{\left(R_t^2 \frac{\hat{\sigma}_{r,t}^2}{\hat{\sigma}_{\hat{r},t}^2} - \mathrm{bias}_t^2\right)} \underbrace{\frac{1}{I_t}\sum_{i=1}^{I_t} w_{i,t}\left(\hat{r}_{i,t|t-1} - \bar{\hat{r}}_{t|t-1}\right)^2}_{\hat{\sigma}_{\hat{r},t}^2} \\
&= \mathrm{bias}_t^2 \hat{\sigma}_{\hat{r},t}^2 + \left(1 - R_t^2\right)\hat{\sigma}_{r,t}^2,
\end{aligned}
$$

which corresponds to Equation (2.27) in main text.

A.3. Derivation of the Optimal Value of θ_t

We can rewrite Equation (2.30) as

$$
\begin{aligned}
\hat{u}_{i,t|t-1}^* &= (r_{i,t} - \bar{r}_t) - \left\{\bar{r}_{t-1} + \theta_t\left(\hat{r}_{i,t|t-1} - \bar{r}_{t-1}\right) - \left[\bar{r}_{t-1} + \theta_t\left(\bar{\hat{r}}_{t|t-1} - \bar{r}_{t-1}\right)\right]\right\} \\
&= (r_{i,t} - \bar{r}_t) - \theta_t\left(\hat{r}_{i,t|t-1} - \bar{\hat{r}}_{t|t-1}\right),
\end{aligned}
$$

so that we can express the cross-sectional value-weighted MSFE for the composite forecast in Equation (2.31) as

$$\begin{aligned}
\frac{1}{I_t}\sum_{i=1}^{I_t} w_{i,t}\big(\hat{u}^{*}_{i,t|t-1}\big)^2 &= \frac{1}{I_t}\sum_{i=1}^{I_t} w_{i,t}\big[(r_{i,t}-\bar{r}_t)-\theta_t\big(\hat{r}_{i,t|t-1}-\bar{\hat{r}}_{t|t-1}\big)\big]^2 \\
&= \frac{1}{I_t}\sum_{i=1}^{I_t} w_{i,t}(r_{i,t}-\bar{r}_t)^2 - 2\theta_t\frac{1}{I_t}\sum_{i=1}^{I_t} w_{i,t}(r_{i,t}-\bar{r}_t)\big(\hat{r}_{i,t|t-1}-\bar{\hat{r}}_{t|t-1}\big) \\
&\quad + \theta_t^2\frac{1}{I_t}\sum_{i=1}^{I_t} w_{i,t}\big(\hat{r}_{i,t|t-1}-\bar{\hat{r}}_{t|t-1}\big)^2.
\end{aligned}$$

Taking the derivative with respect to θ_t,

$$\begin{aligned}
\frac{d}{d\,\theta_t}\left[\frac{1}{I_t}\sum_{i=1}^{I_t} w_{i,t}\big(\hat{u}^{*}_{i,t|t-1}\big)^2\right] &= 2\theta_t\frac{1}{I_t}\sum_{i=1}^{I_t} w_{i,t}\big(\hat{r}_{i,t|t-1}-\bar{\hat{r}}_{t|t-1}\big)^2 \\
&\quad - 2\frac{1}{I_t}\sum_{i=1}^{I_t} w_{i,t}(r_{i,t}-\bar{r}_t)\big(\hat{r}_{i,t|t-1}-\bar{\hat{r}}_{t|t-1}\big).
\end{aligned}$$

Setting the derivative equal to zero and solving for θ_t yields

$$\theta_t^* = \frac{\sum_{i=1}^{I_t} w_{i,t}(r_{i,t}-\bar{r}_t)\big(\hat{r}_{i,t|t-1}-\bar{\hat{r}}_{t|t-1}\big)}{\sum_{i=1}^{I_t} w_{i,t}\big(\hat{r}_{i,t|t-1}-\bar{\hat{r}}_{t|t-1}\big)^2},$$

so that the value of θ_t that minimizes Equation (2.31) is the same as $\hat{\delta}_t$ in Equation (2.25), as stated in the main text.

A.4. Equivalence Between the Optimal Value of η_t and $\hat{b}_t^\dagger$

We begin by rewriting Equation (2.36) as

$$\begin{aligned}
\hat{u}^{\dagger}_{i,t|t-1} &= (r_{i,t}-\bar{r}_t) - \left(\hat{r}^{\dagger}_{i,t|t-1} - \bar{\hat{r}}^{\dagger}_{t|t-1}\right) \\
&= (r_{i,t}-\bar{r}_t) - \left\{\hat{r}^{\mathrm{A}}_{i,t|t-1} + \eta_t\big(\hat{r}^{\mathrm{B}}_{i,t|t-1} - \hat{r}^{\mathrm{A}}_{i,t|t-1}\big) - \big[\bar{\hat{r}}^{\mathrm{A}}_{t|t-1} + \eta_t\big(\bar{\hat{r}}^{\mathrm{B}}_{t|t-1} - \bar{\hat{r}}^{\mathrm{A}}_{t|t-1}\big)\big]\right\} \\
&= (r_{i,t}-\bar{r}_t) - \left\{\big(\hat{r}^{\mathrm{A}}_{i,t|t-1} - \bar{\hat{r}}^{\mathrm{A}}_{t|t-1}\big) + \eta_t\big[\big(\hat{r}^{\mathrm{B}}_{i,t|t-1} - \bar{\hat{r}}^{\mathrm{B}}_{t|t-1}\big) - \big(\hat{r}^{\mathrm{A}}_{i,t|t-1} - \bar{\hat{r}}^{\mathrm{A}}_{t|t-1}\big)\big]\right\} \\
&= \big(r_{i,t} - \hat{r}^{\mathrm{A}}_{i,t|t-1}\big) - \big(\bar{r}_t - \bar{\hat{r}}^{\mathrm{A}}_{t|t-1}\big) - \eta_t\big[\big(\hat{r}^{\mathrm{B}}_{i,t|t-1} - \bar{\hat{r}}^{\mathrm{B}}_{t|t-1}\big) - \big(\hat{r}^{\mathrm{A}}_{i,t|t-1} - \bar{\hat{r}}^{\mathrm{A}}_{t|t-1}\big)\big]
\end{aligned}$$

$$
\begin{aligned}
&+\eta_t[(r_{i,t}-\bar{r}_t)-(r_{i,t}-\bar{r}_t)] \\
&= \underbrace{\left(r_{i,t}-\hat{r}^{\mathrm{A}}_{i,t|t-1}\right)-\left(\bar{r}_t-\bar{\hat{r}}^{\mathrm{A}}_{t|t-1}\right)}_{\hat{u}^{\mathrm{A}}_{i,t|t-1}} + \eta_t\Big\{\underbrace{\left(r_{i,t}-\hat{r}^{\mathrm{B}}_{i,t|t-1}\right)-\left(\bar{r}_t-\bar{\hat{r}}^{\mathrm{B}}_{t|t-1}\right)}_{\hat{u}^{\mathrm{B}}_{i,t|t-1}} \\
&\quad - \underbrace{\left[\left(r_{i,t}-\hat{r}^{\mathrm{A}}_{i,t|t-1}\right)-\left(\bar{r}_t-\bar{\hat{r}}^{\mathrm{A}}_{t|t-1}\right)\right]}_{\hat{u}^{\mathrm{A}}_{i,t|t-1}}\Big\} \\
&= \hat{u}^{\mathrm{A}}_{i,t|t-1}-\eta_t\left(\hat{u}^{\mathrm{A}}_{i,t|t-1}-\hat{u}^{\mathrm{B}}_{i,t|t-1}\right).
\end{aligned}
$$

We can then write the cross-sectional value-weighted MSFE in Equation (2.35) as

$$
\begin{aligned}
\frac{1}{I_t}\sum_{i=1}^{I_t} w_{i,t}\left(\hat{u}^{\dagger}_{i,t|t-1}\right)^2 &= \frac{1}{I_t}\sum_{i=1}^{I_t} w_{i,t}\left[\hat{u}^{\mathrm{A}}_{i,t|t-1}-\eta_t\left(\hat{u}^{\mathrm{A}}_{i,t|t-1}-\hat{u}^{\mathrm{B}}_{i,t|t-1}\right)\right]^2 \\
&= \frac{1}{I_t}\sum_{i=1}^{I_t} w_{i,t}\left(\hat{u}^{\mathrm{A}}_{i,t|t-1}\right)^2 - 2\eta_t\frac{1}{I_t}\sum_{i=1}^{I_t} w_{i,t}\hat{u}^{\mathrm{A}}_{i,t|t-1}\left(\hat{u}^{\mathrm{A}}_{i,t|t-1}-\hat{u}^{\mathrm{B}}_{i,t|t-1}\right) \\
&\quad + \eta_t^2\frac{1}{I_t}\sum_{i=1}^{I_t} w_{i,t}\left(\hat{u}^{\mathrm{A}}_{i,t|t-1}-\hat{u}^{\mathrm{B}}_{i,t|t-1}\right)^2.
\end{aligned}
$$

Taking the derivative with respect to η_t,

$$
\begin{aligned}
\frac{d}{d\,\eta_t}\left[\frac{1}{I_t}\sum_{i=1}^{I_t} w_{i,t}(\hat{u}^{\dagger}_{i,t|t-1})^2\right] &= 2\eta_t\frac{1}{I_t}\sum_{i=1}^{I_t} w_{i,t}\left(\hat{u}^{\mathrm{A}}_{i,t|t-1}-\hat{u}^{\mathrm{B}}_{i,t|t-1}\right)^2 \\
&\quad - 2\frac{1}{I_t}\sum_{i=1}^{I_t} w_{i,t}\hat{u}^{\mathrm{A}}_{i,t|t-1}\left(\hat{u}^{\mathrm{A}}_{i,t|t-1}-\hat{u}^{\mathrm{B}}_{i,t|t-1}\right).
\end{aligned}
$$

Setting the derivative to zero and solving for η_t gives

$$
\eta_t^* = \frac{\sum_{i=1}^{I_t} w_{i,t}\hat{u}^{\mathrm{A}}_{i,t|t-1}\left(\hat{u}^{\mathrm{A}}_{i,t|t-1}-\hat{u}^{\mathrm{B}}_{i,t|t-1}\right)}{\sum_{k=1}^{I_t} w_{i,t}\left(\hat{u}^{\mathrm{A}}_{i,t|t-1}-\hat{u}^{\mathrm{B}}_{i,t|t-1}\right)^2}.
$$

Inspection of η_t^* reveals that it is equivalent to the WLS slope coefficient estimate in a regression of the demeaned forecast error for A on the difference between the demeaned forecast errors for A and B. By the FWL theorem, η_t^* is thus identical to the WLS estimate of $b_t^\dagger$ in Equation (2.37).

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.

Barth, M. E., J. A. Elliott, and M. W. Finn (1999). Market Rewards Associated with Patterns of Increasing Earnings. *Journal of Accounting Research* 37:2, 387–413.

Bickel, P. J., Y. Ritov, and A. B. Tsybakov (2009). Simultaneous Analysis of Lasso and Dantzig Selector. *Annals of Statistics* 37:4, 1705–1732.

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

Cenesizoglu, T. and A. Timmermann (2012). Do Return Prediction Models Add Economic Value? *Journal of Banking and Finance* 36:11, 2974–2987.

Chinco, A., A. D. Clark-Joseph, and M. Ye (forthcoming). Sparse Signals in the Cross-Section of Returns. *Journal of Finance*.

Chong, Y. Y. and D. F. Hendry (1986). Econometric Evaluation of Linear Macro-Economic Models. *Review of Economic Studies* 53:4, 671–690.

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

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

Fair, R. C. and R. J. Shiller (1990). Comparing Information in Forecasts from Economic Models. *American Economic Review* 80:3, 375–389.

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.

— (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.

Feng, G., S. Giglio, and D. Xiu (2017). Taming the Factor Zoo. Manuscript.

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 (2018). Dissecting Characteristics Nonparametrically. Manuscript.

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 (2018). Empirical Asset Pricing via Machine Learning. Manuscript.

Hanna, J. D. and M. J. Ready (2005). Profitable Predictability in the Cross Section of Stock Returns. *Journal of Financial Economics* 78:3, 463–505.

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.

He, A., D. Huang, and G. Zhou (2018). Pricing Error Reversal: A Diagnostic Test of Asset Pricing Models. Manuscript.

Henkel, S. J., J. S. Martin, and F. Nadari (2011). Time-Varying Short-Horizon Predictability. *Journal of Financial Economics* 99:3, 560–580.

Hong, H. and M. Kacperczyk (2009). The Price of Sin: The Effects of Social Norms on Markets. *Journal of Financial Economics* 93:1, 15–36.

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

Huang, J., J. L. Horowitz, and F. Wei (2010). Variable Selection in Nonparametric Additive Models. *Annals of Statistics* 38:4, 2282–2313.

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

Jegadeesh, N. and S. Titman (1993). Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. *Journal of Finance* 48:1, 65–91.

Kozak, S., S. Nagel, and S. Santosh (forthcoming). Shrinking the Cross Section. *Journal of Financial Economics*.

Leitch, G. and J. E. Tanner (1991). Economic Forecast Evaluation: Profits Versus the Conventional Error Measures. *American Economic Review* 81:3, 580–590.

Lev, B. and D. Nissim (2004). Taxable Income, Future Earnings, and Equity Values. *The Accounting Review* 79:4, 1039–1074.

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

Loughran, T. and J. R. Ritter (1995). The New Issues Puzzle. *Journal of Finance* 50:1, 23–51.

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

Meinshausen, N. and B. Yu (2009). Lasso-Type Recovery of Sparse Representations for High-Dimensional Data. *Annals of Statistics* 37:1, 246–270.

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.

Ou, J. A. and S. H. Penman (1989). Financial Statement Analysis and the Prediction of Stock Returns. *Journal of Accounting and Economics* 11:4, 295–329.

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.

Rapach, D. E. and G. Zhou (2013). Forecasting Stock Returns. In *Handbook of Economic Forecasting*. Ed. by G. Elliott and A. Timmermann. Vol. 2A. Amsterdam: Elsevier, pp. 328–383.

Scherbina, A. (2008). Suppressed Negative Information and Future Underperformance. *Review of Finance* 12:1, 533–565.

Soliman, M. T. (2008). The Use of DuPont Analysis by Market Participants. *The Accounting Review* 83:3, 823–853.

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.

Titman, S., K. C. J. Wei, and F. Xie (2004). Capital Investments and Stock Returns. *Journal of Financial and Quantitative Analysis* 39:3, 677–700.

Zhang, C.-H. and J. Huang (2008). The Sparsity and Bias of the Lasso Selection in High-Dimensional Linear Regression. *Annals of Statistics* 36:4, 1567–1594.

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 Characteristic Acronyms and Definitions

(1) Acronym	(2) Definition	(3) Acronym	(4) Definition
absacc	Absolute accruals	mom1m	1-month momentum
acc	Working capital accruals	mom36m	36-month momentum
aeavol	Abnormal earnings announcement volume	ms	Financial statement score
age	# years since first Compustat coverage	mve	Size
agr	Asset growth	mve_ia	Industry-adjusted size
baspread	Bid-ask spread	nanalyst	Number of analysts covering stock
beta	Beta	nincr	Number of earnings increases
bm	Book to market	operprof	Operating profitability
bm_ia	Industry-adjusted book to market	orgcap	Organizational capital
cash	Cash holdings	pchcapx_ia	Industry-adjusted Δ% in capital exps.
cashdebt	Cash flow to debt	pchcurrat	Δ% in current ratio
cashpr	Cash productivity	pchdepr	Δ% in depreciation
cfp	Cash-flow-to-price ratio	pchgm_pchsale	Δ% in gross margin − Δ% in sales
cfp_ia	Industry-adjusted cash-flow-to-price ratio	pchsale_pchinvt	Δ% in sales − Δ% in inventory
chatoia	Industry-adjusted Δ in asset turnover	pchsale_pchrect	Δ% in sales − Δ% in A/R
chcsho	Δ in shares outstanding	pchsale_pchxsga	Δ% change in sales − Δ% in SG&A
chempia	Industry-adjusted change in employees	pchsaleinv	Δ% sales-to-inventory
chfeps	Δ in forecasted EPS	pctacc	Percent accruals
chinv	Δ in inventory	pricedelay	Price delay
chmom	Δ in 6-month momentum	ps	Financial statements score
chnanalyst	Δ in number of analysts	rd	R&D increase
chpmia	Industry-adjusted Δ in profit margin	rd_mve	R&D to market capitalization
chtx	Δ in tax expense	rd_sale	R&D to sales
cinvest	Corporate investment	realestate	Real estate holdings
convind	Convertible debt indicator	retvol	Return volatility
currat	Current ratio	roaq	Return on assets
depr	Depreciation / PP&E	roavol	Earnings volatility
disp	Dispersion in forecasted EPS	roeq	Return on equity
divi	Dividend initiation	roic	Return on invested capital
divo	Dividend omission	rsup	Revenue surprise
dy	Dividend to price	salecash	Sales to cash
ear	Earnings announcement return	saleinv	Sales to inventory
egr	Growth in common shareholder equity	salerec	Sales to receivables
ep	Earnings to price	secured	Secured debt
fgr5yr	Forecasted growth in 5-year EPS	securedind	Secured debt indicator
gma	Gross profitability	sfe	Scaled earnings forecast
grCAPX	Growth in capital expenditures	sgr	Sales growth
grltnoa	Growth in long-term net operating assets	sin	Sin stocks
herf	Industry sales concentration	sp	Sales to price
hire	Employee growth rate	std_dolvol	Volatility of liquidity ($ trading volume)
idiovol	Idiosyncratic return volatility	std_turn	Volatility of liquidity (share turnover)
ill	Illiquidity	stdcf	Cash flow volatility
indmom	Industry momentum	sue	Unexpected quarterly earnings
invest	Capital expenditures	tang	Debt capacity / firm tangibility
IPO	New equity issue	tb	Tax income to book income
lev	Leverage	turn	Share turnover
mom12m	12-month momentum	zerotrade	Zero trading days

The table provide acronyms and definitions for 94 firm characteristics from Green, Hand, and Zhang (2017, Table 1).

Table 2
Fama-MacBeth Regression Results for Cross-Sectional Return Forecasts Based on Individual Firm Characteristics, 1990:01 to 2017:12

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Value Weighted			Equal Weighted excl. Microcap			Equal Weighted		
Characteristic	Coefficient	t-statistic	R^2	Coefficient	t-statistic	R^2	Coefficient	t-statistic	R^2
absacc	0.48	0.79	1.07%	1.25	0.79	0.63%	2.13	1.32*	0.47%
acc	−1.35	−1.09	0.83%	−0.58	−1.09	0.42%	2.40	1.41*	0.25%
aeavol	1.62	1.78**	0.59%	1.84	1.78**	0.22%	−0.95	−1.18	0.07%
age	−0.14	−0.64	1.77%	0.49	0.64	1.01%	1.05	2.13**	0.39%
agr	16.96	0.60	1.00%	−0.68	−0.60	0.65%	4.32	0.97	0.26%
baspread	−4.93	−1.60	4.90%	−2.68	−0.60	4.28%	−0.37	−1.00	2.22%
beta	−1.22	−0.76	6.18%	−1.10	−0.76	4.36%	0.11	0.23	1.94%
bm	4.53	1.78**	1.73%	0.73	0.78	0.89%	0.35	0.79	0.41%
bm_ia	0.90	0.22	0.95%	−7.67	−1.22	0.43%	−0.01	−0.02	0.15%
cash	5.09	1.12	2.48%	1.71	1.12	2.03%	0.38	0.61	0.86%
cashdebt	1.48	0.40	0.79%	−0.21	−0.40	0.62%	1.17	1.70**	0.48%
cashpr	−0.20	−0.16	0.58%	1.14	1.16	0.28%	0.93	1.41*	0.14%
cfp	−2.18	−1.05	1.00%	−0.05	−0.05	0.77%	−13.10	−0.95	0.56%
cfp_ia	−4.72	−1.37	0.90%	−0.89	−1.37	0.50%	−0.24	−0.50	0.21%
chatoia	0.41	0.38	0.45%	−2.20	−1.38	0.16%	0.68	1.33*	0.06%
chcsho	0.75	0.58	0.76%	−2.89	−1.58	0.37%	−0.19	−0.18	0.20%
chempia	−0.14	−0.49	0.58%	−2.01	−0.49	0.25%	−1.28	−1.10	0.10%
chfeps	−1.22	−1.62	0.73%	0.25	0.62	0.31%	0.36	0.82	0.10%
chinv	−8.93	−1.06	0.57%	−3.98	−1.06	0.32%	−1.62	−0.93	0.10%
chmom	−9.92	−1.97	2.16%	−1.43	−0.97	1.15%	−0.88	−0.72	0.44%
chnanalyst	0.33	0.55	0.56%	−0.84	−0.55	0.20%	−0.78	−0.49	0.06%
chpmia	0.50	0.40	0.83%	1.29	1.40*	0.34%	−1.29	−0.89	0.12%
chtx	0.03	0.13	0.86%	0.44	1.13	0.37%	−1.37	−0.80	0.09%
cinvest	−5.01	−0.50	0.14%	1.69	0.50	0.22%	0.25	0.34	0.08%
convind	−4.30	−1.36	0.53%	−1.65	−2.36	0.28%	−0.29	−0.32	0.11%
currat	−0.63	−0.83	0.41%	3.12	1.83**	0.36%	0.48	0.48	0.11%

The table reports Fama-MacBeth regression results for out-of-sample cross-sectional return forecasts based on 94 individual firm characteristics (defined in Table 1). For each characteristic, we first estimate a cross-sectional univariate regression that relates returns in month t to the characteristic values in month $t-1$; we then use the fitted cross-sectional regression and corresponding characteristic values in month t to generate cross-sectional return forecasts for month $t+1$. Using the forecasts for each month, we estimate a Fama-MacBeth cross-sectional univariate regression that relates the realized returns to the forecasted returns. The table reports the time-series averages of the slope coefficients and R^2 statistics for the Fama-MacBeth cross-sectional univariate regressions. The t-statistics are based on Newey and West (1987) standard errors (computed using twelve lags); *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively, for a one-sided, upper-tailed test. "Value-Weighted" indicates that the cross-sectional regressions are estimated via weighted least squares with observations weighted by market value at the end of the preceding month; "Equal Weighted" ("Equal Weighted excl. Microcaps") indicates that the cross-sectional regressions are estimated via ordinary least squares (excluding stocks with market value below the NYSE 20% percentile).

Table 2 (continued)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Value Weighted			Equal Weighted excl. Microcap			Equal Weighted		
Characteristic	Coefficient	t-statistic	R^2	Coefficient	t-statistic	R^2	Coefficient	t-statistic	R^2
depr	−0.28	−0.42	1.07%	−0.21	−0.42	0.63%	−3.06	−0.55	0.38%
disp	−1.02	−1.21	0.80%	−38.81	−1.21	0.47%	−0.95	−1.10	0.15%
divi	−0.13	−0.19	0.29%	1.60	1.34*	0.09%	0.43	0.44	0.05%
divo	−3.72	−1.04	0.19%	0.56	0.10	0.12%	0.27	0.20	0.05%
dy	−44.62	−0.96	2.26%	−7.18	−1.02	1.10%	10.12	1.00	0.29%
ear	2.02	1.40*	0.55%	−0.35	−0.49	0.24%	−7.79	−0.98	0.11%
egr	−2.95	−1.05	0.72%	−0.11	−0.20	0.42%	11.56	1.01	0.15%
ep	−0.47	−0.46	0.89%	−0.08	−0.17	0.97%	1.11	1.48*	0.91%
fgr5yr	0.59	1.05	3.41%	1.05	0.94	2.10%	0.24	0.57	0.58%
gma	−0.32	−0.50	1.76%	2.87	1.19	0.76%	−0.28	−0.60	0.26%
grCAPX	31.91	1.01	0.49%	1.07	1.68**	0.34%	−1.28	−1.19	0.11%
grltnoa	0.23	0.24	0.45%	0.30	1.20	0.25%	4.15	1.66**	0.12%
herf	0.99	0.52	0.54%	1.15	1.11	0.29%	−1.42	−1.50	0.11%
hire	0.92	1.68**	0.96%	−1.45	−0.85	0.58%	−2.12	−1.68	0.20%
idiovol	0.24	0.72	4.06%	−0.45	−1.07	3.67%	−1.23	−1.95	2.01%
ill	−1.17	−0.63	0.02%	−0.08	−0.05	0.16%	−0.21	−0.18	0.41%
indmom	0.72	0.60	2.25%	0.10	0.36	1.73%	1.72	2.01**	0.85%
invest	−0.77	−1.33	0.76%	0.22	0.27	0.52%	−2.46	−0.94	0.20%
IPO	3.89	1.12	0.32%	1.04	0.72	0.39%	−0.04	−0.05	0.14%
lev	0.79	0.88	2.30%	4.31	0.97	1.17%	2.20	0.96	0.59%
mom12m	0.84	0.74	3.33%	−2.66	−1.03	1.97%	1.18	1.04	0.90%
mom1m	−0.40	−0.51	2.17%	0.35	0.47	1.43%	1.40	2.03**	0.87%
mom36m	−0.67	−1.92	2.16%	−3.94	−1.01	0.85%	−0.37	−0.90	0.38%
ms	0.17	0.53	1.35%	1.91	1.65**	0.37%	−1.74	−0.59	0.32%
mve	0.57	1.15	1.91%	0.89	0.54	0.95%	3.60	1.18	0.88%
mve_ia	−25.92	−1.01	1.87%	0.26	0.39	0.47%	4.83	0.71	0.17%
nanalyst	−0.05	−0.09	1.48%	−0.59	−0.39	0.60%	−1.92	−1.27	0.39%
nincr	−4.30	−1.16	0.54%	2.72	1.08	0.25%	−1.60	−0.83	0.09%
operprof	−0.71	−0.55	0.57%	0.07	0.07	0.24%	0.06	0.06	0.10%
orgcap	−0.32	−0.45	0.97%	0.85	0.48	0.48%	0.72	0.61	0.32%
pchcapx_ia	1.62	0.68	0.86%	2.72	1.53*	0.44%	0.88	1.19	0.18%
pchcurrat	−0.12	−0.20	0.25%	9.62	1.25	0.15%	0.50	1.28*	0.07%
pchdepr	−10.97	−1.27	0.44%	3.78	0.95	0.21%	0.74	1.72**	0.13%
pchgm_pchsale	2.31	1.17	0.22%	0.55	0.57	0.19%	2.26	2.02**	0.12%

Table 2 (continued)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Value Weighted			Equal Weighted excl. Microcap			Equal Weighted		
Characteristic	Coefficient	t-statistic	R^2	Coefficient	t-statistic	R^2	Coefficient	t-statistic	R^2
pchsale_pchinvt	1.37	1.16	0.39%	0.40	1.41*	0.11%	2.00	1.36*	0.06%
pchsale_pchrect	1.10	1.05	0.31%	−0.18	−0.26	0.14%	1.21	2.33**	0.05%
pchsale_pchxsga	1.94	1.21	0.51%	1.56	1.48*	0.26%	0.07	0.08	0.09%
pchsaleinv	1.40	1.13	0.33%	2.34	2.03**	0.12%	0.20	0.62	0.05%
pctacc	0.98	0.50	0.39%	0.08	0.17	0.13%	−0.60	−0.59	0.07%
pricedelay	0.33	1.00	0.51%	−0.22	−0.29	0.18%	0.11	0.09	0.09%
ps	1.80	1.32*	0.77%	1.95	1.89**	0.42%	3.40	1.07	0.37%
rd	0.46	0.27	0.39%	0.02	0.09	0.33%	−0.54	−0.52	0.31%
rd_mve	0.19	0.39	0.94%	6.20	1.20	0.46%	−1.05	−1.35	0.43%
rd_sale	0.27	0.61	0.34%	0.24	0.30	0.46%	49.02	1.03	0.23%
realestate	0.05	0.10	0.50%	−0.39	−0.71	0.26%	4.34	1.19	0.14%
retvol	0.75	0.71	4.18%	1.69	1.59*	3.23%	−0.89	−0.83	1.77%
roaq	1.82	1.99**	1.30%	3.49	0.86	1.01%	2.05	0.93	1.01%
roavol	2.80	0.94	1.23%	2.46	2.12**	1.19%	0.66	0.63	0.77%
roeq	0.86	1.44*	0.64%	−0.25	−0.13	0.49%	0.48	0.66	0.56%
roic	−0.26	−0.47	0.64%	0.36	0.61	0.80%	−0.62	−0.85	0.68%
rsup	1.60	1.17	0.52%	605.30	1.02	0.37%	−0.04	−0.06	0.21%
salecash	−2.06	−1.73	0.46%	1.19	1.36*	0.20%	4.15	1.01	0.07%
saleinv	−1.75	−1.48	0.31%	1.47	2.08**	0.13%	−0.03	−0.04	0.05%
salerec	1.06	1.05	0.76%	1.19	0.58	0.37%	1.07	1.60*	0.10%
secured	−1.53	−0.61	0.56%	0.93	1.38*	0.34%	0.82	1.32*	0.16%
securedind	1.01	0.97	0.52%	−2.65	−1.28	0.37%	−0.17	−0.11	0.20%
sfe	1.15	1.03	0.43%	−0.63	−0.64	0.56%	−0.99	−0.41	0.41%
sgr	2.43	2.00**	1.07%	−0.62	−0.29	0.66%	−0.21	−0.31	0.20%
sin	0.51	1.12	0.57%	−14.16	−1.04	0.13%	1.00	0.41	0.04%
sp	0.65	0.40	0.93%	0.47	0.41	0.60%	0.33	0.48	0.30%
std_dolvol	−2.47	−1.87	0.90%	−1.24	−1.29	0.34%	0.37	0.22	0.46%
std_turn	1.22	0.42	2.10%	0.85	1.08	1.49%	−29.46	−1.01	0.72%
stdcf	−0.25	−0.20	0.20%	1.36	1.40*	0.32%	−0.22	−0.09	0.21%
sue	−0.76	−1.37	0.28%	2.56	0.82	0.27%	−0.01	−0.01	0.18%
tang	0.91	0.97	1.24%	−0.97	−0.82	0.99%	−0.12	−0.19	0.39%
tb	0.30	0.78	0.53%	1.12	0.64	0.15%	−6.29	−1.40	0.10%
turn	−0.62	−0.49	3.92%	0.24	0.57	2.73%	0.56	1.10	1.19%
zerotrade	−0.73	−1.45	0.05%	−0.01	−0.03	0.12%	0.24	0.11	0.39%

Table 3
Fama-MacBeth Regression Results for Combination Cross-Sectional Return Forecasts

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Value Weighted			Equal Weighted excl. Microcap			Equal Weighted		
Method	Coefficient	t-statistic	R^2	Coefficient	t-statistic	R^2	Coefficient	t-statistic	R^2
Panel A: Full out-of-sample period (1990:01–2017:12)									
Conventional	0.31	3.10***	1.67%	0.35	3.65***	1.02%	0.67	11.99***	0.77%
Mean	1.56	2.43***	5.37%	2.89	4.08***	4.40%	4.04	5.61***	2.35%
Trimmed mean	2.23	2.44***	5.35%	4.26	4.00***	4.35%	5.88	5.61***	2.28%
LASSO	1.64	3.96***	4.14%	1.61	4.21***	3.11%	2.61	7.47***	1.83%
ENet	1.66	3.95***	4.14%	1.68	4.38***	3.11%	2.61	7.66***	1.84%
Panel B: Pre-2003 out-of-sample period (1990:01–2002:12)									
Conventional	0.63	4.14***	2.11%	0.72	7.41***	1.45%	0.90	25.42***	1.05%
Mean	1.47	1.47*	6.13%	3.15	2.51***	5.85%	5.99	5.92***	2.98%
Trimmed mean	1.97	1.36*	6.24%	4.33	2.28**	5.83%	8.65	5.76***	2.89%
LASSO	1.73	2.72***	4.51%	2.33	4.52***	3.82%	3.39	6.45***	2.25%
ENet	1.76	2.69***	4.54%	2.43	4.76***	3.81%	3.34	6.59***	2.26%
Panel C: Post-2003 out-of-sample period (2004:01–2017:12)									
Conventional	0.05	0.58	1.24%	0.04	0.35	0.64%	0.48	6.73***	0.52%
Mean	1.94	2.40***	4.74%	2.80	3.53***	3.14%	1.91	2.82***	1.77%
Trimmed mean	2.92	2.62***	4.60%	4.24	3.54***	3.08%	2.86	2.97***	1.71%
LASSO	1.55	2.70***	3.89%	1.04	1.98**	2.49%	1.79	4.69***	1.43%
ENet	1.58	2.74***	3.89%	1.08	2.06**	2.49%	1.82	4.76***	1.43%

The table reports Fama-MacBeth regression results for out-of-sample cross-sectional return forecasts based on combination approaches. For each of 94 individual firm characteristics (defined in Table 1), we first estimate a cross-sectional univariate regression that relates returns in month t to the characteristic values in month $t-1$; we then use the fitted cross-sectional regression and corresponding characteristic values in month t to generate cross-sectional return forecasts for month $t+1$. The "Mean" ("Trimmed mean") combination forecast is the arithmetic mean of the return forecasts based on the individual characteristics (after excluding the 5% smallest and 5% largest of the individual return forecasts); the "LASSO" ("ENet") combination forecast is the average of the return forecasts based on the individual characteristics selected by the LASSO (elastic net). "Conventional" refers to a forecast based on a conventional multiple regression approach (e.g., Green, Hand, and Zhang 2017). Using the forecasts for each month, we estimate a Fama-MacBeth cross-sectional univariate regression that relates the realized returns to the forecasted returns. The table reports the time-series averages of the slope coefficients and R^2 statistics for the Fama-MacBeth cross-sectional univariate regressions. The t-statistics are based on Newey and West (1987) standard errors (computed using twelve lags); *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively, for a one-sided, upper-tailed test. "Value-Weighted" indicates that the cross-sectional regressions are estimated via weighted least squares with observations weighted by market value at the end of the preceding month; "Equal Weighted" ("Equal Weighted excl. Microcap") indicates that the cross-sectional regressions are estimated via ordinary least squares (excluding stocks with market value below the NYSE 20% percentile).

Table 4
Encompassing Test Results for Combination Cross-Sectional Return Forecasts

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
	Value Weighted				Equal Weighted excl. Microcap				Equal Weighted			
	$\hat{b}^{\dagger}$		$1-\hat{b}^{\dagger}$		$\hat{b}^{\dagger}$		$1-\hat{b}^{\dagger}$		$\hat{b}^{\dagger}$		$1-\hat{b}^{\dagger}$	
Method	Coeff.	*t*-stat.	Coeff.	*t*-stat.	Coeff.	*t*-stat.	Coeff.	*t*-stat.	Coeff.	*t*-stat.	Coeff.	*t*-stat.
Panel A: Full out-of-sample period (1990:01–2017:12)												
Mean	0.70	6.62***	0.30	2.89***	0.71	7.82***	0.29	3.25***	0.39	6.86***	0.61	10.79***
Trimmed mean	0.70	6.62***	0.30	2.82***	0.70	7.47***	0.30	3.24***	0.37	6.70***	0.63	11.32***
LASSO	0.68	6.57***	0.32	3.14***	0.68	7.38***	0.32	3.45***	0.40	6.26***	0.60	9.43***
ENet	0.68	6.62***	0.32	3.12***	0.68	7.23***	0.32	3.45***	0.40	6.25***	0.60	9.46***
Panel B: Pre-2003 out-of-sample period (1990:01–2002:12)												
Mean	0.36	2.35***	0.64	4.22***	0.40	3.83***	0.60	5.72***	0.19	3.11***	0.81	13.08***
Trimmed mean	0.36	2.34***	0.64	4.12***	0.36	3.63***	0.64	6.47***	0.17	3.25***	0.83	16.29***
LASSO	0.35	2.28**	0.65	4.22***	0.39	2.98***	0.61	4.57***	0.20	2.26**	0.80	9.28***
ENet	0.35	2.32**	0.65	4.23***	0.38	2.85***	0.62	4.56***	0.20	2.28**	0.80	9.21***
Panel C: Post-2003 out-of-sample period (2004:01–2017:12)												
Mean	0.98	10.62***	0.02	0.18	0.96	9.06***	0.04	0.38	0.55	8.12***	0.45	6.57***
Trimmed mean	0.99	10.73***	0.01	0.13	0.98	9.07***	0.02	0.18	0.55	7.98***	0.45	6.66***
LASSO	0.95	10.86***	0.05	0.54	0.92	9.79***	0.08	0.89	0.57	8.35***	0.43	6.43***
ENet	0.95	10.82***	0.05	0.51	0.92	9.83***	0.08	0.87	0.56	8.22***	0.44	6.42***

The table reports encompassing test results for out-of-sample cross-sectional return forecasts based on combination approaches. For each of 94 individual firm characteristics (defined in Table 1), we first estimate a cross-sectional univariate regression that relates returns in month t to the characteristic values in month $t-1$; we then use the fitted cross-sectional regression and corresponding characteristic values in month t to generate cross-sectional return forecasts for month $t+1$. The "Mean" ("Trimmed mean") combination forecast is the arithmetic mean of the return forecasts based on the individual characteristics (after excluding the 5% smallest and 5% largest of the individual return forecasts); the "LASSO" ("ENet") combination forecast is the average of the return forecasts based on the individual characteristics selected by the LASSO (elastic net). The $\hat{b}^{\dagger}$ and $1-\hat{b}^{\dagger}$ estimates correspond to the average coefficients attached to the combination forecast and a conventional multiple regression forecast (e.g., Green, Hand, and Zhang 2017), respectively, in an optimal composite forecast. The t-statistics are based on Newey and West (1987) standard errors (computed using twelve lags); *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively, for a one-sided, upper-tailed test. "Value-Weighted" indicates that the cross-sectional regressions are estimated via weighted least squares with observations weighted by market value at the end of the preceding month; "Equal Weighted" ("Equal Weighted excl. Microcap") indicates that the cross-sectional regressions are estimated via ordinary least squares (excluding stocks with market value below the NYSE 20% percentile).

Table 5
Weighted Elastic Net Selection Frequencies, 1990:01 to 2017:12

(1) Characteristic	(2) Full	(3) Pre 2003	(4) Post 2003	(5) Characteristic	(6) Full	(7) Pre 2003	(8) Post 2003	(9) Characteristic	(10) Full	(11) Pre 2003	(12) Post 2003
mom1m	49%	45%	52%	lev	37%	37%	36%	stdcf	31%	29%	33%
tb	45%	46%	43%	pchdepr	36%	34%	39%	bm_ia	30%	30%	30%
sin	44%	42%	45%	chmom	36%	36%	36%	ep	30%	32%	29%
nincr	43%	42%	45%	sue	36%	37%	35%	ill	29%	33%	26%
salerec	43%	43%	43%	saleinv	35%	38%	33%	pchsale_pchinvt	29%	29%	30%
chnanalyst	43%	45%	40%	aeavol	35%	32%	38%	roeq	29%	26%	33%
cinvest	42%	42%	42%	beta	35%	36%	33%	mve_ia	29%	29%	29%
chpmia	42%	40%	43%	nanalyst	35%	33%	37%	roaq	29%	30%	27%
IPO	41%	42%	40%	rd_mve	35%	37%	33%	cash	28%	22%	34%
herf	40%	42%	39%	chatoia	35%	33%	36%	grCAPX	28%	29%	26%
pchcurrat	40%	42%	39%	mom36m	35%	28%	41%	dy	27%	30%	24%
secured	40%	43%	38%	ps	35%	35%	34%	acc	27%	29%	25%
convind	40%	38%	42%	rd_sale	35%	39%	30%	fgr5yr	27%	26%	28%
mom12m	40%	44%	36%	chinv	34%	28%	40%	turn	27%	26%	27%
ear	40%	42%	38%	salecash	34%	39%	29%	cashdebt	26%	26%	27%
pchcapx_ia	40%	42%	38%	chtx	33%	32%	35%	age	26%	27%	26%
pchsale_pchxsga	40%	37%	43%	sp	33%	36%	31%	egr	26%	20%	32%
securedind	40%	43%	37%	std_dolvol	33%	32%	35%	roavol	26%	29%	23%
orgcap	40%	38%	41%	realestate	33%	27%	38%	ms	26%	27%	24%
pchsale_pchrect	40%	39%	40%	cashpr	32%	34%	31%	absacc	25%	26%	24%
divi	38%	35%	41%	cfp	32%	30%	35%	mve	24%	25%	24%
chfeps	38%	39%	37%	pchsaleinv	32%	34%	31%	invest	24%	23%	24%
divo	38%	38%	38%	pctacc	32%	32%	33%	retvol	24%	24%	23%
indmom	38%	41%	35%	currat	32%	29%	35%	hire	23%	19%	27%
pchgm_pchsale	38%	38%	38%	operprof	32%	33%	31%	roic	22%	19%	25%
rsup	38%	36%	39%	bm	32%	29%	34%	gma	21%	21%	22%
sfe	38%	42%	33%	chempia	32%	25%	38%	agr	21%	16%	25%
chcsho	37%	38%	36%	grltnoa	32%	32%	31%	baspread	20%	21%	20%
pricedelay	37%	39%	35%	depr	31%	26%	37%	sgr	20%	20%	20%
rd	37%	39%	35%	zerotrade	31%	40%	22%	idiovol	18%	18%	19%
tang	37%	35%	39%	disp	31%	35%	27%	std_turn	14%	14%	14%
cfp_ia	37%	33%	40%								

The table reports weighted elastic net selection frequencies for firm characteristics in cross-sectional regressions. For each of 94 individual firm characteristics (defined in Table 1), we first estimate a cross-sectional univariate regression via weighted least squares that relates returns in month t to the characteristic values in month $t-1$; we then use the fitted cross-sectional regression and corresponding characteristic values in month t to generate cross-sectional return forecasts for month $t+1$. Using the forecasts for each month, we estimate a cross-sectional multiple regression via the weighted elastic net that relates the realized returns to the entire set of return forecasts based on the individual firm characteristics. Observations are weighted by market value at the end of the preceding month.

Table 6
Spread Portfolio Summary Statistics

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Value Weighted			Equal Weighted excl. Microcap			Equal Weighted		
Method	Mean	t-statistic	Volatility	Mean	t-statistic	Volatility	Mean	t-statistic	Volatility
Panel A: Full out-of-sample period (1990:01–2017:12)									
Conventional	0.98%	3.63***	4.95%	1.24%	4.43***	5.13%	2.95%	13.07***	4.14%
Mean	0.99%	2.29**	7.88%	1.60%	2.86***	10.27%	1.78%	4.07***	8.03%
Trimmed mean	0.95%	2.29**	7.64%	1.54%	2.81***	10.08%	1.71%	3.95***	7.95%
LASSO	1.16%	2.95***	7.20%	1.38%	2.90***	8.73%	1.93%	4.95***	7.15%
ENet	1.16%	2.96***	7.20%	1.36%	2.86***	8.73%	1.92%	4.88***	7.21%
Panel B: Pre-2003 out-of-sample period (1990:01–2002:12)									
Conventional	2.03%	4.39***	5.79%	2.59%	5.23***	6.19%	4.45%	12.68***	4.39%
Mean	1.21%	1.74*	8.61%	2.30%	2.22**	12.95%	2.41%	3.15***	9.55%
Trimmed mean	1.31%	1.93*	8.49%	2.26%	2.20**	12.80%	2.39%	3.10***	9.62%
LASSO	1.45%	2.24**	8.08%	2.24%	2.53**	11.06%	2.46%	3.63***	8.46%
ENet	1.46%	2.24**	8.11%	2.21%	2.49**	11.04%	2.48%	3.64***	8.53%
Panel C: Post-2003 out-of-sample period (2004:01–2017:12)									
Conventional	0.004%	0.01	3.86%	0.11%	0.39	3.58%	1.68%	6.31***	3.46%
Mean	0.78%	1.41	7.20%	0.93%	1.68*	7.17%	0.99%	2.03**	6.33%
Trimmed mean	0.68%	1.21	6.80%	0.81%	1.53	6.88%	0.88%	1.88*	6.03%
LASSO	0.93%	1.87*	6.42%	0.60%	1.30	5.95%	1.28%	2.93***	5.66%
ENet	0.93%	1.88*	6.41%	0.60%	1.29	5.97%	1.24%	2.81***	5.73%
Panel D: Post 2003 minus pre 2003									
Conventional	−2.03%	−3.68***		−2.48%	−4.38***		−2.77%	−6.28***	
Mean	−0.42%	−0.47		−1.37%	−1.17		−1.42%	−1.57	
Trimmed mean	−0.64%	−0.74		−1.45%	−1.25		−1.51%	−1.68*	
LASSO	−0.52%	−0.64		−1.64%	−1.65*		−1.18%	−1.46	
ENet	−0.53%	−0.65		−1.61%	−1.61		−1.24%	−1.53	

The table reports summary statistics for spread portfolios formed from out-of-sample cross-sectional return forecasts based on combination approaches. For each of 94 individual firm characteristics (defined in Table 1), we first estimate a cross-sectional univariate regression that relates returns in month t to the characteristic values in month $t-1$; we then use the fitted cross-sectional regression and corresponding characteristic values in month t to generate cross-sectional return forecasts for month $t+1$. The "Mean" ("Trimmed mean") combination forecast is the arithmetic mean of the return forecasts based on the individual characteristics (after excluding the 5% smallest and 5% largest of the individual return forecasts); the "LASSO" ("ENet") combination forecast is the average of the return forecasts based on the individual characteristics selected by the LASSO (elastic net). "Conventional" refers to a forecast based on a conventional multiple regression approach (e.g., Green, Hand, and Zhang 2017). Using the forecasts for each month, we form a spread portfolio that goes long (short) the top (bottom) decile of stocks sorted according to the cross-sectional return forecasts. For the t-statistics, *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively. "Value-Weighted" indicates that the cross-sectional regressions are estimated via weighted least squares with observations weighted by market value at the end of the preceding month, and the long and short legs of the spread portfolio are value weighted; "Equal Weighted" ("Equal Weighted excl. Microcap") indicates that the cross-sectional regressions are estimated via ordinary least squares (excluding stocks with market value below the NYSE 20% percentile), and the long and short legs of the spread portfolio are equal weighted.

Table 7
Multifactor Model Estimation Results, 1990:01 to 2017:12

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Conventional		Mean		Trimmed mean		LASSO		ENet	
Factor	Coeff.	t-stat.	Coeff.	t-stat.	Coeff.	t-stat.	Coeff.	t-stat.	Coeff.	t-stat.
Panel A: Carhart (1997) four-factor model										
α	0.62%	2.91***	1.31%	2.98***	1.24%	2.94***	1.20%	3.09***	1.20%	3.09***
MKT	0.16	3.01***	−0.36	−3.31***	−0.35	−3.31***	−0.25	−2.60***	−0.25	−2.55**
SMB	0.03	0.49	−0.06	−0.43	−0.13	−0.97	−0.09	−0.74	−0.09	−0.74
HML	−0.31	−4.23***	−0.21	−1.38	−0.22	−1.51	−0.16	−1.16	−0.16	−1.21
UMD	0.62	13.39***	−0.06	−0.63	0.01	0.07	0.33	3.91***	0.33	3.91***
Panel B: Fama and French (2015) five-factor model										
α	0.93%	3.44***	1.15%	2.58***	1.13%	2.62***	1.33%	3.29***	1.33%	3.29***
MKT	0.04	0.52	−0.26	−2.17**	−0.26	−2.23**	−0.27	−2.53**	−0.27	−2.48**
SMB	0.16	1.69*	−0.13	−0.86	−0.22	−1.43	−0.11	−0.79	−0.11	−0.78
HML	−0.73	−6.04***	−0.44	−2.19**	−0.50	−2.60***	−0.53	−2.93***	−0.54	−2.98***
RMW	0.27	2.21**	−0.08	−0.39	−0.09	−0.45	−0.02	−0.09	−0.01	−0.07
CMA	0.27	1.53	0.67	2.30**	0.76	2.73***	0.65	2.45**	0.65	2.47**
Panel C: Hou, Xue, and Zhang (2015) q-factor model										
α	0.62%	2.26***	1.36%	2.92***	1.30%	2.90***	1.36%	3.20***	1.36%	3.20***
MKT	0.10	1.42	−0.35	−2.96***	−0.34	−2.96***	−0.33	−3.02***	−0.32	−2.98***
ME	0.33	3.92***	−0.13	−0.91	−0.18	−1.28	−0.05	−0.39	−0.05	−0.38
I/A	−0.41	−3.05***	0.17	0.71	0.18	0.82	0.07	0.33	0.06	0.29
ROE	0.70	6.53***	−0.30	−1.67*	−0.23	−1.34	−0.04	−0.25	−0.04	−0.22

The table reports multifactor model estimation results for spread portfolios formed from out-of-sample cross-sectional return forecasts based on combination approaches. For each of 94 individual firm characteristics (defined in Table 1), we first estimate a cross-sectional univariate regression that relates returns in month t to the characteristic values in month $t-1$; we then use the fitted cross-sectional regression and corresponding characteristic values in month t to generate cross-sectional return forecasts for month $t+1$. The "Mean" ("Trimmed mean") combination forecast is the arithmetic mean of the return forecasts based on the individual characteristics (after excluding the 5% smallest and 5% largest of the individual return forecasts); the "LASSO" ("ENet") combination forecast is the average of the return forecasts based on the individual characteristics selected by the LASSO (elastic net). "Conventional" refers to a forecast based on a conventional multiple regression approach (e.g., Green, Hand, and Zhang 2017). Using the forecasts for each month, we form a spread portfolio that goes long (short) the top (bottom) decile of stocks sorted according to the cross-sectional return forecasts. For the t-statistics, *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively. The cross-sectional regressions are estimated via weighted least squares with observations weighted by market value at the end of the preceding month, and the long and short legs of the spread portfolio are value weighted. The factors are defined as follows: MKT = CRSP value-weighted market excess return; SMB (HML) = Fama and French (1993) "small minus big" size ("high minus low" value) factor; UMD = "up minus down" momentum factor; RMW (CMA) = Fama and French (2015) "robust minus weak" profitability ("conservative minus aggressive" investment) factor; ME, I/A, and ROE = Hou, Xue, and Zhang (2015) size, investment, and return on equity factors, respectively. The results in Panel C are for 1990:01 to 2016:12.

Table 8
Risk-Adjusted Average Returns for Spread Portfolios, 1990:01 to 2017:12

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
	Value Weighted				Equal Weighted excl. Microcap				Equal Weighted			
	$\hat{\alpha}$		$\hat{\Delta}$		$\hat{\alpha}$		$\hat{\Delta}$		$\hat{\alpha}$		$\hat{\Delta}$	
Method	Coeff.	t-stat.	Coeff.	t-stat.	Coeff.	t-stat.	Coeff.	t-stat.	Coeff.	t-stat.	Coeff.	t-stat.
Panel A: Carhart (1997) four-factor model												
Conventional	1.41%	4.79***	−1.55%	−3.81***	1.55%	6.10***	−1.55%	−4.39***	4.05%	14.21***	−2.31%	−5.86***
Mean	1.56%	2.52**	−0.51%	−0.59	2.67%	3.32***	−1.39%	−1.25	2.67%	4.26***	−1.47%	−1.69*
Trimmed mean	1.56%	2.60***	−0.61%	−0.74	2.68%	3.39***	−1.51%	−1.38	2.62%	4.21***	−1.53%	−1.78*
LASSO	1.30%	2.37**	−0.21%	−0.27	2.05%	3.08***	−1.22%	−1.32	2.35%	4.38***	−0.91%	−1.23
ENet	1.31%	2.39**	−0.22%	−0.29	2.01%	3.02***	−1.18%	−1.28	2.37%	4.39***	−0.97%	−1.30
Panel B: Fama and French (2015) five-factor model												
Conventional	1.99%	5.42***	−2.07%	−4.14***	2.43%	6.29***	−2.25%	−4.27***	4.36%	14.28***	−2.63%	−6.34***
Mean	1.27%	2.03**	−0.23%	−0.28	2.36%	2.95***	−0.96%	−0.88	2.33%	3.78***	−1.13%	−1.35
Trimmed mean	1.32%	2.19**	−0.37%	−0.45	2.40%	3.06***	−1.08%	−1.02	2.29%	3.76***	−1.20%	−1.45
LASSO	1.50%	2.65***	−0.32%	−0.42	2.09%	3.11***	−1.19%	−1.29	2.47%	4.51***	−0.95%	−1.28
ENet	1.50%	2.65***	−0.33%	−0.43	2.06%	3.06***	−1.16%	−1.26	2.50%	4.53***	−1.01%	−1.35
Panel C: Hou, Xue, and Zhang (2015) q-factor model												
Conventional	1.57%	4.17***	−1.86%	−3.62***	1.84%	4.93***	−1.99%	−3.89***	4.18%	13.46***	−2.45%	−5.78***
Mean	1.60%	2.46**	−0.46%	−0.53	2.81%	3.34***	−1.42%	−1.23	2.46%	3.77***	−1.28%	−1.44
Trimmed mean	1.59%	2.55**	−0.57%	−0.67	2.85%	3.46***	−1.53%	−1.36	2.40%	3.71***	−1.33%	−1.51
LASSO	1.61%	2.72***	−0.49%	−0.61	2.26%	3.19***	−1.46%	−1.50	2.37%	4.11***	−0.99%	−1.25
ENet	1.62%	2.73***	−0.51%	−0.63	2.23%	3.14***	−1.43%	−1.48	2.40%	4.12***	−1.04%	−1.31

The table reports estimates of alpha ($\hat{\alpha}$) and the change in alpha after 2003 ($\hat{\Delta}$) in multifactor models for spread portfolios formed from out-of-sample cross-sectional return forecasts based on combination approaches. For each of 94 individual firm characteristics (defined in Table 1), we first estimate a cross-sectional univariate regression that relates returns in month t to the characteristic values in month $t-1$; we then use the fitted cross-sectional regression and corresponding characteristic values in month t to generate cross-sectional return forecasts for month $t+1$. The "Mean" ("Trimmed mean") combination forecast is the arithmetic mean of the return forecasts based on the individual characteristics (after excluding the 5% smallest and 5% largest of the individual return forecasts); the "LASSO" ("ENet") combination forecast is the average of the return forecasts based on the individual characteristics selected by the LASSO (elastic net). "Conventional" refers to a forecast based on a conventional multiple regression approach (e.g., Green, Hand, and Zhang 2017). Using the forecasts for each month, we form a spread portfolio that goes long (short) the top (bottom) decile of stocks sorted according to the cross-sectional return forecasts. For the t-statistics, *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively. "Value-Weighted" indicates that the cross-sectional regressions are estimated via weighted least squares with observations weighted by market value at the end of the preceding month, and the long and short legs of the spread portfolio are value weighted; "Equal Weighted" ("Equal Weighted excl. Microcap") indicates that the cross-sectional regressions are estimated via ordinary least squares (excluding stocks with market value below the NYSE 20% percentile), and the long and short legs of the spread portfolio are equal weighted. The results in Panel C are for 1990:01 to 2016:12.

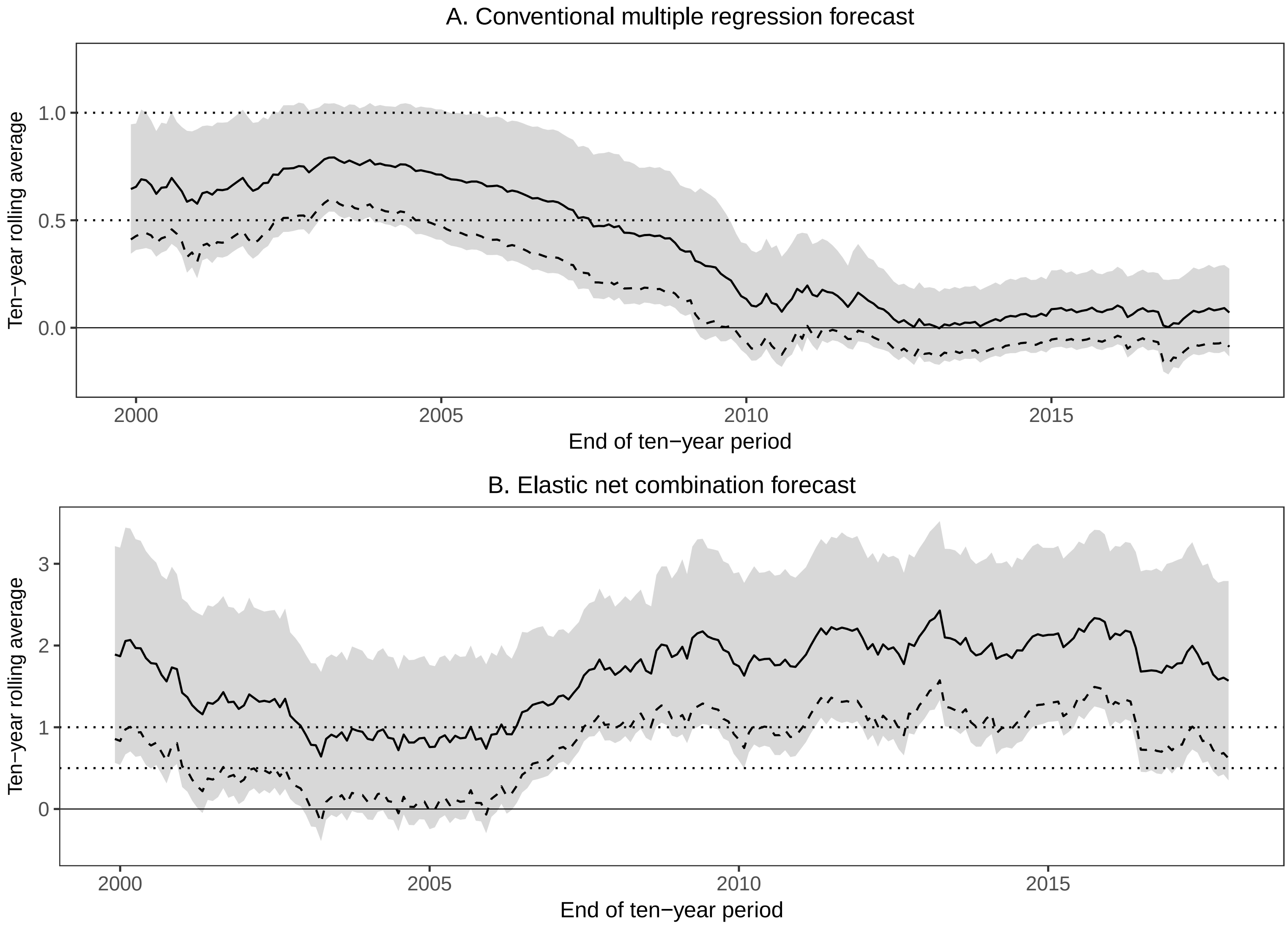

Figure 1. Ten-year rolling averages of slope coefficients for Fama-MacBeth cross-sectional regressions. The solid lines show ten-year moving averages of slope coefficient estimates for Fama-MacBeth cross-sectional univariate regressions that relate month-t realized returns to return forecasts based on two approaches. Panel A (B) refers to a forecast based on a conventional multiple regression (elastic net combination) approach. The cross-sectional univariate regressions are estimated using weighted least squares, where the observations are weighted by market value at the end of the preceding month. Bands delineate two-sided 90% confidence intervals based on Newey and West (1987) standard errors (computed using twelve lags). The dashed lines delineate lower bounds for one-sided, upper-tailed 90% confidence intervals.

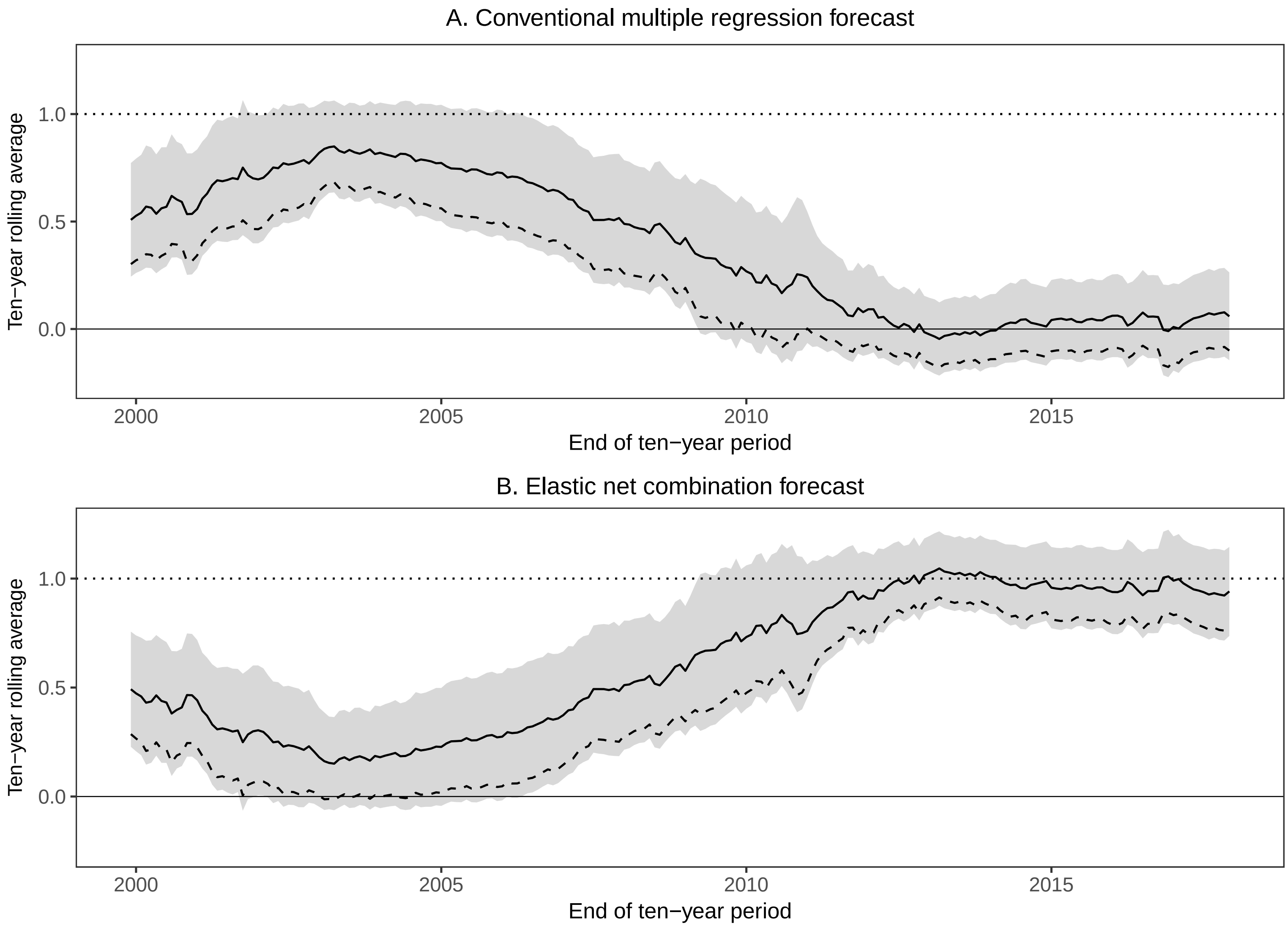

Figure 2. Ten-year rolling averages of optimal composite forecast coefficients.

The figure shows ten-year moving averages of estimated coefficients attached to a conventional multiple regression forecast (Panel A) and elastic net combination forecast (Panel B) in an optimal composite forecast. The coefficients are estimated using a weighted least squares approach, where the observations are weighted by market value at the end of the preceding month. Bands delineate two-sided 90% confidence intervals based on Newey and West (1987) standard errors (computed using twelve lags). The dashed lines delineate lower bounds for one-sided, upper-tailed 90% confidence intervals.

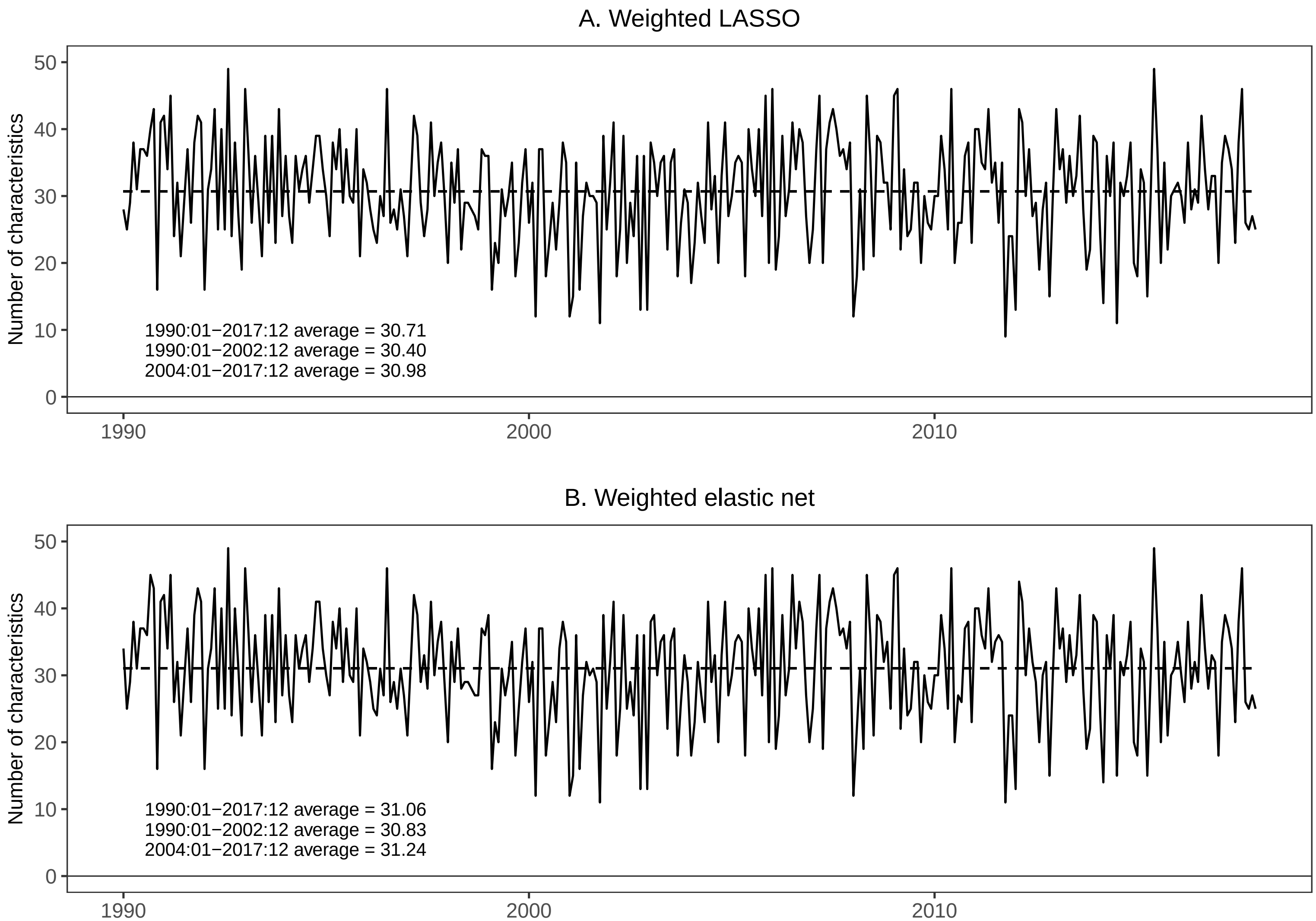

Figure 3. Number of characteristics selected by the LASSO and elastic net.

The figure shows the number of characteristics selected by the weighted LASSO (Panel A) and weighted elastic net (Panel B) for a cross-sectional multiple regression that relates month-t realized returns to the entire set of return forecasts based on 94 individual firm characteristics (defined in Table 1). Observations are weighted in the LASSO and elastic net by market value at the end of the preceding month. Horizontal dashed lines delineate the average number of characteristics selected for the 1990:01 to 2017:12 period.

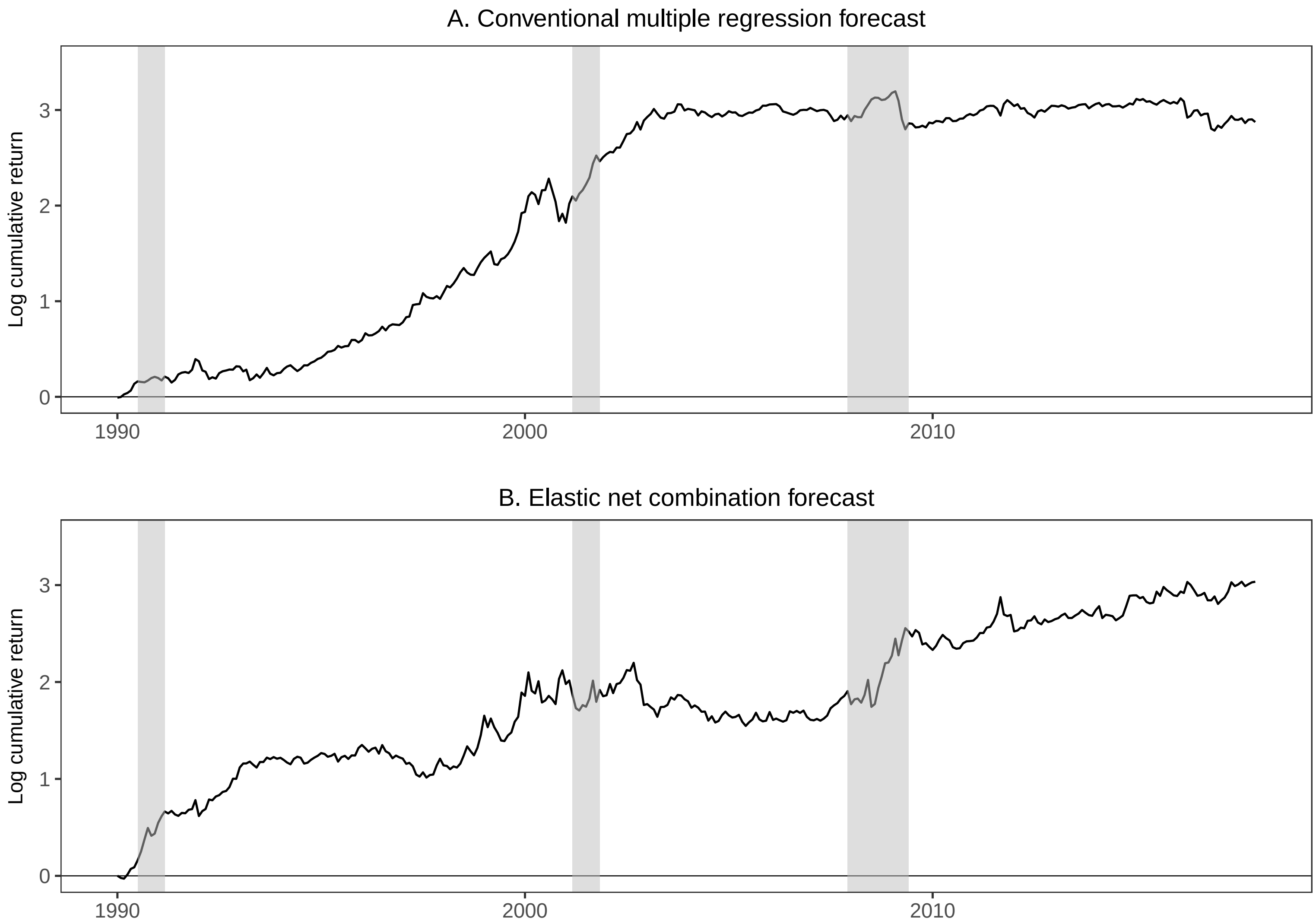

Figure 4. Log cumulative returns for spread portfolios.

The figure shows log cumulative returns for spread portfolios formed from cross-sectional return forecasts based on a conventional multiple regression approach (Panel A) or elastic net combination approach (Panel B). The spread portfolio goes long (short) firms in the top (bottom) decile of sorted cross-sectional return forecasts. The long and short legs of the spread portfolio are value weighted. Vertical bars delineate business-cycle recessions as dated by the National Bureau of Economic Research.