1

Time-series and Cross-sectional Stock Return Forecasting: New Machine Learning Methods

This chapter extends the machine learning methods developed in Han *et al*. (2019) for forecasting cross-sectional stock returns to a time-series context. The methods use the elastic net to refine the simple combination return forecast from Rapach *et al*. (2010). In a time-series application focused on forecasting the US market excess return using a large number of potential predictors, we find that the elastic net refinement substantively improves the simple combination forecast, thereby providing one of the best market excess return forecasts to date. We also discuss the cross-sectional return forecasts developed in Han *et al*. (2019), highlighting how machine learning methods can be used to improve combination forecasts in both the time-series and cross-sectional dimensions. Overall, because many important questions in finance are related to time-series or cross-sectional return forecasts, the machine learning methods discussed in this chapter should provide valuable tools to researchers and practitioners alike.

1.1. Introduction

Researchers in finance increasingly rely on machine learning techniques to analyze Big Data. The initial application of the *least absolute shrinkage and selection operator* (Tibshirani 1996, LASSO) – one of the most popular machine learning techniques – in finance appears to be Rapach *et al*. (2013), who analyze lead-lag relationships among monthly international equity

---

Chapter written by David E. RAPACH and Guofu ZHOU.

Machine Learning for Asset Management: New Developments and Financial Applications, First Edition. Edited by Emmanuel Jurczenko.

returns in a high-dimensional setting. More recently, Gu *et al*. (2019) employ a comprehensive set of machine learning tools, including the LASSO, to analyze the time-series predictability of monthly individual stock returns, while Chinco *et al*. (2019) use the LASSO to predict individual stock returns one minute ahead. Freyberger *et al*. (2019) apply a nonparametric version of the LASSO to accommodate nonlinear relationships between numerous firm characteristics and cross-sectional stock returns. Kozak *et al*. (2019) use the LASSO in a Bayesian context to model the stochastic discount factor based on a large number of firm characteristics. Incorporating insights from Bates and Granger (1969); Diebold and Shin (2019), Han *et al*. (2019) propose procedures for forecasting cross-sectional returns using the information in more than 100 firm characteristics[1].

In this chapter, we show how the approach of Han *et al*. (2019), originally designed for forecasting cross-sectional stock returns, can be modified for time-series forecasting of the market excess return. A voluminous literature investigates market excess return predictability based on a wide variety of predictor variables[2]. In the presence of a large number of potential predictor variables, conventional forecasting methods are susceptible to in-sample overfitting, which often translates into poor out-of-sample performance. Rapach *et al*. (2010) employ forecast combination (Bates and Granger 1969) to incorporate the information in a large number of predictor variables in a manner that guards against overfitting. They find that a simple combination forecast – the average of univariate predictive regression forecasts based on the individual predictor variables – substantially improves out-of-sample forecasts of the US market excess return. Extending the methods of Han *et al*. (2019) to a time-series context along the lines of Diebold and Shin (2019), we describe how the *elastic net* (Zou and Hastie 2005), a well-known variant of the LASSO, can be used to refine the simple combination forecast, resulting in what we call the *combination elastic net* forecast. Intuitively, as explained by Han *et al*. (2019), the elastic net refinement allows us to more efficiently use the information in the predictor variables by selecting the most relevant predictors to include in the combination forecast. In an empirical application, we show that the combination elastic net approach indeed improves the

1 We focus on numerical data in this chapter. For textual analysis see, for example, Tetlock (2007); Loughran and McDonald (2011); Ke *et al*. (2019).

2 See Rapach and Zhou (2013) for a survey of the literature.

accuracy of US market excess return forecasts and provides substantive economic value to a mean-variance investor. Overall, our combination elastic net forecast appears to be among the best market excess return forecasts to date.

The rest of the chapter is organized as follows. Section 1.2 describes the construction of market excess return forecasts, including the combination elastic net forecast. Section 1.3 reports results for an empirical application centered on forecasting the US market excess return, using a variety of predictor variables from the literature. Section 1.4 outlines the construction of the cross-sectional return forecasts proposed by Han *et al*. (2019). Section 1.5 concludes this chapter.

1.2. Time-series return forecasts

1.2.1. *Predictive regression*

Stock market excess return predictability is typically analyzed in the context of a univariate predictive regression model:

$$r_t = \alpha + \beta x_{j,t-1} + \varepsilon_t, \qquad [1.1]$$

where r_t is the period-t return on a broad stock market index in excess of the risk-free return, $x_{j,t}$ is the predictor variable, and ε_t is a zero-mean disturbance term. It is straightforward to use equation [1.1] to generate an out-of-sample forecast of r_{t+1} based on $x_{j,t}$ and data available through period t:

$$\hat{r}^{(j)}_{t+1|t} = \hat{\alpha}^{(j)}_{1:t} + \hat{\beta}^{(j)}_{1:t} x_{j,t}, \qquad [1.2]$$

where $\hat{\alpha}^{(j)}_{1:t}$ and $\hat{\beta}^{(j)}_{1:t}$ are the ordinary least squares (OLS) estimates of α and β, respectively, in equation [1.1] based on data available from the start of the sample through t (i.e. the period of forecast formation).

Because there are a plethora of plausible predictor variables, it is advisable to aggregate information when forecasting the market excess return. The most obvious approach for incorporating information from multiple predictor variables is to specify a multiple predictive regression model:

$$r_t = \alpha + \sum_{j=1}^{J} \beta_j x_{j,t-1} + \varepsilon_t. \qquad [1.3]$$

It is again straightforward to use equation [1.3] to generate an out-of-sample forecast of r_{t+1} based on $x_{j,t}$ for $j = 1, \ldots, J$ and data available through t:

$$\hat{r}_{t+1|t}^{\text{OLS}} = \hat{\alpha}_{1:t}^{\text{OLS}} + \sum_{j=1}^{J} \hat{\beta}_{j,1:t}^{\text{OLS}} x_{j,t}, \qquad [1.4]$$

where $\hat{\alpha}_{1:t}^{\text{OLS}}$ and $\hat{\beta}_{j,1:t}^{\text{OLS}}$ are the OLS estimates of α and β_j, respectively, for $j = 1, \ldots, J$ in equation [1.3] based on data available through t.

Although the out-of-sample market excess return forecasts in equation [1.2] and [1.4] are easy to obtain, Goyal and Welch (2008) find that such forecasts, based on numerous popular predictor variables from the literature, fail to outperform the naive prevailing mean benchmark forecast on a consistent basis over time (as judged by the out-of-sample R^2 statistic, which we define below). The prevailing mean forecast ignores information in any predictor variable; it is simply the historical average excess return based on data available through t:

$$\hat{r}_{t+1|t}^{\text{PM}} = \frac{1}{t} \sum_{s=1}^{t} r_s. \qquad [1.5]$$

The prevailing mean forecast corresponds to the following simple data-generating process for the market excess return:

$$r_t = \alpha + \varepsilon_t, \qquad [1.6]$$

namely, the constant expected excess return (or random walk with drift) model. The Goyal and Welch (2008) findings pose important challenges for out-of-sample return predictability, as they indicate that exploiting the information in popular predictor variables via conventional regression methods does not improve forecast accuracy.

1.2.2. *Forecast combination*

The study of Goyal and Welch (2008) was influential in stimulating thinking about how to better use the information in predictor variables to forecast the market excess return. With respect to the univariate predictive regression forecast in equation [1.2], Rapach *et al*. (2010) argue that it is risky to rely on a single predictor variable, due to factors such as investor learning and structural change. Building on the seminal work of Bates and Granger (1969), Rapach *et al*. (2010) recommend forecast combination as a strategy for incorporating information from a variety of predictor variables. Forecast combination reduces forecast "risk" by diversifying across individual forecasts, similarly to diversifying across assets to reduce portfolio risk (Timmermann 2006). Specifically, Rapach *et al*. (2010) consider a combination forecast that takes the form of a simple average of the univariate predictive regression forecasts, based on $x_{j,t}$ for $j = 1, \ldots, J$ in equation [1.2]:

$$\hat{r}^{\mathrm{C}}_{t+1|t} = \frac{1}{J} \sum_{j=1}^{J} \hat{r}^{(j)}_{t+1|t}. \qquad [1.7]$$

They show that, in contrast to the conventional univariate and multiple predictive regression forecasts in equations [1.2] and [1.4], respectively, the combination forecast in equation [1.7] is able to deliver out-of-sample accuracy gains relative to the prevailing mean forecast, on a much more consistent basis over time.

How is it that – unlike the conventional multiple predictive regression forecast in equation [1.4], which also includes information from $x_{j,t}$ for $j = 1, \ldots, J$ – the combination forecast in equation [1.7] is able to improve out-of-sample performance? Rapach *et al*. (2010) point out that forecast combination is effectively a strong shrinkage estimator. They show that the combination forecast in equation [1.7] makes two adjustments to the conventional multiple predictive regression forecast in equation [1.4]: first, it replaces the OLS multiple regression coefficient estimates with their univariate counterparts, which reduces the role of multi-collinearity in producing imprecise parameter estimates; second, the combination forecast shrinks the univariate slope coefficients by the factor $1/J$, thereby shrinking the forecast to the prevailing mean benchmark.

The usefulness of shrinkage for improving out-of-sample market excess return forecasts stems from a delicate balance required for stock return forecasting. On the one hand, we want to incorporate information from a wide variety of potentially relevant predictor variables, especially since we do not want to neglect relevant information and cannot know *a priori* which predictors are the most relevant. On the other hand, incorporating information from numerous predictors via the multiple prediction regression forecast in equation [1.4] is inadvisable. Equation [1.4] is based on conventional estimation of the multiple predictive regression model in equation [1.3], which is susceptible to overfitting. Conventional OLS estimation maximizes the explanatory ability of the model over the estimation sample, which often leads to poor out-of-sample performance. Overfitting concerns are exacerbated as the number of explanatory variables increases and the signal-to-noise ratio in the data decreases. We encounter both of these challenges when forecasting stock returns: there are numerous plausible predictor variables, and the predictable component in returns is inherently limited. The combination forecast in equation [1.7] apparently provides an effective shrinkage strategy for incorporating information from numerous plausible predictor variables in a manner that avoids overfitting.

1.2.3. *Elastic net*

Machine learning techniques also provide a means for implementing shrinkage. Indeed, the popular LASSO estimator is a penalized regression approach that is explicitly designed to prevent overfitting via shrinkage. To compute a forecast based on the multiple predictive regression model in equation [1.3], instead of the OLS objective function, we estimate the coefficients using the LASSO objective function:

$$\underset{\alpha,\beta_1,\ldots,\beta_J\in\mathbb{R}}{\arg\min}\left[\frac{1}{2t}\sum_{s=1}^{t}\left(r_s-\alpha-\sum_{j=1}^{J}\beta_j x_{j,s-1}\right)^2+\lambda\sum_{j=1}^{J}|\beta_j|\right], \quad [1.8]$$

where $\lambda \geq 0$ is a regularization parameter that controls the degree of shrinkage[3]. The first component of the LASSO objective function is the

3 Following standard practice, the predictor variables are standardized to have zero mean and unit variance before entering equation [1.8]. The final parameter estimates reflect the original scales of the predictor variables.

familiar sum of squared fitted residuals, so that the LASSO and OLS estimators coincide when $\lambda = 0$. The regularization parameter λ shrinks the coefficients towards zero. Unlike rigde regression (Hoerl and Kennard 1970), which relies on an ℓ_2 penalty term, the LASSO employs an ℓ_1 penalty, so that it permits shrinkage to exactly zero (for sufficiently large λ). Shrinkage to zero means that the LASSO also performs variable selection, which facilitates the interpretation of the fitted model.

To implement LASSO estimation, it is necessary to choose the value for λ. The most popular approach is K-fold cross-validation. However, the selection of the number of folds K and construction of the folds are largely arbitrary. The Hurvich and Tsai (1989) corrected version of the Akaike information criterion (Akaike 1973, AIC) provides an alternative to K-fold cross-validation for choosing λ. The corrected AIC is simpler to use in that it does not require arbitrary choices for the number and type of folds. Furthermore, Flynn *et al.* (2013) show that the corrected AIC has good asymptotic and finite-sample properties for choosing λ.

Zou and Hastie (2005) propose the elastic net (ENet) as a refinement to the LASSO that includes both ℓ_1 and ℓ_2 components in the penalty term. The ENet estimator is defined by the following objective function:

$$\underset{\alpha,\beta_1,\ldots,\beta_J\in\mathbb{R}}{\arg\min}\left[\frac{1}{2t}\sum_{s=1}^{t}\left(r_s-\alpha-\sum_{j=1}^{J}\beta_j x_{j,s-1}\right)^2+\lambda P_\delta\left(\beta_1,\ldots,\beta_J\right)\right], \quad [1.9]$$

where

$$P_\delta\left(\beta_1,\ldots,\beta_J\right)=0.5(1-\delta)\sum_{j=1}^{J}\beta_j^2+\delta\sum_{j=1}^{J}|\beta_j| \quad [1.10]$$

and $0 \leq \delta \leq 1$ is a parameter for blending the ℓ_1 and ℓ_2 components. A potential drawback of the LASSO is that it tends to somewhat arbitrarily select one predictor from a group of highly correlated predictors. In contrast, using $\delta = 0.5$ in equation [1.9] results in a stronger tendency to select the highly correlated predictors as a group (Hastie and Qian 2016). The corrected AIC can again be used to choose λ in equation [1.9].

A market excess return forecast based on ENet estimation of the multiple predictive regression model in equation [1.3] is given by:

$$\hat{r}_{t+1|t}^{\mathrm{ENet}} = \hat{\alpha}_{1:t}^{\mathrm{ENet}} + \sum_{j=1}^{J} \beta_{j,1:t}^{\mathrm{ENet}} x_{j,t}, \qquad [1.11]$$

where $\hat{\alpha}_{1:t}^{\mathrm{ENet}}$ and $\hat{\beta}_{j,1:t}^{\mathrm{ENet}}$ are the ENet estimates of α and β_j, respectively, for $j = 1, \ldots, J$ in equation [1.3]. Intuitively, we rely on the shrinkage properties of the ENet to generate a market excess return forecast that incorporates information from a potentially large number of predictor variables in a manner that guards against overfitting. Whether the ENet is an effective shrinkage strategy for forecasting the market excess return is ultimately an empirical issue. We investigate this issue in our empirical application in section 1.3[4].

1.2.4. *Combination elastic net*

Incorporating insights from Diebold and Shin (2019), we can also use machine learning techniques to refine the combination forecast in equation [1.7]. A potential drawback to equation [1.7] is that it may "overshrink" the forecast to the prevailing mean, thereby neglecting substantive relevant information in the predictor variables. In an effort to improve the combination forecast by exploiting more of the relevant information in the predictor variables (while still avoiding overfitting), we consider the following Granger and Ramanathan (1984) regression:

$$r_t = \eta + \sum_{j=1}^{J} \theta_j \hat{r}_{t|t-1}^{(j)} + \varepsilon_t, \qquad [1.12]$$

which we estimate via the elastic net to select the most relevant univariate forecasts to include in the combination forecast[5]. Specifically, to construct the combination elastic net (C-ENet) forecast, we first need to define an initial

4 We focus on the results for the ENet in our empirical application in section 1.3, although the results are qualitatively similar for the LASSO.

5 Again, the results are qualitatively similar in our empirical application in section 1.3 if we use the LASSO to estimate equation [1.12].

in-sample estimation period and corresponding holdout out-of-sample period; let t_1 denote the size of the initial in-sample period. We then proceed in three steps:

Step 1 For each predictor variable, we compute recursive univariate predictive regression forecasts based on equation [1.2] over the holdout out-of-sample period:

$$\hat{r}^{(j)}_{s|s-1} = \hat{\alpha}^{(j)}_{1:s-1} + \hat{\beta}^{(j)}_{1:s-1} x_{j,s-1}, \qquad [1.13]$$

for $s = t_1 + 1, \ldots, t$ and $j = 1, \ldots, J$.

Step 2 We estimate the Granger and Ramanathan (1984) regression in equation [1.12] via the ENet over the holdout out-of-sample period:

$$r_s = \eta + \sum_{j=1}^{J} \theta_j \hat{r}^{(j)}_{s|s-1} + \varepsilon_s, \qquad [1.14]$$

for $s = t_1 + 1, \ldots, t$. Let $\mathcal{J}_t \subseteq \{1, \ldots, J\}$ denote the index set of individual univariate predictive regression forecasts selected by the ENet in equation [1.14]. When estimating equation [1.14], we impose the restriction that $\theta_j \geq 0$ for $j = 1, \ldots, J$. This imposes the economically reasonable requirement that a univariate market excess return forecast be positively related to the realized excess return in order to be selected by the ENet in equation [1.14].

Step 3 We compute the C-ENet forecast as:

$$\hat{r}^{\text{C-ENet}}_{t+1|t} = \frac{1}{|\mathcal{J}_t|} \sum_{j \in \mathcal{J}_t} \hat{r}^{(j)}_{t+1|t}, \qquad [1.15]$$

where $|\mathcal{J}_t|$ is the cardinality of $\mathcal{J}_t$ and $\hat{r}^{(j)}_{t+1|t}$ is given by equation [1.2] for $j = 1, \ldots, J$.

The usefulness of the C-ENet approach for capturing the relevant information in numerous predictor variables, in a manner that guards against overfitting, is again ultimately an empirical issue. In our empirical application in section 1.3, we find that the C-ENet approach is indeed an effective strategy for forecasting the market excess return[6].

6 Observe that all the forecasts that we construct only use data available through t to forecast r_{t+1}, so that the forecasts do not entail "look-ahead" bias.

1.3. Empirical application

1.3.1. *Data*

We investigate the performance of the strategies discussed in section 1.2 for forecasting the monthly S&P 500 excess return. Using data available from Amit Goyal's website[7], we measure the excess return as the CRSP value-weighted S&P 500 return in excess of the risk-free return (based on the Treasury bill rate).

We consider 12 plausible predictor variables, which are illustrative of popular predictors used by academics and practitioners alike:

– *Log dividend-price ratio* (*DP*). Log of the 12-month moving sum of S&P 500 dividends minus the log of the S&P 500 price index.

– *Log earnings-price ratio* (*EP*). Log of the 12-month moving sum of S&P 500 earnings minus the log of the S&P 500 price index.

– *Volatility* (*VOL*). We follow Mele (2007) in measuring the annualized volatility for month t as $\sqrt{\frac{\pi}{2}}\sqrt{12}\hat{\sigma}_t$, where $\hat{\sigma}_t = \frac{1}{12}\sum_{s=1}^{12}|r_{t-(s-1)}|$.

– *Treasury bill yield* (*BILL*). Three-month Treasury bill yield minus the 12-month moving average of the three-month Treasury bill yield.

– *Treasury bond yield* (*BOND*). Ten-year Treasury bond yield minus the 12-month moving average of the ten-year Treasury bond yield.

– *Term spread* (*TERM*). Difference in yields on a ten-year treasury bond and a three-month treasury bill.

– *Credit spread* (*CREDIT*). Difference in yields on a AAA-rated corporate bond and a ten-year treasury bond.

– *Inflation* (*PPIG*). Producer price index (PPI) inflation rate.

– *Industrial production growth* (*IPG*). Growth rate of industrial production.

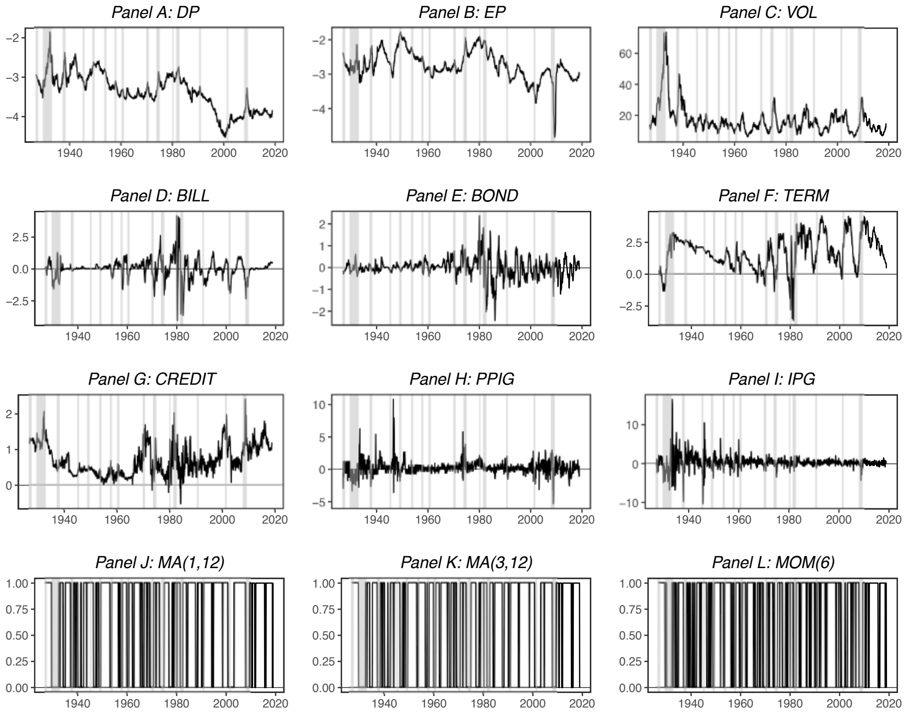
– *MA(1,12) technical signal* [*MA(1,12)*]. An indicator variable that takes a value of one (zero) if the S&P 500 price index is greater than or equal to (less than) the 12-month moving average of the S&P 500 price index.

– *MA(3,12) technical signal* [*MA(3,12)*]. An indicator variable that takes a value of one (zero) if the three-month moving average of the S&P 500 price

7 http://www.hec.unil.ch/agoyal/.

index is greater than or equal to (less than) the 12-month moving average of the S&P 500 price index.

– *Momentum technical signal* [*MOM(6)*]. An indicator variable that takes a value of one (zero) if the S&P 500 price index is greater than or equal to (less than) its value six months ago.



The figure depicts 12 predictor variables for 1927:01 to 2018:12. The predictor variable definitions are provided in section 1.3.1. Vertical bars delineate business-cycle recessions as dated by the National Bureau of Economic Research.

Figure 1.1. *Predictor variables*

The data used to construct the predictor variables are from Amit Goyal's website and the Federal Reserve Bank of St. Louis's Federal Reserve Economic Data (FRED)[8]. We account for the one-month publication lag in *PPIG* and *IPG*. We follow Neely *et al.* (2014) in defining indicator variables to include information from technical signals. Figure 1.1 portrays the 12

8 https://fred.stlouisfed.org/.

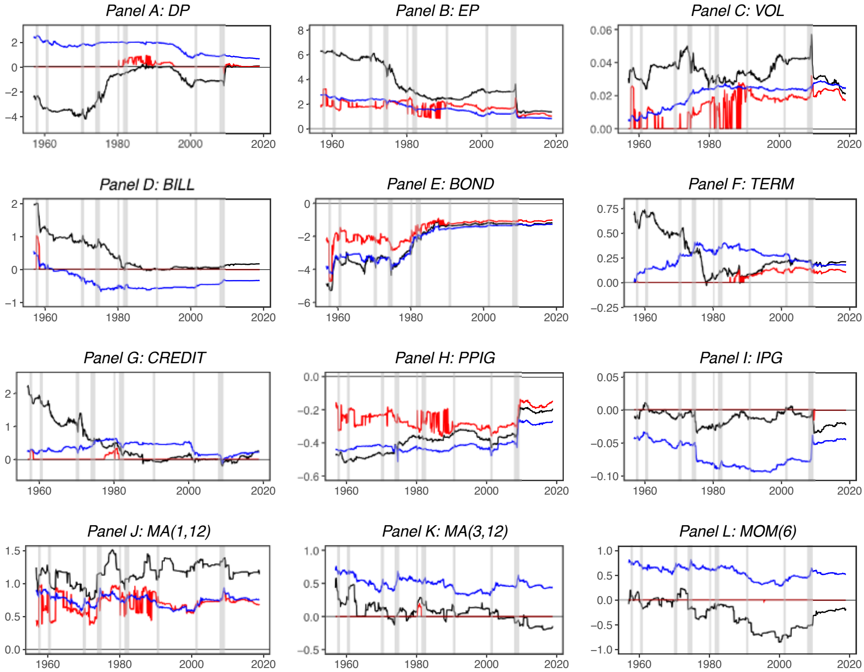
predictor variables for the 1927:01 to 2018:12 sample period. Visually, the predictors represent a variety of information sources.

1.3.2. *Forecasts*

We reserve the first two decades of the sample (1927:01 to 1946:12) as the initial in-sample estimation period. This provides us with an adequate number of observations to reasonably reliably estimate the predictive regression coefficients in equations [1.1] and [1.3]. The initial holdout out-of-sample period for computing the C-ENet forecast covers 1947:01 to 1956:12, so that we evaluate the out-of-sample forecasts for 1957:01 to 2018:12. The out-of-sample forecast evaluation period covers more than six decades, which allows us to analyze return predictability under a variety of economic conditions.

Figure 1.2 depicts the recursive slope coefficient estimates used to compute the predictive regression forecasts in equations [1.2], [1.4] and [1.11]. The black line in each panel delineates the OLS slope coefficient estimates for the multiple predictive regression model in equation [1.3]. The recursive estimates point to problems with the conventional OLS estimates of the multiple predictive regression model slope coefficients – the estimates often have the "wrong" sign (e.g. DP and $BILL$) and reach extreme values, which are manifestations of in-sample overfitting. Overfitting is not surprising when we rely on conventional methods to estimate a relatively high-dimensional predictive regression model in a noisy environment.

The blue line in each panel of Figure 1.2 delineates the recursive OLS slope coefficient estimates for the univariate predictive regression model in equation [1.1]. Compared to the recursive slope coefficient estimates for the multiple predictive regression model, the univariate estimates are generally much more stable. This reflects the increase in estimation precision afforded by the mitigation of multi-collinearity. Of course, the univariate estimates are potentially biased (due to omitted variable bias). However, in light of the bias-efficiency trade-off, the increase in estimation precision can outweigh the cost of the bias for the purpose of out-of-sample forecasting.

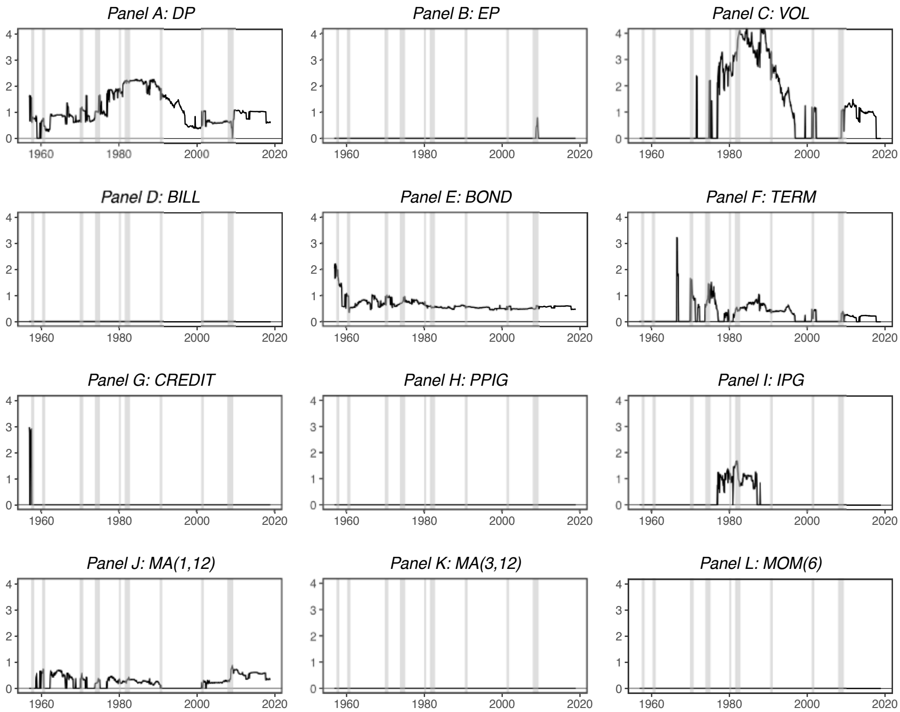


The figure depicts recursive predictive regression model slope coefficient estimates used to compute market excess return forecasts. The black and red lines delineate multiple predictive regression model slope coefficients estimated via ordinary least squares and the elastic net, respectively, for the predictor variable in the panel heading. The blue lines delineate univariate predictive regression model slope coefficients estimated via ordinary least squares. The predictor variable definitions are provided in section 1.3.1. The vertical bars delineate business-cycle recessions as dated by the National Bureau of Economic Research.

Figure 1.2. *Recursive predictive regression model slope coefficient estimates. For a color version of this figure, see www.iste.co.uk/jurczenko/machine.zip*

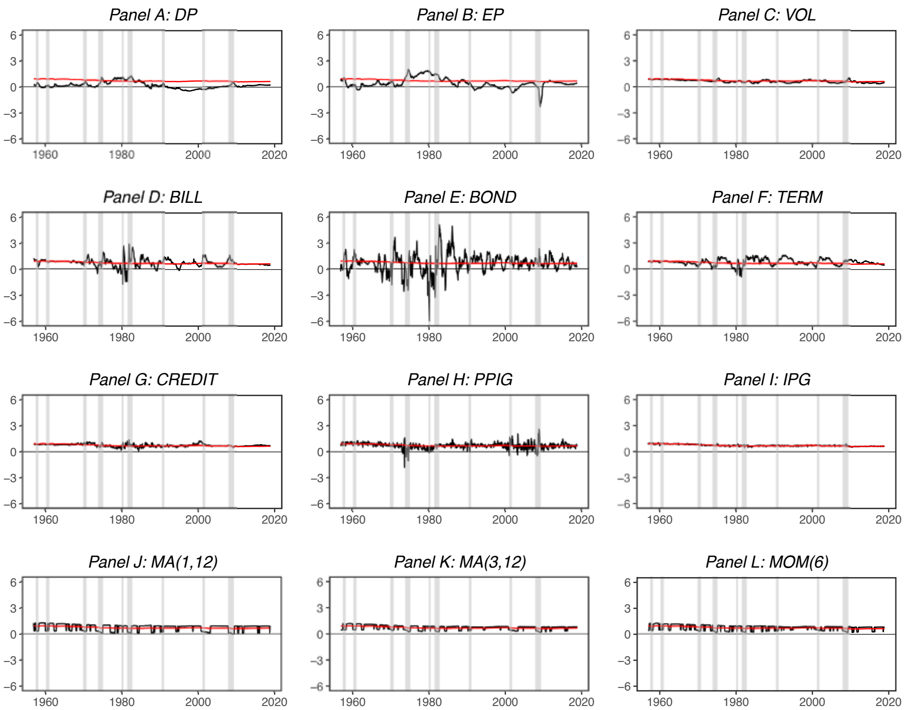
The red line in each panel of Figure 1.2 shows the ENet slope coefficient estimates for the multiple predictive regression model in equation [1.3]. The shrinkage effect of ENet *vis-à-vis* OLS estimation of the multiple regression slope coefficients is clear in Figure 1.2. For some of the predictor variables – such as *BILL*, *IPG*, *MA(3,12)* and *MOM(6)* – the ENet nearly always shrinks the coefficient estimates all the way to zero. The shrinkage induced by ENet estimation of the multiple predictive regression model in equation [1.3] should help at least somewhat in alleviating overfitting.

Figure 1.3 presents the recursive ENet slope coefficient estimates for the Granger and Ramanathan (1984) regression in equation [1.14] used to compute the C-ENet forecast. Recall that the C-ENet forecast is the average of the individual univariate predictive regression forecasts selected by the ENet in equation [1.14]. Figure 1.3 indicates that *BOND* is always selected by the ENet for inclusion in the C-ENet forecast, while *DP* is nearly always selected. *VOL*, *TERM*, and *MA(1,12)* are also frequently selected for inclusion in the C-ENet forecast.



The figure depicts recursive elastic net slope coefficient estimates for a Granger and Ramanathan (1984) regression based on univariate predictive regression forecasts for 12 individual predictor variables. The predictor variable definitions are provided in section 1.3.1. The vertical bars delineate business-cycle recessions as dated by the National Bureau of Economic Research.

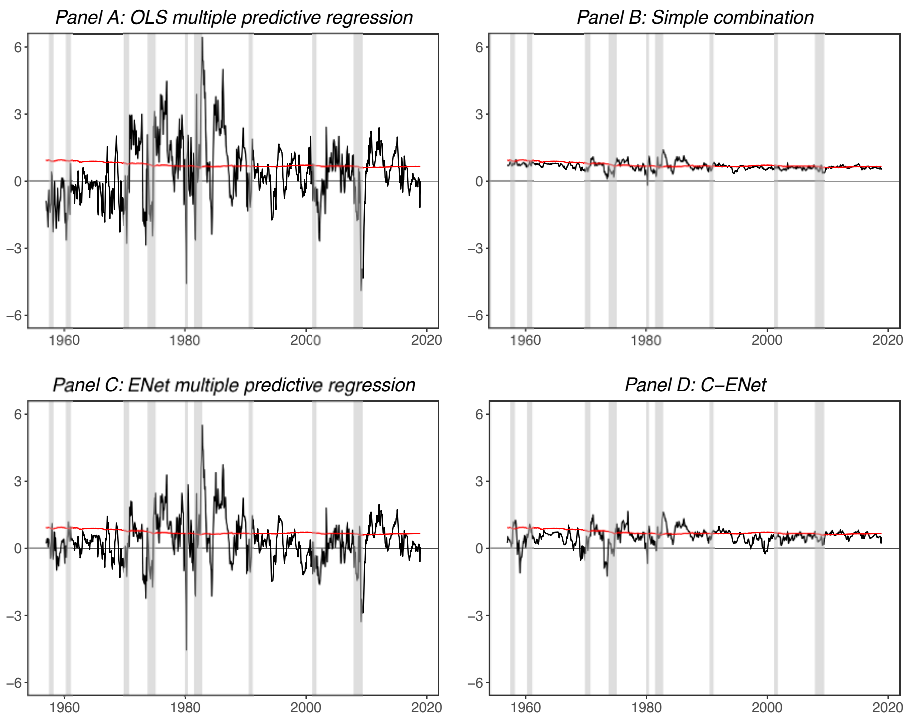
Figure 1.3. *Recursive elastic net Granger and Ramanathan (1984) regression slope coefficient estimates*



The figure depicts out-of-sample market excess return forecasts (in %) for 1957:01 to 2018:12. The black line in each panel delineates a forecast based on ordinary least squares estimation of a univariate predictive regression model with the predictor variable in the panel heading. The predictor variable definitions are provided in section 1.3.1. The red line in each panel delineates the prevailing mean benchmark forecast. The vertical bars delineate business-cycle recessions as dated by the National Bureau of Economic Research.

Figure 1.4. *Market excess return forecasts based on individual predictor variables. For a color version of this figure, see www.iste.co.uk/jurczenko/machine.zip*

Univariate predictive regression forecasts based on equation [1.2] for the 12 individual predictor variables are shown in Figure 1.4, while the forecasts based on multiple predictor variables in equations [1.4], [1.7], [1.11] and [1.15] are presented in Figure 1.5. The red line in each panel delineates the prevailing mean benchmark forecast. A number of the univariate forecasts in Figure 1.4 exhibit distinct behavior over the business cycle. In general, the market excess return forecast lies below the prevailing mean benchmark in the months preceding and early months of a recession; the excess return forecast then increases, so that it moves above the prevailing mean benchmark in the later months of and months immediately following a recession. The forecasts based on the valuation ratios (*DP* and *EP*) also display some longer swings that persist beyond business-cycle frequencies.



The figure depicts out-of-sample market excess return forecasts (in %) for 1957:01 to 2018:12 based on 12 predictor variables using the method in the panel heading. ENet (C-ENet) stands for elastic net (combination elastic net). The vertical bars delineate business-cycle recessions as dated by the National Bureau of Economic Research.

Figure 1.5. *Market excess return forecasts based on multiple predictor variables. For a color version of this figure, see www.iste.co.uk/jurczenko/machine.zip*

Panel A of Figure 1.5 depicts the OLS multiple predictive regression forecast in equation [1.4]. The forecast generally displays the same type of behavior over the business cycle as many of the univariate predictive regression forecasts in Figure 1.4. However, the OLS multiple predictive regression forecast is substantially more volatile and it reaches quite extreme positive and negative values for numerous months over the out-of-sample evaluation period[9]. This is symptomatic of overfitting. Because it maximizes the fit over the estimation sample, conventional estimation is vulnerable to over-responding to noise in the data, especially in a setting with a low signal-to-noise ratio (such as stock return forecasting).

The simple combination forecast in Panel B of Figure 1.5 is the average of the 12 univariate forecasts in Figure 1.4. As discussed in section 1.2.2, forecast combination exerts a strong shrinkage effect, which is immediately evident from the sharp reduction in forecast volatility as we move from Panel A to Panel B in Figure 1.5.

ENet estimation of the multiple predictive regression model in equation [1.3] provides an alternative means for inducing shrinkage in the forecasts. From Panel C of Figure 1.5, we see that ENet estimation does induce some forecast shrinkage relative to the OLS forecast in Panel A. However, the degree of shrinkage is considerably weaker than that induced by the simple combination forecast in Panel B.

The degree of forecast shrinkage induced by the C-ENet forecast in Panel D of Figure 1.5 falls between that of the simple combination forecast in Panel B and ENet multiple predictive regression forecast in Panel C. By not necessarily averaging across all of the univariate forecasts, the C-ENet forecast is able to respond more strongly to fluctuations in the univariate forecasts selected by the ENet in the Granger and Ramanathan (1984) regression, in equation [1.14]. In this fashion, the C-ENet forecast is designed to mitigate the potential overshrinking induced by the simple combination forecast, while avoiding the complete lack of shrinkage in the OLS multiple predictive regression forecast.

1.3.3. *Statistical gains*

Next, we assess the statistical accuracy of the forecasts in Figures 1.4 and 1.5 in terms of mean squared forecast error (MSFE). To this end, we compute the Campbell and Thompson (2008) out-of-sample R^2 statistic:

$$R^2_{\text{OS}} = 1 - \frac{\sum_{s=t_2+1}^{T} \left(\hat{e}_{s|s-1}\right)^2}{\sum_{s=t_2+1}^{T} \left(\hat{e}^{\text{PM}}_{s|s-1}\right)^2}, \qquad [1.16]$$

9 Note that the forecasts are not annualized in Figures 1.4 and 1.5.

where t_2 is the last observation for the initial holdout out-of-sample period, T is the total number of available observations, $\hat{e}_{s|s-1} = r_s - \hat{r}_{s|s-1}$ generically denotes a competing forecast error, and $\hat{e}^{\mathrm{PM}}_{s|s-1} = r_s - \hat{r}^{\mathrm{PM}}_{s|s-1}$ is the prevailing mean forecast error. The R^2_{OS} statistic, which is akin to the familiar in-sample R^2 statistic, measures the proportional reduction in MSFE for a competing forecast *vis-à-vis* the prevailing mean benchmark forecast. Of course, because monthly stock returns inherently contain a limited predictable component, the R^2_{OS} statistic will be "small"[10]. Nevertheless, Campbell and Thompson (2008) argue that a monthly R^2_{OS} statistic of approximately 0.5%, indicates an economically significant degree of market excess return predictability.

To gauge whether a competing forecast provides a statistically significant improvement in MSFE relative to the prevailing mean benchmark forecast, we use the Clark and West (2007) adjusted version of the familiar Diebold and Mariano (1995) and West (1996) statistic. The latter is less informative for comparing forecasts from nested models (Clark and McCracken 2001; McCracken 2007). In particular, we use the Clark and West (2007) MSFE-adj statistic to test the null hypothesis that the prevailing mean MSFE is less than or equal to the competing MSFE, against the alternative that the competing MSFE is less than the prevailing mean MSFE.

Panel A of Table 1.1 reports R^2_{OS} statistics for the univariate predictive regression forecasts. The R^2_{OS} statistics are negative for the majority of the predictor variables, so that most of the forecasts fail to outperform the prevailing mean benchmark in terms of MSFE. The univariate forecasts based on *VOL*, *BILL*, *BOND*, *TERM*, and *MA(1,12)* produce positive R^2_{OS} statistics, and the MSFE-adj statistics indicate that each of the five forecasts provides a statistically significant reduction in MSFE *vis-à-vis* the prevailing mean benchmark (at the 10% level or better)[11]. Among the positive R^2_{OS} statistics, only that for *BOND* is above the Campbell and Thompson (2008) threshold of 0.5%. Furthermore, we cannot know *a priori* which of the 12 predictor variables in Panel A will perform the best.

10 Indeed, if it is "large", it is likely too good to be true!

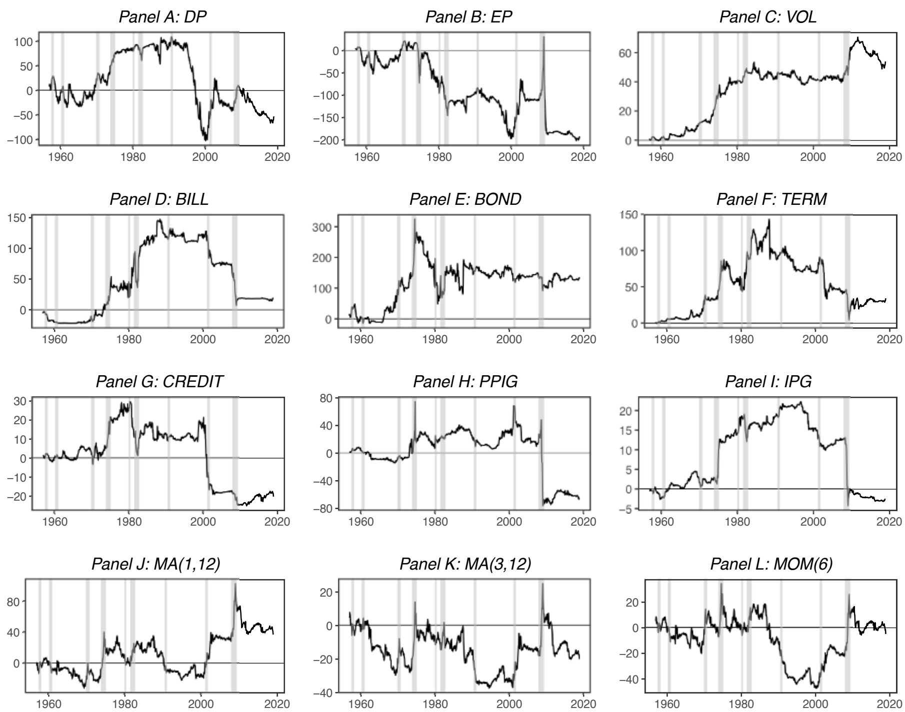
11 Despite having a negative R^2_{OS} statistic, the MSFE-adj statistic for *DP* is significant. Although this result may seem surprising, it can occur when comparing nested forecasts (Clark and West 2007; McCracken 2007).

(1) Forecast	(2) R^2_{OS}	(3) MSFE-adj	(4) Forecast	(5) R^2_{OS}	(6) MSFE-adj
Panel A: Individual predictor variables					
DP	–0.40%	1.81**	*CREDIT*	–0.15%	–0.16
EP	–1.47%	0.79	*PPIG*	–0.50%	0.18
VOL	0.42%	2.58***	*IPG*	–0.02%	–0.04
BILL	0.15%	1.63*	*MA(1,12)*	0.28%	1.38*
BOND	1.04%	3.37***	*MA(3,12)*	–0.15%	0.32
TERM	0.26%	1.59*	*MOM(6)*	–0.04%	0.71
Panel B: Multiple predictor variables					
OLS multiple predictive regression	–4.33%	2.73***	ENet multiple predictive regression	0.22%	3.11***
Simple combination	1.11%	3.70***	C-ENet	2.12%	4.05***

The table reports out-of-sample R^2 (R^2_{OS}) statistics for monthly market excess return forecasts for 1957:01 to 2018:12. The out-of-sample R^2 statistic is the percent reduction in mean squared forecast error (MSFE) for a competing forecast *vis-á-vis* the prevailing mean benchmark forecast. The competing forecasts in Panel A are based on univariate predictive regression models estimated via ordinary least squares. The predictor variable definitions in Panel A are provided in section 1.3.1. The competing forecasts in Panel B use all 12 predictor variables. The OLS (ENet) multiple predictive regression forecast is based on a multiple predictive regression model with all 12 predictor variables estimated via ordinary least squares (the elastic net). The simple combination forecast is the average of the individual univariate predictive regression forecasts in Panel A. The C-ENet forecast is an average of the individual univariate predictive regression forecasts in Panel A selected by the elastic net in a Granger and Ramanathan (1984) regression. MSFE-adj is the Clark and West (2007) statistic for testing the null hypothesis that the benchmark MSFE is less than or equal to the competing MSFE against the alternative hypothesis that the benchmark MSFE is greater than the competing MSFE; *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

Table 1.1. *Forecast accuracy*

To get a sense of the performance of the individual univariate forecasts over time, Figure 1.6 shows the cumulative differences in squared forecast errors for the prevailing mean benchmark relative to each univariate forecast (Goyal and Welch 2003, 2008). The cumulative differences in Figure 1.6 make it straightforward to determine whether a competing forecast is more accurate than the prevailing mean benchmark for any subsample. We simply compare the height of the curve at the start and end of the subsample. If the curve is higher (lower) at the end, then the competing forecast has a lower (higher) MSFE than the prevailing mean benchmark over the subsample. A competing forecast that provides out-of-sample gains on a consistent basis will thus have a predominantly positively sloped curve, while a steeply negatively sloped segment indicates an episode of severe underperformance.



The figure depicts cumulative differences in squared forecast errors between benchmark and competing out-of-sample market excess return forecasts for 1957:01 to 2018:12. The competing forecast is based on ordinary least squares estimation of a univariate predictive regression model with the predictor variable in the panel heading; the benchmark forecast is the prevailing mean. The predictor variable definitions are provided in section 1.3.1. The vertical bars delineate business-cycle recessions as dated by the National Bureau of Economic Research.

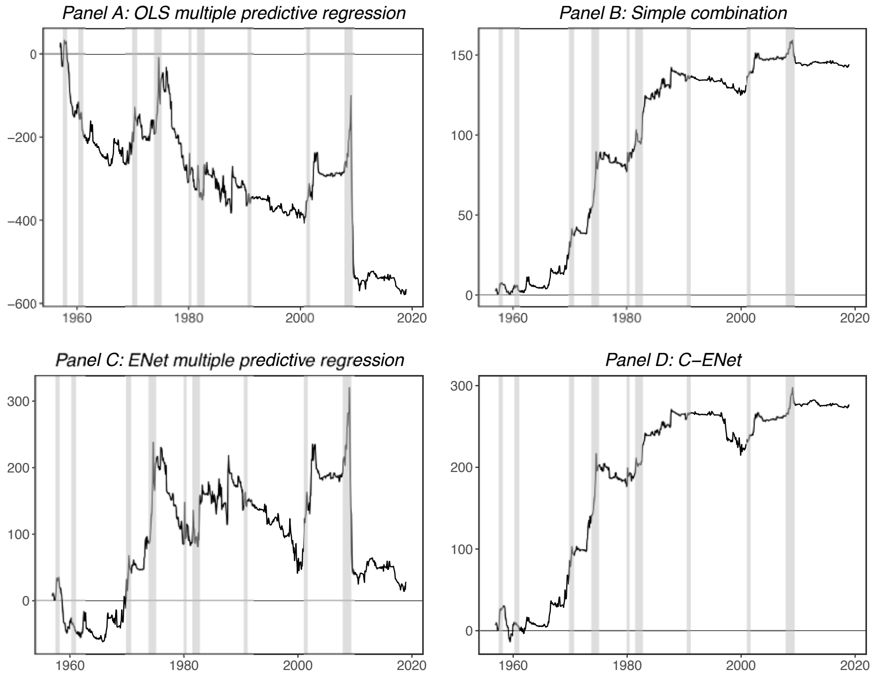
Figure 1.6. *Cumulative differences in squared forecast errors for market excess return forecasts based on individual predictor variables*

With the exception of *VOL*, the univariate forecasts in Figure 1.6 fail to provide accuracy gains on a reasonably consistent basis over time. Although there are times when the univariate forecasts substantively outperform the prevailing mean benchmark, there are also periods when they perform much worse than the naive benchmark. Overall, Figure 1.6 is reminiscent of the findings in Goyal and Welch (2008).

Panel B of Table 1.1 reports R^2_{OS} statistics for the four forecasts that incorporate information from multiple predictor variables, while Figure 1.7 depicts the cumulative differences in squared forecast errors for the prevailing mean benchmark relative to the four forecasts. The R^2_{OS} statistic for the OLS multiple predictive regression forecast is -4.33% in Panel B of Table 1.1, so that the forecast is substantially less accurate than the prevailing mean benchmark over the 1957:01 to 2018:12 evaluation period. Panel A of Figure 1.7 shows that the forecast exhibits prolonged periods of underperformance relative to the naive benchmark. Overall, the OLS multiple predictive regression forecast apparently suffers from considerable overfitting.

The R^2_{OS} statistic for the simple combination forecast is 1.11% in Panel B of Table 1.1 (and its MSFE-adj statistic is significant at the 1% level), which is larger than all the R^2_{OS} statistics for the univariate predictive regression forecasts in Panel A. In addition, Panel B of Figure 1.7 reveals that the simple combination forecast outperforms the prevailing mean benchmark on a reasonably consistent basis over time, as the curve is predominantly positively sloped and displays only limited segments with a negative slope. The strong shrinkage induced by the simple combination forecast is apparently useful for incorporating information from multiple predictor variables in a manner that guards against overfitting.

Although the ENet multiple predictive regression forecast also induces shrinkage, recall from Figure 1.5 that the degree of shrinkage appears limited. The R^2_{OS} statistic for the ENet forecast is positive but quite close to zero (0.22%) in Panel B of Table 1.1. Furthermore, Panel C of Figure 1.7 indicates that the ENet forecast fails to outperform the prevailing mean benchmark consistently over time. Indeed, there are a number of segments where the ENet forecast severely underperforms the prevailing mean. It thus seems that the degree of shrinkage induced by the ENet multiple predictive regression forecast is too weak to effectively guard against overfitting when forecasting the market excess return.



The figure depicts cumulative differences in squared forecast errors between benchmark and competing out-of-sample market excess return forecasts for 1957:01 to 2018:12. The competing forecast is based on 12 predictor variables using the method in the panel heading; the benchmark forecast is the prevailing mean. ENet (C-ENet) stands for elastic net (combination elastic net). The vertical bars delineate business-cycle recessions as dated by the National Bureau of Economic Research.

Figure 1.7. *Cumulative differences in squared forecast errors for market excess return forecasts based on multiple predictor variables*

The simple combination forecast exerts a strong – perhaps too strong – shrinkage effect, while the ENet multiple predictive regression forecast induces insufficient shrinkage. Again recalling Figure 1.5, the degree of shrinkage induced by the C-ENet forecast falls between that induced by the simple combination and ENet forecasts. The R^2_{OS} statistic for the C-ENet forecast is a sizable 2.12% in Panel B of Table 1.1 (and its MSFE-adj statistic is significant at the 1% level). The ENet forecast provides the largest improvement in MSFE among all the competing forecasts relative to the prevailing mean benchmark, and its R^2_{OS} statistic is nearly double that for the simple combination forecast. Panel C of Figure 1.7 confirms that the C-ENet forecast outperforms the prevailing mean quite consistently over time[12].

Overall, the C-ENet forecast appears close to a "Goldilocks" shrinkage technique. It induces stronger shrinkage than the ENet forecast, so that it better guards against overfitting. At the same time, it exerts a weaker shrinkage effect than the simple combination forecast, allowing it to incorporate more information from the most relevant univariate forecasts to further improve out-of-sample performance. Recall from Figure 1.3 that the ENet often selects the univariate predictive regression forecasts based on *DP*, *VOL*, *BOND*, *TERM*, and *MA(1,12)* in the Granger and Ramanathan (1984) regression. This judicious real-time selection of relevant predictor variables to include in the combination forecast demonstrates the value of machine learning for improving out-of-sample market excess return forecasts.

1.3.4. *Economic gains*

In addition to the gains in forecast accuracy documented in section 1.3.3, we next show that the C-ENet forecast provides substantial economic gains for an investor in an asset allocation context. Specifically, we consider a mean-variance investor who allocates across risky equities and risk-free Treasury bills each month. The investor's objective function is given by:

$$\underset{w_{t+1|t}}{\arg\max}\; w_{t+1|t}\hat{r}_{t+1|t} - 0.5\gamma w_{t+1|t}^2 \hat{\sigma}_{t+1|t}^2, \qquad [1.17]$$

where γ is the coefficient of relative risk aversion, $w_{t+1|t}$ $(1 - w_{t+1|t})$ is the investor's allocation to equities (risk-free bills) in month $t + 1$ and $\hat{r}_{t+1|t}$ $(\hat{\sigma}_{t+1|t}^2)$ is the market excess return point (variance) forecast used by the investor. The well-known solution to equation [1.17] takes the form[13]:

$$w_{t+1|t}^* = \left(\frac{1}{\gamma}\right)\left(\frac{\hat{r}_{t+1|t}}{\hat{\sigma}_{t+1|t}^2}\right). \qquad [1.18]$$

12 The out-of-sample gains in Figure 1.7 are often particularly evident during business-cycle recessions. This is a stylized fact in the literature on market excess return predictability (e.g. Rapach *et al*. 2010; Henkel *et al*. 2011; Rapach and Zhou 2013).

13 To prevent what could be construed as impractical allocations, we impose the restriction that $-1 \leq w_{t+1|t} \leq 2$. As shown in Panel A of Figure 1.8, the constraint is rarely binding.

To measure the economic value of return predictability to the investor, we first analyze portfolio performance when the investor relies on the prevailing mean forecast – which ignores return predictability – to determine the optimal equity allocation in equation [1.18]. We then analyze portfolio performance when the investor instead uses the C-ENet forecast to select the optimal allocation. In both cases, the investor uses the sample variance computed over a 60-month rolling window to forecast the variance[14]. We assume that $\gamma = 5$; the results are qualitatively similar for reasonable alternative values for γ.

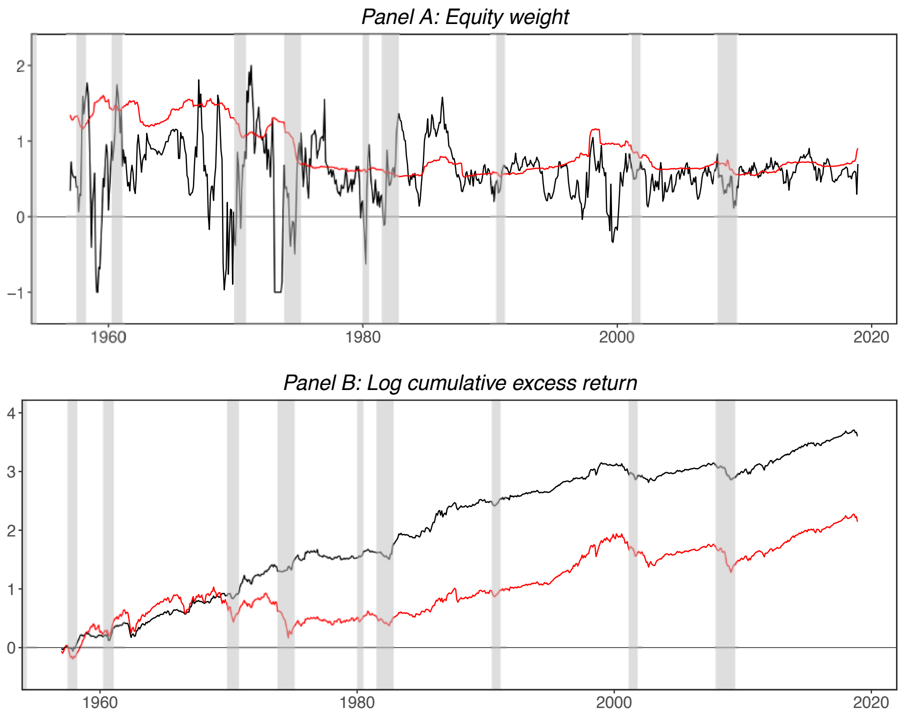
When the investor relies on the prevailing mean benchmark forecast, the optimal portfolio earns an annualized average excess return of 4.30% for the 1957:01 to 2018:12 forecast evaluation period. Together with an annualized volatility of 12.87%, the portfolio's annualized Sharpe ratio is 0.33. If the investor uses the C-ENet forecast in lieu of the prevailing mean in equation [1.18], then the optimal portfolio yields an annualized average excess return of 6.29% and volatility of 9.76%. These translate into a substantive annualized Sharpe ratio of 0.64, which is nearly twice as large as that for the portfolio based on the prevailing mean[15]. Moreover, the gain in certainty equivalent return (CER) indicates that the investor would be willing to pay a hefty annualized management fee of 375 bps to switch from the prevailing mean to the C-ENet forecast.

The red and black lines in Panel A of Figure 1.8 depict the equity allocations for the optimal portfolios based on the prevailing mean and C-ENet forecasts, respectively. There are numerous months when the C-ENet forecast leads to a markedly different allocation than the prevailing mean benchmark. Such reallocations appear quite valuable to the investor, as they substantially improve portfolio performance. This is further illustrated in Panel B of Figure 1.8, which shows the log cumulative excess returns for the two portfolios. For example, the portfolio based on the C-ENet forecast typically suffers smaller drawdowns than the portfolio based on the prevailing

14 In this chapter, we focus on the ability of machine learning techniques to improve expected return forecasts. It would be interesting in future research to explore the usefulness of machine learning techniques for improving forecasts of return volatilities and correlations.

15 The annualized Sharpe ratio for the market excess return for 1957:01 to 2018:12 is 0.42, so that the optimal portfolio based on the C-ENet forecast delivers a Sharpe ratio that is over 50% larger than that for the market portfolio.

mean benchmark; indeed, the maximum drawdown for the former is only half as large as that for the latter (29% and 58%, respectively).



The black and red line in Panel A delineate the equity weight for a mean-variance investor with a coefficient of relative risk aversion of five who uses the combination elastic net (prevailing mean) forecast when allocating between equities and risk-free Treasury bills for 1957:01 to 2018:12. The lines in Panel B show the corresponding log cumulative excess returns for the two portfolios. The vertical bars delineate business-cycle recessions as dated by the National Bureau of Economic Research.

Figure 1.8. *Equity allocations and cumulative excess returns. For a color version of this figure, see www.iste.co.uk/jurczenko/machine.zip*

The optimal portfolio based on the C-ENet forecast also generates a higher Sharpe ratio and annualized CER gain than the optimal portfolios based on any of the 12 univariate predictive regression forecasts, as well as the OLS and ENet multiple predictive regression and simple combination forecasts[16].

16 The complete results are available upon request from the authors.

In sum, the C-ENet forecast delivers the best overall performance in terms of statistical accuracy and investor value.

1.4. Cross-sectional return forecasts

In this section, we outline the cross-sectional return forecasting procedures in Han *et al*. (2019) which parallel the time-series strategies in section 1.2[17]. As Lewellen (2015) argues, the appropriate target is the cross-sectional dispersion in firm returns (and not the cross-sectional average return *per se*); we can simply adjust all of the cross-sectional return forecasts up or down as needed to reflect our forecast of the overall market return.

Consider the following cross-sectional multiple regression model for month t:

$$r_{i,t} = \alpha_t + \sum_{j=1}^{J} \beta_{j,t} z_{i,j,t-1} + \varepsilon_{i,t}, \tag{1.19}$$

for $i = 1, \ldots, I_t$, where $z_{i,j,t}$ is the jth characteristic for firm i in month t and I_t is the number of firm observations available in month t. Analogous to equation [1.4], a conventional cross-sectional return forecast based on equation [1.19] is given by:

$$\hat{r}^{\text{OLS}}_{i,t+1|t} = \hat{\alpha}^{\text{OLS}}_t + \sum_{j=1}^{J} \hat{\beta}^{\text{OLS}}_{j,t} z_{i,j,t}, \tag{1.20}$$

where $\hat{\alpha}^{\text{OLS}}_t$ and $\hat{\beta}^{\text{OLS}}_{j,t}$ are the OLS estimates of α_t and $\beta_{j,t}$, respectively, for $j = 1, \ldots, J$ in equation [1.19]. The literature on cross-sectional returns has investigated many characteristics (e.g. Harvey *et al*. 2016), so that the number of plausible predictors J to include in equation [1.19] is quite large. As in the time-series context, conventional estimation of equation [1.19] is prone to overfitting.

It is common to smooth the OLS coefficient estimates in equation [1.20] over time (e.g. Lewellen 2015; Green *et al*. 2017). However, this appears

17 See Han *et al*. (2019) for further details on the construction of the cross-sectional stock return forecasts discussed in this section.

inadequate for mitigating overfitting. To see this, we use data for 84 firm characteristics starting in 1980:01 to forecast cross-sectional stock returns[18]. The firm characteristics are similar to those used in Lewellen (2015); Green *et al*. (2017) and Freyberger *et al*. (2019). We compute out-of-sample cross-sectional stock returns for 1990:01 to 2018:06 using equation [1.20], although we smooth the OLS coefficient estimates over time using an expanding window before forming the forecasts.

Lewellen (2015) provides an informative predictive slope to assess the ability of a forecast to track cross-sectional expected returns. We compute the predictive slope in two steps. In the first step, we estimate a cross-sectional version of a Mincer and Zarnowitz (1969) regression for month t:

$$r_{i,t} = \phi_t + \psi_t \hat{r}_{i,t|t-1} + \varepsilon_{i,t}, \qquad [1.21]$$

where $\hat{r}_{i,t|t-1}$ generically denotes a cross-sectional return forecast. We estimate ψ_t via OLS for each month over the forecast evaluation period. In the second step, we compute the time-series average of the monthly cross-sectional slope coefficient estimates in equation [1.21]; we denote the average predictive slope estimate by $\hat{\psi}$. As discussed by Lewellen (2015), $\psi = 1$ indicates that the forecasts are unbiased with respect to the cross-sectional dispersion in expected returns: a percentage point increase in the forecast corresponds to a percentage point increase in the realized return on average. If $\psi < 1$ then the cross-sectional forecasts are characterized by overfitting, because a percentage point increase in the forecast corresponds to a less than percentage point increase in the realized return on average. Alternatively, $\psi > 1$ signals that the forecasts are conservative in that a percentage point increase in the forecast coincides with a more than percentage point increase in the actual return on average.

For the OLS multiple regression forecast in equation [1.20], the $\hat{\psi}$ estimate is 0.10 while the average cross-sectional R^2 statistic in equation [1.21] is 0.61%. Based on its standard error, the $\hat{\psi}$ estimate is significantly greater than zero. However, it is also significantly below unity, so that the conventional OLS forecast exhibits significant cross-sectional overfitting, as anticipated.

18 To minimize the effects of micro-cap stocks, we omit stocks with market capitalization below the NYSE median.

Analogous to equation [1.7], Han *et al.* (2019) propose a simple combination strategy for forecasting cross-sectional returns:

$$\hat{r}^{\mathrm{C}}_{i,t+1|t} = \frac{1}{J}\sum_{j=1}^{J}\hat{r}^{(j)}_{i,t+1|t}, \qquad [1.22]$$

where

$$\hat{r}^{(j)}_{i,t+1|t} = \hat{\alpha}^{(j)}_t + \hat{\beta}^{(j)}_t z_{i,j,t}, \qquad [1.23]$$

and $\hat{\alpha}^{(j)}_t$ and $\hat{\beta}^{(j)}_t$ are the OLS estimates of the intercept and slope coefficients, respectively, in the following cross-sectional univariate regression model:

$$r_{i,t} = \alpha_t + \beta_t z_{i,j,t-1} + \varepsilon_{i,t}, \qquad [1.24]$$

for $j = 1, \ldots, J$. Again like the time-series case, the simple combination forecast in equation [1.22] exerts a strong shrinkage effect[19]. The $\hat{\psi}$ estimate is 2.27 for the combination forecast (with an average cross-sectional R^2 of 4.85%), which is significantly greater than both zero and unity. The simple combination approach thus appears to overshrink the forecast, rendering it overly conservative on average.

Han *et al.* (2019) recommend refining the simple combination forecast in equation [1.22] using machine learning techniques. Specifically, they suggest selecting the individual characteristics to include in the combination forecast by using the LASSO or ENet to estimate a cross-sectional Granger and Ramanathan (1984) regression, that relates month-t realized returns to the univariate forecasts in equation [1.23]. The cross-sectional C-ENet forecast is given by:

$$\hat{r}^{\text{C-ENet}}_{i,t+1|t} = \frac{1}{|\mathcal{J}_t|}\sum_{j\in\mathcal{J}_t}\hat{r}^{(j)}_{i,t+1|t}, \qquad [1.25]$$

19 Observe that we do not smooth the OLS coefficient estimates over time in equation [1.23] when forming the cross-sectional forecast as the simple combination forecast already induces strong shrinkage.

where $\mathcal{J}_t \subseteq \{1, \ldots, J\}$ is the index set of cross-sectional univariate forecasts selected by the ENet in the month-t Granger and Ramanathan (1984) regression. As in the time-series case, the refinement proves efficacious: the $\hat{\psi}$ estimate for the C-ENet forecast is 1.25 (with an average cross-sectional R^2 statistic of 3.65%) which is significantly greater than zero but insignificantly different from unity.

Han *et al*. (2019) also develop a cross-sectional forecast that blends the conventional OLS multiple regression and C-ENet forecasts using the notion of forecast encompassing (Harvey *et al*. 1998). The $\hat{\psi}$ estimate for their encompassing C-ENet forecast is 1.10 (with an average cross-sectional R^2 statistic of 2.69%) which is even closer to unity. The $\hat{\psi}$ estimate is significantly greater than zero and insignificantly different from unity. In an extensive empirical application involving more than 100 firm characteristics, Han *et al*. (2019) show that their encompassing C-ENet approach provides the most accurate forecasts to date of the cross-sectional dispersion in expected stock returns.

1.5. Conclusion

Extending the cross-sectional return forecasting procedures developed by Han *et al*. (2019), this chapter introduces some new machine learning methods for time-series stock return forecasting. Our empirical application focuses on forecasting the US market excess return, a central issue in finance. Despite evidence of in-sample predictability, Goyal and Welch (2008) show that conventional forecasts based on many popular predictor variables from the literature fail to provide out-of-sample gains on a consistent basis over time. Using simple forecast combination, Rapach *et al*. (2010) are seemingly the first to provide evidence of consistent out-of-sample market excess return predictability. The methods proposed in this chapter use the elastic net to refine the simple combination forecast and we find that the elastic net refinement, embodied in our combination elastic net forecast, indeed generates substantive further improvements in out-of-sample market excess return predictability.

Machine learning techniques are often criticized for being “black boxes”. However, by performing variable selection, the elastic net (and LASSO) is a machine learning technique that facilitates economic interpretation. In our empirical application, our combination elastic net approach consistently identifies the dividend-price ratio, volatility, Treasury bond yield, term spread and a popular technical signal, as relevant market excess return predictors. The identification of the most relevant out-of-sample market excess return predictors from among the plethora of predictors from the literature provides a useful guide for researchers in constructing theoretical asset pricing models. As analyzed by Han *et al*. (2019) in a cross-sectional context, the combination elastic net approach can also be used to identify the most relevant firm characteristics for explaining cross-sectional expected returns. More generally, since countless questions in asset pricing and corporate finance are related to either time-series or cross-sectional return forecasting, the combination elastic net approach discussed in this chapter should provide a valuable resource for researchers and practitioners alike.

1.6. Acknowledgements

We thank Emmanuel Jurczenko, the editor, for helpful comments on an earlier draft.

1.7. References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In Petrov, B.N. and Csaki, F. (eds). *Proceedings of the 2nd International Symposium on Information Theory*, Akadémiai Kiadó, Budapest, 267–281.

Bates, J.M. and Granger, C.W.J. (1969). The combination of forecasts. *Journal of the Operational Research Society*, 20(4), 451–468.

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

Chinco, A., Clark-Joseph, A.D., and Ye, M. (2019). Sparse signals in the cross-section of returns. *Journal of Finance*, 74(1), 449–492.

Clark, T.E. and McCracken, M.W. (2001). Tests of equal forecast accuracy and ecompassing for nested models. *Journal of Econometrics*, 105(1), 85–110.

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

Diebold, F.X. and Mariano, R.S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13(3), 253–263.

Diebold, F.X. and Shin, M. (2019). Machine learning for regularized survey forecast combination: Partially-egalitarian lasso and its derivatives. *International Journal of Forecasting*, 35(4), 1679–1691.

Flynn, C.J., Hurvich, C.M., and Simonoff, J.S. (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., Neuhierl, A., and Weber, M. (2019). Dissecting characteristics nonparametrically. *Review of Financial Studies*.

Goyal, A. and Welch, I. (2003). Predicting the equity premium with dividend ratios. *Management Science*, 49(5), 639–654.

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

Granger, C.W.J. and Ramanathan, R. (1984). Improved methods of combining forecasts. *Journal of Forecasting*, 3(2), 197–204.

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

Gu, S., Kelly, B.T., and Xiu, D. (2019). Empirical asset pricing via machine learning. *Review of Financial Studies*.

Han, Y., He, A., Rapach, D.E., and Zhou, G. (2019). Firm characteristics and expected stock returns. Working Paper.

Harvey, C.R., Liu, Y., and Zhu, H. (2016).... and the cross-section of expected returns. *Review of Financial Studies*, 29(1), 5–68.

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

Hastie, T. and Qian, J. (2016). Glmnet vignette. Working Paper.

Henkel, S.J., Martin, J.S., and Nadari, F. (2011). Time-varying short-horizon predictability. *Journal of Financial Economics*, 99(3), 560–580.

Hoerl, A.E. and Kennard, R.W. (1970). Ridge regression: Applications to nonorthogonal problems. *Technometrics*, 12(1), 69–82.

Hurvich, C.M. and Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika*, 76(2), 297–307.

Ke, Z.T., Kelly, B.T., and Xiu, D. (2019). Predicting returns with text data. Working Paper.

Kozak, S., Nagel, S., and Santosh, S. (2019). Shrinking the cross section. *Journal of Financial Economics*.

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

Loughran, T. and McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries and 10-Ks. *Journal of Finance*, 66(1), 35–65.

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

Mele, A. (2007). Asymmetric stock market volatility and the cyclical behavior of expected returns. *Journal of Financial Economics*, 86(2), 446–478.

Mincer, J.A. and Zarnowitz, V. (1969). The evaluation of economic forecasts. In Mincer, J.A. (ed.). *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*, National Bureau of Economic Research. Columbia University Press, New York.

Neely, C.J., Rapach, D.E., Tu, J., and Zhou, G. (2014). Forecasting the equity risk premium: The role of technical indicators. *Management Science*, 60(7), 1772–1791.

Rapach, D.E., Strauss, J.K., and Zhou, G. (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., Strauss, J.K., and Zhou, G. (2013). International stock return predictability: What is the role of the United States? *Journal of Finance*, 68(4), 1633–1662.

Rapach, D.E. and Zhou, G. (2013). Forecasting stock returns. In Elliott, G. Timmermann, A. (eds). *Handbook of Economic Forecasting*, Volume 2A. Elsevier, Amsterdam.

Tetlock, P.C. (2007). Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance*, 62(3), 1139–1168.

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 Elliott, G., Granger, C.W.J., Timmermann, A. (eds). *Handbook of Economic Forecasting*, Volume 1. Elsevier, Amsterdam.

West, K.D. (1996). Asymptotic inference about predictive ability. *Econometrica*, 64(5), 1067–1084.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2), 301–320.