Asset Pricing: Time-Series Predictability

David E. Rapach, Chaifetz School of Business, Saint Louis University

Email: david.rapach@slu.edu

Guofu Zhou, Olin Business School, Washington University in St. Louis

Email: zhou@wustl.edu

March 24, 2022

Summary

Asset returns change with fundamentals and other factors, such as technical information and sentiment over time. In modeling time-varying expected returns, this article focuses on the out-of-sample predictability of the aggregate stock market return via extensions of the conventional predictive regression approach.

The extensions are designed to improve out-of-sample performance in realistic environments characterized by large information sets and noisy data. Large information sets are relevant because there are a plethora of plausible stock return predictors. The information sets include variables typically associated with a rational time-varying market risk premium, as well as variables more likely to reflect market inefficiencies resulting from behavioral influences and information frictions. Noisy data stem from the intrinsically large unpredictable component in stock returns. When forecasting with large information sets and noisy data, it is vital to employ methods that incorporate the relevant information in the large set of predictors in a manner that guards against overfitting the data.

Methods that improve out-of-sample market return prediction include forecast combination, principal component regression, partial least squares, the LASSO and elastic net from ma-

chine learning, and a newly developed C-ENet approach that relies on the elastic net to refine the simple combination forecast. Employing these methods, a number of studies provide statistically and economically significant evidence that the aggregate market return is predictable on an out-of-sample basis. Out-of-sample market return predictability based on a rich set of predictors thus appears to be a well-established empirical result in asset pricing.

Keywords: market excess return, out-of-sample tests, utility gains, forecast combination, principal component regression, partial least squares, LASSO, elastic net

Subject: Financial Economics

Introduction

Since the influential review by Fama (1970) of the theoretical and empirical literature at the time, the efficient market hypothesis (EMH) has become well known: a security's price equals its "fundamental value"—no abnormal (i.e., risk-adjusted) returns can be made relative to one of the information sets (price history, all public information, and all public as well as private information). There is often confusion surrounding the EMH and asset return predictability. It is commonly believed that if the EMH is true, then asset returns are not predictable. This is incorrect. As long as return predictability reflects compensation for taking on risk, then return predictability is consistent with the EMH. This does not imply that the return predictability found in the literature and documented in this article's section "Alternative Methods" is necessarily consistent with the EMH. In essence, there are two potential explanations for the return predictability found in the literature: rational risk-based explanations that are consistent with the EMH and behavioral influences and various types of information frictions that give rise to market inefficiencies (i.e., security mispricing). Sometimes it is difficult to squarely place an economic explanation in one of the two categories, and return predictability can reflect a combination of efficient and inefficient influences.

Theoretically, under general conditions in a frictionless market where all investors have access to the same information and process it optimally, assets will be priced in equilibrium by a stochastic discount factor (SDF):

$$P_t = E_t[M_{t+1}V_{t+1}], \tag{1}$$

where E_t is the expectation operator conditional on information available through period t, P_t is the asset price, M_{t+1} is the SDF common to all assets, and V_{t+1} is the asset's future payoff (see Cochrane 2004). Equation (1) implies that

$$E_t[R_{t+1}] - R_t^f = -\frac{\text{cov}_t(R_{t+1}, M_{t+1})}{E_t[M_{t+1}]}, \tag{2}$$

where R_{t+1} is the gross asset return and R_t^f is the gross risk-free return, so that any economic variable that impacts the conditional covariance between the return and SDF, as well as the SDF itself, will impact the future expected excess return on the asset. In other words, changing economic conditions can affect expected excess returns, which, in the context of Equation (2), is fully consistent with rational risk-based asset pricing.

The Campbell and Shiller (1988) present-value decomposition, which is a special case of Equation (2), is one of the earliest economic devices for justifying predictability. In this framework, deviations in the dividend-price ratio from its long-term mean signal changes in expected future dividend growth rates and/or expected future stock returns. Changes in the latter represent time-varying discount rates and thus return predictability. Campbell and Cochrane (1999) and Bansal and Yaron (2004) develop well-known theoretical models offering rational explanations of market excess return predictability based on habit formation in consumption and a persistent component in consumption growth in conjunction with fluctuations in the conditional volatility of future growth rates, respectively.[1]

Asset pricing models in behavioral finance involve psychological influences, such as under- and/or over-reaction to information, which can generate momentum and other predictable price patterns. This line of reasoning is the foundation of technical analysis, which primarily employs past price (as well as volume) data to predict future returns. This type of return predictability generally appears to be inconsistent with the EMH, as the theoretical basis and patterns of return predictability largely appear inconsistent with rational time-varying risk premia.

Han, Zhou, and Zhu (2016) provide a short survey of theoretical models that justify the use of technical analysis. Because of differences in the timing of the receiving of information, differences in information processing, behavior biases, and/or feedback trading, Treynor and Ferguson (1985), Brown and Jennings (1989), Hong and Stein (1999), Cespa and Vives (2012), and Edmans, Goldstein, and Jiang (2015), among others, show that past stock prices can predict future returns. In practice, moving averages (MAs) of past asset prices are the most widely used technical indicators. Zhu and Zhou (2009) provide the first theoretical basis for the efficacy of MAs, while Han, Zhou, and Zhu (2016) show the effects in an equilibrium model based on the work of Wang (1993). Detzel et al. (2021) recently propose a model in which efficacious technical analysis can arise endogenously via rational learning.

In the spirit of behavioral finance, investor sentiment can also generate return predictability. For example, DeLong et al. (1990) show theoretically that, due to limits to arbitrage, noise trader risk, which is associated with investor sentiment, can make asset prices predictable, even in the absence of fundamental risk. Baker and Wurgler (2006) propose an investor

[1] Theory also limits the degree of return predictability, as shown by Ross (2005), Zhou (2010), and Huang and Zhou (2017).

sentiment index and show that it can explain returns on stocks that are difficult to value and costly to arbitrage.

Lo (2004, 2005) offers the adaptive market hypothesis in an effort to explain behavioral biases. He argues that many examples of apparently irrational behavior, such as loss aversion and overreaction, can be consistent with an evolutionary model of individuals who adapt rationally to a changing environment based on simple heuristics.

The remainder of this article focuses on methods for generating and testing out-of-sample stock return forecasts, which are generally regarded as the most rigorous and informative for assessing stock return predictability (see Goyal and Welch 2008; Martin and Nagel forthcoming). The article concentrates on aggregate market excess return predictability, which is the subject of a voluminous literature. The section "In-Sample Tests of Return Predictability" provides background for analyzing return predictability, while the section "Out-of-Sample Tests of Return Predictability" discusses methods for testing the statistical and economic significance of out-of-sample return forecasts. The section "Alternative Methods" describes approaches for extending the conventional predictive regression framework to substantially improve out-of-sample return forecasts; the section also discusses empirical results from the literature and provides updated results for a group of well-known studies.

In-Sample Tests of Return Predictability

This section describes popular in-sample tests of return predictability. The tests provide useful background for the discussion of out-of-sample tests.

Variance Ratio Tests

Variance ratio tests, proposed by Lo and MacKinlay (1988), analyze the null hypothesis that asset returns are not predictable. Of course, any econometric test requires the specification of a data-generating process (DGP), either parametric or non-parametric; hence, any test of return predictability is a joint test of the null and the assumed DGP.

Early studies of market efficiency focus on the random walk (with drift) model of stock prices:

$$p_t = \mu + p_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2), \tag{3}$$

where $p_t = \log(P_t)$ is the period-t log stock price. Equation (3) says that the current log price is the previous period's log price plus a drift term μ and a normally distributed noise shock ε_t (or that the continuous return is normally distributed with mean μ and variance σ^2). This is the lognormal assumption underlying the Black-Scholes formula for option pricing. If the random walk model in Equation (3) is true, then the market must be efficient; however, if the market is efficient, then the random walk model is not necessarily true.

Equation (3) says that the time series $p_t - p_{t-1}$ is independently and identically distributed, so that the mean and variance are estimated consistently by their sample analogs:

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T}(p_t - p_{t-1}), \tag{4}$$

$$\hat{\sigma}_a^2 = \frac{1}{T}\sum_{t=1}^{T}[(p_t - p_{t-1}) - \hat{\mu}]^2, \tag{5}$$

respectively, where T is the sample size.

To test Equation (3), note that it implies

$$p_t = 2\mu + p_{t-2} + \varepsilon_t + \varepsilon_{t-1}, \tag{6}$$

so that the sample variance of $p_t - 2\mu - p_{t-2}$ should estimate $2\sigma^2$, or (dividing the result by two) produces an alternative variance estimator:

$$\hat{\sigma}_b^2 = \frac{1}{T}\sum_{k=1}^{T/2}(p_{2k} - p_{2k-2} - 2\hat{\mu})^2. \tag{7}$$

Intuitively, if Equation (3) is true, both variance estimators in Equations (5) and (7) should converge to σ^2, and hence their ratio,

$$J_r = \frac{\hat{\sigma}_b^2}{\hat{\sigma}_a^2}, \tag{8}$$

should converge to one. Indeed, Lo and MacKinlay (1988) show that

$$\sqrt{T} J_r \overset{\text{asy}}{\sim} N(1, 2), \tag{9}$$

which says that the variance ratio in Equation (8) scaled by $\sqrt{T}$ is asymptotically normally distributed with mean one and variance two. Since J_r is based on the ratio of two variances, it is known as the variance-ratio test.

If one finds from actual data that $\sqrt{T} J_r$ is significantly different from one as judged by Equation (9), then one can reject the null hypothesis that Equation (3) is true. Lo and MacKinlay (1988) reject the random walk hypothesis for US stock market indices. If a stock return or the market return is a random walk, then there is no predictability. The rejection by Lo and MacKinlay (1988) of the random walk hypothesis opened the door for studying stock return predictability.

Predictive Regressions

A simple linear regression of an asset return on one or a few lagged predictors of interest is the most popular econometric approach for testing for return predictability. For simplicity, consider a univariate predictive regression of the period-$(t+1)$ stock market return r_{t+1} on a single predictor variable x_t:

$$r_{t+1} = \alpha + \beta x_t + \varepsilon_{t+1} \quad \text{for } t = 1, \ldots, T-1, \tag{10}$$

where ε_{t+1} is a zero-mean, unpredictable disturbance term. When x_t is the inflation rate, dividend yield, book-to-market ratio, an interest rate, or a function of interest rates (e.g., the term spread), Nelson (1976), Fama and Schwert (1977), Rozeff (1984), Keim and Stambaugh (1986), Campbell (1987), Fama and French (1988), Kothari and Shanken (1997), and Pontiff and Schall (1998), among others, find that estimates of β are significantly different from zero; that is, there is in-sample evidence of stock market return predictability. There are two primary reasons for the widespread use of the simple predictive regression in Equation (10). First, it is straightforward to implement, with the results intuitively understandable in a familiar regression framework. Second, it should capture some of the return predictability, even if the true DGP is more complex.

It should be noted that the R^2 statistic for the predictive regression in Equation (10) is usually quite small; for example, less than 5% for monthly stock returns. This simply indicates that stock returns (and asset returns more generally) contain an intrinsically large unpredictable component, so that—at the risk of stating the obvious—it is extremely difficult to predict returns (also see Footnote 1). In addition, it is worth noting that the conventional ordinary least squares (OLS) estimator of β in Equation (10) is generally biased, due to persistence in the predictor x_t and correlation between the disturbance term ε_{t+1} and the innovation to x_t. Stambaugh (1999) provides the econometric theory for understanding the bias in predictive regressions. The persistence in the predictor x_t can create some thorny econometric issues for testing the statistical significance of β in Equation (10). Kostakis, Magdalinos, and Stamatogiannis (2015) propose a powerful Wald test that is robust to the regressor's degree of persistence. Alternatively, we can use a confidence interval for the R^2 statistic to test for significant evidence of predictability. The construction of confidence intervals is not analytically tractable for general distributions, but it can be computed via a bootstrap procedure (e.g., Huang et al. 2020).

To better understand a predictive regression, it is useful to contrast it with an explanatory regression that regresses a current variable y_t on another current variable z_t:

$$y_t = \alpha + \beta z_t + \varepsilon_t. \tag{11}$$

For example, the capital asset pricing model or market model regression uses the current market excess return to explain the current excess return on an individual stock (or portfolio of stocks). Although such a regression typically has a high R^2 statistic, around 80% on a monthly basis for large stocks, the regression is of little use for forecasting the excess return on a stock unless one can forecast the market excess return.

Out-of-Sample Tests of Return Predictability

How do we assess the existence and degree of predictability? Traditionally, we examine the statistical significance of the slope coefficient and/or R^2 statistic in Equation (10) using all of the data, that is, by running the regression from the beginning to the end of the available sample period. However, this can be misleading in that using all the data leads to a "look-ahead" bias, because an investor in real time does not have have access to all the sample data. As a result, the traditional in-sample approach cannot be used to make forecasts that mimic the situation of an investor in real time.

Especially since the influential study of Goyal and Welch (2008), researchers focus more on out-of-sample tests of return predictability, which are viewed as more relevant and rigorous. The idea is to compare two competing out-of-sample forecasts. The first incorporates information from a predictor variable. Consider, for example, the univariate predictive regression in Equation (10). An out-of-sample forecast of r_{t+1} based on the predictor x_t and information available through period t is given by

$$\hat{r}_{t+1|t} = \hat{\alpha}_t + \hat{\beta}_t x_t, \tag{12}$$

where $\hat{\alpha}_t$ and $\hat{\beta}_t$ are the OLS estimates of α and β, respectively, in Equation (10) based on data available through t. The forecast in Equation (12) uses only information available through t, so that it avoids look-ahead bias and mimics the situation of a forecaster in real time.

The parameters estimates $\hat{\alpha}_t$ and $\hat{\beta}_t$ can be based on either an expanding or a rolling estimation window. The former uses observations from the start of the available sample, so that the estimation sample increases in size as additional forecasts are computed. The latter drops earlier observations as additional forecasts are computed, so that the size of the estimation sample remains constant over time. Intuitively, a rolling window appears better able to accommodate changes in the parameters over time, although this comes at the cost of a shorter estimation sample and thus less precise parameter estimates. For out-of-sample return prediction, an expanding window often works better in practice, a manifestation of the bias-efficiency trade-off.

The second forecast—the benchmark—is based on the following DGP, which assumes that the return is not predictable:

$$r_{t+1} = \mu + \varepsilon_{t+1}. \tag{13}$$

The forecast corresponding to Equation (13) is straightforwardly given by the prevailing mean (also known as the historical average) computed based on data through t:

$$\bar{r}_{t+1|t} = \frac{1}{t}\sum_{s=1}^{t} r_s. \tag{14}$$

The idea is to compare the out-of-sample mean squared forecast error (MSFE) for the forecast in Equation (12) to that for the prevailing mean in Equation (14). If Equation (12)—which incorporates information from the predictor x_t—delivers a lower MSFE than Equation (14)—which ignores return predictability—then we have out-of-sample evidence of return predictability.

A popular and convenient measure for comparing MSFEs for competing return forecasts is the Campbell and Thompson (2008) out-of-sample R^2 statistic:

$$R^2_{\mathrm{OS}} = 1 - \frac{\sum_{t=T_1}^{T}\left(r_t - \hat{r}_{t|t-1}\right)^2}{\sum_{t=T_1}^{T}\left(r_t - \bar{r}_{t|t-1}\right)^2}, \tag{15}$$

where T_1 is the first observation in the out-of-sample period used for forecast evaluation. Equation (15) measures the proportional reduction in MSFE for the forecast that utilizes the information in the predictor variable vis-à-vis the naïve benchmark forecast that assumes that returns are unpredictable. Again, because returns inherently contain a large unpredictable component, the R^2_{OS} statistic in Equation (15) will necessarily be small.

Statistical Significance

It is also of interest to determine whether a competing forecast can deliver a statistically significant improvement in MSFE relative to the prevailing mean benchmark. This is equivalent to testing H_0: $R^2_{\text{OS}} \leq 0$ versus H_A: $R^2_{\text{OS}} > 0$. This is often done via the Clark and West (2007) test (see, for example, Rapach, Strauss, and Zhou 2010, p. 828). As shown by Clark and McCracken (2001) and McCracken (2007), the popular Diebold and Mariano (1995) and West (1996) test can be severely undersized when comparing forecasts from nested models, as is the case for Equations (12) and (14). This can lead to much lower power to detect return predictability when it exists. Clark and West (2007) adjust the Diebold and Mariano (1995) and West (1996) statistic so that it is well approximated asymptotically by the standard normal distribution.

Economic Value

A result can be statistically significant but not economically significant. In practice, an investor is obviously keenly interested in the economic value of return predictability. Hence, for a given degree of return predictability, an important issue is whether it generates significant economic value.

How can asset allocation benefit from return predictability? The work of Kandel and Stambaugh (1996) and Barberis (2000) are early examples of this line of research, which finds that there are often substantive economic gains associated with return predictability. Of course, the size of the gains will vary across applications.

Consider a mean-variance investor who allocates their wealth between a broad stock market index and a risk-free asset (typically proxied by Treasury bills). How can the investor benefit from the stock return forecast? The investor's optimal allocation to stocks for period $t + 1$

based on information through t is given by

$$w_{t+1|t} = \left(\frac{1}{\gamma}\right)\left(\frac{\hat{r}_{t+1|t}}{\hat{\sigma}^2_{t+1|t}}\right), \tag{16}$$

where γ is the investor's coefficient of relative risk aversion, $\hat{r}_{t+1|t}$ is a forecast of the market excess return based on a predictor or set of predictors, and $\hat{\sigma}^2_{t+1|t}$ is a forecast of the variance of the market excess return. The variance is often predicted using the sample variance and a rolling estimation window, but any method (e.g., the RiskMetrics model) can be used. In practice, it is common to restrict $w_{t+1|t}$ to lie between -0.5 and 1.5, which imposes realistic portfolio constraints and produces better-behaved portfolio weights given the well-known sensitivity of mean-variance optimal weights to return forecasts.

The investor's realized average utility or certainty equivalent return (CER) is given by

$$\text{CER} = \bar{r}_p - 0.5\gamma\sigma^2_p, \tag{17}$$

where $\bar{r}_p$ and σ^2_p are the mean and variance, respectively, of the portfolio return over the forecast evaluation period. The CER is the risk-free rate of return that an investor would be willing to accept in lieu of holding the risky portfolio.

We then repeat the asset allocation exercise assuming that the investor uses the prevailing mean benchmark forecast $\bar{r}_{t+1|t}$ instead of $\hat{r}_{t+1|t}$ in Equation (16). We assume that the investor uses the same variance forecast. Let CER_0 denote the certainty equivalent return over the forecast evaluation period when the investor uses the prevailing mean forecast. The average utility gain corresponding to using $\hat{r}_{t+1|t}$ instead of $\bar{r}_{t+1|t}$ to guide asset allocation is then given by

$$\text{Gain} = \text{CER} - \text{CER}_0. \tag{18}$$

Equation (18) is the gain in CER for the investor when they assume that returns are predictable compared to the case when they assume that they are not. The CER gain is typically annualized, and it can be interpreted as the annual portfolio management fee that the investor would be willing to pay to have access to the information in the predictors relative to ignoring the information. This is a common measure of the economic value of return predictability. McCracken and Valente (2018) provide methods, including bootstrap procedures, for testing the significance of the CER gain in Equation (18).

Alternative Methods

Using univariate predictive regressions forecasts like Equation (12) to predict the US equity premium, Goyal and Welch (2008) find that a lengthy list of popular predictors from the literature—which typically evince significant in-sample evidence of return predictability—are unable to outperform the prevailing mean benchmark forecast on an out-of-sample basis in terms of MSFE. The influential study by Goyal and Welch (2008) called into question the out-of-sample predictability of the US market excess return. However, Rapach, Strauss, and Zhou (2010) and subsequent studies offer evidence in support of out-of-sample US equity premium predictability, provided that methods are used to address the challenges posed by stock return forecasting.

This section discusses extensions of the univariate predictive regression in Equation (10), which are designed to improve out-of-sample performance. They accommodate a large number of potential predictors, which is the relevant case in practice, as a plethora of plausible predictors exist for stock returns. They also recognize that the large unpredictable component in stock returns means that we must contend with noisy data when estimating predictive models.

Forecast Combination

As previously mentioned, Goyal and Welch (2008) find that a number of individual popular market return predictors from the literature fail to outperform the naïve prevailing mean benchmark on an out-of-sample basis. Rapach, Strauss, and Zhou (2010) confirm their finding. They argue that the inability of many individual predictors to consistently generate out-of-sample gains is not surprising, as individual predictors may perform well during certain periods but poorly during others. It is thus risky to rely on an individual predictor, similarly to relying on a single asset in a portfolio. To improve out-of-sample performance, Rapach, Strauss, and Zhou (2010) suggest forecast combination, which incorporates information from a large number of predictors in a manner that guards against overfitting. In essence, this allows for forecast diversification, as some predictors perform well when others are performing poorly, which is similar to the benefits of portfolio diversification (Timmermann 2006; Chen and Maung 2020; Gospodinov and Maasoumi 2021).

The most straightforward approach for incorporating information from a large number of predictors is a multiple predictive regression:

$$r_{t+1} = \alpha + \sum_{i=1}^{n} \beta_i x_{i,t} + \varepsilon_{t+1}, \tag{19}$$

where $x_{i,t}$ is the ith predictor and n is the number of predictors. An obvious out-of-sample forecast corresponding to Equation (19) is given by

$$\hat{r}_{t+1|t} = \hat{\alpha}_t + \sum_{i=1}^{n} \hat{\beta}_{i,t} x_{i,t}, \tag{20}$$

where $\hat{\alpha}_t$ and $\hat{\beta}_{i,t}$ are the OLS estimates of α and β_i, respectively, in Equation (19) based on data through t. However, the forecast in Equation (20) is highly problematic for forecasting stock returns. When n is large, the regression is high dimensional, which can substantially increase the variance of the coefficient estimates. In addition, stock returns inherently contain a large unpredictable component, so that the data are very noisy. These considerations make the OLS forecast in Equation (20) highly susceptible to in-sample overfitting, which can lead to poor out-of-sample performance. Not surprisingly, Goyal and Welch (2008) and Rapach, Strauss, and Zhou (2010) find that the forecast in Equation (20) is substantially outperformed by the prevailing mean benchmark, a manifestation of overfitting.

Forecast combination proceeds in two steps. First, instead of computing a forecast based on the high-dimensional regression in Equation (19), one begins with univariate predictive regressions based on the n individual predictors (considered in turn):

$$r_{t+1} = \alpha_i + \beta_i x_{i,t} + \varepsilon_{i,t+1} \quad \text{for } i = 1, \ldots, n. \tag{21}$$

A return forecast is then computed based on each of the individual univariate predictive regressions:

$$\hat{r}_{t+1|t}^{(i)} = \hat{\alpha}_{i,t} + \hat{\beta}_{i,t} x_{i,t} \quad \text{for } i = 1, \ldots, n, \tag{22}$$

where $\hat{\alpha}_{i,t}$ and $\hat{\beta}_{i,t}$ are the OLS estimates of α_i and β_i, respectively, in Equation (21) based on data through t. In the second step, a combination forecast is formed by taking the arithmetic mean of the univariate forecasts in Equation (22):

$$\hat{r}_{t+1|t}^{\text{C-Mean}} = \frac{1}{n} \sum_{i=1}^{n} \hat{r}_{t+1|t}^{(i)}. \tag{23}$$

Relative to the multiple predictive regression forecast in Equation (20), Rapach, Strauss, and Zhou (2010) show that the combination mean (C-Mean) forecast in Equation (23) makes two adjustments. First, it replaces the multiple regression slope coefficient estimates with their univariate counterparts, which reduces the variance of the parameter estimates and improves out-of-sample performance in light of the bias-variance trade-off. Second, it shrinks the slope coefficient estimates toward zero by the factor $1/n$. These adjustments induce a strong shrinkage effect that allows for the incorporation of information from a large number of predictors in a manner that guards against overfitting. Rapach, Strauss, and Zhou (2010) find that a combination forecast like Equation (23) based on the popular predictors used by Goyal and Welch (2008) provides statistically and economically significant out-of-sample gains vis-à-vis the prevailing mean benchmark.

Since the seminal paper by Bates and Granger (1969), it has known that a combination of forecasts often performs better than a single forecast in various domains. Rapach, Strauss, and Zhou (2010) show that the benefits of forecast combination also apply to predicting the US market excess return.

The C-Mean forecast in Equation (23) uses an equal-weighted average of the individual forecasts. It can be beneficial to "tilt" the weights toward individual forecasts that appear to be more accurate. More generally, a combination forecast can be expressed as

$$\hat{r}_{t+1|t}^{\text{Combine}} = \sum_{i=1}^{n} \omega_{i,t+1|t} \hat{r}_{t+1|t}^{(i)}, \tag{24}$$

where $\omega_{i,t+1|t} \geq 0$ for $i = 1, \ldots, n$ are the combining weights and $\sum_{i=1}^{n} \omega_{i,t+1|t} = 1$. The C-Mean forecast in Equation (23) sets $\omega_{i,t+1|t} = 1/n$ for $i = 1, \ldots, n$. Rapach, Strauss, and Zhou (2010) also consider a discount MSFE (DMSFE) approach (Stock and Watson 2004) that places greater weight on individual forecasts that evince lower MSFE over a holdout out-of-sample period (for details, see Rapach, Strauss, and Zhou 2010, p. 827).

Table 1 reports updated out-of-sample results for the monthly market excess return and individual univariate predictive regression forecasts in Equation (22) based on 14 popular predictors from Goyal and Welch (2008) as well as C-Mean and DMSFE combination fore-

Table 1: Individual and Combination Forecast Results

The table reports monthly out-of-sample results for 14 univariate predictive regression and two combination forecasts of the US market excess return. The out-of-sample period is 1957:01–2020:12. Each univariate predictive regression forecast is based on the predictor variable in the first column. C-Mean is a combination forecast based on the arithmetic mean of the individual univariate predictive regression forecasts. DMSFE is a combination forecast that attaches more weight to individual univariate predictive regression forecasts with lower MSFE over a holdout out-of-sample period. R^2_{OS} is the Campbell and Thompson (2008) out-of-sample R^2 statistic. It measures the proportional reduction in MSFE for the univariate predictive regression or combination forecast via-à-vis the prevailing mean benchmark. For the positive R^2_{OS} statistics, *, **, and *** indicate that the reduction in MSFE is significant at the 10%, 5%, or 1% level, respectively, according to the Clark and West (2007) test. Gain is the annualized increase in CER for a mean-variance investor with a relative risk aversion coefficient of five who uses the univariate predictive regression forecast or combination forecast instead of the prevailing mean benchmark to allocate between stocks and risk-free Treasury bills.

(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Overall		Expansion		Recession	
Predictor	R^2_{OS} (%)	Gain (%)	R^2_{OS} (%)	Gain (%)	R^2_{OS} (%)	Gain (%)
log(DP)	−0.36	0.32	−1.53	−1.84	2.29***	12.51
log(DY)	−0.75	0.46	−2.55	−2.42	3.34***	16.85
log(EP)	−1.92	0.24	−2.29	−0.76	−1.05	5.69
log(DE)	−1.75	−0.41	−1.02	0.00	−3.41	−2.66
SVAR	−0.44	−0.19	0.03	−0.29	−1.50	0.45
BM	−1.93	−1.17	−2.79	−2.48	0.02	6.03
NTIS	−0.60	−0.05	0.71	0.88	−3.56	−5.49
TBL	0.21*	1.47	−0.41	0.49	1.62*	6.94
LTY	−0.83	1.15	−1.72	0.04	1.20	7.36
LTR	−0.08	0.49	−0.70	−0.14	1.33*	3.81
TMS	0.02	1.07	−0.42	−0.01	0.99*	7.13
DFY	−0.03	0.23	−0.06	−0.10	0.02	1.89
DFR	−0.07	0.74	0.35*	0.53	−1.03*	1.99
INFL	−0.03	0.36	0.15	0.20	−0.42	1.34
C-Mean	0.33**	1.04	0.11	0.31	0.84**	5.13
DMSFE	0.39**	1.27	0.08	0.30	1.10**	6.74

casts based on pooling the 14 univariate predictive regression forecasts.[2] The in-sample

[2]The univariate predictive regression forecasts are based on an expanding estimation window. As in Campbell and Thompson (2008), we forecast the variance in Equation (16) using the sample variance and a 60-month rolling estimation window.

estimation period begins in 1926:12, while the out-of-sample period spans 1957:01–2020:12. The 14 individual predictors are defined as follows:

1. Log dividend-price ratio [log (DP)]: log of 12-month moving sum of dividends on the S&P 500 index minus the log of the S&P 500 index.

2. Log dividend yield [log (DY)]: log of 12-month moving sum of dividends minus the log of the lagged S&P 500 index.

3. Log earnings-price ratio [log (EP)]: log of 12-month moving sum of earnings on the S&P 500 index minus the log of the S&P 500 index.

4. Log dividend-payout ratio [log (DE)]: log of 12-month moving sum of dividends minus the log of 12-month moving sum of earnings on the S&P 500 index.

5. Stock variance (SVAR): monthly sum of squared daily returns on the S&P 500 index.

6. Book-to-market ratio (BM): book-to-market value ratio for the DJIA.

7. Net equity expansion (NTIS): ratio of 12-month moving sum of net equity issues by NYSE-listed stocks to the total end-of-year market capitalization of NYSE stocks.

8. Treasury bill rate (TBL): three-month Treasury bill yield (secondary market).

9. Long-term yield (LTY): long-term government bond yield.

10. Long-term return (LTR): return on long-term government bonds.

11. Term spread (TMS): long-term government bond yield minus the Treasury bill yield.

12. Default yield spread (DFY): difference between BAA- and AAA-rated corporate bond yields.

13. Default return spread (DFR): long-term corporate bond return minus the long-term government bond return.

14. Inflation (INFL): calculated from the CPI (all urban consumers); $x_{i,t-1}$ is used in Equations (21) and (22) for inflation to account for the delay in CPI releases.

The updated data are from Amit Goyal's website.[3]

The results in Table 1 are reminiscent of those in the work by Rapach, Strauss, and Zhou (2010). According to the second column, the individual predictors generally fare quite poorly for the full 1957:01–2020:12 out-of-sample period. For the 14 individual predictors, only two of the R^2_{OS} statistics are positive: 0.21% for TBL and 0.02% for TMS, and only the former is significant at the 10% level.[4] Forecast combination improves out-of-sample performance, as both the C-Mean and DMSFE combination forecasts deliver positive R^2_{OS} statistics of 0.33% and 0.39%, respectively, both of which are significant at the 5% level and greater than the largest R^2_{OS} statistic for the individual predictors. The C-Mean and DMSFE forecasts also provide sizable annualized CER gains of 104 and 127 basis points, respectively.

Table 1 illustrates another finding in the study by Rapach, Strauss, and Zhou (2010) and a number of subsequent studies (e.g., Henkel, Martin, and Nardari 2011; Dangl and Halling 2012): out-of-sample market return predictability tends to be substantially stronger in business-cycle recessions than expansions.[5] As shown by comparing columns four and five with six and seven, a number of the individual predictors perform markedly better during recessions than during expansions. Focusing on the combination forecasts, the C-Mean and DMSFE forecasts generate positive R^2_{OS} statistics of 0.11% and 0.08%, respectively, during expansions (neither of which is significant at conventional levels); the statistics increase substantively to 0.84% and 1.10% during recessions, each of which is significant at the 5% level (despite the reduced number of observations). A similar pattern holds for the annualized

[3]The authors thank Amit Goyal for generously providing updated data for popular predictors on a regular basis.

[4]As expected, the forecast based on the high-dimensional predictive regression in Equation (20) performs poorly ($R_{\mathrm{OS}} = -8.04\%$), a clear manifestation of overfitting.

[5]Cujean and Hasler (2017) provide a theoretical explanation for this finding.

CER gains, which increase from 31 (30) basis points during expansions to 513 (674) basis points during recessions for the C-Mean (DMSFE) forecast.

Recently, Dong et al. (2022) investigate links between long-short anomaly portfolio returns from the cross-sectional literature and the US market excess returns. Specifically, they use 100 representative anomaly portfolio returns to forecast the monthly market excess return. For the 1985:01–2017:12 out-of-sample period, a C-Mean forecast based on the 100 anomalies generates an R^2_{OS} statistic of 0.89% (significant at the 1% level) and an annualized CER gain of 259 basis points for a mean-variance investor with a relative risk aversion coefficient of three. Economically, Dong et al. (2022) attribute the predictive power of anomaly portfolio returns for the market excess return to asymmetric limits of arbitrage (Shleifer and Vishny 1997) and overpricing correction persistence.

Principal Component Regression

Another strategy for dealing with a large number of potential predictors while guarding against overfitting is to employ dimension reduction via principal component analysis (PCA). Specifically, PCA is initially used to extract the first or first few principal components from the n individual predictors, where each principal component is a linear combination of the underlying variables. The principal components then serve as the predictors in a low-dimensional predictive regression:

$$r_{t+1} = \alpha + \sum_{j=1}^{k} \beta_j z_{j,t} + \varepsilon_{t+1}, \tag{25}$$

where $z_{j,t}$ for $j = 1, \ldots, k$ are the first k principal components and $k \ll n$. The individual predictors are typically standardized to have zero mean and unit variance before the principal components are computed, and the components are uncorrelated by construction. The forecast corresponding to Equation (25) is given by

$$\hat{r}^{\text{PC}}_{t+1|t} = \hat{\alpha}_t + \sum_{j=1}^{k} \hat{\beta}_{j,t} \hat{z}_{i,t}, \tag{26}$$

where $\hat{\alpha}_t$ and $\hat{\beta}_{j,t}$ are the OLS estimates of α and β_j, respectively, in Equation (25) based on data through t, and $\hat{z}_{j,t}$ is the jth principal component, again based on data through t, so that there is no look-ahead bias in the forecast. Intuitively, by computing the first few principal components, much of the noise in the individual predictors is removed to obtain a more reliable predictive signal in a low-dimensional setting.

Ludvigson and Ng (2007, 2009) apply principal component regression to forecast stock and bond returns, respectively, based on a large set of macroeconomic variables. They find that forecasts that include principal components based on macroeconomic variables outperform those that ignore the macroeconomic variables in terms of MSFE. Neely et al. (2014) construct principal component regression forecasts of the monthly market excess return based on a set of 14 economic variables from Goyal and Welch (2008) and 14 technical indicators. They find that predictive regression forecasts based on principal components extracted from the economic variables and technical indicators significantly outperform the prevailing mean benchmark with respect to the market excess return and provide substantive economic value to a mean-variance investor. Furthermore, they show that the information in technical indicators complements that in economic variables, with technical indicators being especially adept at predicting the relatively low market return typically realized near business-cycle peaks.

Table 2 reports updated results from Neely et al. (2014). The same 14 economic variables in Table 1 from Goyal and Welch (2008) are used.[6] The technical indicators are based on popular signals among trend-following traders, including long-short moving averages, momentum

[6] The exception is stock variance, SVAR, which uses a modified version, RVOL, due to Mele (2007).

signals, and on-balance volume. The technical indicators appear as indicator values that take of value of one (zero) in the case of a buy (sell) signal (for details on constructing the technical indicators, see Neely et al. 2014, p. 1775). Based on data availability, the sample begins in 1950:12. The out-of-sample period spans 1966:01–2020:12. Considering a maximum value of four, the adjusted R^2 is used to determine the number of principal components (k) to include in the forecast in Equation (26).

Table 2: Individual and Principal Component Regression Forecast Results

The table reports monthly Campbell and Thompson (2008) R^2_{OS} statistics in percent for 14 univariate predictive regressions based on economic variables, 14 univarite predictive regressions based on technical indicators, and a principal component regression forecast of the US market excess return. The out-of-sample period is 1966:01–2020:12. Each univariate predictive regression forecast is based on the predictor variable in the first or fifth column. PCR is the principal component regression forecast based on the first k principal components extracted from the entire set of 28 predictors, where k is determined by the adjusted R^2 statistic using data available at the time of forecast formation. The R^2_{OS} statistic measures the proportional reduction in MSFE for the univariate predictive regression or principal component regression forecast via-à-vis the prevailing mean benchmark. For the positive R^2_{OS} statistics in the second and sixth columns, *, **, and *** indicate that the reduction in MSFE is significant at the 10%, 5%, or 1% level, respectively, according to the Clark and West (2007) test.

(1) Predictor	(2) Overall	(3) Expansion	(4) Recession	(5) Predictor	(6) Overall	(7) Expansion	(8) Recession
log(DP)	−0.30	−0.30	0.86	MA(1,12)	0.36*	−0.82	2.27
log(DY)	−0.62	−1.62	1.58	MA(2,9)	0.11	−0.62	2.74
log(EP)	−0.76	−0.76	−1.25	MA(2,12)	0.43*	−0.66	1.97
log(DE)	0.21**	−1.21	0.20	MA(3,9)	0.04	−0.36	2.36
RVOL	−1.18	−0.18	0.72	MA(3,12)	−0.15	−0.65	1.73
BM	−0.82	−0.82	−3.03	MOM(9)	−0.10	−0.47	0.62
NTIS	−0.61	−0.61	−2.72	MOM(12)	−0.06	−0.49	0.84
TBL	−0.32	−1.32	0.86	MA(1,9)	0.08	−0.44	0.86
LTY	−0.55	−0.55	0.26	VOL(1,9)	0.30	−0.49	2.22
LTR	0.31**	−1.31	5.18	VOL(1,12)	0.39*	−0.23	1.90
TMS	−0.87	−2.87	3.46	VOL(2,9)	0.07	−0.34	1.05
DFY	−0.54	−0.54	−0.74	VOL(2,12)	0.04	−0.03	0.20
DFR	−0.54	0.54	−2.03	VOL(3,9)	−0.10	−0.29	0.39
INFL	−0.34	0.34	−1.24	VOL(3,12)	0.43*	0.01	1.44
PCR	1.19***	−2.07	9.14				

The overall results are similar to those in Neely et al. (2014) and demonstrate the usefulness of the principal component regression approach. Only two of the economic variables, log(DE) and LTR, deliver positive R^2_{OS} statistics of 0.21% and 0.31%, respectively (both of which are significant at the 5% level). The technical indicators perform somewhat better overall, with ten of the 14 R^2_{OS} statistics being positive. The positive R^2_{OS} statistics range from 0.04% to 0.43% (four are significant at the 10% level). The principal component regression forecast produces an R^2_{OS} statistic of 1.19% (significant at the 1% level), which is substantially greater than the largest statistic for the univariate predictive regression forecasts. Similarly to Table 1, the results in Table 2 indicate that market return predictability tends to be concentrated in recessions. This is especially evident for the technical indicators and principal component regression forecast.

Dong et al. (2022) extract the first principal component from 100 long-short anomaly portfolio returns to forecast the monthly market excess return. The principal component regression forecast provides an R^2_{OS} statistic of 1.25% (significant at the 5% level). It also provides an annualized CER gain of 328 basis points.

Partial Least Squares

Conventional PCA aims to explain as much of the total variation as possible in the predictor variables per se, so that it ignores information in the target variable that is the object of prediction (in this case, the asset return). The partial least squares (PLS) method pioneered by Wold (1966) and extended by Kelly and Pruitt (2013, 2015) takes the target variable into account by constructing linear combinations of the underlying predictors that are maximally correlated with the target.

Following Hastie, Tibshirani, and Friedman (2009, Section 3.5), the idea is similar to PCA in that the goal is to reduce the dimension of the original predictors in the forecasting equation by using $k \ll n$ predictors:

$$r_{t+1} = \alpha + \sum_{j=1}^{k} \beta_j z_{j,t}^* + \varepsilon_{t+1}, \tag{27}$$

where $z_{j,t}^*$ is a linear combination of $\{x_{i,t}\}_{i=1}^n$, each $x_{i,t}$ is standardized to have zero mean and unit variance, and $\left\{z_{j,t}^*\right\}_{j=1}^k$ are constructed to be uncorrelated. The first PLS predictor (or target-relevant factor) is given by

$$z_{1,t}^* = \sum_{i=1}^{n} \phi_{1,i} x_{i,t}, \tag{28}$$

where $\{\phi_{1,i}\}_{i=1}^n$ are linear combination coefficients to be determined. Unlike in conventional PCA, the information in the target r_{t+1} is used. A simple and intuitive way to do this is to weight each $x_{i,t}$ by its covariance with r_{t+1}:

$$\phi_{1,i} = \text{cov}(x_{i,t}, r_{t+1}), \tag{29}$$

which is easily estimated by the sample covariance between $x_{i,t}$ and r_{t+1}. Then, $z_{1,t}^*$ in Equation (28) can be computed, and the one-factor PLS regression is given by

$$r_{t+1} = \alpha + \beta_1 z_{1,t}^* + \varepsilon_{t+1}, \tag{30}$$

which can be straightforwardly estimated via OLS. The forecast based on Equation (30) is given by

$$\hat{r}_{t+1}^{\text{PLS}} = \hat{\alpha}_t + \hat{\beta}_{1,t} \hat{z}_{1,t}^*, \tag{31}$$

where $\hat{\alpha}_t$ and $\hat{\beta}_{1,t}$ are the OLS estimates of α and β_1, respectively, in Equation (30) based on data through t, and $\hat{z}_{1,t}^*$ is the estimated target-relevant factor, which is again based on data through t.

Based on Algorithm 3.3 in Hastie, Tibshirani, and Friedman (2009), the following procedure can be used to compute a forecast when $k > 1$.

1. Set $r_{t+1}^{(0)} = \bar{r}$, where $\bar{r}$ is the sample mean of r_{t+1} and $x_{i,t}^{(0)} = x_{i,t}$ for $i = 1, \ldots, n$.
2. For $j = 1, \ldots, k$:
   (a) $\hat{z}_{j,t}^{*} = \sum_{i=1}^{n} \hat{\phi}_{j,i} x_{i,t}^{(j-1)}$, where $\hat{\phi}_{j,i} = \widehat{\text{cov}}\left(x_{i,t}^{(j-1)}, r_{t+1}\right)$ and $\widehat{\text{cov}}(\cdot,\cdot)$ denotes the sample covariance.
   (b) $\hat{\beta}_j = \widehat{\text{cov}}(\hat{z}_{j,t}, r_{t+1})/\widehat{\text{var}}(\hat{z}_{j,t})$, where $\widehat{\text{var}}(\cdot)$ denotes the sample variance.
   (c) $\hat{r}_{t+1}^{(j)} = \hat{r}_{t+1}^{(j-1)} + \hat{\beta}_j \hat{z}_{j,t}$.
   (d) Compute $x_{i,t}^{(j)} = x_{i,t}^{(j-1)} - \left[\widehat{\text{cov}}\left(\hat{z}_{j,t}^{*}, x_{i,t}^{(j-1)}\right)/\widehat{\text{var}}\left(\hat{z}_{j,t}^{*}\right)\right]\hat{z}_{j,t}^{*}$.

Of course, when computing an out-of-sample forecast of r_{t+1} based on information available through t, all of the computations in the algorithm should use only data through t.

Theoretically, Helland and Almoy (1994) provide asymptotic theory for PLS with n fixed while T goes to infinity. Kelly and Pruitt (2013, 2015) extend PLS and provide asymptotic theory for both n and T going to infinity. Cook and Forzani (2019) address various asymptotic issues relating to PLS, while Cook and Forzani (2021) provide a nonlinear extension of PLS.

PLS has proven quite useful for forecasting the US market excess return. For example, consider using investor sentiment to predict the market return. There are a number of proxies for investor sentiment. The best predictor cannot be known a priori, and multiple predictors can contain relevant information. Huang et al. (2015) show that a target-relevant factor derived from the Baker and Wurgler (2006) sentiment proxies outperforms the individual proxies as well as the first conventional principal component extracted from the sentiment proxies. Jiang et al. (2019) and Chen et al. (2020) show that manager and employee sentiment, respectively, can also predict the market return in the context of PLS.

Consider updated results for Huang et al. (2015) for forecasting the monthly market excess return. Six market sentiment proxies from Baker and Wurgler (2006) are used: closed-end fund discount, share turnover, number of IPOs, monthly average of first-day returns of IPOs, dividend premium, and equity share in new issues. The updated sample spans 1965:07–2020:12, and the out-of-sample period covers 1985:01–2020:12. The results are similar to those in Huang et al. (2015). The principal component regression forecast based on the first principal component extracted from the six sentiment proxies fails to outperform the prevailing mean benchmark ($R^2_{\mathrm{OS}} = -0.10\%$). The C-Mean forecast in Equation (23) constructed from the six proxies outperforms the prevailing mean, with an R^2_{OS} statistic of 0.50% (significant at the 1% level). The PLS forecast based on the first target-relevant factor extracted from the proxies performs even better, producing an R^2_{OS} statistic of 1.00% (significant at the 1% level).

Providing further evidence of the efficacy of PLS for out-of-sample market excess return prediction, Kelly and Pruitt (2013) show that a target-relevant factor extracted from a cross-section of book-to-market ratios based on size- and value-sorted portfolios significantly outperforms the prevailing mean benchmark for forecasting the market excess return. In addition, Dong et al. (2022) extract a target-relevant factor from 100 long-short anomaly portfolio returns and find that it generates substantial statistical and economic out-of-sample gains for forecasting the monthly market excess return, with an R^2_{OS} statistic of 2.06% (significant at the 1% level) and a massive annualized CER gain of 638 basis points.

LASSO and Elastic Net

Consider fitting the multiple predictive regression in Equation (19). As previously discussed, conventional OLS estimation of Equation (19) is prone to in-sample overfitting, especially when n is large relative to the number of available time-series observations; indeed, if the number of predictors is greater than the number of available observations, then the OLS estimator cannot be computed. The problem of overfitting is exacerbated when the data are noisy.

Tibshirani (1996) proposes the least absolute shrinkage and selection operator (LASSO), which has become one of the most popular machine-learning techniques for improving estimation of high-dimensional regressions. The LASSO is a shrinkage device that permits shrinkage to zero for one or more slope coefficients, so that it also performs variable selection, which facilitates model interpretation. Given the plethora of plausible return predictors that exist, the LASSO can be a valuable tool for forecasting stock returns.

In the context of Equation (19), the LASSO solves the following optimization problem:

$$\min_{\alpha,\beta_1,\ldots,\beta_n} \sum_{t=1}^{T-1}\left(r_{t+1}-\alpha-\sum_{i=1}^{n}\beta_i x_{i,t}\right)^2 \quad \text{subject to} \quad \sum_{i=1}^{n}|\beta_i| \leq c, \tag{32}$$

where $c \geq 0$. The first part of Equation (32) is the OLS objective function, while the constraint induces shrinkage in the slope parameters. As c becomes smaller, more shrinkage is induced. Mathematically, Equation (32) is equivalent to the Lagrangian form:

$$\min_{\alpha,\beta_1,\ldots,\beta_n} \left[\sum_{t=1}^{T-1}\left(r_{t+1}-\alpha-\sum_{i=1}^{n}\beta_i x_{i,t}\right)^2 + \lambda\sum_{i=1}^{n}|\beta_i|\right], \tag{33}$$

where $\lambda \geq 0$ is a hyperparameter governing the degree of shrinkage in the penalty term. When $\lambda = 0$, Equation (33) reduces to conventional OLS estimation. As λ increases, more shrinkage is induced. Because the penalty term is based on the ℓ_1 norm, the constraint permits shrinkage to zero (for sufficiently large λ). There is no analytical solution in general to Equation (33), but powerful algorithms are available to efficiently solve the problem and are readily available in software packages like Matlab, R, and Python.

To gain some intuition, consider a special case of a univariate predictive regression without an intercept and with the predictor x_t standardized:

$$r_{t+1} = \beta x_t + \varepsilon_{t+1} \quad \text{for } t = 1, \ldots, T-1. \tag{34}$$

In this case, the goal is to select β to minimize

$$\sum_{t=1}^{T-1} (r_{t+1} - \beta x_t)^2 + \lambda |\beta|. \tag{35}$$

The first-order condition is

$$-\sum_{t=1}^{T-1} (r_{t+1} - \beta x_t) x_t + \lambda \, \mathrm{sign}(\beta) = 0, \tag{36}$$

where $\mathrm{sign}(\cdot)$ is the sign function, so that $\mathrm{sign}(\beta) = 1$ or -1 if $\beta > 0$ or < 0, respectively. Assuming $\beta > 0$, solving from above yields

$$\hat{\beta}_{\mathrm{LASSO}} = \hat{\beta} - \lambda \left(\sum_{t=1}^{T-1} x_t^2 \right)^{-1}, \tag{37}$$

where $\hat{\beta}$ is the convetional OLS estimator (i.e., without the shrinkage constraint):

$$\hat{\beta} = \sum_{t=1}^{T-1} r_{t+1} x_t / \sum_{t=1}^{T-1} x_t^2. \tag{38}$$

If $\lambda = 0$, then $\hat{\beta}_{\mathrm{LASSO}} = \hat{\beta}$. As λ increases to $\hat{\beta} \sum_{t=1}^{T-1} x_t^2$ and beyond, the LASSO shrinks $\hat{\beta}$ to zero. ($\hat{\beta}_{\mathrm{LASSO}}$ is defined as zero if the right-hand-side of Equation (37) is negative because that equation is solved by assuming $\beta > 0$.) In this simple case, the LASSO estimator is a piecewise linear function of λ. This result extends to the general case.

The LASSO optimization problem in Equation (33) employs the ℓ_1 norm in the penalty term. When the ℓ_1 norm is replaced with the ℓ_2 norm, we obtain the well-known ridge objective function (Hoerl and Kennard 1970):

$$\min_{\alpha,\beta_1,\ldots,\beta_n} \left[\sum_{t=1}^{T-1} \left(r_{t+1} - \alpha - \sum_{i=1}^{n} \beta_i x_{i,t} \right)^2 + \lambda \sum_{i=1}^{n} \beta_i^2 \right], \tag{39}$$

where $\lambda \geq 0$ is again a hyperparameter governing the degree of shrinkage. Although Equation (39) differs from Equation (33) only by the replacement of $|\beta_i|$ with β_i^2 in the penalty term, the behaviors of the estimators are quite different.

To understand ridge estimation, it is useful to consider a regression model expressed in matrix notation, where we exclude the intercept term for notational convenience:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{y}$ is the T-vector of dependent variable observations, $\boldsymbol{X}$ is the T-by-n data matrix, $\boldsymbol{\beta}$ is the n-vector of slope coefficients, and $\boldsymbol{\varepsilon}$ is the T-vector of disturbance terms. The ridge estimator is solved explicitly as

$$\hat{\boldsymbol{\beta}}_{\text{Ridge}} = (\boldsymbol{X}'\boldsymbol{X} + \lambda \boldsymbol{I}_n)^{-1} \boldsymbol{X}'\boldsymbol{y}, \tag{40}$$

where $\boldsymbol{I}_n$ is the n-dimensional identity matrix. According to Equation (40), the ridge estimator becomes the conventional OLS estimator when $\lambda = 0$, and it generally induces more shrinkage toward zero in the coefficient estimates as λ increases. Note that the OLS estimator inverts $\boldsymbol{X}'\boldsymbol{X}$, while the ridge estimator inverts $\boldsymbol{X}'\boldsymbol{X}$ plus some positive number along the main diagonal, so that the ridge estimator is particularly useful when $\boldsymbol{X}'\boldsymbol{X}$ is close to being singular. It thus works well when n is large and there is a concern about multicollinearity among the regressors. Although the ridge estimator can shrink the coefficients, unlike the LASSO, it cannot shrink them to exactly zero. Hence, it cannot be used to effectively reduce the dimensionality of the regression or to perform variable selection.

A potential drawback to the LASSO is that it tends to arbitrarily select one predictor from a group of correlated predictors and to shrink the coefficients for the other variables to zero. Zou and Hastie (2005) develop the elastic net (ENet), which refines the LASSO by combining the LASSO and ridge to leverage the relative advantages of both approaches. For the multiple predictive regression in Equation (19), the penalty term in the ENet objective function includes both ℓ_1 and ℓ_2 components:

$$\min_{\alpha,\beta_1,\ldots,\beta_n} \left\{ \sum_{t=1}^{T-1} \left(r_{t+1} - \alpha - \sum_{i=1}^{n} \beta_i x_{i,t} \right)^2 + \lambda \sum_{i=1}^{n} \left[0.5(1-\delta)\beta_i^2 + \delta|\beta_i| \right] \right\}, \tag{41}$$

where $0 \leq \delta \leq 1$ is a hyperparameter for blending the ℓ_1 and ℓ_2 components in the penalty term. When $\delta = 1$ ($\delta = 0$), ENet estimation reduces to LASSO (ridge) estimation. Since its introduction, the ENet has been widely used to implement shrinkage estimation.

To operationalize the LASSO or ENet, it is necessary to select (or "tune") the hyperparameter λ, which governs the degree of shrinkage. The challenge is to achieve the proper balance in light of the bias-variance trade-off. The goal is to induce sufficient shrinkage to prevent overfitting without sacrificing too much of the relevant information in the predictors. The most popular way to tune λ is K-fold cross-validation. However, selection of the number and composition of the folds is somewhat arbitrary. Alternatively, information criteria can be used. For example, Flynn, Hurvich, and Simonoff (2013) show that the corrected Akaike (1973) information criterion (Hurvich and Tsai 1989) has desirable asymptotic properties and performs well in finite-sample simulations for tuning λ. It is also necessary to tune δ, and cross-validation or an information criterion can again be used. More simply, Hastie and Qian (2016) recommend setting $\delta = 0.5$.

In an early application in finance, Rapach, Strauss, and Zhou (2013) use the ENet to fit multiple predictive regressions for individual monthly country stock returns where lagged returns for a large number of countries serve as predictors. They find evidence that the

US market return leads returns in numerous other countries. Chinco, Clark-Joseph, and Ye (2019) use the LASSO to predict high-frequency individual stock returns using lagged returns from across the market. They find significant evidence of out-of-sample predictability for one-minute-ahead returns. Dong et al. (2022) find that a monthly US market excess return forecast based on ENet estimation of a multiple predictive regression with 100 long-short anomaly portfolio returns as predictors generates an R^2_{OS} statistic of 2.03% (significant at the 5% level) and a substantive CER gain of 626 basis points.

C-ENet

Incorporating insights from Diebold and Shin (2019), Rapach and Zhou (2020) use the elastic net to refine the C-Mean forecast in Equation (23), producing a combination ENet (C-ENet) forecast. They proceed in three steps. Like the C-Mean approach, univariate predictive regression forecasts based on each of the individual predictors (considered in turn) are first computed, as in Equation (22). In the second step, a Granger and Ramanathan (1984) regression is estimated over a holdout out-of-sample period via the ENet:

$$r_s = \eta + \sum_{i=1}^{n} \theta_i r_{s|s-1}^{(i)} + \varepsilon_s \quad \text{for } s = t_1 + 1, \ldots, t, \tag{42}$$

where t_1 is the size of the initial in-sample period.[7] Let $\mathcal{I}_t$ be the set of univariate predictive regression forecasts selected by the ENet in Equation (42). In the final step, instead of averaging across all of the individual predictive regression forecasts, as in Equation (23), the average is taken across the individual forecasts selected by the ENet in Equation (42):

$$\hat{r}_{t+1|t}^{\text{C-ENet}} = \frac{1}{|\mathcal{I}_t|} \sum_{i \in \mathcal{I}_t} \hat{r}_{t+1|t}^{(i)}, \tag{43}$$

[7]When fitting Equation (42) via the ENet, we impose the restriction that $\theta_i \geq 0$, which is a reasonable condition for a forecast to be informative.

where $|\mathcal{I}_t|$ is the cardinality of $\mathcal{I}_t$. Intuitively, the C-Mean forecast is refined by including only the predictors deemed relevant by the ENet in Equation (42) when forming the combination forecast.[8]

Considering a set of twelve economic variables and technical indicators, Rapach and Zhou (2020) compute monthly market excess return forecasts for the 1957:01–2018:12 out-of-sample period. The C-Mean forecast produces an R^2_{OS} statistic of 1.11% (significant at the 1% level), while the R^2_{OS} statistic for the C-ENet forecast is nearly twice as large (2.12%, significant at the 1% level). The C-ENet forecast also generates an annualized CER gain of 375 basis points for a mean-variance investors with a relative risk aversion coefficient of five. When forecasting the monthly market excess return with 100 anomaly portfolio returns, Dong et al. (2022) find that the C-ENet forecast delivers a sizable R^2_{OS} statistic of 2.81% (significant at the 1% level) and a large annualized CER gain of 606 basis points.[9]

Conclusion

This article covers advances in the out-of-sample forecasting of asset returns, focusing on the US market excess return. Because the market return contains an intrinsically large unpredictable component, out-of-sample prediction will necessarily be a challenging venture, to say the least. Nevertheless, the literature indicates that the US market excess return is predictable to a statistically and economically significant extent on an out-of-sample basis. As emphasized in this article, the key to improving out-of-sample performance is to move beyond conventional OLS estimation of predictive regressions, which is susceptible to in-sample overfitting, and to utilize alternative methods that are better designed to handle

[8] Han et al. (2021) apply the same basic idea when computing combination forecasts of stock returns in a cross-sectional setting.

[9] Deep learning approaches are not reviewed here because the typical sample sizes for monthly data are relatively limited for time-series forecasting, making them difficult to apply. However, deep learning techniques can be effectively applied in cross-sectional stock return forecasting, where the number of observations is much larger (see, e.g., Gu, Kelly, and Xiu 2020).

large information sets and noisy data. These methods employ techniques like shrinkage and dimension reduction to improve out-of-sample performance in light of the bias-variance trade-off. The methods reviewed include forecast combination, principal component regression, PLS, the LASSO, ENet, and C-ENet. These appear to be valuable tools for forecasting the market return.

Much of the earlier literature on predicting the market excess return relies on financial fundamentals, especially valuation ratios (e.g., the dividend yield and price-earnings ratio) and interest rates (including functions of interest rates, such as the term spread). Market return predictability relating to these variables is generally believed to reflect a rational time-varying equity premium consistent with market efficiency. A number of more recent studies provide significant evidence of out-of-sample market return predictability based on a variety of new variables, including the following:

- short interest (Rapach, Ringgenberg, and Zhou 2016; Chen, Da, and Huang forthcoming),
- options (Bollerslev, Tauchen, and Zhou 2009; Martin 2017; Liu, Tang, and Zhou 2022),
- ESG and corporate activity (Chang et al. 2021; Lie et al. 2021),
- technical indicators, such as long-short MAs, momentum signals, and on-balance volume, which are popular with many traders (Neely et al. 2014),
- investor, manager, employee, and music sentiment (Huang et al. 2015; Jiang et al. 2019; Chen et al. 2020; Edmans et al. forthcoming), and
- long-short anomaly portfolio returns from the cross-sectional literature (Dong et al. 2022).

The out-of-sample predictive ability of many of these variables appears more difficult to square with market efficiency, and they point to, among other things, significant behavioral biases, information frictions, and limits to arbitrage in the equity market. Because there is some evidence that return predictability diminishes with the publication of academic studies (Schwert 2003; McLean and Pontiff 2016), it will be interesting to see the extent to which the new variables used in recent studies retain their out-of-sample predictive ability going forward.[10]

Forecasting the market return is of considerable interest to academics and practitioners alike. Accordingly, there is a vast literature on the topic. This article shows that forecasting the market return remains an exciting area of research, with new methods and predictors being proposed to improve out-of-sample performance.

Acknowledgments

The authors thank Dashan Huang, Fuwei Jiang, Christopher Neely, Matthew Ringgenberg, Jun Tu, and two anonymous referees for helpful discussions and Songrun He for outstanding research assistance.

Further Reading

Koijen, R. and S. Van Nieuwerburgh (2011). Predictability of Returns and Cash Flows. *Annual Review of Financial Economics* 3, 467–491.

[10]Using an updated out-of-sample period for 1990:01–2020:12, the short interest index from Rapach, Ringgenberg, and Zhou (2016) produces R^2_{OS} statistics of 1.35%, 4.40%, 8.14%, and 6.82% for horizons of one, three, six, and twelve months, respectively, all of which are significant at the 1% level and reasonably close to the values reported in the original study for an out-of-sample period ending in 2014:12.

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.

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.

Baker, M. and J. Wurgler (2006). Investor Sentiment and the Cross-Section of Stock Returns. *Journal of Finance* 61:4, 1645–1680.

Bansal, R. and A. Yaron (2004). Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles. *Journal of Finance* 59:4, 1481–1590.

Barberis, N. (2000). Investing for the Long Run When Returns Are Predictable. *Journal of Finance* 55:1, 225–264.

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

Bollerslev, T., G. Tauchen, and H. Zhou (2009). Expected Stock Returns and Variance Risk Premia. *Review of Financial Studies* 22:11, 4463–4492.

Brown, D. P. and R. H. Jennings (1989). On Technical Analysis. *Review of Financial Studies* 2:4, 527–551.

Campbell, J. Y. (1987). Stock Returns and the Term Structure. *Journal of Financial Economics* 18:2, 373–399.

Campbell, J. Y. and J. H. Cochrane (1999). By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior. *Journal of Political Economy* 107:2, 205–251.

Campbell, J. Y. and R. J. Shiller (1988). The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors. *Review of Financial Studies* 1:3, 195–228.

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

Cespa, G. and X. Vives (2012). Dynamic Trading and Asset Prices: Keynes vs. Hayek. *Review of Economic Studies* 79:2, 539–580.

Chang, R., L. Chu, J. Tu, B. Zhang, and G. Zhou (2021). ESG and the Market Return. Working Paper.

Chen, B. and K. Maung (2020). Time-varying Forecast Combination for High-Dimensional Data. Working Paper.

Chen, J., G. Tang, J. Yang, and G. Zhou (2020). Employee Sentiment and Stock Returns. Working Paper.

Chen, Y., Z. Da, and D. Huang (forthcoming). Short Selling Efficiency. *Journal of Financial Economics.*

Chinco, A., A. D. Clark-Joseph, and M. Ye (2019). Sparse Signals in the Cross-Section of Returns. *Journal of Finance* 74:1, 449–492.

Clark, T. E. and M. W. McCracken (2001). Tests of Equal Forecast Accuracy and Encompassing for Nested Models. *Journal of Econometrics* 105:1, 85–110.

Clark, T. E. and K. D. West (2007). Approximately Normal Tests for Equal Predictive Accuracy in Nested Models. *Journal of Econometrics* 138:1, 291–311.

Cochrane, J. H. (2004). *Asset Pricing: Revised Edition*. Princeton, NJ: Princeton University Press.

Cook, R. D. and L. Forzani (2019). Partial Least Squares Prediction in High-Dimensional Regression. *Annals of Statistics* 47:2, 884–908.

Cook, R. D. and L. Forzani (2021). PLS Regression Algorithms in the Presence of Nonlinearity. *Chemometrics and Intelligent Laboratory Systems* 213:1, 15 June 2021.

Cujean, J. and M. Hasler (2017). Why Does Return Predictability Concentrate in Bad Times? *Journal of Finance* 72:6, 2717–2757.

Dangl, T. and M. Halling (2012). Predictive regressions with time-varying coefficients. *Journal of Financial Economics* 106:1, 157–181.

DeLong, J. B., A. Shleifer, L. H. Summers, and R. J. Waldmann (1990). Noise Trader Risk in Financial Markets. *Journal of Political Economy* 98:4, 703–738.

Detzel, A. L., H. Liu, J. Strauss, G. Zhou, and Y. Zhu (2021). Learning and Predictability via Technical Analysis: Evidence from Bitcoin and Stocks with Hard-to-Value Fundamentals. *Financial Management* 50:1, 107–137.

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

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

Dong, X., Y. Li, D. E. Rapach, and G. Zhou (2022). Anomalies and the Expected Market Return. *Journal of Finance* 77:1, 639–681.

Edmans, A., A. Fernandez-Perez, A. Garel, and I. Indriawan (forthcoming). Music Sentiment and Stock Returns Around the World. *Journal of Financial Economics*.

Edmans, A., I. Goldstein, and W. Jiang (2015). Feedback Effects, Asymmetric Trading, and the Limits to Arbitrage. *American Economic Review* 105:12, 3766–3797.

Fama, E. F. (1970). Efficient Capital Markets: A review of Theory and Empirical Work. *Journal of Finance* 25:2, 383–417.

Fama, E. F. and K. R. French (1988). Dividend Yields and Expected Stock Returns. *Journal of Financial Economics* 21:1, 3–25.

Fama, E. F. and G. W. Schwert (1977). Asset Returns and Inflation. *Journal of Financial Economics* 5:2, 115–146.

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.

Gospodinov, N. and E. Maasoumi (2021). Generalized Aggregation of Misspecified Models: With an Application to Asset Pricing. *Journal of Econometrics* 222:1, 451–467.

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

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

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

Han, Y., A. He, D. E. Rapach, and G. Zhou (2021). Expected Stock Returns and Firm Characteristics: E-LASSO, Assessment, and Implications. Working Paper.

Han, Y., G. Zhou, and Y. Zhu (2016). A Trend Factor: Any Economic Gains From Using Information Over Investment Horizons? *Journal of Financial Economics* 122:2, 352–375.

Hastie, T. and J. Qian (2016). Glmnet Vignette. Manuscript.

Hastie, T., R. Tibshirani, and J. Friedman (2009). *Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*. New York: Springer.

Helland, I. S. and T. Almoy (1994). Comparison of Prediction Methods When Only a Few Components are Relevant. *Journal of the American Statistical Association* 89:426, 583–592.

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

Hoerl, A. E. and R. W. Kennard (1970). Ridge Regression: Applications to Nonorthogonal Problems. *Technometrics* 12:1, 69–82.

Hong, H. and J. C. Stein (1999). A Unified Theory of Underreaction, Momentum Trading, and Overreaction in Asset Markets. *Journal of Finance* 54:6, 2143–2184.

Huang, D., F. Jiang, J. Tu, and G. Zhou (2015). Investor Sentiment Aligned: A Powerful Predictor of Stock Returns. *Review of Financial Studies* 28:3, 791–837.

Huang, D., J. Li, L. Wang, and G. Zhou (2020). Time Series Momentum: Is It There? *Journal of Financial Economics* 135:3, 774–794.

Huang, D. and G. Zhou (2017). Upper Bounds on Return Predictability. *Journal of Financial and Quantitative Analysis* 52:2, 401–425.

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

Jiang, F., J. Lee, X. Martin, and G. Zhou (2019). Manager Sentiment and Stock Returns. *Journal of Financial Economics* 132:1, 126–149.

Kandel, S. and R. F. Stambaugh (1996). On the Predictability of Stock Returns: An Asset-Allocation Perspective. *Journal of Finance* 51:2, 385–424.

Keim, D. B. and R. F. Stambaugh (1986). Predicting Returns in the Stock and Bond Markets. *Journal of Financial Economics* 17:2, 357–390.

Kelly, B. and S. Pruitt (2013). Market Expectations in the Cross-Section of Present Values. *Journal of Finance* 68:5, 1721–1756.

Kelly, B. and S. Pruitt (2015). The Three-Pass Regression Filter: A New Approach to Forecasting Using Many Predictors. *Journal of Econometrics* 186:2, 294–316.

Kostakis, A., T. Magdalinos, and M. P. Stamatogiannis (2015). Robust Econometric Inference for Stock Return Predictability. *Review of Financial Studies* 28:5, 1506–1553.

Kothari, S. P. and J. Shanken (1997). Book-to-Market, Dividend Yield, and Expected Market Returns: A Time-Series Analysis. *Journal of Financial Economics* 44:2, 169–203.

Lie, E., B. Meng, Y. Qian, and G. Zhou (2021). Corporate Activities and the Market Risk Premium. Working Paper.

Liu, H., X. Tang, and G. Zhou (2022). Recovering the FOMC Risk Premium. Working Paper.

Lo, A. W. (2004). The Adaptive Markets Hypothesis. *Journal of Portfolio Management* 30:5, 15–29.

Lo, A. W. (2005). Reconciling Efficient Markets with Behavioral Finance: The Adaptive Markets Hypothesis. *Journal of Investment Consulting* 7:2, 21–44.

Lo, A. W. and A. C. MacKinlay (1988). Stock Market Prices Do Not Follow Random Walks: Evidence From a Simple Specification Test. *Review of Financial Studies* 1:1, 41–66.

Ludvigson, S. C. and S. Ng (2007). The Empirical Risk-Return Relation: A Factor Analysis Approach. *Journal of Financial Economics* 83:1, 171–222.

Ludvigson, S. C. and S. Ng (2009). Macro Factors in Bond Risk Premia. *Review of Financial Studies* 22:12, 5027–5067.

Martin, I. (2017). What is the Expected Return on the Market? *Quarterly Journal of Economics* 132:1, 367–433.

Martin, I. and S. Nagel (forthcoming). Market Efficiency in the Age of Big Data. *Journal of Financial Economics*.

McCracken, M. W. (2007). Asymptotics for Out of Sample Tests of Granger Causality. *Journal of Econometrics* 140:2, 719–752.

McCracken, M. W. and G. Valente (2018). Asymptotic Inference for Performance Fees and the Predictability of Asset Returns. *Journal of Business and Economic Statistics* 36:3, 426–437.

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

Mele, A. (2007). Asymmetric Stock Market Volatility and the Cyclical Behavior of Expected Returns. *Journal of Financial Economics* 86:2, 446–478.

Neely, C. J., D. E. Rapach, J. Tu, and G. Zhou (2014). Forecasting the Equity Risk Premium: The role of Technical Indicators. *Management Science* 60:7, 1772–1791.

Nelson, C. R. (1976). Inflation and Rates of Return on Common Stocks. *Journal of Finance* 31:2, 471–483.

Pontiff, J. and L. D. Schall (1998). Book-to-Market Ratios as Predictors of Market Returns. *Journal of Financial Economics* 49:2, 141–160.

Rapach, D. E., M. C. Ringgenberg, and G. Zhou (2016). Short Interest and Aggregate Stock Returns. *Journal of Financial Economics* 121:1, 46–65.

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., J. K. Strauss, and G. Zhou (2013). International Stock Return Predictability: What is the Role of the United States? *Journal of Finance* 68:4, 1633–1662.

Rapach, D. E. and G. Zhou (2020). Time-Series and Cross-Sectional Stock Return Forecasting: New Machine Learning Methods. In *Machine Learning for Asset Management: New Developments and Financial Applications*. Ed. by E. Jurczenko. Hoboken, NJ: Wiley, pp. 1–34.

Ross, S. A. (2005). *Neoclassical Finance*. Princeton, NJ: Princeton University Press.

Rozeff, M. S. (1984). Dividend Yields and Equity Risk Premiums. *Journal of Portfolio Management* 11:1, 68–75.

Schwert, G. W. (2003). Anomalies and Market Efficiency. In *Handbook of of the Economics of Finance*. Ed. by G. M. Constantinides, M. Harris, and R. M. Stulz. Vol. 1B. Amsterdam: Elsevier, pp. 939–974.

Shleifer, A. and R. W. Vishny (1997). The Limits of Arbitrage. *Journal of Finance* 52:1, 35–55.

Stambaugh, R. F. (1999). Predictive Regressions. *Journal of Financial Economics* 54:3, 375–421.

Stock, J. H. and M. W. Watson (2004). Combination Forecasts of Output Growth in a Seven-Country Data Set. *Journal of Forecasting* 23:6, 405–430.

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.

Treynor, J. L. and R. Ferguson (1985). In Defense of Technical Analysis. *Journal of Finance* 40:3, 757–773.

Wang, J. (1993). A Model of Intertemporal Asset Prices Under Asymmetric Information. *Review of Economic Studies* 60:2, 249–282.

West, K. D. (1996). Asymptotic Inference About Predictive Ability. *Econometrica* 64:5, 1067–1084.

Wold, H. (1966). Estimation of Principal Components and Related Models by Iterative Least Squares. In *Multivariate Analysis*. Ed. by P. R. Krishnajah. New York: Academic Press, pp. 391–420.

Zhou, G. (2010). How Much Stock Return Predictability Can We Expect From an Asset Pricing Model? *Economics Letters* 108:2, 184–186.

Zhu, Y. and G. Zhou (2009). Technical Analysis: An Asset Allocation Perspective on the Use of Moving Averages. *Journal of Financial Economics* 97:3, 519–544.

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.