

Forecasting the Market Equity Premium: Does Nonlinearity Matter?

Anwen Yin¹

¹ A. R. Sanchez, Jr. School of Business, Texas A&M International University, Laredo, Texas, USA

Correspondence: Anwen Yin, WHT 217B, A. R. Sanchez, Jr. School of Business, Texas A&M International University, Laredo, TX. 78041, USA. Tel: 1-956-326-2513. E-mail: anwen.yin@tamiu.edu

Received: February 21, 2021

Accepted: March 20, 2021

Online Published: April 10, 2021

doi:10.5539/ijef.v13n5p9

URL: <https://doi.org/10.5539/ijef.v13n5p9>

Abstract

We propose using the nonlinear method of smoothing splines in conjunction with forecast combination to predict the market equity premium. The smooth splines are flexible enough to capture the possible nonlinear relationship between the equity premium and predictive variables while controlling for complexity, overcoming the difficulties often attached to nonlinear methods such as computational cost, overfitting and interpretation. Our empirical results show that when used with forecast combination, the smoothing spline forecasts outperform many competing methods such as the adaptive combinations, shrinkage estimators and technical indicators, in delivering statistical and economic gains consistently.

Keywords: equity premium, forecast combination, smoothing spline, nonlinear method

1. Introduction

Accurate predictions of the aggregate equity premium are vitally important in empirical finance, as they are critical inputs into the construction of optimal portfolios and other investment decisions. As a result, a wealth of predictive variables, such as the dividend-price ratio and dividend-yield, have been proposed and shown with empirical evidence to possess valuable predictive content for the market equity premium, see Campbell (1987) and, Fama and French (1988) for example. Despite the abundance of literature supporting the predictability of the equity premium via exogenous predictors, some studies have questioned if the documented evidence of in-sample predictability can carry over to meaningful out-of-sample predictive gains on a consistent basis. One such work is Welch and Goyal (2008), in which the authors show that many predictors with documented evidence of in-sample predictability fail to forecast the market equity premium out-of-sample. The random walk benchmark, which simply applies the historical average to forecast the future equity premium, outperforms most variable-based predictive regression models.

Nonetheless, the view expressed in Welch and Goyal (2008) has been constantly challenged since its publication. In response, some studies have discovered new predictive variables in economics, accounting and finance accompanied with evidence corroborating the predictability of the equity premium. For example, Li et al. (2013) show that the implied cost of capital has valuable predictive content for excess stock returns. Jiang et al. (2019) construct a new composite index, the manager sentiment index, and show that it contains genuine predictive content for the market equity premium beyond those embedded in typical sentiment indexes in behavior finance. Another strand of the literature focuses on using alternative predictive models or estimation methods other than the ordinary least squares (OLS) considered in Welch and Goyal (2008) to uncover or restore the predictive content of many exogenous variables in the out-of-sample context. To illustrate, Campbell and Thompson (2008) show that one can uncover the predictive content of many variables for the equity premium after imposing economic theory implied restrictions on the linear predictive model, such as the sign constraints on forecasts and model slope coefficients. Given the well documented evidences of pervasive structural breaks in financial time series, such as those documented in Paye and Timmermann (2006) and Rapach and Wohar (2006), Rapach et al. (2010) propose using the simple forecast combination via equal weights to better manage the risks inherent in model selection when forecasting stock returns. Pettenuzzo et al. (2014) further extends the forecast combination analysis in a Bayesian setting under various economic restrictions.

Against this backdrop, we make three contributions to the literature of forecasting the market equity premium out-of-sample. First, we propose using the smoothing splines in the framework of nonparametric regressions to relate the aggregate equity premium to a particular exogenous variable, such as the dividend-ratio and the stock

market variance. Most studies in the extent literature center on linear models or methods. Although nonlinear methods have been considered in the forecasting stock returns literature, such as the regime-switching model outlined in Rapach and Zhou (2013), it is, however, not clear how they compare with other methods which have been shown effective, such as forecast combinations and restricted forecasts. In addition, nonlinear methods such as regime-switching, TAR and STAR, are highly parametric in nature by imposing a particular relationship between the forecast target and the predictor, which may not hold in reality. As discussed in White (2006), nonlinear predictive methods have to overcome three difficulties in practice: computational cost, dangers of overfitting and ease of interpretation. Given the substantial progress made in computational statistics, the first hurdle can be easily removed when estimating smoothing splines with modern computing facilities. Smoothing splines permit flexibility enough to capture various possible nonlinear relationships between the equity premium and exogenous predictors while penalizing overfitting. Moreover, smoothing splines, which can be embedded in the framework of generalized additive models, affords the form of additivity in model specification helpful for interpreting estimation results. Therefore, we argue that smoothing splines have the potential to uncover the genuine predictive content embedded in many variables for the market equity premium, and their use in univariate regressions helps interpret estimation results.

The second contribution we make is that we show the smoothing splines can be used in conjunction with forecast combinations to better manage the risk of model selection. Given the documented evidence of model instability, as well as the elusive nature of stock return predictability shown in Timmermann (2008), combining smoothing spline forecasts from diverse sources of univariate models could help mitigate the concerns over choosing a single best model among a large pool of candidates. Our empirical results show that the combined forecasts from smoothing splines outperform many competing methods, such as the simple combination of Rapach et al. (2010) and the adaptive combination of Timmermann (2008). For example, the combined forecasts from smoothing splines with three effective degrees of freedom report a statistically significant out-of-sample R_{OOS}^2 value of 3.062%, exceeding the values of 1.010% and 2.751% obtained from equal-weight and adaptive combinations, respectively.

In our last contribution, we show that the smoothing spline forecasts of the equity premium can deliver material predictive gains to the investor who uses them to guide optimal portfolio investment decisions on a consistent basis. To illustrate, without transaction costs, the optimal portfolio guided by smoothing spline forecasts with three effective degrees of freedom deliver a certainty equivalent return (CER) gains of 3.228% to the investor who adopts this strategy over the historical mean benchmark. In contrast, the investment strategies of equal-weight and adaptive combinations exhibit CER gains of 0.395% and 2.938%, respectively.

The remainder of this paper is organized as follows. Section two describes the baseline predictive model, the smoothing splines, forecast combinations, and various completing methods in generating the equity premium forecasts. Section three presents the data, and reports results evaluating statistical and economic performance of the equity premium forecasts. Section four concludes.

2. Econometric Methodology

In this section we describe the baseline univariate predictive model employed to construct the equity premium forecasts, which are subsequently used in forecast combination to produce the final combined prediction. We also outline the nonlinear modeling strategy of smoothing splines, and discuss its advantages and possible limitations in practice. Finally, we briefly present various alternative predictive models and methods for forecasting the equity premium, and the associated measures evaluating forecasting performance in terms of statistical and economic gains.

2.1 Baseline Predictive Model

Our baseline univariate predictive model takes the following form:

$$y_{t+1} = \beta_0 + f(x_{k,t}) + \varepsilon_t, \quad (1)$$

where y_{t+1} is the market equity premium at period $t + 1$, $x_{k,t}$ is the predictor k used at time t , f is a function relating the predictor to the equity premium, and ε_t is the error term. This baseline model is general enough to accommodate both linear and nonlinear models relating the forecast target and the predictor. For example, when $f(x_{k,t}) = \beta_1 x_{k,t}$, then the baseline model becomes the linear predictive regression model considered in Welch and Goyal (2008), Rapach et al. (2010) and Pettenuzzo et al. (2014). In Welch and Goyal (2008) and Rapach et al. (2010), the linear baseline model is estimated by OLS while Pettenuzzo et al. (2014) apply a Bayesian estimator.

2.2 Forecast Combination

The baseline model is specified linking the equity premium with a particular variable under examination. In practice, many variables could be available to the forecaster who use them to build predictive models. As a result, to better manage the risk inherent in model selection, the method of forecast combination has been proven useful. See Rapach et al. (2010) and Pettenuzzo et al. (2014) for example.

We construct the combined forecast from all baseline models according to the following:

$$\hat{y}_{t+1} = \frac{1}{K} \sum_{k=1}^K w_k \hat{y}_{k,t+1}. \quad (2)$$

where $\hat{y}_{k,t+1}$ is the forecast made by the baseline model with predictor x_k , w_k is the weight assigned to the baseline predictive model k when constructing the average forecast, and K is the total number of baseline models available. In practice, the weight is often restricted to be nonnegative and the sum of all weights is unity. The simple but effective weighting scheme is equal-weighting, in which each model in Eq.(2) receives a constant weight of $1/K$.

While alternative weighting schemes such as the discounted mean squared forecast error and the approximate Bayesian model averaging are available, empirically they do not improve upon the equally weighted forecast combination.

2.3 Smoothing Splines

In the framework of regression splines, we first specify a set of knots which can be used subsequently to produce a series of basis functions, then we typically use least squares to estimate the spline coefficients. In fitting a smooth curve to a time series dataset, what we would like to achieve is to find some function, say $f(x)$, that fits the observed data points well while controlling for overfitting. Put differently, we would like to obtain a curve that is flexible enough to capture the nonlinear relationship between the equity premium and the exogenous variables, but we do not want to end up with a wiggly curve that interpolate all the training sample observations.

A natural approach to achieve the search objective mentioned above is to find the function f which minimizes the following loss criterion:

$$\sum_{i=1}^T [y_{t+1} - f(x_t)]^2 + \lambda \int f''(z)^2 dz, \quad (3)$$

where $\lambda \geq 0$ is a tuning parameter controlling for the degree of smoothness, and $f''(z)$ indicates the second-order derivative of the function f evaluated at z . The function f that minimizes Eq.(3) is called a smoothing spline.

Note that the loss function for the smoothing splines takes the typical form of penalized information criterion formulation that often seen in the literature of dimension reduction methods, such as lasso and ridge regressions. The first term in Eq.(3) encourages f to fit the time series data well, while the second term is a penalty term that punishes the variability in f . The larger the value of λ , the smoother the f would be. When $\lambda = 0$, then f would be reduced to a polynomial flexible enough to interpolate all the data points in the training sample. In contrast, when $\lambda = \infty$, f would become extremely smooth, reducing to a straight least squares line. For an intermediate value of λ , f will approximate the nonlinear relationship between the forecast target and the predictors while being smooth to some degree. In sum, we can see that the tuning parameter λ essentially manages the bias-variance trade-off of the regression spline, and the value of λ can be adjusted via the effective degrees of freedom.

The global minimizer f of Eq.(3) is a piecewise cubic polynomial with knots at the unique values of predictors, and with continuous first and second order derivatives at each knot. In the optimization process, λ controls the degree of roughness of the smoothing spline, which is termed the effective degrees of freedom (*edf*) in the nonlinear modeling literature. Note that, generally in statistics, the degree of freedom refers to the number of free parameters to be estimated. While a smoothing spline nominally has T parameters, they are heavily constrained towards zero. Thus, the effective degrees of freedom essentially measures the flexibility of the smoothing spline, the greater its value, the more flexible the spline would be. In our empirical applications, we consider two values for the effective degrees of freedom, namely, $edf = 2$ and $edf = 3$, taking into account the trade-off between flexibility and complexity. Their forecasts in the context of forecast combination are labeled SS2 and SS3, respectively.

The main advantages of smoothing spline are: it allows us to fit a flexible curve to data relating the equity premium with the available predictors without the need to specify a particular type of relationship such as STAR; the smoothness of the spline can be adjusted via the effective degrees of freedom to combat overfitting, further

extending its applicability in general settings; its usage in the framework of univariate predictive regressions renders its estimation results interpretable relative to alternative complex nonlinear methods. A primary limitation to the smoothing splines is that it may miss the possible interactions between predictors when used in multivariate predictive regressions. We refer interested readers to Green and Silverman (1994) for further details regarding smoothing splines and related nonparametric regressions.

2.4 Alternative Models and Methods

The efficient market hypothesis inspired random walk model, which simply assumes a constant expected equity premium, is frequently used in empirical finance as the benchmark model. Specifically,

$$y_{t+1} = \beta_0 + \varepsilon_t. \quad (4)$$

Despite being simple in structure, the random walk model outperforms most univariate regression models considered in Welch and Goyal (2008) when forecasting the equity premium out-of-sample. Eq.(4) is also called the historical average or the prevailing mean in the literature of forecasting stock returns.

In addition to the random walk benchmark, we also consider the following models and methods proposed in the literature with which we compare the smoothing spline forecasts. Campbell and Thompson (2008) show that the predictive content of many financial and economic variables can be restored once various economic theory implied restrictions are imposed on linear predictive models, such as the positive forecast restriction. Rapach et al. (2010) argue that using the equally weighted combination to average forecasts from univariate linear models can address the issue of parameter instability inherent in the predictive relationship, leading to superior forecasting gains. In addition to equal weighting, Stock and Watson (2004) also consider the discounted mean squared forecast error (DMSFE) scheme and weights based on the shrinkage estimator to combine forecasts from diverse sources. Neely et al. (2014) argue that technical indicators, such as the moving average and momentum, contains significant predictive power for the market equity premium when combined across various parameter configurations. Given the elusive nature of stock returns predictability, Timmermann (2008) considers the previous best, adaptive combination, and the Bates-Granger least squares weights to combine baseline equity premium forecasts. Li and Tsiakas (2017) demonstrate that shrinkage estimators, such as the Lasso, can be used to estimate the kitchen-sink regression model comprising all variables when forecasting stock returns. Ferreira and Santa-Clara (2011) propose a novel SOP method to forecast the aggregate equity premium out-of-sample. For brevity, we refer interested readers to the articles cited above for details regarding these alternative methods.

2.5 Statistical Evaluation

Campbell and Thompson (2008) propose an out-of-sample R_{OOS}^2 statistic to estimate the average predictive gains against the benchmark over the entire evaluation sample. Specifically, the R_{OOS}^2 can be constructed according to the following:

$$R_{OOS}^2 = 100 \times \left(1 - \frac{\sum_{t=1}^{T-1} (y_{t+1} - \hat{y}_{t+1})^2}{\sum_{t=1}^{T-1} (y_{t+1} - \bar{y}_{t+1})^2} \right), \quad (5)$$

where \bar{y}_{t+1} and \hat{y}_{t+1} are one-step ahead point forecasts from the benchmark and the alternative model, respectively, and y_{t+1} represents the realized equity premium. Intuitively, the R_{OOS}^2 measures the percentage reduction in terms of the mean squared forecast error (MSFE) for the predictive model relative to the benchmark. The greater the R_{OOS}^2 value, the more the predictive gains would be. The simplicity and ease of interpretation of the out-of-sample R_{OOS}^2 explain its popularity among financial economists evaluating stock return forecasts.

Given that the R_{OOS}^2 is a point estimate of relative predictive accuracy, we assess its statistical significance according to the MSFE-adjusted t-statistic (MSFE-t) proposed in Clark and West (2007). The Clark and West (2007) test tests the null hypothesis of equal predictive accuracy between two competing models against the one-sided alternative that the benchmark model is inferior. In practice, the MSFE-t statistic can be conveniently constructed by first creating a new variable via the following equation:

$$f_{t+1} = (y_{t+1} - \bar{y}_{t+1})^2 - [(y_{t+1} - \hat{y}_{t+1})^2 - (\bar{y}_{t+1} - \hat{y}_{t+1})^2]. \quad (6)$$

Next, we regress f_{t+1} on a constant term then compute the t-statistic for the intercept. While MSFE-t statistic is not asymptotically normal, Clark and West (2007) show that the standard normal distribution provides a good approximation in simulations when the sample size is sufficiently large.

In addition to a measure summarizing the average performance over the full sample, we are interested in investigating how the equity premium forecasts perform during specific episodes within the out-of-sample window. To this end, we create a graphical device following Welch and Goyal (2008) to gain a dynamic perspective on how predictive models work. Specifically, we construct a new time series termed the cumulative

differences in squared forecast error between the benchmark and the predictive model (CDSFE) according to the following:

$$CDSFE_t = \sum_{s=1}^t (y_{s+1} - \bar{y}_{s+1})^2 - \sum_{s=1}^t (y_{s+1} - \hat{y}_{s+1})^2 \quad (7)$$

where \bar{y}_{s+1} and \hat{y}_{s+1} are forecasts from the benchmark and the predictive model, respectively.

At any time point t within the forecast evaluation window, if $CDSFE_t > 0$, it indicates that the predictive model under examination outperforms the benchmark. The time series plot of the CDSFE can be conveniently used to ascertain if the model has a smaller MSFE value than the benchmark for any episode by simply comparing the heights of the curve at the starting and end points of the segment corresponding to the period of evaluation. A model which exceeds the benchmark would have a CDSFE slope that is positive everywhere. The closer the CDSFE plot is to this ideal, the greater the predictive gains would be.

2.6 Economic Evaluation

It is reasonable to expect that a model with superior statistical performance in forecast evaluation would deliver material economic gains to investors who use its predictions to make optimal investment decisions. However, as discussed in Cenesizoglu and Timmermann (2012), statistical measures of forecasting gains, such as the R_{OOS}^2 of Campbell and Thompson (2008), may not necessarily lead to economic gains. This possible disparity between statistical and economic performances can be ascribed to the fact that large forecast errors are penalized more substantially by convex loss functions in statistical measures such as the MSFE relative to economic loss functions. Therefore, measures evaluating the economic value of forecasts complement the statistical gauges shown in the previous subsection.

Specifically, we consider the optimal portfolio decision for a mean-variance investor who allocates funds between equities and risk-free bills. At the end of each period t , the investor allocates an optimal share w_t of funds to equities for the subsequent period according to the following rule:

$$w_t = \frac{1}{\gamma} \frac{\hat{y}_{t+1}}{\hat{\sigma}_{t+1}^2}, \quad (8)$$

where γ is the coefficient of relative risk aversion (CRRA), \hat{y}_{t+1} is the one-step ahead point forecast of the equity premium, and $\hat{\sigma}_{t+1}^2$ is the estimated variance of the equity premium. Following Rapach et al. (2016), we estimate $\hat{\sigma}_{t+1}^2$ with a 10-year rolling window. Furthermore, we require that w_t fall into the interval $[-0.5, 1.5]$, which permits realistic short selling and leveraging activities as suggested in Rapach et al. (2016).

The investor who optimally allocates funds according to Eq.(8) then realizes an average certainty equivalent return (CER) of

$$CER = \bar{R}_p - \frac{\gamma}{2} \sigma_p^2, \quad (9)$$

where \bar{R}_p and σ_p^2 are the sample mean and variance of the optimal portfolio returns, respectively.

The CER can be understood as the risk-free return that a mean-variance investor with a CRRA value of γ would consider equivalent to investing following the risky strategy. Similarly, we compute the CER value for the investor if he or she uses benchmark forecasts to guide portfolio decision. We then calculate the CER gain (ΔCER) by taking difference between the two CER values. In our empirical results, we report the annualized CER gain in percentage, so it can be understood as the annual portfolio management fee in percentage that an investor would be willing to pay to access the regression model forecasts instead of the benchmark predictions.

In addition, we utilize the Sharpe ratio (SR) to assess the economic value of equity premium forecasts. The Sharpe ratio is the sample mean portfolio return in excess of the risk-free rate divided by the sample standard deviation of the portfolio returns. Both sample statistics are estimated over the full forecast evaluation sample. In keeping with the certainty equivalent return, we report annualized Sharpe ratio gains (ΔSR) in percentage comparing forecasts.

3. Empirical Results

In this section we first describe our data and monikers used to indicate predictive models. We then present results evaluating and comparing forecasting performance.

3.1 Data

We take updated monthly data on the aggregate U.S. equity premium along with a set of 14 predictive variables covering the period from January 1927 to December 2017 from Amit Goyal's website. The equity premium (e.ret)

is constructed from the S&P 500 index including dividends minus the 3-month Treasury bill rate. The set of financial and economic predictive variables comprises: the dividend-price ratio (dp); the dividend-yield (dy); earnings-price ratio (ep); dividend-payout ratio (de); the stock market variance (svar); book-to-market ratio (bm); net equity expansion (ntis); Treasury bill rate (tbl); long-term yield (lty); long-term return (ltr); term spread (tms); default yield spread (dfy); default return spread (dfr); inflation (infl). For brevity, we refer the interested readers to Welch and Goyal (2008) for details regarding the identity and construction of all variables.

We reserve the first 40 years of data as the initial training sample to estimate model parameters. All out-of-sample forecasts are made for January 1967 - December 2017 with the recursive estimation window, that is, at each forecast origin, one more recently available observation is included with historical data to update model parameter estimates before making predictions. Our empirical results remain qualitatively the same under the rolling estimation window, thus, for brevity, we do not report rolling window results here. We refer interested readers to Clark and McCracken (2013) for details regarding the out-of-sample framework.

3.2 Model Identities

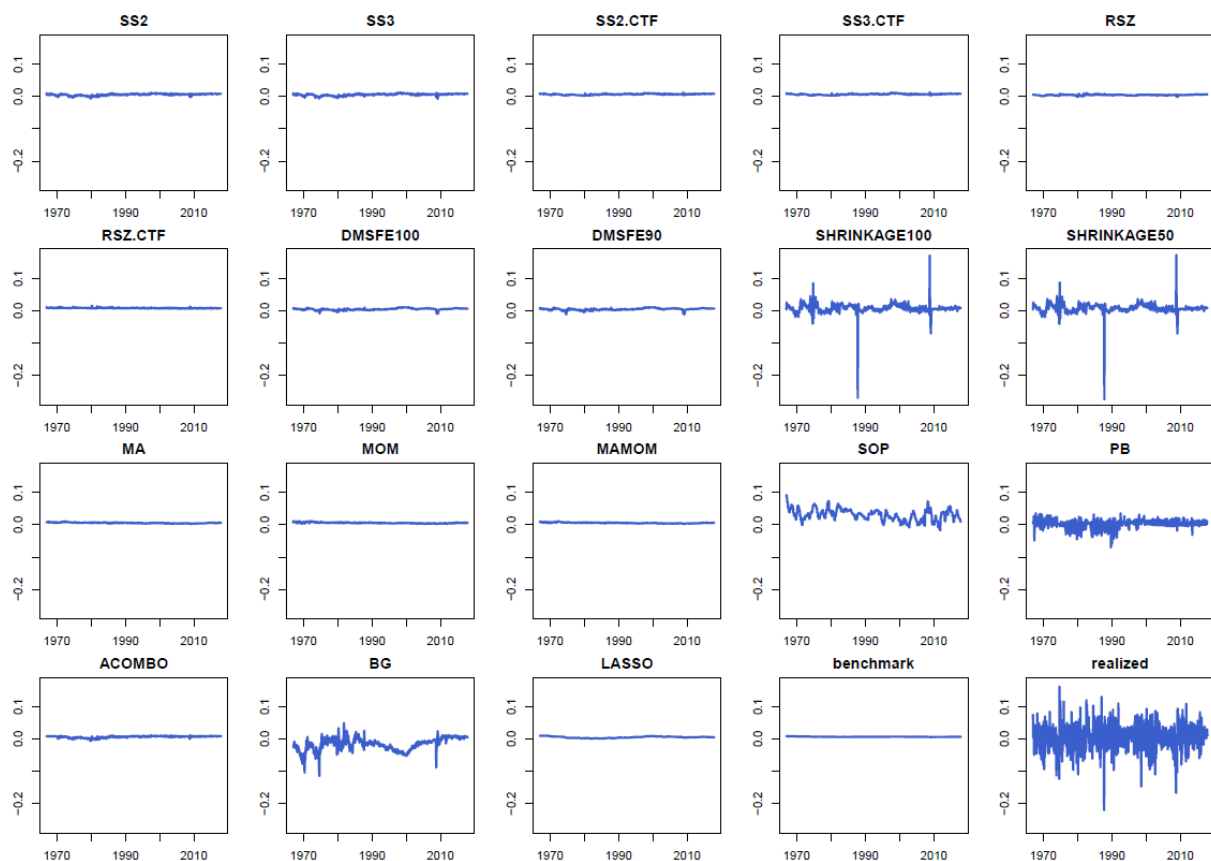


Figure 1. Out-of-Sample equity premium forecasts

Note. This figure presents the monthly equity premium forecasts for various predictive models, as well as the benchmark forecasts and the realized equity premium over 1967-2017.

The identities of all models and methods under consideration in this section are: SS2, the combined forecasts via equal weights from smoothing splines with two effective degrees of freedom; SS3, the combined forecasts via equal weights from smoothing splines with three effective degrees of freedom; SS2.CTF, the restricted combined forecasts via equal weights from smoothing splines with two effective degrees of freedom; SS3.CTF, the restricted combined forecasts via equal weights from smoothing splines with three effective degrees of freedom; RSZ, the combined forecasts via equal weights from baseline regression models estimated by OLS; RSZ.CTF, the restricted combined forecasts via equal weights from baseline regression models estimated by OLS; DMSFE100, the combined forecasts under weights assigned according to the past discounted mean squared forecast error with a discount factor of 100%; DMSFE90, the combined forecasts under weights assigned

according to the past discounted mean squared forecast error with a discount factor of 90%; SHRINKAGE100, the combined forecasts under weights assigned according to the shrinkage estimator of Stock and Watson (2004) with a shrinkage parameter of 100%; SHRINKAGE50, the combined forecasts under weights assigned according to the shrinkage estimator of Stock and Watson (2004) with a shrinkage parameter of 50%; MA, the combined forecasts from the moving average technical indicator according to configurations considered in Neely et al. (2014); MOM, the combined forecasts from the momentum technical indicator according to configurations considered in Neely et al. (2014); MAMOM, the combined forecasts from all moving average and momentum technical indicators according to configurations considered in Neely et al. (2014); SOP, forecasts according to the baseline SOP method proposed in Ferreira and Santa-Clara (2011); PB, the previous best forecasts according to Timmermann (2008); ACOMBO, the adaptive combination approach proposed in Timmermann (2008); BG, the Bates-Granger combination described in Elliott and Timmermann (2016); and LASSO, the lasso forecasts considered in Li and Tsiakas (2017).

3.3 Statistical Forecasting Performance

We begin by providing a matrix plot of forecasts over 1967-2017 for all models in Figure 1. The title of each panel in Figure 1 shows the name of the method or model employed to generate the equity premium predictions, with the exception of the two panels in the lower-right corner, which are reserved for forecasts from the random walk benchmark and the realized equity premium, respectively. Figure 1 shows that models such as SS2, SS3 and RSZ tend to generate stable and smooth forecasts similar to the benchmark while methods such as SOP and BG are prone to produce volatile predictions.

Table 1. Statistical performance

	R_{00s}^2	MSFE-t	p-value
SS2	2.776	4.533	0.000
SS3	3.062	4.194	0.000
SS2.CTF	1.359	4.095	0.000
SS3.CTF	1.393	4.026	0.000
RSZ	1.010	2.613	0.004
RSZ.CTF	0.532	1.640	0.050
DMSFE100	0.442	1.157	0.124
DMSFE90	0.462	1.124	0.131
SHRINKAGE100	-10.481	-0.834	0.798
SHRINKAGE50	-10.884	-0.833	0.798
MA	-0.217	-0.421	0.663
MOM	-0.249	-0.379	0.648
MAMOM	-0.214	-0.430	0.667
SOP	-5.114	-0.661	0.746
PB	-7.997	-0.385	0.650
ACOMBO	2.751	4.926	0.000
BG	-23.196	-3.678	0.999
LASSO	-0.634	-0.646	0.741

Note. This table reports the R_{00s}^2 values in percentage evaluating statistical performance. A positive R_{00s}^2 value indicates better performance than the random walk benchmark. Higher R_{00s}^2 value indicates better forecasting performance. The statistical significance of R_{00s}^2 is assessed via the MSFE-t statistic, with the associated p-values shown in the last column.

Regarding the average forecasting performance over the entire evaluation sample, Table 1 reports the R_{00s}^2 values in percentage assessing forecasts from various methods against those from the random walk benchmark. The first column in Table 1 shows the name of all predictive models. The second column reports the out-of-sample R_{00s}^2 of Campbell and Thompson (2008), which measures the percentage reduction in the mean squared forecast error relative to the benchmark, with positive values indicating better performance than the historical mean benchmark. The higher the R_{00s}^2 value, the more the predictive gains would be. The statistical significance of the R_{00s}^2 is assessed via the MSFE-t statistic of Clark and West (2007), whose values are shown in the third column with the associated p values reported in the fourth column.

We make several observations from Table 1. First, the combined forecasts from smoothing splines clearly dominate other methods in terms of statistical gains, with the SS3 model reporting the largest significant R_{00s}^2

value of 3.062%. Second, imposing the restrictions proposed in Campbell and Thompson (2008) on smoothing splines does not improve predictive performance. Third, the adaptive combination, ACOMBO, performs the best among the remaining models with a significant R_{00S}^2 value of 2.751%, confirming the results shown in Timmermann (2008). Fourth, the simple unrestricted forecast combination, RSZ, indeed restores the predictive power of many predictors examined in Welch and Goyal (2008) by reporting a significant but modest R_{00S}^2 value of 1.010%, confirming the primary message conveyed in Rapach et al. (2010). Fifth, the DMSFE combinations report modest but statistically insignificant predictive gains, explaining in part why they are not widely adopted in the forecasting stock returns literature unlike in macroeconomic forecasting. Finally, technical indicators, least squares combination and shrinkage estimators do not work well in this exercise, as they largely report negative predictive gains.

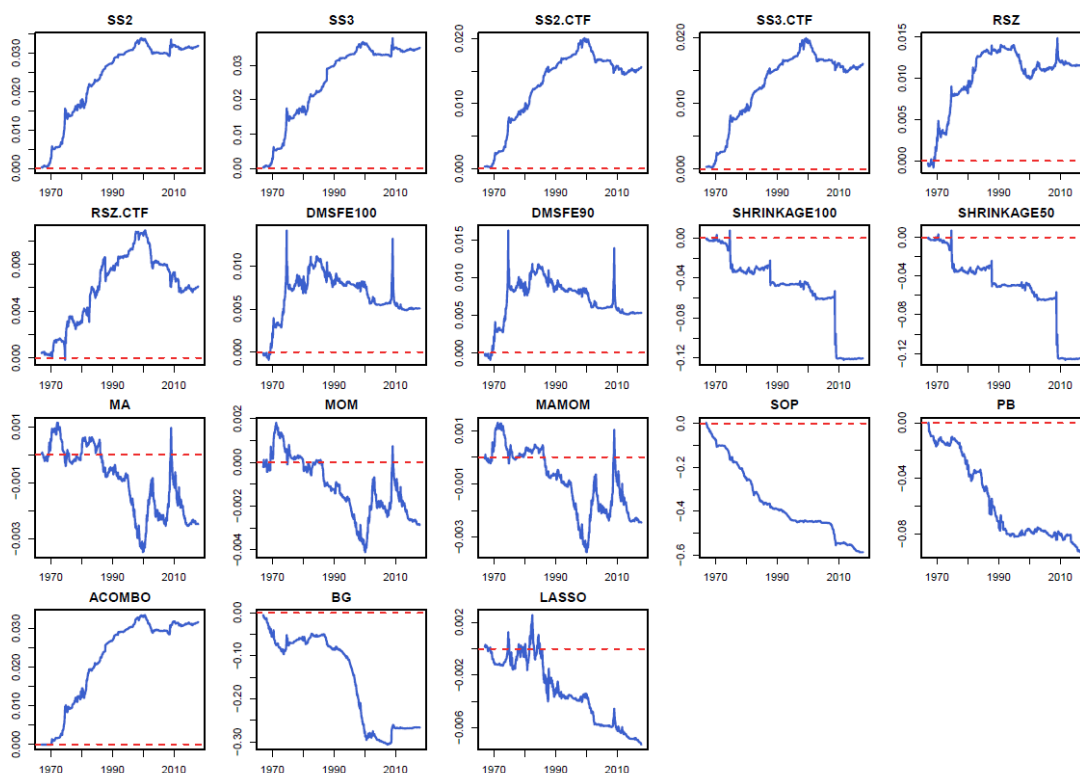


Figure 2. Cumulative differences of squared forecast errors

Note. This figure presents the time series plots of cumulative differences of squared forecast errors between the random walk benchmark and various predictive models over 1967-2017.

Despite being widely adopted for forecast evaluation, the R_{00S}^2 is merely a point estimate of the average relative forecasting performance over the full sample. To gain a dynamic perspective on how each model fares at a particular time window during the evaluation sample, following the empirical device proposed in Welch and Goyal (2008), we plot the time series of the cumulative differences of the squared forecast error between the benchmark and alternative predictive models (CDSFE) in Figure 2. A positive slope of the CDSFE curve indicates predictive gains for the model under examination against the random walk benchmark, while a negative slope suggests otherwise. Models such as SS2, SS3 and ACOMBO have CDSFE curves being positively sloped almost everywhere throughout the out-of-sample, implying robust and consistent gains against the benchmark. In contrast, the simple forecast combination, RSZ, has an upward-sloping CDSFE curve until the late 1990s, then it remains largely flat thereafter, suggesting that most predictive gains of the RSZ model arise from the early half of the out-of-sample.

In sum, we see that the forecast combinations from smoothing splines can uncover the predictive content embedded in many financial and economic variables for the market equity premium, further improving upon methods such as adaptive and simple combinations which have been shown effective in the related literature.

3.4 Regime-Dependent Evaluation

Given the elusive nature and unstable performance of many predictors illustrated in Timmermann (2008), in this subsection, we are interested in investigating how our models perform under different market conditions, highlighting the importance of regime-dependent evaluation for the equity risk premium advocated in Baltas and Karyampas (2018). Specifically, we consider six regimes: economic expansions and recessions defined by the National Bureau of Economic Research (NBER); bullish and bearish market sentiments according to the signs of the realized equity premium; high- and low-volatility regimes separated by above or below the sample average of the stock market variance. Table 2 reports regime-dependent forecast evaluation results. In Table 2, the first column shows the names of all predictive models, while the second through the seventh columns report the R_{OOS}^2 value in percentage across models under various market regimes. Asterisks, ***, ** and *, indicate statistical significance of the R_{OOS}^2 at nominal levels of 1%, 5% and 10%, respectively.

Table 2. Regime-dependent evaluation

	Expansion	Recession	Bull	Bear	High	Low
SS2	2.017***	4.592***	-0.924	5.538***	2.055**	3.693***
SS3	2.457***	4.512***	-1.618	6.555***	2.602**	3.648***
SS2.CTF	1.186***	1.774***	2.718***	0.345	0.260	2.757***
SS3.CTF	1.252***	1.729***	2.811***	0.333	0.310	2.769***
RSZ	0.176	3.008***	-5.878	6.152***	1.774***	0.039
RSZ.CTF	0.410*	0.825	6.930***	-4.243	-0.487	1.829***
DMSFE100	-0.350	2.339	-9.285	7.703***	0.804	-0.019
DMSFE90	-0.420	2.575	-9.701	8.048***	0.877	-0.066
SHRINKAGE100	-5.726	-21.872	-4.240	-15.139	-16.327	-3.042
SHRINKAGE50	-6.602	-22.436	-4.302	-15.797	-16.871	-3.266
MA	-0.676	0.884**	-3.422	2.176***	0.514*	-1.146
MOM	-0.740	0.927**	-3.644	2.285***	0.704**	-1.462
MAMOM	-0.682	0.907**	-3.480	2.225***	0.587**	-1.233
SOP	-4.740	-6.001	3.002***	-11.167	-3.221	-7.516
PB	-13.760	5.810**	-26.489	5.806***	-0.051	-18.108
ACOMBO	2.063***	4.400***	0.909**	4.126***	2.066***	3.622***
BG	-4.067	1.868***	-12.497	5.277***	2.035***	-6.039
LASSO	-0.963	0.154	-6.723	3.911***	-0.043	-1.385

Note. This table reports the R_{OOS}^2 values in percentage evaluating statistical performance under different market regimes. A positive R_{OOS}^2 value indicates better performance than the random walk benchmark. Higher R_{OOS}^2 value indicates better forecasting performance. The superscripts ***, ** and * denote statistical significance at levels of 1%, 5% and 10%, respectively.

A thorough examination of Table 2 reveals several interesting patterns. First, the unrestricted smoothing splines forecast particularly well during down markets, such as recessions. This observation closely aligns with those made in studies such as Rapach et al. (2010) and Rapach et al. (2016) that the predictability of stock returns are more evident during recessions. Second, smoothing splines deliver more predictive gains when the market is in a low-volatility regime, so does the adaptive combination of Timmermann (2008). Other models, except for the restricted simple combination, do not work well in low-volatility regimes. Finally, the technical indicators appear to work well only in down markets. For example, all technical indicator models report significant gains against the random walk benchmark in recessions and bearish markets. However, they do not improve upon the benchmark in expansions or when the market sentiment is bullish.

To summarize, our regime-dependent evaluation broadly supports the conclusion made in the previous subsection: the smoothing splines used in conjunction with forecast combination consistently lead many predictive models which have been shown effective in empirical finance when forecasting the equity premium out-of-sample.

3.4 Economic Value of Forecasts

In the equity premium prediction literature, the statistical performance of predictive models may not be closely aligned with the economic value delivered to investors who use them to guide portfolio investment decisions. This possible disparity between two measures can be attributed to the fact that the economic value of forecasts

are typically evaluated by a loss function drastically different from the quadratic loss often used in statistical evaluation. Against this backdrop, in this subsection we are interested in investigating the economic value of the smoothing spline forecasts delivered to investors who use them to make optimal portfolio decisions.

Table 3. Economic value of equity premium forecasts

	No Transaction Cost		50 bps Transaction Cost	
	ΔCER	ΔSR	ΔCER	ΔSR
SS2	2.808	1.102	1.103	0.363
SS3	3.228	1.285	1.334	0.452
SS2.CTF	1.437	0.485	0.152	0.040
SS3.CTF	1.536	0.515	0.190	0.053
RSZ	0.395	0.404	0.229	-0.002
RSZ.CTF	0.761	0.194	-0.442	-0.089
DMSFE100	0.006	0.166	-0.097	-0.047
DMSFE90	0.077	0.194	0.031	-0.032
SHRINKAGE100	2.052	0.462	0.519	0.229
SHRINKAGE50	2.060	0.463	0.519	0.230
MA	-0.580	-0.111	-0.469	-0.270
MOM	-0.513	-0.075	-0.449	-0.270
MAMOM	-0.549	-0.094	-0.458	-0.270
SOP	-0.463	0.035	-2.704	-0.159
PB	-0.422	0.017	-2.968	-0.183
ACOMBO	2.938	1.079	1.070	0.347
BG	-0.942	-0.296	-1.653	-0.548
LASSO	-1.088	-0.309	-0.579	-0.337

Note. This table reports the annualized CER and SR gains in percentage evaluating economic performance of the equity premium forecasts from various models. The second and third columns show results without taking into transaction costs, while the last two columns report results assuming a 50 bps transaction cost when re-balancing portfolios in each period.

Following studies such as Rapach et al. (2016), we measure economic value via the annualized certainty equivalent return (CER) and Sharp ratio (SR) gains in percentage over the prevailing mean benchmark, with a coefficient of relative risk-aversion (CRRA) value of three. Table 3 reports results assessing economic performance. In Table 3, the first column shows the names of all models under consideration. The second and third columns report annualized CER and SR gains in percentage without taking into account transaction costs when re-balancing the optimal portfolio. The last two columns report annualized CER and SR gains in percentage assuming a transaction cost of 50 bps following the suggestion in Rapach et al. (2016). Overall, the patterns revealed in Table 3 largely support our conclusion drawn from statistical evaluation that the smoothing spline forecasts outperform competing methods in consistently delivering economic gains to investors. To illustrate, without considering transaction costs, the SS3 model delivers a CER gain of 3.228%, indicating that the investor is willing to pay 3.228% more in annual portfolio management fees to gain access to the SS3 forecasts relative to the random walk benchmark predictions.

In addition to evaluating economic performance via the CER and SR gains, to compare economic value from a dynamic perspective, we plot in Figure 3 the log cumulative wealth for a number of portfolios named by the predictive models used when constructing the equity premium forecasts. Without loss of generality, we assume that the investor starts with \$1 and reinvests all proceeds from January 1967 to December 2017. To facilitate comparison and highlight results, we exclusively use solid lines to mark the two unrestricted smoothing spline portfolios, while the rest is denoted in various colored-dashed lines. Figure 3 clearly demonstrates that the smoothing spline forecasts can deliver sizable economic gains to investors who use them to make optimal portfolio decisions, as the two smoothing spline portfolios discernibly lead the rest in generating cumulative wealth.

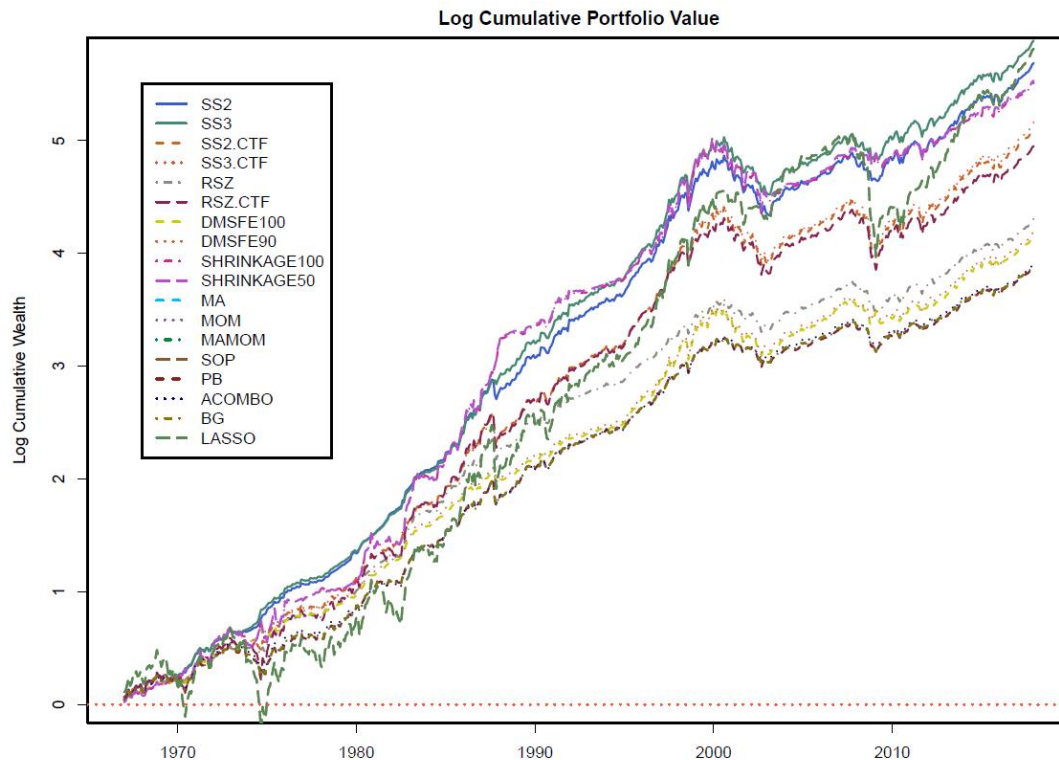


Figure 3. Log cumulative portfolio wealth growth

Note. This figure delineates the log cumulative wealth growth for a mean-variance investor with relative risk coefficient of three, assuming that he or she starts with \$1 and reinvests all proceeds over 1967-2017.

3.5 Discussion

With a comprehensive dataset, Welch and Goyal (2008) show that many aggregate financial and economic predictors fail to beat the simple random walk benchmark in terms of generating superior equity premium forecasts on a consistent basis. However, the primary conclusion drawn in Welch and Goyal (2008) is based on simple bivariate linear regressions estimated via the ordinary least squares, with various econometric issues associated with data features being overlooked in the estimation process. For example, by using a novel structural break test via nonparametric regression, Chen and Hong (2012) show that most predictive regressions considered in Welch and Goyal (2008) are subject to some form of structural break. As a result, the estimation methodology of ordinary least squares may not prove effective addressing the issue of parameter instability. Against this backdrop, in this paper we demonstrate that the nonlinear method of smoothing splines proves capable of taking into account parameter instability while maintaining simplicity in model structure and interpretation. Our empirical results show that smoothing splines can uncover the valuable predictive content for the equity premium embedded in many financial and economic variables, leading to superior predictive gains relative to the random walk benchmark and the simple linear forecast combination.

In addition to investigating the econometric issues related to the weak forecasting performance documented in Welch and Goyal (2008), recent developments in the literature of forecasting stock returns include searching for new predictive variables which may possess genuine predictive content for the aggregate excess returns. To illustrate, Rapach et al. (2016) argue that the short interest rate is a powerful predictor for the equity premium. Jiang et al. (2019) build a new composite index measuring market sentiment, the manager sentiment index, and show that it contains genuine predictive content for the market equity premium beyond those embedded in typical sentiment indexes in behavior finance. Ma et al. (2019) propose a new predictor, MADP, a moving-average momentum strategy based on daily prices, and show that it outperforms the historical mean benchmark as well as various standard moving-average momentum strategies based on monthly prices.

4. Conclusion

While accurate forecasts of the market equity premium are vitally important in empirical finance, the predictability of the equity premium is subject to contentious debate in the academic literature. Welch and Goyal

(2008) show that many financial and economic variables with previously documented in-sample evidence of predictability fail to forecast the equity premium out-of-sample. As a result, in the last decade various new predictors and models are proposed to support the out-of-sample predictability of stock returns, challenging the primary conclusion drawn in Welch and Goyal (2008). Despite the abundance of new variables and models, an important econometric issue has been overlooked in this strand of literature: does nonlinearity matter in the predictive model?

In this paper we propose using the smoothing splines to estimate the baseline univariate predictive models originally considered in Welch and Goyal (2008), then averaging forecasts from these models to form a combined forecast for the equity premium. We show that the smoothing splines can overcome the three major difficulties outlined in White (2006) when forecasting with nonlinear methods: computational cost, overfitting and ease of interpretation. The smoothing splines are flexible enough to capture the possible nonlinear relationship between the equity premium and predictors, while controlling for overfitting. Moreover, they can be used in conjunction with forecast combinations to better manage the model selection risk. In our empirical exercises forecasting the U.S. market equity premium, we show that the combined smoothing spline forecasts outperform many models, such as simple and adaptive combinations, shrinkage methods and technical indicators, in delivering statistical and economic gains on a consistent basis.

Acknowledgments

The author is grateful to the editor, the anonymous referee, and all seminar participants at Texas A&M International University for helpful comments and suggestions. All errors are mine.

References

- Baltas, N., & Karyampas, D. (2018). Forecasting the equity risk premium: The importance of regime-dependent evaluation. *The Journal of Financial Markets*, 38, 83-102. <https://doi.org/10.1016/j.finmar.2017.11.002>
- Campbell, J. Y. (1987). Stock returns and the term structure. *Journal of Financial Economics*, 18, 373-399. [https://doi.org/10.1016/0304-405X\(87\)90045-6](https://doi.org/10.1016/0304-405X(87)90045-6)
- Campbell, J. Y., & Thompson, S. B. (2008). Predicting excess stock returns out of sample: can anything beat the historical average? *Review of Financial Studies*, 21(04), 1509-1531. <https://doi.org/10.1093/rfs/hhm055>
- Cenesizoglu, T., & Timmermann, A. (2012). Do return prediction models add economic value? *Journal of Banking & Finance*, 36(11), 2974-2987. <https://doi.org/10.1016/j.jbankfin.2012.06.008>
- Chen, B., & Hong, Y. (2012). Testing for smooth structural changes in time series models via nonparametric regression. *Econometrica*, 80, 1157-1183. <https://doi.org/10.3982/ECTA7990>
- Clark, T. E., & McCracken, M. W. (2013). Advances in forecast evaluation. *Handbook of Economic Forecasting*, 2, 1107-1201. <https://doi.org/10.1016/B978-0-444-62731-5.00020-8>
- Clark, T. E., & West, K. D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, 138, 291-311. <https://doi.org/10.1016/j.jeconom.2006.05.023>
- Elliott, G., & Timmermann, A. (2016). *Economic Forecasting*. Princeton, NJ: Princeton University Press.
- Fama, E. F., & French, K. R. (1988). Dividend yields and expected stock returns. *Journal of Financial Economics*, 22, 3-14. [https://doi.org/10.1016/0304-405X\(88\)90020-7](https://doi.org/10.1016/0304-405X(88)90020-7)
- Ferreira, M. A., & Santa-Clara, P. (2011). Forecasting stock market returns: the sum of the parts is more than the whole. *Journal of Financial Economics*, 100, 514-537. <https://doi.org/10.1016/j.jfineco.2011.02.003>
- Green, P., & Silverman, B. (1994). *Nonparametric Regression and Generalized Linear Models: A roughness penalty approach*. London, UK: Chapman and Hall.
- Welch, I., & Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, 21(04), 1455-1508. <https://doi.org/10.1093/rfs/hhm014>
- Jiang, F., Lee, J., Martin, X., & Zhou, G. (2019). Manager sentiment and stock returns. *Journal of Financial Economics*, 132, 126-149. <https://doi.org/10.1016/j.jfineco.2018.10.001>
- Li, Y., Ng, D., & Swaminathan, B. (2013). Predicting market returns using aggregate implied cost of capital. *Journal of Financial Economics*, 110, 419-436. <https://doi.org/10.1016/j.jfineco.2013.06.006>
- Li, J., & Tsiakas, I. (2017). Equity premium prediction: the role of economic and statistical constraints. *Journal of Financial Markets*, 36, 56-75. <https://doi.org/10.1016/j.finmar.2016.09.001>
- Ma, C., Wen, D., Wang, G., & Jiang, Y. (2019). Further mining the predictability of moving averages: evidence

- from the US stock market. *International Review of Finance*, 19, 413-433. <https://doi.org/10.1111/irfi.12166>
- Neely, C. J., Rapach, D. E., Tu, J., & Zhou, G. (2019). Forecasting the equity risk premium: the role of technical indicators. *Management Science*, 60, 1772-1791. <https://doi.org/10.1287/mnsc.2013.1838>
- Paye, B., & Timmermann, A. (2006). Instability of return prediction models. *Journal of Empirical Finance*, 13(03), 274-315. <https://doi.org/10.1016/j.jempfin.2005.11.001>
- Pettenuzzo, D., Timmermann, A., & Rossen, V. (2014). Forecasting stock returns under economic constraints. *Journal of Financial Economics*, 114, 517-553. <https://doi.org/10.1016/j.jfineco.2014.07.015>
- Rapach, D. E., Strauss, J. K., & Zhou, G. (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *Review of Financial Studies*, 23, 821-862. <https://doi.org/10.1093/rfs/hhp063>
- Rapach, D. E., & Wohar, M. E. (2006). Structural breaks and predictive regression models of aggregate us stock returns. *Journal of Financial Econometrics*, 4, 238-274. <https://doi.org/10.1093/jjfinec/nbj008>
- Rapach, D. E., & Zhou, G. (2013). Forecasting stock returns. *Handbook of Economic Forecasting*, 2, 328-383. <https://doi.org/10.1016/B978-0-444-53683-9.00006-2>
- Rapach, D. E., Ringgenberg, M. C., & Zhou, G. (2016). Short interest and aggregate stock returns. *Journal of Financial Economics*, 121, 46-65. <https://doi.org/10.1016/j.jfineco.2016.03.004>
- Stock, J. H., & Watson, M. W. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, 23, 405-430. <https://doi.org/10.1002/for.928>
- Timmermann, A. (2008). Elusive return predictability. *International Journal of Forecasting*, 24, 1-18. <https://doi.org/10.1016/j.ijforecast.2007.07.008>
- White, H. (2006). Approximate nonlinear forecasting methods. *Handbook of Economic Forecasting*, 1, 459-512. [https://doi.org/10.1016/S1574-0706\(05\)01009-8](https://doi.org/10.1016/S1574-0706(05)01009-8)

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).