Center for the Study of Western Hemispheric Trade

Texas A&M International University

Working Paper Series

WP 2021-006

July 2021

Robust Forecast Combination for Elusive Return Predictability

Anwen Yin

Texas A&M International University

William Procasky

Texas A&M University - Kingsville

Robust Forecast Combination for Elusive Return Predictability

ANWEN YIN
Division of International Banking & Finance Studies
A. R. Sanchez, Jr. School of Business
Texas A&M International University
Laredo, Texas, USA

WILLIAM PROCASKY
Department of Accounting & Finance
College of Business
Texas A&M University-Kingsville
Kingsville, Texas, USA

We propose using the methodology of robust forecast combination to predict the equity premium out-of-sample in the presence of model instability. When averaging across models, the robust combinations alleviate the impact of over-penalizing an otherwise outperforming model for the occurrence of outliers owing to model instability, thus providing a theoretical foundation for the benefits of combining forecasts in unstable environments. Our empirical results on forecasting the US aggregate equity premium demonstrate the superior performance of the robust forecast combination relative to not only competing averaging methods such as equal weighting and time-varying adaptive weighting, but also information-pooling methods such as principal components and elastic-net.

KEYWORDS Equity Premium; Forecast Combination; Model Instability; Outlier

JEL CLASSIFICATION C53, C58, G11, G17

* Address correspondence to Anwen Yin, A. R. Sanchez, Jr. School of Business, Texas A&M International University, 5201 University Boulevard, WHT 217B, Laredo, Texas 78041. Email: anwen.yin@tamiu.edu.

I. Introduction

Forecasting stock returns plays an important role in empirical finance as the equity predictions are often vital inputs into portfolio management and investment decisions. However, the predictability of the aggregate equity premium has been subject to contentious debate. Historically, a multitude of financial and macroeconomic variables such as the dividend-price ratio and various measures of interest rates have been proposed in the academic literature to forecast stock returns. Extensive evidence of in-sample predictability for a variety of variables is provided in studies such as Campbell and Shiller (1988). However, Goyal and Welch (2008) show that many predictors with previously documented evidence of in-sample predictability fail to beat the simple historical average model when forecasting the aggregate equity premium out-of-sample. In light of the weak forecasting performance, Goyal and Welch (2008) argue that the market equity premium cannot be meaningfully predicted with exogenous variables on a consistent basis.

The view expressed in Goyal and Welch (2008) has been frequently challenged since its publication. Studies such as Rapach and Wohar (2006) and Rapach, Strauss, and Zhou (2010) show that the predictive content of many variables may be negatively affected by the presence of structural breaks or parameter instability. Furthermore, Timmermann (2008) extends the concept of parameter instability to model instability, which could arguably be the cause of the elusive return predictability. In light of their empirical findings, Timmermann (2008) and Rapach, Strauss, and Zhou (2010) show that the methodology of forecast combination can be used to uncover the genuine predictive content embedded in many predictive variables while accounting for the presence of parameter or model instability.

Our contribution to the literature consists of using the robust forecast combination method to forecast the aggregate equity premium out-of-sample. Rather than modeling the structural break process, we look for outliers in the historical forecast errors when constructing weights to combine models as they could be indicative of the occurrence of instability. Typically, the weights used in forecast combination are obtained via minimizing a quadratic risk function. As a result, an otherwise outperforming model may be over-penalized by the quadratic loss for the unusually large forecast errors it generates when assigning weights, leading to compromised performance of the combined forecast. To mitigate this issue, in our robust forecast combinations, we use the risk function based on either the absolute forecast error (L1 norm) or the Huber loss to construct weights. In contrast to the simple combination considered in Rapach, Strauss, and Zhou (2010) and the adaptive combination proposed in Timmermann (2008), our robust combination methods are characterized by a solid theoretical foundation supporting the linkage between combination weights and model instability. When forecasting the US aggregate equity premium out-of-sample, we show that the robust forecast combinations not only outperform competing weighting schemes such as the simple and adaptive combinations, but also beat information-pooling methods such as principal component and elastic-net regressions, in terms of both statistical and economic gains on a consistent basis.

The remainder of this paper is structured as follows: Section 2 describes the data and outlines the econometric methods used in subsequent analysis. Section 3 presents and discusses our main empirical results. Section 4 concludes.

II. Data and econometric methods

We begin by providing an overview of baseline linear predictive models for the market equity premium. Next, we discuss in detail the robust forecast combination when constructing out-of-sample forecasts. Finally, common statistical and economic measures evaluating forecasts are discussed. We focus on the one-step ahead point forecast of the market equity premium.

Data, baseline forecasts, and predictive model instability

We conduct empirical analysis using updated monthly data on the aggregate U.S equity premium along with a set of 14 predictive variables originally analyzed in Goyal and Welch (2008). Our dataset spans the period from January 1927 to December 2017.The equity premium (e.ret) is calculated from the S&P 500 index including dividends minus the 3-month Treasury bill rate. The set of predictors consists of: 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 Goyal and Welch (2008) for details regarding the identity and construction of these predictive variables.[1]

The baseline forecasts of the equity premium are obtained from the bivariate model considered in Goyal and Welch (2008):

$$y_{t+1} = \beta_0 + \beta_1 x_{j,t} + \varepsilon_t, \tag{1}$$

where y_{t+1} is the equity premium, x_{jt} is the predictor j at time t, and ε_t is the error term. This linear bivariate predictive model is simple to interpret and is often estimated via ordinary least squares (OLS).

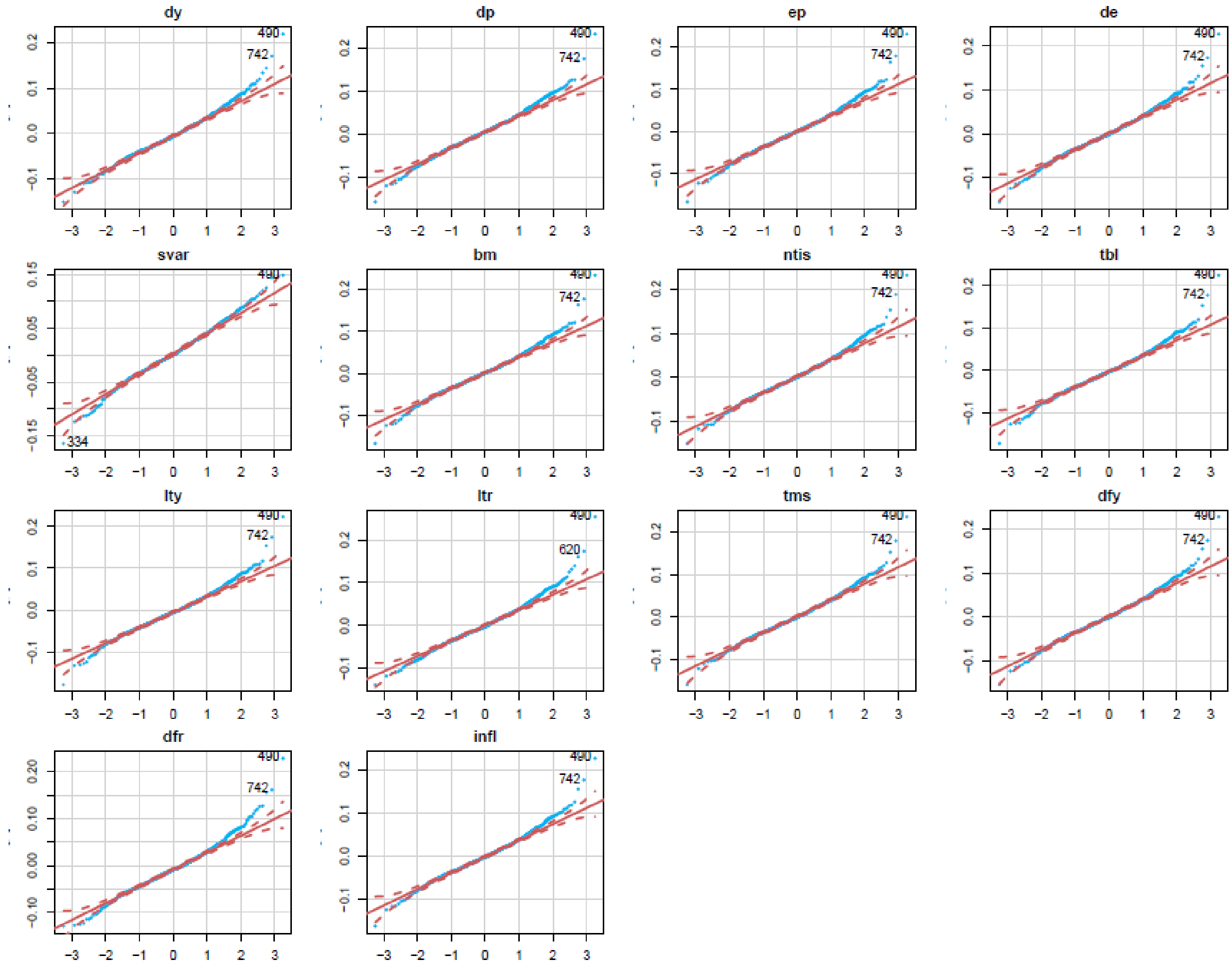
To shed light on the weak performance documented in Goyal and Welch (2008), in Figure 1 we present a panel of quantile-quantile (QQ) plots for all forecast errors from the bivariate models in Eq. (1) using the 14 predictors listed above against a normal distribution. Each baseline predictive model is named after the predictor x_j it contains in Eq. (1). Outliers are clearly visible for almost all forecasts in the upper-right corner of each plot, indicating the presence of model instability. Therefore, we expect that forecast combination methods which account for the impact of the outliers owing to model instability may deliver better predictive accuracy.

It is worth emphasizing that our study differs from many works in closely related literature in that we do not view instability as structural breaks in predictive model coefficients. Conventionally, researchers tend to interpret instability as unstable model coefficients. As a result, various tests have been utilized to detect such breaks in model parameters. However, the rejection of the null hypothesis of stability with any break test statistic does not inform us on the specific form of the breaking process. For example, the breaking process may take the form of large and rare discrete breaks, or it can take the form of small, frequent, and clustered breaks. Therefore, the advantage of treating the instability as forecast outliers instead of structural changes affords us an approach which is robust to the uncertainty on the nature of the breaking process. Next, we describe the weighting strategy in robust combinations which take into account forecast outliers.

---

[1] The dataset is maintained by Amit Goyal at http://www.hec.unil.ch/agoyal.

Figure 1. QQ Plot for forecasts from bivariate predictive models.



Econometric methods

Following the theoretical results in Wei and Yang (2012), we use the absolute forecast error loss or L1 loss, and the Huber loss to construct two separate sets of combination weights. Under the L1 loss, to construct a combined forecast for period *t+1*, the weight assigned to forecast *j* is:

$$w_{j,t+1}^{L1} = \frac{\prod_{s=1}^{t} \hat{d}_{j,s}^{-1} exp(-\lambda \sum_{s=1}^{t} |y_s - \hat{y}_{j,s}| / \hat{d}_{j,s})}{\sum_{j=1}^{M} \prod_{s=1}^{t} \hat{d}_{j,s}^{-1} exp(-\lambda \sum_{s=1}^{t} |y_s - \hat{y}_{j,s}| / \hat{d}_{j,s})}, \quad (2)$$

where y_s is the realized equity premium at time s, $\hat{y}_{j,s}$ is the forecast for the time s equity premium from model j, λ is a tuning parameter, M is the number of baseline forecasts, and $\hat{d}_{j,s}$ is the estimated mean absolute forecast error of model j at time s.[2]

In addition to the L1 loss, we also use the Huber loss to construct a combined forecast for period $t+1$. In this case, the weight assigned to forecast j is:

[2] In the empirical results section, we set all tuning parameters in the robust combination to the values suggested in the simulations reported in Wei and Yang (2012).

$$w_{j,t+1}^{Huber} = \frac{\prod_{s=1}^{t} \hat{v}_{j,s}^{-1/2} exp\left(-\lambda \sum_{s=1}^{t} \phi_k\left(\left(y_s - \hat{y}_{j,s}\right)/\sqrt{2\hat{v}_{j,s}}\right)\right)}{\sum_{j=1}^{M} \prod_{s=1}^{t} \hat{v}_{j,s}^{-1/2} exp\left(-\lambda \sum_{s=1}^{t} \phi_k\left(\left(y_s - \hat{y}_{j,s}\right)/\sqrt{2\hat{v}_{j,s}}\right)\right)}, \quad (3)$$

where y_s is the realized equity premium at time s, $\hat{y}_{j,s}$ is the forecast for the time s equity premium from model j, λ is a tuning parameter, M is the number of baseline forecasts, $\hat{v}_{j,s}$ is a variance estimate of y from model j as of time s, and $\phi_k(x)$ is the Huber loss function with:

$$\phi_k(x) = \begin{cases} x^2 & if -1 < x < k, \\ 2kx - k^2 & if\ x > k, \\ -2x - 1 & otherwise. \end{cases} \quad (4)$$

Note that the robust combination weights under the Huber loss take into account the impact of asymmetry compared with the weights constructed under the L1 loss. Put differently, the Huber loss function is an asymmetric loss function while the L1 loss is symmetric.

In addition to the two robust forecast combination schemes, to evaluate forecasts and compare performance in the empirical results section of this paper, we also consider the following alternative combination schemes: AFTER, weights according to the standard AFTER algorithm in Yang (2004); RSZ, simple combination based on equal weighting in Rapach, Strauss, and Zhou (2010); GAM, equal weighting applied to generalized additive models of the original bivariate regressions in Rapach, Strauss, and Zhou (2010); RSZ+CT, simple combination based on equal weighting in Rapach, Strauss, and Zhou (2010) combined with the restrictions in Campbell and Thompson (2008); GAM+CT, equal weighting applied to generalized additive models combined with the restrictions in Campbell and Thompson (2008); AFC, adaptive combination in Timmermann (2008); GR, forecast combination based on estimates of historical precision; and PBest, previous best forecast.

In recent years, dimension-reduction methods in the field of economic forecasting have been receiving growing attention, especially for situations where the forecaster has access to a large number of predictive variables without clear guidance on variable selection. Methods such as lasso and ridge regressions automatically perform variable selection in the model estimation stage, with influential variables receiving greater weights while coefficients for unimportant variables shrink towards zero. The success of shrinkage estimators in empirical finance has been shown in studies such as Li and Tsiakas (2017).

Against this backdrop, in addition to the various alternative combination schemes, we also consider the following dimension-reduction methods which pool information to generate out-of-sample forecasts: PCR, principal component regression; LASSO, lasso regression; RIDGE, ridge regression; ENET, the elastic-net considered in Li and Tsiakas (2017); Horseshoe, forecasts via the horseshoe estimator in Carvalho, Polson and Scott (2010); and GLASSO, the grouped lasso in Yuan and Lin (2006). For brevity, we refer interested readers to the articles cited above regarding the details of alternative models and methods.

Forecast evaluation

It is common practice in the literature of forecasting equity returns to compare the predictive accuracy of various models and methods with that of the random walk benchmark. The efficient market hypothesis inspired random walk model takes the following form:

$$y_{t+1} = \beta_0 + \varepsilon_t. \tag{5}$$

Intuitively, the random walk benchmark assumes that the expected value of the equity premium remains constant. Despite its simplicity, the random walk model proves difficult to beat in empirical studies related to forecasting stock returns and foreign exchange rates. For example, using a comprehensive dataset, Goyal and Welch (2008) show that most predictive models based on various aggregate economic and financial indicators fail to beat the random walk benchmark forecasting the equity premium in terms of statistical gains out-of-sample.

As a result, in the literature of forecasting the equity premium, researchers often use the out-of-sample R^2 statistic (OOS-R^2) proposed in Campbell and Thompson (2008), as a statistical measure for forecast evaluation. Intuitively, the OOS-R^2 measures the percentage reduction in mean squared forecast error for a predictive model under examination relative to that of the random walk benchmark. A positive value of the OOS-R^2 indicates better forecasting performance for the predictive model relative to the random walk, while a negative value suggests otherwise. The higher the OOS-R^2 value, the more predictive gains would be.

Since the OOS-R^2 is a point estimate of the relative predictive accuracy, we assess its statistical significance via the equal predictive accuracy test proposed in Diebold and Mariano (1995), which tests the null hypothesis that the predictive model under examination and the random walk benchmark forecast equally well against the one-sided alternative that the predictive model exceeds the benchmark. It is worth emphasizing that the Diebold and Mariano (1995) test is created to compare forecasts instead of predictive models; thus, it can be applied broadly in empirical works. On the contrary, some forecast evaluation test statistics widely used in empirical finance are created under strong assumptions which rule out instability in the underlying data generating process. Since we allow for instability in our data, applying the Diebold and Mariano (1995) test for forecast evaluation ensures the validity of our empirical results.

Despite its convenience and ease of interpretation, the OOS-R^2 merely tells us how predictive models preform on average over the entire forecast evaluation period. To see how models perform over the entire out-of-sample path from a dynamic perspective, following Goyal and Welch (2008), we construct a time series variable called the cumulative sum of the squared forecast errors between the random walk benchmark and the predictive model under examination (CDSFE), then plot it in a graph. Over any time window in the out-of-sample, if the CDSFE curve moves up, it indicates that the predictive model under evaluation outperforms the random walk benchmark. Therefore, a predictive model which dominates the random walk would have a CDSFE curve being positively sloped everywhere over the entire out-of-sample. The closer to this ideal, the more predictive gains would be.

In addition to statistical measures evaluating forecasts, we also examine the economic value delivered to investors who use the equity premium forecasts to guide portfolio decisions. Specifically, we use the annualized certainty equivalent return (CER) and the Sharp ratio (SR) gains to gauge the economic value of equity premium forecasts.

III. Empirical results

We use the monthly equity premium forecasts from the 14 bivariate models via Eq. (1) from

January 1947 to December 2017, and a ten-year rolling window to construct the averaged forecasts for the robust combinations. Hence, our first equity premium forecast from the robust combination is made for January 1957.[3]

Forecasting performance

Our main empirical results comparing forecasting performance are reported in Table 1. All entries in the table reflect the out-of-sample OOS-R^2 statistic proposed in Campbell and Thompson (2008), with the random walk chosen as the benchmark. A higher value of OOS-R^2 indicates better forecasting performance. The names of all predictive methods are shown in the first column, while the first row reports all forecast evaluation samples considered. In addition to the full evaluation sample of 1957 to 2017, we consider five subsamples with 12 years of data in each subsample except for the last one which contains 13 years of monthly observations. Statistical significance of the OOS-R^2 via the Diebold and Mariano (1995) statistic at levels of 1%, 5%, and 10% is denoted by ***, **, and *, respectively.

Several interesting observations can be made from an examination of Table 1. First, the L1 forecasts perform the best over the full evaluation sample and in two of the five subsamples. Second, the L1 forecasts beat the random walk benchmark regardless of subsamples. Third, while the Huber forecasts deliver the second best performance over the full sample, subsample evaluation suggests that their predictive gains are primarily obtained in the first half of the sample.

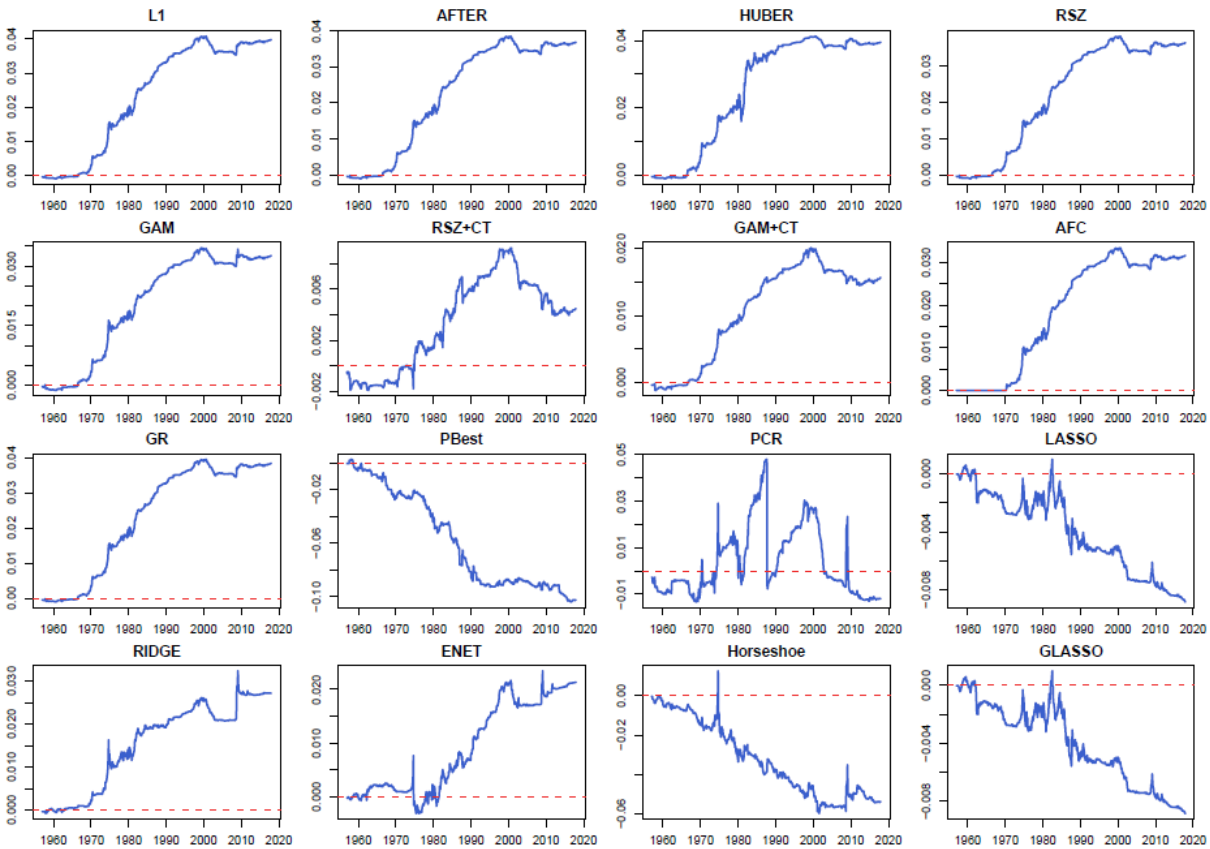
Table 1. Forecasting performance.

	1957-2017	**1957-1968**	**1969-1980**	**1981-1992**	**1993-2004**	**2005-2017**
L1	3.10***	0.48**	5.50***	5.79***	0.25*	1.38***
AFTER	2.87***	0.72**	5.14***	5.48***	0.20*	0.90**
HUBER	3.08***	1.13**	4.87***	7.03***	-0.03	0.35**
RSZ	2.84***	0.74**	5.09***	5.45***	0.19*	0.83**
GAM	2.54***	0.74**	4.92***	4.45***	0.21*	0.71**
RSZ+CT	0.35	-0.80	1.19*	1.50*	-0.16	-0.88
GAM+CT	1.22**	0.17**	2.88**	2.49***	-0.05	-0.48
AFC	2.46***	0.00	3.94**	5.45***	0.19*	0.83**
GR	2.99***	0.64**	5.40***	5.58***	0.21*	1.13***
PBest	-7.97	-14.21	-7.76	-13.92	0.89**	-5.87
PCR	-0.93	-7.85	2.63**	5.75***	-6.13	-3.80
LASSO	-0.69	-1.01	-0.38	-0.78	-0.80	-0.63
RIDGE	2.12**	0.48**	3.58**	3.13***	-0.24	1.16***
ENET	1.66*	0.25	-1.06	4.58***	0.77**	1.19***
Horseshoe	-4.21	-6.11	-6.07	-3.44	-5.38	0.63**
GLASSO	-0.69	-1.01	-0.38	-0.78	-0.80	-0.63

Notes: This table reports the values in percentage of the OOS-R^2 statistic. A positive OOS-R^2 value indicates better performance than the random walk benchmark. The superscripts ***, **, and * denote statistical significance at 1%, 5%, and 10%, respectively. All results are based on the forecast evaluation period of 1957 to 2017.

[3] Our empirical results remain qualitatively the same under the recursive window.

Figure 2. Cumulative differences in squared forecast error for monthly equity premium forecasts against the random walk.



Notes: The title of each plot indicates the name of the method used to generate forecasts. For any time window in each plot, a positively sloped curve indicates that the underlying method outperforms the benchmark, while the opposite holds when the curve falls.

Finally, while methods such as AFTER, RSZ, AFC, and ENET also report statistical gains over the benchmark, confirming the results documented in studies such as Rapach, Strauss, and Zhou (2010) and Li and Tsiakas (2017), their performances are dominated by the robust forecast combinations, particularly the L1 forecasts.

Moreover, in Figure 2, we plot the time series of the cumulative differences of the squared forecast error (CDSFE) for all 16 methods over the 1957 to 2017 sample. For any CDSFE plot, a positive slope indicates that the forecasts from a method under examination outperforms those from the random walk benchmark, while a negative slope suggests otherwise. Overall, all CDSFE plots in Figure 2 largely support the conclusions drawn from Table 1.

Forecasts and the business cycles

As a robustness check, we are interested in examining how robust combinations perform during economic expansions and recessions. Following the approach in Rapach, Strauss, and Zhou (2010), we report the OOS-R^2 values separately for economic expansions and recessions designated by the NBER over the full sample in Table 2. Our results shown in Table 2 broadly

Table 2. Forecasting performance and the business cycles.

	Economic Expansions	**Economic Recessions**
L1	2.32***	5.09***
AFTER	2.18***	4.63***
HUBER	1.44*	7.21***
RSZ	2.15***	4.57***
GAM	1.89**	4.18***
RSZ+CT	0.27	0.55
GAM+CT	1.10*	1.53**
AFC	1.82**	4.08***
GR	2.24***	4.89***
PBest	-13.39	5.71***
PCR	-5.53	1.07
LASSO	-1.04	0.20
RIDGE	1.01*	4.93***
ENET	1.92**	1.00
Horseshoe	-6.90	2.58**
GLASSO	-1.04	0.20

Notes: This table reports the values in percentage of the OOS-R^2 statistic across the business cycles of economic expansion and recession defined by the NBER. A positive OOS-R^2 value indicates better performance than the random walk benchmark. The superscripts ***, **, and * denote statistical significance at 1%, 5%, and 10%, respectively. All results are based on the forecast evaluation period of 1957 to 2017.

support the conclusion drawn in Rapach, Strauss, and Zhou (2010) and Guidolin, McMillan, and Wohar (2013) that the evidence of return predictability is stronger during recessions. Furthermore, Table 2 reveals that robust combination forecasts dominate other methods, with the L1 leading the rest during expansions while the Huber dominates during recessions.

Economic value of forecasts

In addition to statistical evaluations, predictive methods are often assessed according to the economic value delivered to investors who use their forecasts to guide optimal portfolio decisions. The reason why economic evaluation matters in the literature of forecasting stock returns lies in the fact that economic measures penalize forecast errors differently compared with statistical measures. For example, statistical measures such as the OOS-R^2 evaluate forecasts according to a quadratic loss function while economic measures often take into account nonlinearity and asymmetry. Therefore, a seemingly small statistical gain could be translated into sizable economic gain to investors. As a result, following closely related literature, we report the annualized certainty equivalent return (CER) and Sharp ratio (SR) gains in percentage over the random walk benchmark for all 16 methods in Table 3. Overall, Table 3 shows that our robust combination forecasts deliver the largest economic gains to investors among all methods considered.

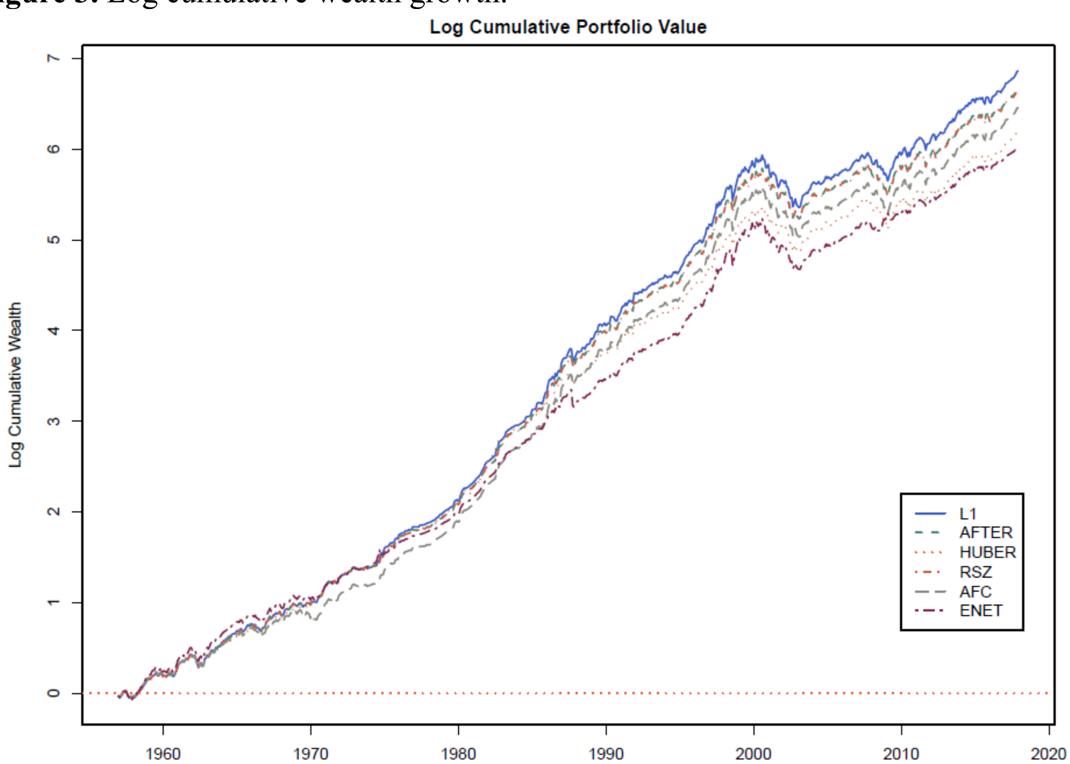
In Figure 3, we plot the log cumulative wealth for six portfolios named by the methods used when constructing forecasts. Without loss of generality, we assume that the investor starts with $1 and reinvests all proceeds over the period from 1957 to 2017. For ease of comparison, in Figure 3, the L1 portfolio is designated by a solid line while other portfolios are denoted in dashed lines in various colors. Figure 3 reveals that the superior predictive accuracy of the L1 forecasts can be

Table 3. Economic value of forecasts.

	ΔCER	ΔSR
L1	3.19	1.14
AFTER	2.91	1.07
HUBER	2.28	0.97
RSZ	2.88	1.06
GAM	2.54	0.98
RSZ+CT	0.75	0.14
GAM+CT	1.40	0.43
AFC	2.54	0.92
GR	3.03	1.11
PBest	0.00	-0.06
PCR	1.50	0.58
LASSO	-0.72	-0.23
RIDGE	1.96	0.80
ENET	2.27	0.81
Horseshoe	0.79	0.14
GLASSO	-0.72	-0.23

Notes: This table reports the annualized certainty equivalent return (CER) and Sharpe ratio (SR) gains in percentage for a mean-variance investor with a relative risk aversion coefficient of five who optimally allocates funds between equities and 3-month Treasury bills on the basis of equity premium forecasts. The portfolio weight on equities is constrained to fall in the interval of [-0.5, 1.5]. Results are based on the evaluation period of 1957 to 2017.

Figure 3. Log cumulative wealth growth.



Notes: This figure delineates the log cumulative wealth for a portfolio investor assuming that he or she starts with $1 and reinvests all proceeds from 1957 to 2017. Each portfolio is named after the method it uses to construct forecasts.

translated into sizable economic gains, as the L1 portfolio clearly leads the rest in generating cumulative wealth to the investor.

Discussion

Interestingly, in all methods, significantly less predictive power is documented during the 1993 to 2004 timeframe, irrespective of their calibration. We attribute this general reduction to the unusually powerful bull market that occurred during the time period and related effect on investor psychology. To illustrate, from a level of 295.46 on October 11th, 1990, the S&P 500 rose 417% to peak at 1,527.46 on March 24th, 2000, or 546% on a total return basis. Concurrent with these outsized gains, investors began paying less and less attention to the fundamental factors which make up our set of predictors. In fact, this effect became so pronounced that then Fed chairman Alan Greenspan described it as "irrational exuberance."

While investors did subsequently return to fundamental analysis after the bursting of the dot.com bubble, the historically and persistently low interest rate environment which ensued from 2008 onwards once again impacted investor attention to fundamental factors, although this time not as extreme. This more muted departure was driven by the need for investors to "chase yield" in order to realize adequate portfolio returns, an activity necessitating the move into higher returning asset classes like equities. As a result, while predictive power recovered during the 2005 to 2017 timeframe, it did not return to pre-1993 levels.

Finally and related to the above, we attribute the general reduction in predictive power over time to the comparatively lower frequency and shorter duration of recessions in recent history. Given this dynamic coupled with the models' documented better performance during periods of recession, the out-of-sample forecasting performance at the back end of our sample is not as strong as in the front end.

IV. Conclusion

In this paper, we show that the robust forecast combinations can further improve upon methods such as equal weighting, adaptive combination, and elastic-net when forecasting the equity premium out-of-sample. Since return predictability tends to be elusive as characterized in Timmermann (2008) due to model instability, using combination weights based on L1 or Huber loss might alleviate the concern of over-penalizing an otherwise outperforming model for a few outliers. Our empirical results show that robust combinations outperform many popular alternatives pooling forecasts or information in terms of both statistical and economic gains consistently.

References

Campbell, J. Y., and R. J. Shiller. 1988. "The dividend-price ratio and expectations of future dividends and discount factors." *The 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?" *The Review of Financial Studies* 21 (4):1509–1531.

Carvalho, C. M., N. G. Polson, and J. G. Scott. 2010. "The horseshoe estimator for sparse signals."

Biometrika 97 (02): 465–480.
Diebold, F. X., and R. S. Mariano. 1995. “Comparing predictive accuracy.” *Journal of Business and Economic Statistics* 13 (3): 253–263.
Goyal, A., and I. Welch. 2008. “A comprehensive look at the empirical performance of equity premium prediction.” *The Review of Financial Studies* 21 (4): 1455–1508.
Guidolin, M., D. G. McMillan, and M. E. Wohar. 2013. “Time varying stock return predictability: Evidence from us sectors.” *Finance Research Letters* 10 (1): 34–40.
Li, J., and I. Tsiakas. 2017. “Equity premium prediction: the role of economic and statistical constraints.” *Journal of Financial Markets* 36: 56–75.
Rapach, D., J. Strauss, and G. Zhou. 2010. “Out-of-sample equity premium prediction: Combination forecasts and links to the real economy.” *The Review of Financial Studies* 23 (2): 821–862.
Rapach, D. E., and M. E. Wohar. 2006. “Structural breaks and predictive regression models of aggregate US stock returns.” *Journal of Financial Econometrics* 4 (2): 238–274.
Timmermann, A. 2008. “Elusive return predictability.” *International Journal of Forecasting* 24 (1): 1–18.
Wei, X., and Y. Yang. 2012. “Robust forecast combinations.” *Journal of Econometrics* 166 (2): 224–236.
Yang, Y. 2004. “Combining forecasting procedures: Some theoretical results.” *Econometric Theory* 20 (1): 176–222.
Yuan, M., and Y. Lin. 2006. “Model selection and estimation in regression with grouped variables.” *Journal of the Royal Statistical Society: Series B* 68 (1): 49–67.