

# Equity Premium Prediction and the State of the Economy\*

Jiahan Li

University of Notre Dame

jiahan.li0@gmail.com

Ilias Tsiakas

University of Guelph

itsiakas@uoguelph.ca

January 2017

## Abstract

We detect cyclical variation in the predictive information of economic fundamentals, which can be used to substantially improve and simplify out-of-sample equity premium prediction. Economic fundamentals based on stock-specific information (notably the dividend yield) deliver better predictions in expansions. Economic fundamentals based on aggregate information (notably the short rate) deliver better predictions in recessions. A simple forecast combination of the dividend yield and the short rate generates statistically significant and economically valuable equity premium predictions in both expansions and recessions. This is an effective prediction strategy because it provides strong diversification gains. A strategy timing the business cycle performs better.

*Keywords:* Equity Premium; Out-of-Sample Prediction; Economic Fundamentals; Business Cycle; Diversification.

*JEL Classification:* G11; G14; G17.

---

\*The authors are grateful for useful comments to Angelo Melino and seminar participants at the University of Toronto and the 2016 RCEF Conference in Waterloo. *Corresponding author:* Ilias Tsiakas, Department of Economics and Finance, University of Guelph, Guelph, Ontario N1G 2W1, Canada. Tel: 519-824-4120 ext. 53054. Fax: 519-763-8497. Email: itsiakas@uoguelph.ca.

# 1 Introduction

Does the state of the economy matter for equity premium prediction? According to a recent study by Kacperczyk, Van Nieuwerburgh, and Veldkamp (2016), the state of the economy changes the way investors process information. In recessions, aggregate risk is significantly higher and consequently investors care more about aggregate shocks. In expansions, aggregate risk is lower and consequently investors care more about idiosyncratic shocks. In this framework, it is possible that the predictive information of economic fundamentals is linked to the observed state of the business cycle.

Consider, for example, the case where certain economic fundamentals generate equity premium forecasts that perform well in good times, whereas others generate forecasts that perform well in bad times. In the spirit of diversification, we could form a forecast combination of cyclical and countercyclical forecasts. Such a combination can be thought of as a portfolio of forecasts, which ideally achieves two desirable properties: diversification, which implies that the combined forecast performs better than any individual forecast alone; and insurance, which implies that the combined forecast performs well in every state of the world.

Motivated by these ideas, we propose a simple way to predict the equity premium out of sample: form an equally weighted combination of the forecasts generated by the dividend yield and the short rate. We will show that this simple approach has solid theoretical foundations, empirically it achieves both diversification and insurance, and performs much better than standard forecast combinations based on all economic fundamentals.

The dividend yield and the short rate are perhaps the two most prominent predictors in the literature. They are also theoretically motivated by the present value framework of Campbell and Shiller (1998). More importantly, we show that these two predictors generate forecasts that perform well in different states of the economy: the dividend yield performs well only in expansions, whereas the short rate only in recessions. Then, the equally weighted combination takes advantage of this cyclical pattern by exploiting the negative correlation between the two sets of forecasts, which leads to diversification gains and insurance. In short, the equally weighted forecast combination of the dividend yield and short rate turns out to be a simple, yet powerful out-of-sample predictor of the equity premium that is hard to improve upon.

This result is consistent with the Kacperczyk, Van Nieuwerburgh, and Veldkamp (2016) findings mentioned earlier. The dividend yield is a measure of idiosyncratic information because it is based on stock-specific dividends and prices. As a result, it is expected to be more informative in expansions. In contrast, the short rate is a measure of aggregate information because it is an economy-wide variable that affects the future returns and cash

flows of all firms. As a result, it is expected to be more informative in recessions.

Our empirical findings are based on a standard predictive framework. We employ monthly equity premium returns and the 14 monthly economic fundamentals of Welch and Goyal (2008) for a sample period ranging from January 1927 to December 2015. We estimate simple predictive regressions and impose the economic constraints of Campbell and Thompson (2008) on the sign of the regression coefficients and return forecasts. The constraints impose economic theory on the predictive regressions and almost universally improve performance. The analysis is performed purely out-of-sample in order to inform real-time investment decisions. Our evidence on out-of-sample equity premium predictability is an important finding because, from a conceptual point of view, return predictability is crucial for rationalizing the observed variation in stock prices (Cochrane, 2008, 2011). It is also important because, from a practical point of view, out-of-sample predictability allows investors and firms to make better real-time investment decisions.

In particular, we find that the equally weighted combination of the dividend yield and the short rate forecasts delivers a statistically significant monthly out-of-sample  $R^2$  of 1.3%. This is almost double the value of the  $R^2$  of the best individual model. It also outperforms a large set of alternative models examined in the literature. More importantly, the simple combination delivers good performance in both expansions ( $R^2 = 1.1\%$ ) and recessions ( $R^2 = 1.7\%$ ). The equally weighted combination also delivers high economic value in the context of a dynamic mean-variance strategy. The certainty equivalent return is about 1.7% per year over and above the historical mean benchmark for the full sample, 1% per year in expansions, and 5% per year in recessions.

The forecast combination we propose has two features that make it simple and attractive: equal weights and using only two predictors. Both of these features are based on solid foundations. The equal weights on the two predictors is a simple choice that works well. It is also an optimal choice since we compute the optimal weights that maximize the diversification gains and find that they are very close to equal weights. This is due to the fact that, on average, the mean squared errors of the two sets of forecasts are similar in value.

Using only two predictors is also a good choice because: (1) the forecasts generated by the two predictors are negatively correlated with each other; and (2) each forecast generated by a given predictor is highly positively correlated with the forecasts of all other predictors that perform well in the same state of the world. For example, the performance of the dividend yield is highly correlated with that of the earnings-price ratio. Therefore, adding the latter in a forecast combination adds very little to the performance of the combination. Similarly, the performance of the short rate is highly correlated to that of the long rate or the term spread. Again, adding any of the latter two predictors in a forecast combination does not

improve much the performance of the combination.

These results imply that a combination of the dividend yield and the short rate is not uniquely powerful. We could replace the dividend yield by the earnings-price ratio, or the short rate by the term spread, and the resulting two-predictor combination would remain similarly powerful. In this paper, we propose a combination based on the dividend yield and the short rate because, as mentioned earlier, these two are the most popular predictors with solid theoretical foundations. A combination of these two predictors is hard to beat. Having said that, we show that one can also combine say the earnings-price ratio with the term spread and achieve similar diversification and insurance results.

Our findings also provide an explanation for why standard forecast combinations of all 14 predictors typically work well. Since the contribution of Rapach, Strauss, and Zhou (2010), forecast combinations of all available predictors have become quite popular, partly because they are simple. By naively combining all predictors, the standard approach includes some predictors that work well in expansions, some that work well in recessions and some that work well in neither. We effectively show that by excluding from the combination these predictors that work well in neither state of the economy in addition to these predictors that do not add any further diversification gains, we can form simpler and yet superior combinations.

Finally, we consider a formal rule for ex-ante timing of the business cycle. The rule is based on the Markov-switching dynamic factor model of Chauvet (1998), which produces a probability that the economy is in expansion or recession every month. We establish turning point dates by converting the probability into a dummy variable for forecasting which state the economy is in at a given point in time. We find that, overall, the ex-ante timing strategy performs better than the simple forecast combination. However, all the performance gains in the ex-ante timing strategy are due to recessions, which is to be expected because the main feature of the strategy is that it forecasts recessions. In expansions, the simple combination of the dividend yield and the short rate is hard to beat.

This paper contributes to a long line of research on the out-of-sample predictability of the equity premium. We briefly discuss a short list of papers that are closely related to this study. Welch and Goyal (2008) evaluate the predictive ability of a number of individual predictors and most of the ensuing literature has been using the same data and empirical setup. Campbell and Thompson (2008) demonstrate the improvement in out-of-sample predictability by imposing economic constraints on the sign of regression coefficients and return forecasts. The role of the dividend yield in forecasting returns is emphasized by Cochrane (2008, 2011).

Similar to our approach, Rapach, Strauss, and Zhou (2010) estimate several predictive regressions each conditioning on one predictor, and then combine the individual forecasts

into one forecast combination. Whereas this approach is designed to pool forecasts, Li and Tsiakas (2016) provide a way of pooling information using a single predictive regression that conditions on a large number of predictors. This “kitchen sink” regression is estimated with shrinkage, which improves performance by reducing the effect of less informative predictors in out-of-sample forecasting.

Dangl and Halling (2012) allow for time variation in the regression coefficients and use Bayesian model averaging to combine forecasts. Pettenuzzo, Timmermann, and Valkanov (2014) also use a Bayesian framework to assess the role of economic constraints on the equity premium and the Sharpe ratio. Finally, Neely, Rapach, Tu and Zhou (2014) use principal components to extract the predictive information from a large set of technical indicators.

The remainder of the paper is organized as follows. In the next section, we examine the theoretical foundations of our analysis. The data on economic fundamentals are described in Section 3. In Section 4, we discuss the predictive framework and the empirical results. In Section 5, we present the dynamic mean-variance strategy for evaluating the portfolio performance of equity premium predictability. We evaluate the performance of a strategy timing the business cycle in Section 6. Finally, we conclude in Section 7.

## 2 Theoretical foundations

In this section, we discuss the theoretical foundations of our predictive approach by answering three questions. First, why do we focus on the dividend yield and the short rate as our main predictors of the equity premium? Second, why does the state of the business cycle matter? Finally, third, why do simple forecast combinations provide superior results in equity premium prediction?

### 2.1 Motivating the choice of predictors

We use the present value framework of Campbell and Shiller (1998) to motivate the choice of the dividend yield (or equivalently the dividend-price ratio) and the short rate as predictors of the future equity premium. We also show that similar arguments can be used to motivate replacing the dividend-price ratio by the earnings-price ratio and the short rate by the long rate.

Campbell and Shiller (1998) show that the following difference equation holds approximately:

$$dp_t = -k + \rho dp_{t+1} + r_{t+1} - \Delta d_{t+1}, \quad (1)$$

where  $dp_t$  is the log dividend-price ratio,  $k$  is a known constant,  $\rho$  is also a known constant

that typically has a value of slightly less than one,  $r_{t+1}$  is the one-period stock return including dividends, and  $\Delta d_{t+1}$  is the log dividend growth rate.

Iterating forward and taking the time  $t$  expectation of both sides, we arrive at the equation:

$$dp_t = -\frac{k}{1-\rho} + E_t \sum_{j=0}^{\infty} \rho^j (r_{t+j+1} - \Delta d_{t+j+1}). \quad (2)$$

Equation (2) states that the current log dividend-price ratio is directly related to the expected present value of all future returns minus all future dividend growth rates. Therefore, changes in stock prices are either due to news about future returns or news about future dividends. This equation motivates the use of dividend-price ratio (or the dividend yield) as a predictor of future returns.

Equation (2) can be re-written in terms of future excess returns as follows:

$$dp_t - r_{f,t} = -\frac{k}{1-\rho} + E_t \sum_{j=0}^{\infty} \rho^j (r_{t+j+1}^e - \Delta d_{t+j+1}) + E_t \sum_{j=1}^{\infty} \rho^j r_{f,t+j}, \quad (3)$$

where  $r_{t+1}^e = r_{t+1} - r_{f,t}$ , and  $r_{f,t}$  is the one-period risk-free rate. This equation implies that the current short rate is also a relevant predictor of future excess returns.

Equation (3) can have two further forms that are relevant for our analysis. First, we can replace the short rate by the long rate by further assuming the pure expectations hypothesis of the term structure:  $y_{n,t} = \frac{1}{n} \sum_{j=0}^{n-1} E_t (r_{f,t+j})$ , where  $y_{n,t}$  is the yield of an  $n$ -period bond. Then, equation (3) becomes:

$$dp_t - ny_{n,t} = -\frac{k}{1-\rho} + E_t \sum_{j=0}^{\infty} \rho^j (r_{t+j+1}^e - \Delta d_{t+j+1}) + E_t \sum_{j=n}^{\infty} \rho^j r_{f,t+j}. \quad (4)$$

Hence both the short and the long rate can be relevant predictors of future excess returns.

Second, we return to equation (1) and add the log earnings growth rate  $\Delta e_{t+1}$  to both sides. Solving forward, we can then show that the following equation holds for the log earnings-price ratio,  $de_t$ :

$$de_t - ny_{n,t} = -\frac{\kappa}{1-\rho} + E_t \sum_{j=0}^{\infty} \rho^j (r_{t+j+1}^e - \Delta e_{t+j+1} - (1-\rho) de_{t+j+1}) + E_t \sum_{j=n}^{\infty} \rho^j r_{f,t+j}. \quad (5)$$

This equation, therefore, shows that  $de_t$  contains similar information to  $dp_t$  for predicting future excess returns.

In summary, equation (2) shows that the current dividend-price ratio (or equivalently

the dividend yield) is related to future returns. Given that we focus on the equity premium (which is an excess return), equation (3) suggests that the current short rate is also related to future excess returns. Equations equation (4) and (5) further suggest that we can replace the dividend-price ratio by the earnings-price ratio or the short rate by the long rate.<sup>2</sup>

## 2.2 Why does the business cycle matter?

The state of the business cycle is an important variable in predicting stock returns. It is abundantly clear in the asset pricing literature that during recessions we observe three effects: (i) returns are unexpectedly low, (ii) returns are more volatile, and (iii) the price of risk is higher, partly driven by higher risk aversion (e.g., Fama and French, 1989). This is also true in our sample since the equity premium during recessions is on average negative ( $-6.4\%$  annually) and its volatility is high ( $29\%$  annually). One implication of these observations is that using the historical mean forecast (i.e., the benchmark) is a poor way of forecasting the equity premium in recessions. This is because the historical mean moves very slowly and tends to be much higher than realized returns during recessions.

More importantly, however, the state of the economy changes the way investors process information. In a recent paper, Kacperczyk, Van Nieuwerburgh, and Veldkamp (2016) propose a theoretical model that explains how the optimal attention allocation of investors depends on the business cycle. They also provide empirical evidence suggesting that recessions have different information content from expansions. The evidence shows that aggregate risk is significantly higher in recessions. As a result, skilled investors (e.g., mutual funds) care more about aggregate shocks in recessions. To be more specific, the covariance between portfolio weights and aggregate shocks rises in recessions. This contributes to better timing ability by skilled investors in recessions.

In contrast, the empirical evidence shows that idiosyncratic risk is essentially the same in expansions and recessions. In expansions, when aggregate risk is lower, skilled investors care more about idiosyncratic shocks. In other words, the covariance between portfolio weights and idiosyncratic shocks rises in expansions. This contributes to better picking ability by skilled investors in expansions.

The intuition behind these empirical findings is simple. Information is most valuable when risk and uncertainty is higher. Since aggregate risk is higher in recessions, skilled investors will allocate their attention towards aggregate risk. Learning more about aggregate risk is the most efficient way to resolve uncertainty and reduce portfolio risk. In expansions, when aggregate risk is low, they will allocate their attention to idiosyncratic risk. In short,

---

<sup>2</sup>The derivation of equations (1)-(5) is straightforward using the approach of Campbell and Shiller (1998). Some of the details for these derivations can be found in Maio (2013).

therefore, this theoretical model and the supporting empirical evidence provide a framework for linking information choices (aggregate vs. idiosyncratic) to the observed state of the business cycle.

The Kacperczyk, Van Nieuwerburgh, and Veldkamp (2016) framework is consistent with the basic premise of our paper: the dividend yield is more informative in expansions and the short rate is more informative in recessions. The dividend yield is a measure of idiosyncratic information because it is based on stock-specific dividends and prices. In contrast, the short rate is a measure of aggregate information because it is an economy-wide variable that affects the future returns and cash flows of all firms. According to the theory and evidence of cyclical variation of attention allocation, the short rate reflects aggregate risk that is more informative in recessions, whereas the dividend yield reflects idiosyncratic risk that is more informative in expansions.

### 2.3 Diversification gains of combined forecasts

It is well known that a portfolio investing in two negatively correlated assets exhibits large diversification gains: the volatility of the portfolio return is much lower than the average return volatility of the two assets. We will show that this diversification result has a direct analogue in forecasting. A simple forecast combination of two negatively correlated forecasts can be thought of as a portfolio of forecasts that can also deliver diversification gains: the mean squared error (MSE) of the combination is much lower than the average MSE of the individual forecasts.

To better understand this claim, consider the simple linear predictive regression:

$$r_{t+1}^e = \alpha + \beta x_t + \varepsilon_{t+1}, \quad (6)$$

where  $r_{t+1}^e = r_{t+1} - r_{f,t}$  is the equity premium at time  $t + 1$ ,  $r_{t+1}$  is the total return on the S&P 500 Index at time  $t + 1$ ,  $r_{f,t}$  is the Treasury bill rate,  $x_t$  is a predictor at time  $t$ ,  $\alpha$  and  $\beta$  are estimated with ordinary least squares (OLS), and  $\varepsilon_{t+1}$  is a normal error term.

We form forecasts  $\widehat{r}_{t+1}^e = \widehat{\alpha} + \widehat{\beta}x_t$ , where  $\widehat{\alpha}$  and  $\widehat{\beta}$  are the OLS estimates. The forecast error  $\widehat{\varepsilon}_{t+1} = r_{t+1}^e - \widehat{r}_{t+1}^e$  is unbiased with  $E[\varepsilon_t] = 0$  and  $Var[\varepsilon_t] = \sigma^2$ . The predictive performance of the model, captured by the MSE of the forecasts, is directly linked to the variance of the forecast error:

$$MSE = E \left[ (r_{t+1}^e - \widehat{r}_{t+1}^e)^2 \right] = E [\varepsilon_{t+1}^2] = \sigma^2. \quad (7)$$



### 2.3.1 The case of equal weights

Suppose that two models with the same structure as above generate two forecasts  $\widehat{r}_{1,t+1}$  and  $\widehat{r}_{2,t+1}$ . We form an equally-weighted combined forecast:  $\widehat{r}_{c,t+1} = \frac{1}{2}\widehat{r}_{1,t+1} + \frac{1}{2}\widehat{r}_{2,t+1}$ . Then, it is straightforward to show that the MSE of the combined forecast is given by:

$$MSE_c = \sigma_c^2 = \frac{1}{4}\sigma_1^2 + \frac{1}{4}\sigma_2^2 + \frac{1}{2}\sigma_1\sigma_2\rho_{12}, \quad (8)$$

where  $\rho_{12}$  is the correlation between the forecast errors of the two models.

This equation shows that the variance of the combined forecast error,  $\sigma_c^2$ , has the same structure as the variance of the portfolio return in an asset allocation exercise. Hence we can use a diversification argument for forecasts the same way we use it for assets. In the case of forming a portfolio of two forecasts with equal weights, the diversification condition is:

$$\sigma_c < \frac{1}{2}\sigma_1 + \frac{1}{2}\sigma_2. \quad (9)$$

Equation (9) states that the MSE of the combined forecasts is lower than the weighted average of the individual MSEs. This condition is directly analogous to standard portfolio diversification and holds as long as the two forecast errors are imperfectly correlated.<sup>3</sup>

For forecasting purposes, however, it is of greater interest if  $\rho_{12}$  is “low enough” such that the “strong” diversification condition holds:

$$\sigma_c < \min(\sigma_1, \sigma_2). \quad (10)$$

If this condition holds, then the forecast combination will produce superior predictive ability to any of the two models individually in terms of a lower MSE and hence higher  $R^2$ .

### 2.3.2 The general case

Instead of assuming equal weights, we can generalize this framework by choosing the weights  $w_1$  and  $w_2$  in a way that minimizes the objective function:

$$\sigma_c^2 = w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\sigma_1\sigma_2\rho_{12}. \quad (11)$$

---

<sup>3</sup>To see this consider  $\rho_{12} = 1$ . Then,  $\sigma_c^2 = \frac{1}{4}\sigma_1^2 + \frac{1}{4}\sigma_2^2 + \frac{1}{2}\sigma_1\sigma_2 = \frac{1}{4}(\sigma_1 + \sigma_2)^2$ . Hence,  $\sigma_c = \frac{1}{2}\sigma_1 + \frac{1}{2}\sigma_2$  and there is no diversification benefit from combining forecasts. If instead  $\rho_{12} < 1$ , the diversification condition holds.

Solving for the optimal weights gives:

$$w_1^* = \frac{\sigma_2^2 - \sigma_1\sigma_2\rho_{12}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}} \quad (12)$$

$$w_2^* = \frac{\sigma_1^2 - \sigma_1\sigma_2\rho_{12}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}}. \quad (13)$$

The weights are higher for models that produce more precise forecasts with lower forecast error variances. Moreover, equal weights will be optimal if  $\sigma_1^2 = \sigma_2^2$ .

Inserting the optimal weights into the objective function (11), we get the MSE associated with the optimal weights:

$$\sigma_c^{2*} = \frac{\sigma_1^2\sigma_2^2(1 - \rho_{12}^2)}{\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}}. \quad (14)$$

It is straightforward to show that  $\sigma_c^* < \min(\sigma_1, \sigma_2)$ , except for the following cases:  $\sigma_1 = \sigma_2$  and  $\rho_{12} = 1$ ;  $\sigma_1 = 0$  or  $\sigma_2 = 0$ ; or  $\rho_{12} = \sigma_1/\sigma_2$ . In summary, our analysis so far implies that if  $\sigma_1 = \sigma_2$  and  $\rho_{12} < 1$ , then equal weights are optimal and the strong diversification condition holds.

Finally, to quantify the diversification gain, it is natural to compare  $\sigma_c^*$  to  $\min(\sigma_1, \sigma_2)$ . Suppose that  $\sigma_1 > \sigma_2$  and define  $\kappa = \frac{\sigma_2}{\sigma_1}$  so that  $\kappa < 1$ . Then, the diversification gain is:

$$\frac{\sigma_c^{2*}}{\sigma_2^2} = \frac{(1 - \rho_{12}^2)}{1 + \kappa^2 - 2\kappa\rho_{12}}. \quad (15)$$

To see how the efficiency gain changes with  $\rho_{12}$ , we compute:

$$\frac{\partial(\sigma_c^{2*}/\sigma_2^2)}{\partial\rho_{12}} \propto \kappa\rho_{12}^2 - (1 + \kappa^2)\rho_{12} + \kappa. \quad (16)$$

This is a second order polynomial in  $\rho_{12}$ , where the efficiency gain has a positive relation with  $\rho_{12}$  in the interval  $[-1, \kappa]$ , and a negative relation in the interval  $[\kappa, +1]$ . Only when  $\kappa = 1$  (i.e.,  $\sigma_1 = \sigma_2$ ), the relation is always positive since:

$$\frac{\partial(\sigma_c^{2*}/\sigma)}{\partial\rho_{12}} \propto (\rho_{12} - 1)^2 > 0. \quad (17)$$

In this case, the diversification gain  $\sigma_c^{2*}/\sigma$  is always a positive function of the correlation  $\rho_{12}$ , so that the lower the correlation, the lower the  $\sigma_c^{2*}/\sigma$ , and hence the higher the diversification gain. In short, when  $\sigma_1 = \sigma_2$ , we attain maximum diversification gains, i.e., the largest reduction in  $\sigma_c^{2*}/\sigma$  for a given change in  $\rho_{12}$ . For further details, see Timmermann (2006).<sup>4</sup>

---

<sup>4</sup>Note that in the extreme case where  $\sigma_1 = \sigma_2$  and  $\rho_{12} = -1$ ,  $\sigma_c^{2*} = 0$ , which implies perfect forecasting

To apply this framework to our forecasting exercise, consider the two individual predictive models that condition on only the dividend yield or only the short rate. Suppose that both models exhibit a positive out-of-sample  $R^2$ , and their forecasts are negatively correlated: the dividend yield performs well in expansions and the short rate performs well in recessions. The negative correlation among the forecasts implies that the forecast errors are imperfectly correlated ( $\rho_{12} < 1$ ) such that the standard diversification condition holds. If, in addition  $\sigma_1^2 \approx \sigma_2^2$ , then the strong diversification condition holds, and hence the simple combined forecast substantially outperforms both individual models. Even better,  $\sigma_1^2 \approx \sigma_2^2$  leads to maximum diversification gains for a given  $\rho_{12} < 1$ . Finally,  $\sigma_1^2 \approx \sigma_2^2$  implies that equal weights are optimal. In short, therefore, by carefully selecting individual forecasts which overall perform close to equally well but also perform better in different states of the world, then the equally-weighted combination can deliver maximum diversification gains and superior predictive performance.

### 3 Data on economic fundamentals

We use a set of monthly economic fundamentals for predicting the monthly equity premium for the period of January 1927 to December 2015. All data are taken from Amit Goyal's website. These are the same data used in Welch and Goyal (2008), Campbell and Thompson (2008), Dangl and Halling (2012), Neely et al. (2014), and Pettenuzzo, Timmermann, and Valkanov (2014) extended to 2015.

Most of our results are reported for the full sample and the two subsamples of expansions and recessions. We measure recessions using the definition of the National Bureau of Economic Research (NBER) business cycle dating committee. The start of a recession is the peak of economic activity and its end is the trough.

The equity premium is the continuously compounded return on the S&P 500 Index including dividends obtained from CRSP minus the Treasury bill rate (defined below). For the full sample, the equity premium in annualized terms exhibits a mean return of 7.8%, volatility of 19%, and a Sharpe ratio of 0.41. For expansions, the annualized Sharpe ratio of the equity premium rises to 0.72 and for recessions it falls to  $-0.22$ . Recessions correspond to about 20% of the full sample and are known ex post. The full set of descriptive statistics is reported in the Internet Appendix.

The economic fundamentals include the following 14 monthly predictors:

1. Dividend yield ( $dy$ ) is the difference between the log of dividends and the log of lagged

---

ability for the equally-weighted combination.

prices.

2. Dividend-price ratio ( $dpr$ ) is the difference between the log of dividends and the log of current prices.
3. Earnings-price ratio ( $epr$ ) is the difference between the log of earnings and the log of prices.
4. Dividend payout ratio ( $dpayr$ ) is the difference between the log of dividends and the log of earnings.
5. Book-to-market ratio ( $bm$ ) is the ratio of book value to market value for the Dow Jones Industrial Average.
6. Net equity expansion ( $ntis$ ) is the ratio of twelve-month moving sums of net issues by NYSE-listed stocks divided by the total market capitalization of NYSE stocks.
7. Stock variance ( $svar$ ) is the sum of squared daily returns on the S&P 500.
8. Treasury bill rate ( $tbl$ ) is the 3-month rate.
9. Long-term yield ( $lty$ ) on government bonds.
10. Term spread ( $tms$ ) is the difference between  $lty$  and  $tbl$ .
11. Long-term rate of return ( $ltr$ ) for government bonds.
12. Default yield spread ( $dfy$ ) is the difference between BAA- and AAA-rated corporate bond yields.
13. Default return spread ( $dfr$ ) is the difference between the return on long-term corporate bonds and the return on long-term government bonds.
14. Inflation ( $infl$ ). Note that inflation information is released in the following month. Therefore, we lag inflation by an additional month in the predictive regressions.

The cross-correlations between the predictors are reported in the Internet Appendix. The values of the cross-correlations range from  $-0.522$  to  $0.995$ , with an average value of  $0.015$ .

## 4 Models for predicting the equity premium

### 4.1 Our predictive framework

We generate forecasts for the equity premium using the simple predictive regression in equation (6). Our main analysis is based on conditioning on a single predictor, such as the dividend yield. Hence, we report results from 14 regressions each conditioning on one of the Welch and Goyal (2008) predictors. We then form simple forecast combinations, which aim at exploiting the diversification gains discussed earlier.

Following Campbell and Thompson (2008), we also impose two constraints motivated by economic theory. First, we constrain the equity premium forecast to be positive in every time period. We do so by replacing negative forecasts with zero. Campbell and Thompson (2008) argue that a reasonable investor would not have used a model to forecast a negative equity premium. The positive forecast constraint is also implemented more recently by Pettenuzzo, Timmermann, and Valkanov (2014), who motivate this constraint by arguing that risk-averse investors would not hold stocks if their expected excess return was negative.

Second, we constrain the sign of the slope coefficients to be consistent with economic theory. Campbell and Thompson (2008, p. 1516) explain that “[a] regression estimated over a short sample period can easily generate perverse results, such as a negative coefficient when theory suggests that the coefficient should be positive... In practice, an investor would not use a perverse coefficient but would likely conclude that the coefficient is zero, in effect imposing prior knowledge on the output of the regression.” We implement this constraint by setting a value of zero for a coefficient that does not have the theoretically motivated sign of Campbell and Thompson (2008). To be more specific, the slope constraint is positive for all predictors except for *dpayr*, *ntis*, *tbl*, *lty*, and *infl*. As we report later, these economic constraints considerably improve the performance of the models.

### 4.2 Other predictive regressions

#### 4.2.1 The benchmark

The benchmark against which we compare all models is the historical mean for the equity premium. This is the prevalent benchmark in the literature and corresponds to the case of  $\beta = 0$  in equation (6). In other words, the historical mean benchmark reflects the view that the expected equity premium is constant and hence it is not predictable when conditioning on economic fundamentals.

### 4.2.2 Forecast combinations

A natural alternative to our predictive framework is to form standard forecast combinations using all 14 forecasts generated by the predictive regressions that condition on one predictor at a time. Following Rapach, Strauss, and Zhou (2010), we implement two approaches to standard forecast combination. First, we simply compute the equally-weighted average of all forecasts at each point in time. We refer to this as the “mean” combination. Second, we compute a weighted average of the individual forecasts at each point in time using as weights the inverse of the discounted MSE of each model up to that point. The discount factor is set to 0.90 as in Rapach, Strauss, and Zhou (2010). We refer to this as the “MSE” combination.

### 4.2.3 Combinations of predictors

Next we consider the “model selection” (MS) approach of Welch and Goyal (2008). In the MS regressions, we include all possible combinations of predictors. Then, at each time period, we select the one forecast that has performed the best by displaying the lowest cumulative MSE up to that point. The MS approach corresponds to choosing one among  $2^N$  models, which in our case is 2,048 models.

We also consider predictive regressions based on principal component analysis (PCA), which has recently been implemented in equity premium prediction by Neely et al. (2014). This approach involves estimating a set of principal components that parsimoniously incorporate information from the 14 predictors. As in Neely et al. (2014), at each time period we select the number of principal components that give the highest adjusted  $R^2$  using data up to that point.

Finally, we estimate a kitchen-sink regression that conditions on all predictors.<sup>5</sup> The kitchen-sink regression is first estimated with OLS. However, it is well known that the out-of-sample performance of the kitchen-sink regression estimated with OLS is very poor relative to the historical mean benchmark (see Welch and Goyal, 2008).

For this reason, we follow Li and Tsiakas (2016) in also estimating the kitchen-sink regression using three prominent shrinkage estimators: ridge regression, lasso, and the elastic net. These are designed to improve performance by reducing the effect of less informative predictors in out-of-sample prediction. Shrinkage estimation produces biased parameter es-

---

<sup>5</sup>Note that the kitchen-sink regression conditions on 11 of the 14 predictors. We exclude three predictors in order to avoid perfect multicollinearity: the dividend-price ratio, which is almost perfectly correlated with the dividend yield; the dividend payout ratio, which is a combination of the dividend-price ratio and the earnings-price ratio; and the long-term yield, which is a combination of the Treasury bill rate and the term spread. For the same reason, we also use a maximum of 11 predictors for the model selection approach.

timates by shrinking all estimates towards zero, which is the value implied by the benchmark historical mean model. This is done in a way that improves the bias-variance tradeoff in estimation and reduces the mean squared error of the forecasts, leading to a higher degree of predictive accuracy. Note that the kitchen sink regression is a way of pooling information because it directly conditions on a large number of predictors. In contrast, combined forecasts are designed to pool forecasts.

### 4.3 Out-of-sample analysis

All empirical models are evaluated out-of-sample relative to the historical mean benchmark. We generate out-of-sample forecasts with rolling predictive regressions using a 20-year estimation window such that the first forecast is for January 1947 and the last for December 2015. We adopt a rolling window approach to be consistent with Welch and Goyal (2008) and the ensuing literature.

The main statistical criterion for evaluating the out-of-sample predictive ability of the models is the Campbell and Thompson (2008) and Welch and Goyal (2008)  $R_{oos}^2$  statistic. The  $R_{oos}^2$  compares the unconditional one-month ahead forecasts  $\bar{r}_{t+1|t}^e$  of the historical mean benchmark to the conditional forecasts  $\hat{r}_{t+1|t}^e$  of the alternative model, and is defined as follows:

$$R_{oos}^2 = 1 - \frac{MSE\left(\hat{r}_{t+1|t}^e\right)}{MSE\left(\bar{r}_{t+1|t}^e\right)} = 1 - \frac{\sum_{t=1}^{T-1} \left(r_{t+1}^e - \hat{r}_{t+1|t}^e\right)^2}{\sum_{t=1}^{T-1} \left(r_{t+1}^e - \bar{r}_{t+1|t}^e\right)^2}. \quad (18)$$

A positive  $R_{oos}^2$  implies that the alternative model outperforms the benchmark by means of lower MSE.

We assess the statistical significance of the  $R_{oos}^2$  statistic by applying the Clark and West (2006, 2007) testing procedure. This is a test of the null hypothesis of equal predictive ability between the benchmark and the alternative model. The Clark and West (2006, 2007) procedure accounts for the fact that, under the null, the MSE of the benchmark is expected to be lower. This is because the alternative models estimate a parameter vector that, under the null, is not helpful in prediction thus introducing noise into the forecasting process. Clark and West (2006, 2007) adjust the MSE as follows:

$$MSE_{adj} = \frac{1}{T-1} \sum_{t=1}^{T-1} \left(r_{t+1}^e - \hat{r}_{t+1|t}^e\right)^2 - \frac{1}{T-1} \sum_{t=1}^{T-1} \left(\bar{r}_{t+1|t}^e - \hat{r}_{t+1|t}^e\right)^2. \quad (19)$$

Then, we define:

$$\widehat{test}_{t+1} = (r_{t+1}^e - \bar{r}_{t+1|t}^e)^2 - [(r_{t+1}^e - \widehat{r}_{t+1|t}^e)^2 - (\bar{r}_{t+1|t}^e - \widehat{r}_{t+1|t}^e)^2], \quad (20)$$

and regress  $\widehat{test}_{t+1}$  on a constant, using the  $t$ -statistic for a zero coefficient. Even though the asymptotic distribution of this test is non-standard (e.g., McCracken, 2007), Clark and West (2006, 2007) show that standard normal critical values provide a good approximation, and therefore recommend to reject the null of equal predictive ability if the statistic is greater than +1.282 (for a one-sided 0.10 test) or +1.645 (for a one-sided 0.05 test) or +2.326 (for a one-sided 0.01 test).

## 4.4 Empirical results

### 4.4.1 Individual Predictors

We begin our analysis by focusing on the out-of-sample performance of the 14 individual predictors. This allows us to identify which predictors perform well, whether the economic constraints improve performance, and whether the performance of the predictors depends on the state of the economy. We report results for the full sample as well as NBER-dated expansions and recessions. The results for expansions and recessions are based on estimating the models each month over the full forecasting period and then separating the forecasting errors ex post across the two subsamples as in Neely et al. (2014).

Table 1 reports the  $R_{oos}^2$ . In the base case of no economic constraints, the only two predictors that exhibit a positive and statistically significant  $R_{oos}^2$  are the dividend yield ( $dy$ ) with an  $R_{oos}^2$  of 0.42% and the dividend-price ratio ( $dpr$ ) with an  $R_{oos}^2$  of 0.14%.<sup>6</sup> The other 12 predictors display a negative  $R_{oos}^2$ . For example, the short rate ( $tbl$ ) exhibits an  $R_{oos}^2$  of -1.75%. Therefore, in the absence of constraints, only dividends have predictive information for the equity premium and all other predictors appear to be devoid of any predictive information.

This finding changes considerably with the imposition of the two economic constraints as they almost universally improve performance: the  $R_{oos}^2$  rises for 13 of the 14 predictors, the single exception being the dividend payout ratio ( $dpayr$ ). Under the constraints, six more predictors exhibit a positive and significant  $R_{oos}^2$ . Notably, the  $R_{oos}^2$  of the dividend yield rises to 0.69%, and is significant at 1%, whereas the  $R_{oos}^2$  of the short rate rises to a positive value (0.33%) and is significant at 5%. In short, therefore, economic constraints are instrumental in revealing the predictive information of economic fundamentals. For this reason, the rest

---

<sup>6</sup>These two predictors ( $dy$  and  $dpr$ ) are virtually the same by definition and their correlation exceeds 99.5%. Our discussion will therefore focus on  $dy$  as it is more popular in the literature than  $dpr$ .



of our analysis imposes the constraints. Figure 1 illustrates the time variation of the out-of-sample slope estimates of the 14 individual models and hence shows when the slopes are set to zero by the constraints.

The most important finding in Table 1 is that certain economic fundamentals perform well in expansions, whereas others perform well in recessions. Indeed, none of the 14 predictors perform well in both states of the economy. In expansions, predictors associated with stock dividends and earnings, such as the dividend yield and the earnings-price ratio, display a significant  $R_{oos}^2$  that is higher than 1%. In recessions, predictors associated with interest rates, such as the short rate, the long rate and the term spread, display a significant  $R_{oos}^2$  that is higher than 2%. Therefore, predictors that capture stock-specific information on dividends and earnings have good predictive power in good times. In contrast, predictors that capture aggregate economy-wide information on interest rates have good predictive power in bad times. This finding is consistent with the theory and empirical evidence of Kacperczyk, Van Nieuwerburgh, and Veldkamp (2016) on the cyclical variation of investor attention allocation. It also suggests that combinations of cyclical and countercyclical forecasts may potentially lead to substantial performance gains.

#### 4.4.2 Combinations of Predictors

We turn to this next by forming an equally-weighted combination of the dividend yield and the short rate. This is the simplest possible combination we can form: two predictors, equal weights, one predictor performs well in expansions and one in recessions. In what follows, we will show that this is a powerful predictor that is hard to beat.

From a practical point of view, this is a combination of the two most prominent predictors in the literature. From a theoretical point of view, it is a combination that is motivated by the present value framework of Campbell and Shiller (1998). From a diversification point of view, it is a combination that exploits the negative correlation between the two sets of forecasts, which leads to an imperfect correlation between the forecast errors, and therefore delivers diversification gains.

To quantify the diversification gains consider the results of Table 2. The dividend yield alone delivers an  $R_{oos}^2$  of 0.69% and the short rate alone an  $R_{oos}^2$  of 0.33%. A simple combination of the two delivers an  $R_{oos}^2$  of 1.29%, which is almost double the  $R_{oos}^2$  of the better model. This implies that the MSE of the combined forecasts is much lower than the lowest MSE of the two predictors. Hence this result satisfies the strong diversification condition of equation (10). Even better, the simple combination delivers good performance in both expansions ( $R_{oos}^2 = 1.13\%$ ) and recessions ( $R_{oos}^2 = 1.70\%$ ), a result that is statistically significant. In conclusion, therefore, a simple combination of the dividend yield and the short rate com-

bines popular predictors, which are theoretically motivated, delivers strong diversification gains and performs well in both expansions and recessions.

This forecast combination has two features that make it simple and attractive: the equal weights and using only two predictors. We will now explain why both of these features make sense. The equal weights on the two predictors is not just a simple choice that happens to work well. It is also an optimal choice. When computing the optimal diversification weights using equations (12) and (13), the result is a weight of 52% on the dividend yield and 48% on the short rate. Therefore, as far as forecast combinations go, in our sample equal weights are practically as good as it gets.

Using only two predictors is also a good choice because: (1) the two predictors are negatively correlated with each other; and (2) each predictor is highly positively correlated with the other predictors that perform well in the same state of the world. In other words, the performance of  $dy$  is highly correlated with that of  $dpr$  and  $epr$ . Therefore, adding any of the latter two does not add much to the performance of the first one. Similarly, the performance of the short rate ( $tbl$ ) is highly correlated to that of the long rate ( $lty$ ), the term spread ( $tms$ ), and the long-term return ( $ltr$ ). Again, adding any of the latter three does not improve much the performance of the first one.

This finding is illustrated in Table 3, which reports on the performance of more combinations. The table shows that the performance of the mean combination remains almost the same when: (1) we replace  $dy$  by  $dpr$  or  $epr$ , (2) we replace  $tbl$  by  $tms$ ,  $lty$  or  $ltr$ ; (3) we add with equal weights to the  $dy+tbl$  combination any one of the following:  $dpr$ ,  $epr$ ,  $tms$ ,  $lty$  or  $ltr$ ; and (4) we form a grand mean combination of all seven predictors that work well in one state of the economy:  $dy$ ,  $dpr$ ,  $epr$ ,  $tbl$ ,  $tms$ ,  $lty$  and  $ltr$ . Arguably, the grand combination of the seven predictors has slightly better performance than the simple combination of  $dy$  and  $tbl$ , but the difference is not substantial. Overall, in all these cases, the results are similar in the sense all these models deliver a positive and significant  $R_{oos}^2$  that is higher than 1%, which remains positive and significant in expansions and recessions.

These results also imply that the combination of  $dy$  and  $tbl$  is not unique in delivering a good performance in both states of the economy. We could replace  $dy$  by  $epr$ , or  $tbl$  by  $tms$ , and the resulting two-predictor combination would have a similar performance. Our proposed combination is based on  $dy$  and  $tbl$  because, as mentioned earlier, these two are the most popular predictors with solid theoretical foundations. A combination of these two predictors is hard to improve on. Having said that, we show that one can also combine say  $epr$  with  $tms$  and achieve similar diversification and insurance results.

Furthermore, the simple mean combination of  $dy$  and  $tbl$  significantly outperforms the standard forecast combinations of all 14 predictors. Since Rapach, Strauss, and Zhou (2010)

forecast combinations of all available predictors have become quite popular, partly due to their simplicity. We show here that much simpler combinations can perform much better if we carefully select the predictors to combine in a way that achieves diversification gains. Indeed, our simple approach not only outperforms standard combinations but also justifies why standard combinations work: by naively combining all predictors, we combine some that work well in expansions, some that work well in recessions and some that work well in neither. Then, by excluding from the combination the predictors that work well in neither state of the economy, we can form simpler and yet superior combinations.

To illustrate these results, in Figure 2 we plot the out-of-sample performance of selected models over time. Following Welch and Goyal (2008), the figure shows the difference of the cumulative squared error of the null (historical mean) minus the cumulative squared error of the alternative. The figure shows that simple mean combination of *dy* and *tbl* displays a consistent upward trend in its out-of-sample performance over time. Hence the good performance of this model is not due to a particular subsample but is systematic over a long sample spanning the full postwar period. The figure also shows that the simple combination outperforms the standard mean forecast combination based on all 14 predictors. Finally, the simple combination strongly outperforms the performance of its two individual components, the dividend yield and the short rate. It is worth noting that, although the performance of the latter two predictors is not consistently better than the benchmark over the long sample, the equally-weighted combination is.

Finally, consider the strategy perfectly timing the business cycle using the true recession dummy. This strategy can only be implemented ex post and hence it is not directly comparable to the out-of-sample strategies discussed above. We use this ex-post timing strategy just to illustrate that the diversification gains of the simple mean combination are due to the cyclical performance of the predictors. We find that the  $R_{oos}^2$  of the ex-post timing strategy is only marginally better than the simple combination. This indicates that the simple combination of the dividend yield and the short rate captures most of the gains of exploiting predictors that work well in different states of the world.

## 5 Predictability and asset allocation

### 5.1 A mean-variance trading strategy

Following Campbell and Thompson (2008), we assess the economic value of equity premium predictability using a dynamic asset allocation strategy. The strategy involves monthly rebalancing of a portfolio that invests in the S&P 500 Index (the risky asset) and the Treasury

bill (the riskless asset). We consider a mean-variance efficient investor with a one-month ahead horizon, who determines the optimal weights by implementing a maximum expected utility rule as follows:

$$\begin{aligned}
\max_{w_t} \quad & E_t [U(r_{p,t+1})] = r_{p,t+1|t} - \frac{\gamma}{2} \sigma_{p,t+1|t}^2 \\
\text{s.t.} \quad & r_{p,t+1|t} = w_t r_{t+1|t} + (1 - w_t) r_f, \\
& \sigma_{p,t+1|t}^2 = w_t^2 \sigma_{t+1|t}^2,
\end{aligned} \tag{21}$$

where  $r_{p,t+1|t}$  is the  $t + 1$  forecast of the portfolio return conditional on time  $t$  information,  $\gamma$  is the investor's degree of relative risk aversion,  $\sigma_{p,t+1|t}^2$  is the  $t + 1$  forecast of the portfolio variance made at time  $t$ ,  $r_{t+1|t}$  is the  $t + 1$  forecast of the S&P 500 Index return made at time  $t$ ,  $r_f$  is the risk-free rate of return, and  $\sigma_{t+1|t}^2$  is the  $t + 1$  forecast of the variance to the S&P 500 Index return made at time  $t$ . Note that we forecast  $\sigma_{t+1|t}^2$  using a five-year rolling average of the variance of past monthly returns as in Campbell and Thompson (2008). We also set  $\gamma = 5$  as in Neely et al. (2014).

The solution to the maximum expected utility rule delivers the risky asset weight:

$$w_t = \frac{1}{\gamma} \frac{r_{t+1|t} - r_f}{\sigma_{t+1|t}^2}. \tag{22}$$

Consistent with the literature (e.g., Campbell and Thompson, 2008; Neely et al., 2014), we constrain the weight on the risky asset by imposing  $w_t \in [0, 1.5]$ . In other words, we do not allow short-selling and leverage is limited to no more than 50%.

We evaluate the performance of portfolios generated by a given set of equity premium forecasts using the Sharpe ratio ( $SR$ ) and the certainty equivalent return ( $CER$ ). The  $SR$  is perhaps the most commonly used performance measure and is defined as the average excess return of a portfolio divided by the standard deviation of the portfolio returns. We assess statistical significance using the Ledoit and Wolf (2008) bootstrap two-sided test of whether the  $SR$  of the alternative model is different from the benchmark.

The certainty equivalent return is defined as:

$$CER = \left( \bar{r}_p - \frac{\gamma}{2} \bar{\sigma}_p^2 \right), \tag{23}$$

where  $\bar{r}_p$  is the mean portfolio return and  $\bar{\sigma}_p^2$  is the portfolio variance over the forecast evaluation period. The  $CER$  can be interpreted as the performance fee the risk-averse investor is willing to pay for switching from the riskless asset to the risky portfolio. We focus on the difference in  $CER$  ( $\Delta CER$ ), which is equal to the  $CER$  of the portfolio generated by the forecasts of the alternative model minus the  $CER$  of the portfolio generated by the

historical mean benchmark.  $\Delta CER$  measures the performance fee the risk-averse investor is willing to pay for switching from the risky portfolio generated by the benchmark model to the risky portfolio generated by the alternative model. To provide a realistic assessment of the profitability of dynamic trading strategies, we also take into account the effect of transaction costs. In particular, we compute the  $\Delta CER$  net of proportional transaction costs equal to 50 bps per month as in Neely et al. (2014).

Finally, we compute the average turnover of each trading strategy, which is defined as follows:

$$Turnover = \frac{1}{T-1} \sum_{t=1}^{T-1} (|w_{t+1} - w_{t+1}^-|), \quad (24)$$

where  $T-1$  is the number of trading periods,  $w_{t+1}$  is the weight on the risky asset at time  $t+1$ , and  $w_{t+1}^- = w_t \frac{1+r_{t+1}}{1+r_{p,t+1}}$  is the weight on the risky asset right before rebalancing at time  $t+1$ . This turnover measure represents the average monthly trading volume. We report the average relative turnover, which is the ratio of the average turnover of the alternative model divided by the average turnover of the benchmark model.

## 5.2 Portfolio performance

We assess the performance of dynamically rebalanced portfolios generated by the monthly forecasts of the predictive models. Table 4 reports the empirical findings. The first result to note is that the historical mean benchmark delivers a  $CER = 5.70\%$  per year relative to riskless investing. The  $CER$  rises to  $8.14\%$  in expansions but falls to  $-6.81\%$  in recessions. Clearly, using the historical mean is a poor predictor of the equity premium during recessions.

Similar to our statistical findings, our main result here is that the simple mean combination of the dividend yield and the short rate performs very well for the full sample: better than the historical mean benchmark, better than the two individual predictors, and better than the standard forecast combination. The  $\Delta CER$  of the simple mean combination relative to the historical mean benchmark is  $1.66\%$  per year. More importantly, the  $\Delta CER$  remains positive for both expansions ( $0.99\%$ ) and recessions ( $4.98\%$ ). Net of transaction costs, it still retains a positive  $\Delta CER$  of  $1.44\%$  per year. Moreover, the annualized  $SR$  is  $0.56$ , which is significantly higher than the  $0.45$  of the benchmark strategy according to the Ledoit and Wolf (2008) bootstrap two-sided test.

Overall, our evidence shows that, in the context of a dynamic mean-variance strategy, a simple mean combination of the dividend yield and the short rate has high economic value in predicting the equity premium. The economic gains of this approach can be summarized into a performance fee of approximately  $1.7\%$  per year overall,  $1\%$  per year in expansions,

5% per year in recessions, together with a statistically significant increase in the  $SR$  from 0.45 to 0.56.

Finally, the ex-post timing strategy using the true recession dummy performs even better by generating a performance fee of about 2.7% per year and Sharpe ratio of about 0.67. This is further evidence that there is high economic value in timing recessions.

### 5.3 Comparing statistical and economic gains

In this section, we relate the  $R^2$  of the models to the  $SR$  of the strategies. Following Campbell and Thompson (2008, p. 1525), “the correct way to judge the magnitude of  $R^2$  is to compare it with the squared Sharpe ratio” since the proportional increase in the expected return is approximately equal to the ratio of  $\frac{R_{oos}^2}{SR^2}$ , where  $SR$  is the unconditional  $SR$  of the risky asset.<sup>7</sup>

For example, recall that the simple forecast combination delivers an  $R_{oos}^2$  of 1.29%. Over the same forecasting period (1947-2015), the risky asset has a squared monthly  $SR$  of  $0.1501^2 = 0.0225$ . Then, the proportional increase in the expected return is  $0.0129/0.0225 = 0.5726$ . In other words, the simple combination forecasts will increase the average monthly portfolio return by a factor of 57.26%. Campbell and Thompson (2008) show that this corresponds to an actual increase in the expected return of  $\left(\frac{1}{\gamma}\right) \left(\frac{R_{oos}^2}{1-R_{oos}^2}\right) (1 + SR^2) = 0.27\%$  per month or 3.21% per year. In conclusion, modest predictive ability for the equity premium can plausibly generate large economic gains.

## 6 Timing the business cycle

The phase of the business cycle can only be known ex post and hence cannot be used to inform ex-ante prediction. In fact, the business cycle peak and trough dates are determined by the NBER’s business cycle dating committee with a substantial lag, often more than a year later. Therefore, instead of using the actual NBER dates, we turn to a popular formal rule for ex-ante timing of the business cycle. This rule is based on the Markov-switching dynamic factor model (DFMS) of Chauvet (1998), which produces a probability that the economy is in an expansion or a recession every month. The Chauvet (1998) DFMS model uses information from the four coincident economic variables highlighted by the NBER in establishing turning point dates: (1) nonfarm payroll employment, (2) industrial production,

---

<sup>7</sup>Specifically, when moving from the unconditional forecast of the expected return to a conditional forecast, the proportional increase in the expected return is  $\left(\frac{R_{oos}^2}{1-R_{oos}^2}\right) \left(\frac{1+SR^2}{SR^2}\right)$ , which is approximately equal to  $\frac{R_{oos}^2}{SR^2}$ , when  $R_{oos}^2$  and  $SR^2$  are both small.

(3) real manufacturing and trade sales, and (4) real personal income excluding transfer payments.

We time the business cycle using the smoothed recession probabilities based on the Chauvet (1998) DFMS model, which are downloaded from the FRED database of the Federal Reserve Bank of St. Louis. Then, following Chauvet and Piger (2008), we identify the first month of a recession as the first month for which the probability rises above 50%. Similarly, an expansion begins on the first month for which the probability falls below 50%. This formal rule defines our forecasted recession dummy.<sup>8</sup> Chauvet and Piger (2008) find that this approach is very accurate and identifies peaks and troughs much sooner than the NBER. More importantly, it does not produce any “false positives,” that is turning points that were established in real time, but did not correspond to an actual NBER turning point date.

Note that the recession probability is known with a lag of two months in addition to the one-month forecasting lag. For example, the first available probability in the sample is for June 1967. This probability is known to us in August 1967 and we use it to forecast the equity premium in September 1967. Our analysis fully accounts for these lags so there is no look-ahead bias.<sup>9</sup>

The statistical performance of the ex-ante timing strategy is presented in Table 5. For the full sample, the ex-ante timing strategy ( $R_{oos}^2 = 1.24\%$ ) performs better than the simple forecast combination ( $R_{oos}^2 = 0.79\%$ ). However, all the performance gains in the ex-ante timing strategy are due to recessions, when it achieves an  $R_{oos}^2 = 4.49\%$  compared to  $R_{oos}^2 = 1.84\%$  of the simple forecast combination. In expansions, the simple forecast combination outperforms the ex-ante timing strategy.

The results are similar for the portfolio performance of the ex-ante timing strategy, which is presented in Table 6. The economic gains of this approach can be summarized into a performance fee of 2.92% per year overall, 0.33% per year in expansions, 23.10% per year in recessions, together with a statistically significant increase in the  $SR$  from 0.31 to 0.57. Notably, this performance is essentially the same as the ex-post timing strategy that uses the true recession dummy. This suggests that the rule for dating recessions based on the recession probability works very well. Furthermore, the performance of the ex-ante timing strategy is much better than the simple forecast combination but all the gains are exclusively due to recessions. To conclude, therefore, the ex-ante timing strategy forecasts business cycles well

---

<sup>8</sup>Note that our results are not very sensitive to the choice of the 50% probability threshold. The performance of the predictive models remains qualitatively the same when using a similar threshold.

<sup>9</sup>Note that the smoothed recession probabilities are not real-time probabilities: although the probability at a given point in time is estimated using the four macroeconomic variables available up to that point, these variables may have been subject to revision. However, our results are not very sensitive to the choice of the probability threshold of 50% and, therefore, this issue is not likely to affect our findings.

and delivers a superior performance during recessions. In expansions, however, the simple forecast combination performs better.

## 7 Conclusion

We propose a simple way to predict the equity premium out of sample based on forming an equally-weighted combination of the forecasts generated by the dividend yield and the short rate. This combination achieves two highly desirable properties: strong diversification and insurance. Strong diversification occurs by combining two negatively correlated forecasts, which individually display a positive  $R_{oos}^2$ , to create a new set of forecasts that have a substantially higher  $R_{oos}^2$  than either of the two individual models. Insurance occurs when the combined forecasts perform well in both states of the economy. These results stem from the empirical observation that the dividend yield generates forecasts that perform well in good times and the short rate generates forecasts that perform well in bad times. In the end, a simple combination of these two forecasts with equal weights performs equally well or better than any other combination based on economic fundamentals.

These findings are not surprising. The dividend yield and the short rate have been popular predictors that have solid theoretical foundations. In this paper, we simply uncover the cyclical variation in their predictive information and take advantage of it by forming a combination. The result is consistent with a recent theory on investor attention allocation that links information choices to the business cycle. According to this theory, it is natural for investors to care about aggregate information in recessions, such as interest rates, and stock-specific information in expansions, such as dividends. Overall, our analysis highlights the predictive information of widely-used economic fundamentals and demonstrates that there is a simple and powerful way to predict the equity premium out of sample.



## References

- 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**, 195–227.
- 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**, 1509–1531.
- Chauvet, M. (1998). “An Econometric Characterization of Business Cycle Dynamics with Factor Structure and Regime Switching,” *International Economic Review* **39**, 969–996.
- Chauvet, M., and J. Piger (2008). “A Comparison of the Real-Time Performance of Business Cycle Dating Methods,” *Journal of Business and Economic Statistics* **26**, 42–49.
- Claessens, S., M.A. Kose, and M.E. Terrones (2012). “How do Business and Financial Cycles Interact?” *Journal of International Economics* **87**, 178–190.
- Clark, T.E., and K.D. West (2006). “Using Out-of-Sample Mean Squared Prediction Errors to Test the Martingale Difference Hypothesis,” *Journal of Econometrics* **135**, 155–186.
- Clark, T.E., and K.D. West (2007). “Approximately Normal Tests for Equal Predictive Accuracy,” *Journal of Econometrics* **138**, 291–311.
- Cochrane, J. (2008). “The Dog that Did Not Bark: A Defense of Return Predictability,” *Review of Financial Studies* **21**, 1533–1575.
- Cochrane, J. (2011). “Presidential Address: Discount Rates,” *Journal of Finance* **66**, 1047–1108.
- Dangl, T., and M. Halling (2012). “Predictive Regressions with Time-Varying Coefficients,” *Journal of Financial Economics* **106**, 157–181.
- Fama, E.F., and K.R. French (1989). “Business Conditions and Expected Returns on Stocks and Bonds,” *Journal of Financial Economics* **25**, 23–49.
- Kacperczyk, M., S. Van Nieuwerburgh, and L. Veldkamp (2016). “A Rational Theory of Mutual Funds’ Attention Allocation,” *Econometrica* **84**, 571–626.
- Ledoit, O., and M. Wolf (2008). “Robust Performance Hypothesis Testing with the Sharpe Ratio,” *Journal of Empirical Finance* **15**, 850–859.
- Li, J., and I. Tsiakas (2016). “Equity Premium Prediction: The Role of Economic and Statistical Constraints” *Journal of Financial Markets* (forthcoming).

- Maio, P. (2013). “The ‘Fed Model’ and the Predictability of Stock Returns,” *Review of Finance* **17**, 1489–1533.
- McCracken, M.W. (2007). “Asymptotics for Out of Sample Tests of Granger Causality,” *Journal of Econometrics* **140**, 719–752.
- 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**, 1772–1791.
- Pettenuzzo, D., A. Timmermann, and R. Valkanov (2014). “Forecasting Stock Returns under Economic Constraints,” *Journal of Financial Economics* **114**, 517–553.
- 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**, 821–862.
- Timmermann, A. (2006). “Forecast Combinations,” in Elliott, G., C.W.J. Granger, and A. Timmermann (eds.), *Handbook of Economic Forecasting*, Volume 1, Amsterdam: North-Holland.
- Welch, I., and A. Goyal (2008). “A Comprehensive Look at the Empirical Performance of Equity Premium Prediction,” *Review of Financial Studies* **21**, 1455–1508.

**Table 1. Equity Premium Prediction: Individual Predictors**

The table displays the out-of-sample  $R_{oos}^2$  in percent for predictive models of the monthly equity premium against the null of the historical mean. In the last three columns, the  $R_{oos}^2$  is for models that impose a sign constraint on the slope coefficients and a positivity constraint on the forecasts. The out-of-sample monthly forecasts are obtained using a 20-year rolling window for the sample period of January 1927 to December 2015. Expansions and recessions are according to the NBER. \*, \*\*, and \*\*\* denote statistical significance at the 10%, 5%, and 1% level, respectively, using the Clark and West (2006, 2007) one-sided  $t$ -statistic.

<i>Predictor</i>	$R_{oos}^2$ (%)					
	<i>No Economic Constraint</i>			<i>Both Economic Constraints</i>		
	<i>Full Sample</i>	<i>Expansion</i>	<i>Recession</i>	<i>Full Sample</i>	<i>Expansion</i>	<i>Recession</i>
<i>dy</i>	0.42**	0.63***	-0.09	0.69***	1.02***	-0.11
<i>dpr</i>	0.14**	0.54***	-0.81	0.44**	0.97***	-0.84
<i>epr</i>	-1.63	0.69**	-7.25	0.04**	1.11***	-2.58
<i>dpayr</i>	-1.10	-1.00	-1.34	-11.20	-9.32	-15.74
<i>bm</i>	-1.42	-0.36*	-3.99	0.04*	0.09**	-0.08
<i>ntis</i>	-0.71	-1.16	0.38	-0.25*	-0.30*	-0.11
<i>svar</i>	-3.16	-3.27	-2.92	-2.40	-2.63	-1.83
<i>tbl</i>	-1.75*	-3.23	1.83*	0.33**	-0.44	2.20**
<i>lty</i>	-0.58*	-1.33	1.24	0.39**	-0.52	2.58***
<i>ltr</i>	-0.18**	-1.73	3.58***	0.40**	-0.54	2.67**
<i>tms</i>	-0.62**	-2.17	3.16**	0.07***	-1.11*	2.94***
<i>dfy</i>	-1.83	-2.38	-0.50	-1.64	-1.15	-2.84
<i>dfr</i>	-1.47	-0.22	-4.50	-0.76	-0.18	-2.16
<i>infl</i>	-0.31	-1.03	1.42	-0.26	-0.59	0.51

**Table 2. Equity Premium Prediction: Simple Combinations**

The table displays the out-of-sample  $R_{oos}^2$  in percent for predictive models of the monthly equity premium against the null of the historical mean. The  $R_{oos}^2$  is for models that impose a sign constraint on the slope coefficients and a positivity constraint on the forecasts. The out-of-sample monthly forecasts are obtained using a 20-year rolling window for the sample period of January 1927 to December 2015. Expansions and recessions are according to the NBER. \*, \*\*, and \*\*\* denote statistical significance at the 10%, 5%, and 1% level, respectively, using the Clark and West (2006, 2007) one-sided  $t$ -statistic.

<i>Predictor</i>	$R_{oos}^2$ (%)		
	<i>Full Sample</i>	<i>Expansion</i>	<i>Recession</i>
<i>Two individual predictors</i>			
<i>dy</i>	0.69***	1.02***	-0.11
<i>tbl</i>	0.33**	-0.44	2.20**
<i>Mean combination with two predictors</i>			
<i>dy + tbl</i>	1.29***	1.13***	1.70**
<i>Ex-post timing strategy using the true recession dummy</i>			
<i>dy + tbl</i>	1.36***	1.02***	2.20**
<i>Standard forecast combinations: all 14 predictors</i>			
<i>Mean</i>	0.77***	0.96***	0.30
<i>MSE</i>	0.73**	0.95***	0.21

**Table 3. Equity Premium Prediction: More Combinations**

The table displays the out-of-sample  $R_{oos}^2$  in percent for predictive models of the monthly equity premium against the null of the historical mean. The  $R_{oos}^2$  is for models that impose a sign constraint on the slope coefficients and a positivity constraint on the forecasts. The out-of-sample monthly forecasts are obtained using a 20-year rolling window for the sample period of January 1927 to December 2015. Expansions and recessions are according to the NBER. \*, \*\*, and \*\*\* denote statistical significance at the 10%, 5%, and 1% level, respectively, using the Clark and West (2006, 2007) one-sided  $t$ -statistic.

<i>Predictor</i>	$R_{oos}^2$ (%)		
	<i>Full Sample</i>	<i>Expansion</i>	<i>Recession</i>
<i>Mean combinations with two predictors</i>			
<i>dy + tbl</i>	1.29***	1.13***	1.70**
<i>dy + tms</i>	1.34***	1.08***	1.96**
<i>dy + lty</i>	1.23***	1.01***	1.78**
<i>dy + ltr</i>	1.27***	1.04***	1.83*
<i>dpr + tbl</i>	1.06***	0.99***	1.23*
<i>dpr + tms</i>	1.18***	1.04***	1.51*
<i>dpr + lty</i>	1.03***	0.90***	1.34*
<i>dpr + ltr</i>	1.11***	0.97***	1.46*
<i>epr + tbl</i>	0.90***	1.05***	0.55
<i>epr + tms</i>	1.17***	1.23***	1.03*
<i>epr + lty</i>	0.73***	0.84***	0.47
<i>epr + ltr</i>	0.95***	0.95***	0.97*
<i>Mean combinations with three predictors</i>			
<i>dy + tbl + dpr</i>	1.15***	1.22***	0.97
<i>dy + tbl + epr</i>	1.21***	1.43***	0.67
<i>dy + tbl + tms</i>	1.44***	1.09***	2.30***
<i>dy + tbl + lty</i>	1.21***	0.83***	2.12***
<i>dy + tbl + ltr</i>	1.46***	1.06***	2.43***
<i>Mean combinations with seven predictors</i>			
<i>dy + dpr + epr + tbl + tms + lty + ltr</i>	1.49***	1.34***	1.85***

**Table 4. Portfolio Performance**

The table shows the out-of-sample portfolio performance for a mean-variance investor, who each month rebalances her portfolio by investing in one risky asset (S&P 500) and the riskless rate (T-bill). The investor has a degree of relative risk aversion equal to 5 and follows a maximum utility strategy.  $\Delta CER$  is the gain in the percent annualized Certainty Equivalent Return ( $CER$ ) for switching from the forecasts of the benchmark to the forecasts generated by the alternative model.  $SR$  is the annualized Sharpe ratio.  $Turnover$  is the ratio of the average turnover of the portfolio generated by the forecasts of the alternative model divided by the average turnover of the portfolio generated by the benchmark. For the historical mean benchmark, we report the level of  $CER$  and the average turnover for that portfolio. The superscripts \*, \*\*, and \*\*\* denote statistical significance at the 10%, 5%, and 1% level, respectively, using the Ledoit and Wolf (2008) bootstrap two-sided test of whether the Sharpe ratio of a model is different from that of the benchmark. The last column is the annual percent  $\Delta CER$  assuming a proportional transaction cost of 50 basis points per month. The out-of-sample monthly forecasts are obtained using a 20-year rolling window for the sample period of January 1927 to December 2015.

<i>Predictor</i>	$\Delta CER$ (%)			$SR$	$Turnover$	$\Delta CER$ (%)
	<i>Full</i>	<i>Expansion</i>	<i>Recession</i>	<i>Full</i>	<i>Full</i>	<i>c = 50 bps</i>
	<i>Sample</i>			<i>Sample</i>	<i>Sample</i>	<i>Full Sample</i>
<i>Historical mean</i>	5.70	8.14	-6.81	0.45	0.04	5.56
<i>Two individual predictors</i>						
<i>dy</i>	0.47	1.05	-2.70	0.47	2.44	0.09
<i>tbl</i>	1.27	-0.64	11.06	0.53	1.46	1.11
<i>Mean combination with two predictors</i>						
<i>dy + tbl</i>	1.66	0.99	4.98	0.56*	1.71	1.41
<i>Ex-post timing strategy using the true recession dummy</i>						
<i>dy + tbl</i>	2.69	1.05	11.06	0.67**	2.60	2.26
<i>Standard forecast combinations: all 14 predictors</i>						
<i>Mean</i>	0.16	0.59	-2.11	0.47	1.63	-0.02
<i>MSE</i>	0.15	0.60	-2.20	0.47	2.01	-0.14

**Table 5. Timing the Business Cycle**

The table displays the out-of-sample  $R_{oos}^2$  in percent for predictive models of the monthly equity premium against the null of the historical mean. The  $R_{oos}^2$  is for models that impose a sign constraint on the slope coefficients and a positivity constraint on the forecasts. The out-of-sample monthly forecasts are obtained using a 20-year rolling window for the sample period of August 1967 to December 2015. Expansions and recessions are according to the NBER. \*, \*\*, and \*\*\* denote statistical significance at the 10%, 5%, and 1% level, respectively, using the Clark and West (2006, 2007) one-sided  $t$ -statistic.

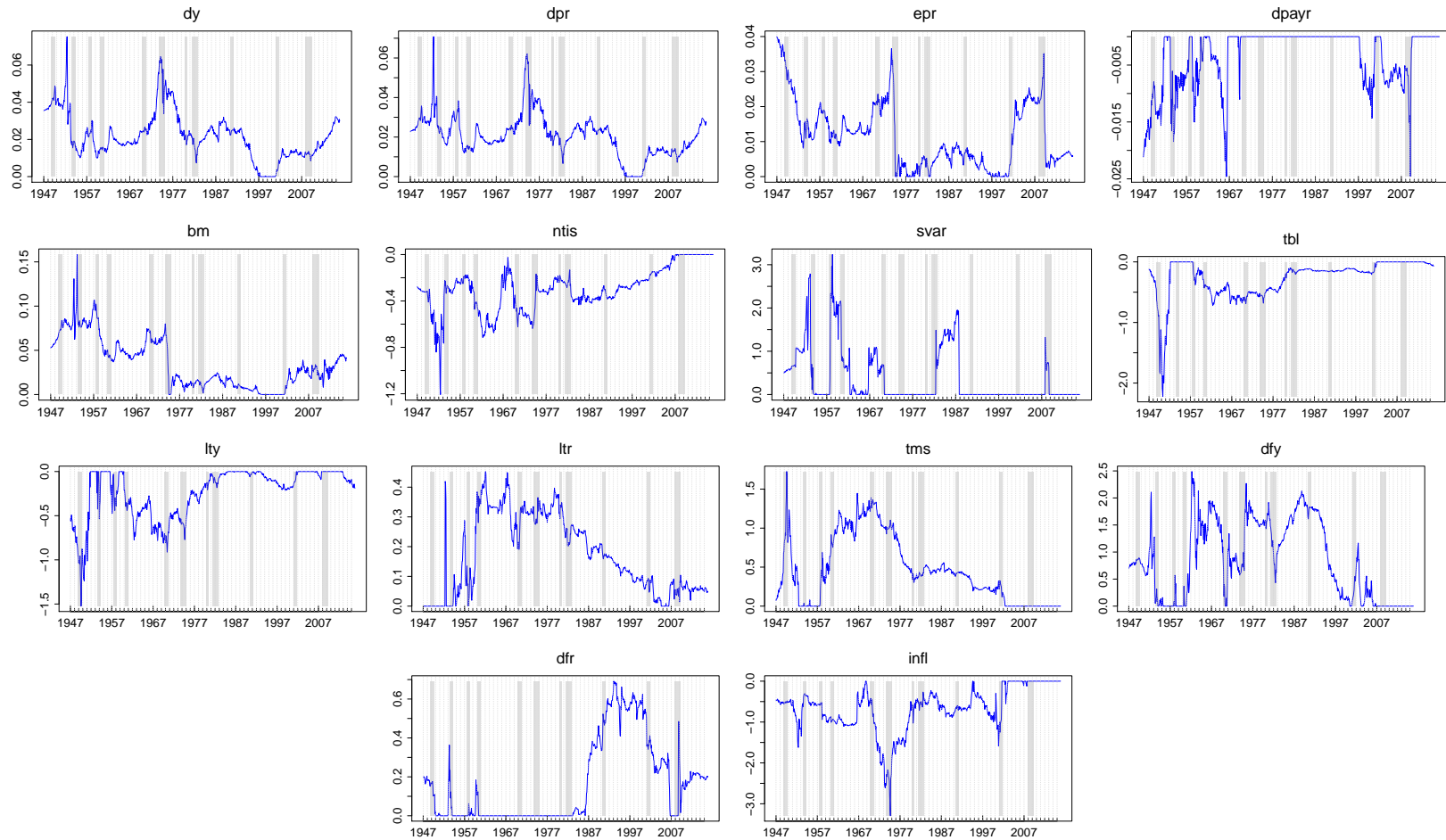
<i>Predictor</i>	$R_{oos}^2$ (%)		
	<i>Full Sample</i>	<i>Expansion</i>	<i>Recession</i>
<i>Two individual predictors</i>			
<i>dy</i>	0.72*	0.64*	0.95
<i>tbl</i>	-0.16	-0.91	2.29*
<i>Mean combination with two predictors</i>			
<i>dy + tbl</i>	0.79**	0.46	1.84***
<i>Ex-ante timing strategy using the forecasted recession dummy</i>			
<i>dy + tbl</i>	1.24**	0.24	4.49***
<i>Ex-post timing strategy using the true recession dummy</i>			
<i>dy + tbl</i>	1.03**	0.64*	2.29*
<i>Standard forecast combinations: all 14 predictors</i>			
<i>Mean</i>	-0.01	0.09	-0.33
<i>MSE</i>	0.21	0.31	-0.14

**Table 6. Portfolio Performance: Timing the Business Cycle**

The table shows the out-of-sample portfolio performance for a mean-variance investor, who each month rebalances her portfolio by investing in one risky asset (S&P 500) and the riskless rate (T-bill). The investor has a degree of relative risk aversion equal to 5 and follows a maximum utility strategy.  $\Delta CER$  is the gain in the percent annualized Certainty Equivalent Return ( $CER$ ) for switching from the forecasts of the benchmark to the forecasts generated by the alternative model.  $SR$  is the annualized Sharpe ratio.  $Turnover$  is the ratio of the average turnover of the portfolio generated by the forecasts of the alternative model divided by the average turnover of the portfolio generated by the benchmark. For the historical mean benchmark, we report the level of  $CER$  and the average turnover for that portfolio. The superscripts \*, \*\*, and \*\*\* denote statistical significance at the 10%, 5%, and 1% level, respectively, using the Ledoit and Wolf (2008) bootstrap two-sided test of whether the Sharpe ratio of a model is different from that of the benchmark. The last column is the annual percent  $\Delta CER$  assuming a proportional transaction cost of 50 basis points per month. The out-of-sample monthly forecasts are obtained using a 20-year rolling window for the sample period of August 1967 to December 2015.

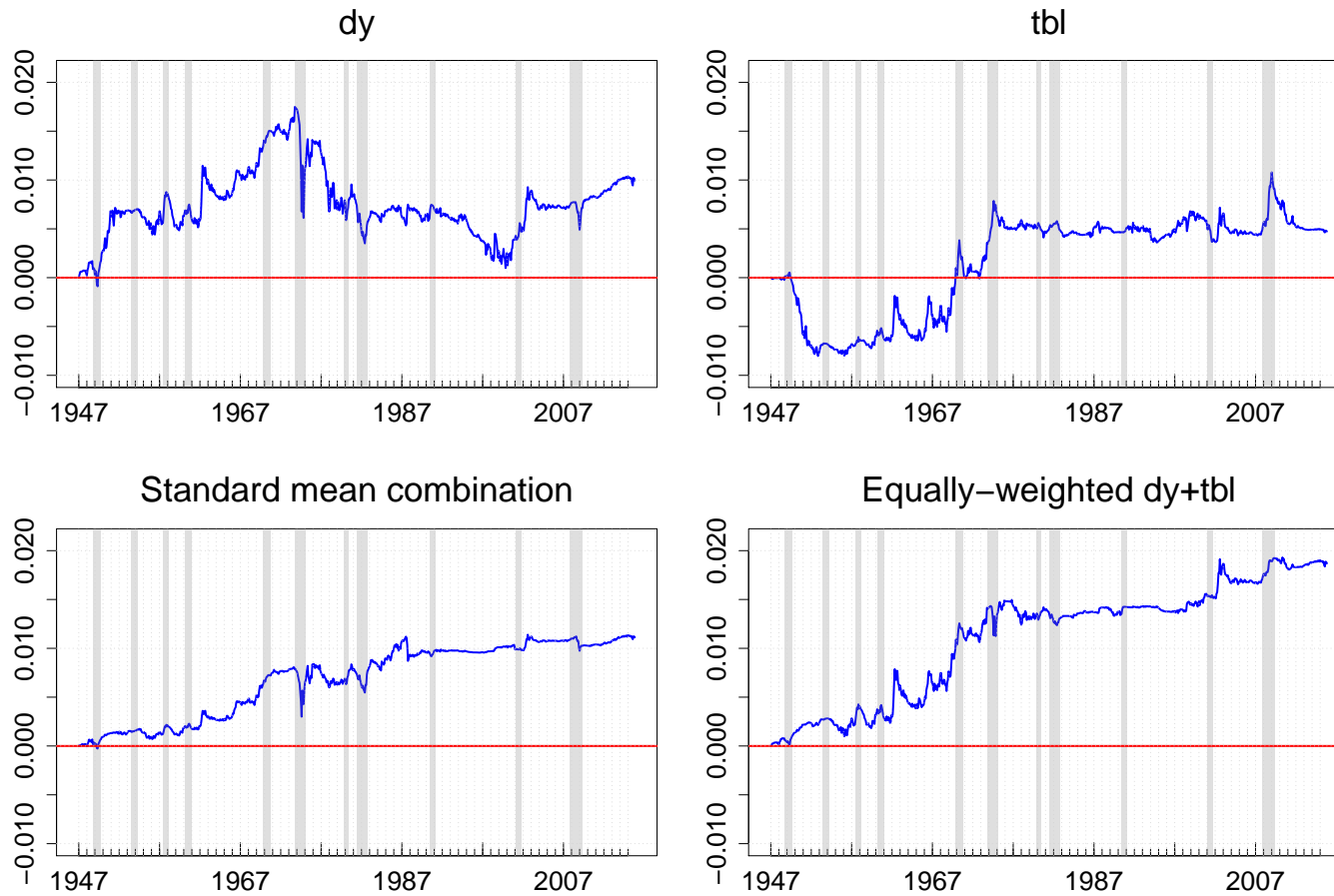
<i>Predictor</i>	$\Delta CER$ (%)			$SR$	$Turnover$	$\Delta CER$ (%)
	<i>Full Sample</i>	<i>Expansion</i>	<i>Recession</i>	<i>Full Sample</i>	<i>Full Sample</i>	<i>c = 50 bps Full Sample</i>
<i>Historical mean</i>	3.24	7.04	-26.63	0.31	0.03	3.07
<i>Two individual predictors</i>						
<i>dy</i>	0.19	0.71	-4.38	0.28	2.37	-0.07
<i>tbl</i>	1.66	-0.71	20.10	0.44	1.68	1.56
<i>Mean combination with two predictors</i>						
<i>dy + tbl</i>	1.43	0.47	8.82	0.40	1.65	1.29
<i>Ex-ante timing strategy using the forecasted recession dummy</i>						
<i>dy + tbl</i>	2.92	0.33	23.10	0.57	2.40	2.65
<i>Ex-post timing strategy using the true recession dummy</i>						
<i>dy + tbl</i>	2.93	0.71	20.10	0.57	2.48	2.64
<i>Standard forecast combinations: all 14 predictors</i>						
<i>Mean</i>	-0.61	-0.07	-4.78	0.29	1.47	-0.70
<i>MSE</i>	-0.23	0.28	-4.17	0.31	1.77	-0.37





**Figure 1. Out-of-sample slope estimates over time**

The figure displays the out-of-sample slope estimates for each of the 14 individual equity premium predictors. The slopes are constrained to have the sign implied by economic theory. This is based on rolling regressions using a window of 20 years for the sample period of January 1927 to December 2015. The grey areas indicate NBER-defined recessions.



**Figure 2. Out-of-Sample predictive performance over time**

The figure displays the performance of selected empirical models in out-of-sample equity premium prediction. Each graph shows the difference of the cumulative squared error of the null (historical mean) minus the cumulative squared error of the alternative. The grey areas indicate NBER-defined recessions. The out-of-sample monthly forecasts are obtained with rolling regressions using a window of 20 years for the sample period of January 1927 to December 2015.