ELSEVIER

Equity premium prediction and the state of the economy☆

Ilias Tsiakas [a,*], Jiahan Li [b], Haibin Zhang [a]

[a] *University of Guelph, Canada*
[b] *GMO LLC, United States of America*

ARTICLE INFO

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. Accordingly, a simple forecast combination of one predictor that generates cyclical forecasts and one predictor that generates countercyclical forecasts can deliver statistically significant and economically valuable equity premium predictions in both expansions and recessions. A prominent two-predictor forecast combination that performs well is the dividend yield and the short rate. Strategies designed for ex-ante timing of the business cycle can provide additional economic gains in equity premium prediction.

1. Introduction

Does the state of the economy matter for equity premium prediction? According to a recent study by Kacperczyk et al. (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 significantly lower and, consequently, investors care more about idiosyncratic shocks. This framework implies that the predictive information of economic fundamentals may be linked to the observed state of the business cycle.

In this paper, we find that the state of the economy does matter for equity premium prediction. The empirical evidence indicates that certain economic fundamentals (notably the dividend yield) generate equity premium forecasts that perform well in good times, whereas other economic fundamentals (notably the short rate) generate forecasts that perform well in bad times. Motivated by this finding, we propose a simple way to predict the equity premium out of sample: form an equally weighted forecast combination of one predictor that generates cyclical forecasts and one predictor that generates countercyclical forecasts. For example, we can form a simple forecast combination of the dividend yield and the short rate. We show that this approach has solid theoretical foundations and achieves two desirable properties: diversification, which implies that the combined forecast performs better than any individual forecast; and insurance, which implies that the combined forecast performs well in every state of the world.

Our main analysis focuses on the dividend yield and the short rate but we also consider other two-predictor forecast combinations. The dividend yield and the short rate are perhaps the two most prominent predictors for stock returns in the literature. They

---

are also theoretically motivated by the present value framework of Campbell and Shiller (1988). The dividend yield is a measure of idiosyncratic information because it is based on stock-specific dividends and prices. Therefore, consistent with the Kacperczyk et al. (2016) framework, 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. It is no surprise, therefore, that our analysis provides strong empirical support for the cyclical predictive ability of the dividend yield and the countercyclical predictive ability of the short rate. This, in turn, motivates the equally weighted forecast combination of the two predictors as a simple, yet powerful, out-of-sample approach for equity premium prediction.

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 2017. The Welch and Goyal (2008) data set focuses on the excess return to the S&P 500 index. All predictors are US variables used to predict the S&P 500 index excess returns. 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 main empirical finding is that the equally weighted forecast combination of the dividend yield and the short rate delivers a statistically significant monthly out-of-sample R^2 of 1.5%. This is considerably higher than the value of: (1) the R^2 of the best individual model; and (2) the standard forecast combination based on all 14 predictors. Therefore, the simple forecast combination achieves diversification. More importantly, the equally weighted forecast combination of the dividend yield and the short rate delivers good performance in both expansions ($R^2 = 1.6\%$) and recessions ($R^2 = 1.3\%$). Therefore, the simple forecast combination also achieves insurance, a property that is not achieved by any of the individual models. Finally, it also delivers high economic value in the context of a dynamic mean–variance strategy: the certainty equivalent return is about 2.1% per year over and above the historical mean benchmark for the full sample, 1.8% per year in expansions, and 3.6% 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 choice of equal weights on the two predictors is not only a simple choice that works well, but it is also an optimal choice. This is because, for the data set we employ, the optimal weights that maximize the diversification gains of this forecast combination are close to equal weights.

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) the forecasts generated by one of the two predictors are 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.

It is important to emphasize that a combination of the dividend yield and the short rate is not uniquely powerful. There are several two-predictor combinations that perform well. For example, 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 perform similarly well. Two-predictor forecast combinations based on one cyclical forecast and one countercyclical forecast have the potential to perform well. We propose the specific combination because the dividend yield and the short rate are popular predictors in the literature, they are theoretically motivated by the present value framework of Campbell and Shiller (1988), one captures systematic risk while the other one idiosyncratic risk, and they perform well in different states of the economy. Therefore, this two-predictor combination is a natural choice to be the main focus of our analysis.

In this context, we formally evaluate the ex-ante performance of several two-predictor forecast combinations. We take the point of view of an econometrician who, at a given point in time, must decide which model to use for forecasting based on the out-of-sample performance of each model up to that point. We find that the dividend yield and short rate combination is consistently one of the top combinations, and often the top combination, at every point in time during the out-of-sample period. Therefore, an econometrician who considers the ex-ante evidence up to a given date is likely to use this combination for forecasting.

Our findings provide an explanation for why standard forecast combinations of all 14 predictors typically work well. Since the contribution of Rapach et al. (2010), forecast combinations of all available predictors have become popular primarily because they are simple and work well. By naively combining the forecasts of all predictors, the standard approach includes some forecasts that work well in expansions, some that work well in recessions and some that work well in neither. We show that by excluding from the combination the forecasts that work well in neither state of the economy, in addition to those that add no further diversification gains, we can form simpler and yet superior combinations.

We also consider two formal rules for ex-ante timing of the business cycle. The first 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. The second rule is based on the Hodrick and Prescott (1980) filter for estimating the output gap. Following Colacito et al. (2019), the sign of the output gap can be used to forecast the future state of the economy.

We find that, overall, the ex-ante timing strategies perform very well, especially in recessions. The best ex-ante timing strategy is based on whether the output gap is increasing or decreasing at a given point in time. For example, whereas the simple mean combination provides a certainty equivalent return of about 2%, for the timing strategies it rises to about 3%. To conclude, we find that the ex-ante timing strategies provide a more sophisticated and better performing alternative to the simple mean combination.

In addition, we assess the cyclical variation of equity premium predictions around another cycle: the financial cycle. Following Claessens et al. (2012), we use data on aggregate credit from the Bank for International Settlements to identify the phases of the financial cycle: the recovery phase (from trough to peak) called the "upturn", and the contraction phase (from peak to trough) called the "downturn". We find that our results remain qualitatively the same for the business cycle, the financial cycle and the combination of the two cycles. The equally weighted forecast combination of the dividend yield and the short rate has essentially the same performance in either of the two cycles and delivers a positive and significant out-of-sample R^2 in both expansions/upturns and recessions/downturns. Therefore, our main finding is robust to determining the state of the economy around either the business cycle or the financial cycle.

Finally, we report international evidence for four additional countries: Canada, UK, Germany and Japan. For these countries the sample period is substantially shorter than for the US. We find that for three of the countries (Canada, UK and Japan) our main result holds: the dividend yield is a powerful predictor in expansions, whereas the short rate is a powerful predictor in recessions. However, the performance of the simple mean combination is not as strong as in the US. This is because the dividend yield and the short rate perform very well in one state of the economy but very poorly in other state of the economy. For this reason, the ex-post and ex-ante timing strategies perform better than the simple mean combination in the international context.

Our analysis adds to the well-known theoretical and empirical finding that stock return predictability concentrates in bad times. The intuition behind this finding is primarily based on countercyclical risk premiums (e.g., Campbell and Cochrane, 1999; Menzly et al., 2004; Bekaert et al., 2009) and, more recently, on countercyclical investor disagreement about the state of the economy (Cujean and Hasler, 2017). The empirical evidence on this finding is reported in Henkel et al. (2011), who use a regime-switching vector autoregression framework to show that standard predictors of the equity premium are effective primarily in recessions. Our analysis extends the Henkel et al. (2011) findings by focusing on out-of-sample prediction to show that there are both cyclical and countercyclical forecasts so that a simple combination of the two delivers better predictions in good and bad times.

Overall, this paper contributes to a long line of research on the out-of-sample predictability of the equity premium. Rapach and Zhou (2013) review this research and propose the following groupings for the forecasting models that deliver statistically and economically significant out-of-sample gains: (1) economically motivated model restrictions (e.g., Campbell and Thompson, 2008; Ferreira and Santa-Clara, 2011; Pettenuzzo et al., 2014; Li and Tsiakas, 2017); (2) forecast combination (e.g., Rapach et al., 2010); (3) diffusion indices (e.g., Ludvigson and Ng, 2007; Kelly and Pruitt, 2013; Neely et al., 2014); and (4) regime shifts and structural breaks (e.g., Paye and Timmermann, 2006; Guidolin and Timmermann, 2007; Lettau and Van Nieuwerburgh, 2008; Henkel et al., 2011; Dangl and Halling, 2012). In this setting, the contribution of our paper is best described as a refinement of a forecast combination with economically motivated model restrictions.

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 provide an empirical analysis of the ex-ante selection of several two-predictor forecast combinations. The dynamic mean–variance strategy for evaluating the portfolio performance of equity premium predictability is described in Section 6. We evaluate the performance of strategies timing the business cycle in Section 7. In Section 8, we evaluate the performance of the models around financial cycles and in Section 9 we report some international evidence. Finally, we conclude in Section 10.

2. Theoretical foundations

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

2.1. Motivating the choice of predictors

We use the present value framework of Campbell and Shiller (1988) 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 or the short rate by the long rate.

Campbell and Shiller (1988) show that the following difference equation holds approximately:

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

where dp_t is the log dividend–price ratio, k is a known constant, $\rho < 1$ is also a known constant, r_{t+1} is the one-period stock return including dividends, and Δd_{t+1} is the log dividend growth rate.

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

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

This equation relates the dividend–price ratio (or equivalently the dividend yield) with future returns, and hence motivates the dividend yield as a predictor of the equity premium.[1]

---

[1] See also Cochrane (2008, 2011) on the prominent role of the dividend yield in stock return predictability.

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

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

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

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

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

This equation suggests that the long rate can also be a relevant predictor of future excess returns.

Finally, we return to Eq. (1) and add the log earnings growth rate Δe_{t+1} to both sides. Solving forward, we can then show that the following equation holds for the log earnings–price ratio, de_t:

$$de_t - r_{f,t} = -\frac{k}{1-\rho} + E_t \sum_{j=0}^{\infty} \rho^j \left(r_{t+j+1}^e - \Delta e_{t+j+1} - (1-\rho) de_{t+j+1} \right) + E_t \sum_{j=1}^{\infty} \rho^j r_{f,t+j}. \tag{5}$$

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

In summary, Eq. (2) shows that the current dividend–price ratio (or equivalently the dividend yield) is related to future returns. Given our focus on the equity premium, which is defined as an excess return, Eq. (3) suggests that the current short rate is also related to future excess returns. Therefore, the dividend yield and the short rate are directly motivated by the present value framework of Campbell and Shiller (1988). In addition, Eqs. (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.[2] Consequently, there are alternative two-predictor combinations motivated by this framework and the choice of which one to use remains an empirical issue to be examined later.

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: (1) returns are unexpectedly low, (2) returns are more volatile, and (3) 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 (−5.7% annually) and its volatility is high (30% annually).

More importantly, Kacperczyk et al. (2016) show that the state of the economy changes the way investors process information. They propose a theoretical model, which describes how the optimal attention allocation of investors depends on the business cycle. They also provide empirical evidence, which strongly indicates that aggregate risk is significantly higher in recessions. As a result, skilled investors (e.g., mutual funds) care more about aggregate shocks in recessions, which is revealed by a higher covariance between portfolio weights and aggregate shocks. In other words, skilled investors have better timing ability in recessions.

The empirical evidence of Kacperczyk et al. (2016) also 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, which is revealed by a higher covariance between portfolio weights and idiosyncratic shocks. In other words, skilled investors have better picking ability in expansions.

The intuition behind these empirical findings is straightforward. Information is most valuable when uncertainty is higher. In recessions, when aggregate risk is higher, 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 towards idiosyncratic risk.

Overall, this framework links information choices (aggregate vs. idiosyncratic) to the observed state of the business cycle and motivates our findings in the following way: interest rates (such as the short rate) are economy-wide variables that reflect aggregate risk and hence they are more informative in recessions; in contrast, variables such as the dividend yield, the dividend–price ratio and the earnings–price ratio reflect idiosyncratic risk, since they are based on stock-specific dividends, earnings and prices, and hence they are 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 diversification gains: the return variance of the portfolio is lower than the average return variance of the two assets. In this section, we show that this diversification result has a direct analogue in forecasting. A forecast combination of two negatively correlated forecasts can be thought of as a portfolio of forecasts that also delivers diversification gains: the variance of the forecast error of the combination is lower than the average variance of the forecast error of the individual models.

---

[2] The derivation of Eqs. (1)–(5) is straightforward using the approach of Campbell and Shiller (1988). Further details on these derivations can be found in Maio (2013).

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

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

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, α and β are estimated with ordinary least squares (OLS), and ε_{t+1} is a normal error term.

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

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

2.3.1. The case of equal weights

Consider two models that generate forecasts $\hat{r}_{1,t+1}$ and $\hat{r}_{2,t+1}$. We form an equally weighted combined forecast: $\hat{r}_{c,t+1} = \frac{1}{2}\hat{r}_{1,t+1} + \frac{1}{2}\hat{r}_{2,t+1}$. Suppose that the correlation ρ_{12} between the forecast errors of the two models is less than one. Then, it is straightforward to show that the diversification condition holds:

$$\sigma_c < \frac{1}{2}\sigma_1 + \frac{1}{2}\sigma_2.\tag{8}$$

Eq. (8) states that the MSE of the combined forecasts (σ_c) is lower than the weighted average of the individual MSEs (σ_1, σ_2). This condition is directly analogous to standard portfolio diversification.[3]

For forecasting purposes, it is of greater interest if ρ_{12} is "low enough" so that the "strong" diversification condition holds:

$$\sigma_c < \min(\sigma_1, \sigma_2).\tag{9}$$

If this condition holds, then the forecast combination will produce superior predictive ability compared to any of the two models individually. To describe the conditions under which the strong diversification holds, we turn to the more general case of optimal weights.

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 that minimize the objective function:

$$\sigma_c^2 = w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1 w_2 \sigma_1 \sigma_2 \rho_{12}.\tag{10}$$

Solving for the optimal weights gives:

$$w_1^* = \frac{\sigma_2^2 - \sigma_1\sigma_2\rho_{12}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}},\tag{11}$$

$$w_2^* = \frac{\sigma_1^2 - \sigma_1\sigma_2\rho_{12}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}}.\tag{12}$$

Inserting the optimal weights into the objective function (10), we get the MSE associated with the optimal weights:

$$\sigma_c^{2*} = \frac{\sigma_1^2 \sigma_2^2 \left(1 - \rho_{12}^2\right)}{\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}}.\tag{13}$$

In this framework, Timmermann (2006) shows that the following results hold: if $\sigma_1 = \sigma_2 = \sigma$ and $\rho_{12} < 1$, then equal weights are optimal and the strong diversification condition holds. Moreover, we attain *maximum* diversification gains, that is the largest reduction in σ_c^{2*}/σ for a given change in ρ_{12}.[4] These results are crucially important for our empirical exercise. By carefully selecting two individual forecasts with similar performance (i.e., $\sigma_1 \approx \sigma_2$), where each performs better in a different state of the world (i.e., $\rho_{12} < 1$), then the equally weighted combination can deliver maximum diversification gains and superior predictive performance. In the following sections, our empirical analysis shows that this is indeed the case for the equally weighted forecast combination of two predictors such as the dividend yield and the short rate.

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 2017.[5] 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 et al. (2014) extended to 2017.

---

[3] To see this consider $\rho_{12} = 1$. Then, $\sigma_c^2 = \frac{1}{4}\sigma_1^2 + \frac{1}{4}\sigma_2^2 + \frac{1}{2}\sigma_1\sigma_2 = \frac{1}{4}\left(\sigma_1 + \sigma_2\right)^2$. 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.

[4] Note that in the extreme case where $\sigma_1 = \sigma_2$ and $\rho_{12} = -1$, $\sigma_c^{2*} = 0$, which implies perfect forecasting ability for the equally weighted combination.

[5] Throughout the analysis, we fix both the frequency and the horizon to be monthly for three reasons: (i) monthly predictability is notoriously difficult to achieve; (ii) given that recessions are less than 20% of the sample, monthly data provides the most recession observations; and (iii) long horizons complicate the analysis because for a long horizon some months in the horizon may be in recession and some in expansion.

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 period following 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 7.9%, volatility of 19%, and a Sharpe ratio of 0.42. For expansions, the annualized Sharpe ratio of the equity premium rises to 0.71 and for recessions it falls to −0.20. Recessions correspond to about 18% of the full sample and are known ex post.

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 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-rated 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 range from −0.466 to 0.993 , with an average value of 0.095.

4. Models for predicting the equity premium

4.1. Our predictive framework

We generate forecasts for the equity premium by OLS estimation of the predictive regression in Eq. (6). Our analysis is based on conditioning on a single predictor every time we estimate the predictive regression. Hence, we report results from 14 regressions each conditioning on one of the Welch and Goyal (2008) predictors. We then form an equally weighted combination of just the *dy* and *tbl* forecasts, which is the focal point of our analysis. We refer to this combined forecast as the "simple mean combination."

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 by Pettenuzzo et al. (2014).

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. Standard 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 et al. (2010), we compute the equally weighted average of all forecasts at each point in time. We refer to this as the "standard mean combination" to be distinguished from the "simple mean combination" described above.[6]

---

[6] We have also considered other combinations of the 14 forecasts, such as the combination based on the most recent MSE of each model. However, the performance of the additional standard combinations is very similar to the standard mean combination. For this reason, these results remain unreported but available upon request.

4.3. 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 Eq. (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.

4.4. 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 2017. 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^2_{oos} statistic. The R^2_{oos} compares the unconditional one-month ahead forecasts $\bar{r}^e_{t+1|t}$ of the historical mean benchmark to the conditional forecasts $\hat{r}^e_{t+1|t}$ of the alternative model, and is defined as follows:

$$R^2_{oos} = 1 - \frac{MSE\left(\hat{r}^e_{t+1|t}\right)}{MSE\left(\bar{r}^e_{t+1|t}\right)} = 1 - \frac{\sum_{t=1}^{T-1}\left(r^e_{t+1} - \hat{r}^e_{t+1|t}\right)^2}{\sum_{t=1}^{T-1}\left(r^e_{t+1} - \bar{r}^e_{t+1|t}\right)^2}. \tag{14}$$

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

We assess the statistical significance of the R^2_{oos} 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:

$$MSE_{adj} = \frac{1}{T-1}\sum_{t=1}^{T-1}(r^e_{t+1} - \hat{r}^e_{t+1|t})^2 - \frac{1}{T-1}\sum_{t=1}^{T-1}(\bar{r}^e_{t+1|t} - \hat{r}^e_{t+1|t})^2. \tag{15}$$

Then, we define:

$$\widehat{test}_{t+1} = (r^e_{t+1} - \bar{r}^e_{t+1|t})^2 - [(r^e_{t+1} - \hat{r}^e_{t+1|t})^2 - (\bar{r}^e_{t+1|t} - \hat{r}^e_{t+1|t})^2], \tag{16}$$

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).

4.5. Empirical results

4.5.1. 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 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^2_{oos}. In the case of no economic constraints, the only two predictors that exhibit a positive and statistically significant R^2_{oos} are the dividend yield (*dy*) with an R^2_{oos} of 0.53% and the dividend–price ratio (*dpr*) with an R^2_{oos} of 0.26%.[7] The other 12 predictors display a negative R^2_{oos}. For example, the short rate (*tbl*) exhibits an R^2_{oos} of -1.64%. In the absence of constraints, therefore, only *dy* and *dpr* beat the historical mean, while all other predictors appear to be devoid of any predictive information over and above the benchmark.

This finding changes considerably with the imposition of the two economic constraints as they universally improve performance: the R^2_{oos} rises for 13 of the 14 predictors, the single exception being the dividend payout ratio (*dpayr*). Under the constraints, seven predictors exhibit a positive and significant R^2_{oos}. Notably, the R^2_{oos} of the dividend yield rises to 1.07%, and is significant at 1%, whereas the R^2_{oos} of the short rate rises to a positive value (0.46%) 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 of our analysis imposes 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 performs 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 positive and statistically

---

[7] These two predictors (*dy* and *dpr*) are virtually the same by definition and their correlation exceeds 99%. Our discussion will therefore focus on *dy* as it is more popular in the literature than *dpr*.

Table 1

Equity premium prediction: Individual predictors. The table displays the out-of-sample R^2_{oos} in percent for predictive models of the monthly equity premium against the null of the historical mean. In the first three columns, there are no constraints on slope coefficients and forecasts. In the last three columns, the R^2_{oos} 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 2017. 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.

| | R^2_{oos} (%) | | | | | |
| | No Economic Constraint | | | Both Economic Constraints | | |
Predictor	Full Sample	Expansion	Recession	Full Sample	Expansion	Recession
dy	0.53***	0.93***	−0.60	1.07***	1.68***	−0.62
dpr	0.26**	0.89***	−1.52	0.79***	1.61***	−1.51
epr	−1.49	0.17**	−6.17	0.15**	0.68**	−1.35
dpayr	−1.08	−0.93	−1.51	−11.17	−10.13	−14.07
bm	−1.32	−0.41	−3.88	−0.03	−0.03	−0.02
ntis	−1.31	−1.76	−0.03	−0.43	−0.39	−0.54
svar	−3.23	−3.11	−3.60	−2.43	−2.57	−2.03
tbl	−1.64	−2.60	1.04	0.46**	−0.01	1.80**
lty	−0.44	−0.76	0.46	0.35**	−0.33	2.27**
ltr	−0.10	−1.89	4.90***	0.46**	−0.71	3.71***
tms	−0.57	−1.69	2.58**	0.12***	−0.72	2.48**
dfy	−1.85	−2.27	−0.68	−1.64	−1.09	−3.17
dfr	−1.46	−0.48	−4.23	−0.81	−0.24	−2.43
infl	−0.35	−1.14	1.88	−0.34	−0.66	0.54

Table 2

Equity premium prediction: Simple forecast combinations. The table reports the out-of-sample R^2_{oos} in percent for predictive models of the monthly equity premium against the null of the historical mean. The R^2_{oos} 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 2017. 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.

| | R^2_{oos} (%) | | |
Predictor	Full Sample	Expansion	Recession
Two individual predictors			
dy	1.07***	1.68***	−0.62
tbl	0.46**	−0.01	1.80**
Simple mean combination with two predictors			
dy + tbl	1.54***	1.63***	1.28*
Ex-post timing using the true recession dummy			
dy + tbl	1.71***	1.68***	1.80**
Standard forecast combination			
all 14 predictors	0.83***	0.99***	0.39

significant R^2_{oos}. In recessions, predictors associated with interest rates, such as the short rate, the long rate and the term spread, also display a positive and statistically significant R^2_{oos}. 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 et al. (2016) on the cyclical variation of investor attention allocation.

4.5.2. The simple mean combination

Next we turn to the main model of our analysis: the equally weighted forecast combination of the dividend yield and the short rate. As mentioned earlier, we refer to this as the "simple mean combination." Indeed, this is the simplest possible combination we can form: two predictors with equal weights, where one predictor performs well in expansions and one in recessions. In what follows, we will show that this is a powerful forecast combination that is hard to beat.

To quantify the diversification gains of the simple mean combination consider the results of Table 2. The dividend yield alone delivers an R^2_{oos} of 1.07% and the short rate alone an R^2_{oos} of 0.46%. An equally weighted combination of the two delivers an R^2_{oos} of 1.54%, which is considerably higher than the R^2_{oos} of the better model. This implies that the MSE of the simple mean combination is much lower than the lowest MSE of the two predictors. Hence the simple mean combination satisfies the strong diversification condition of Eq. (9).

Moreover, the simple mean combination provides insurance since it delivers good performance in both expansions ($R^2_{oos} = 1.63\%$) and recessions ($R^2_{oos} = 1.28\%$), where in both cases the R^2_{oos} is statistically significant. In conclusion, therefore, the simple mean combination of the forecasts generated by the dividend yield and the short rate delivers strong diversification gains and insurance.

Table 3

Equity premium prediction: More forecast combinations. The table displays the out-of-sample R^2_{oos} in percent for predictive models of the monthly equity premium against the null of the historical mean. The formation of forecasts is as described in Table 2. The sample period ranges from January 1927 to December 2017. 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.

Predictor	R^2_{oos} (%)		
	Full Sample	Expansion	Recession
Mean combinations with two predictors			
dy + tbl	1.54***	1.63***	1.28*
dy + tms	1.54***	1.57***	1.47*
dy + lty	1.40***	1.41***	1.39*
dy + ltr	1.51***	1.29***	2.13*
dpr + tbl	1.30***	1.51***	0.72
dpr + tms	1.37***	1.52***	0.95
dpr + lty	1.19***	1.30***	0.87
dpr + ltr	1.34***	1.21***	1.68*
epr + tbl	1.02***	1.03***	0.99*
epr + tms	1.23***	1.16***	1.43**
epr + lty	0.78***	0.73**	0.93*
epr + ltr	1.05***	0.64**	2.19***
Mean combinations with three predictors			
dy + tbl + dpr	1.43***	1.77***	0.47
dy + tbl + epr	1.41***	1.63***	0.81*
dy + tbl + tms	1.62***	1.53***	1.86**
dy + tbl + lty	1.36***	1.23***	1.74**
dy + tbl + ltr	1.66***	1.35***	2.53***
Mean combination for typical idiosyncratic predictors			
dy + epr	1.04***	1.59***	−0.49
Mean combination for typical systematic predictors			
tbl + tms + ltr	1.33***	0.61**	3.35***
Mean combination for idiosyncratic + systematic predictors			
dy + epr + tbl + tms + ltr	1.68***	1.48***	2.24***

The simple mean combination has two features that make it simple and attractive: equal weights and using only two predictors. We will now explain why both of these features make sense. In addition to being a simple choice, setting equal weights also turns out to be an optimal choice: when computing the optimal diversification weights using Eqs. (11) and (12), the result is a weight of 54% on the dividend yield and 46% on the short rate. Therefore, from the point of view of diversification, using equal weights is practically as good as it gets. It also means that the simple mean combination achieves maximum diversification gains for two reasons: (1) $\sigma_1 \approx \sigma_2$, i.e., the similar MSE values of the two models; and (2) $\rho_{12} < 1$ since one set of forecasts is cyclical and one countercyclical. Consequently, the simple mean combination achieves the maximum diversification gains defined by Timmermann (2006) as the largest reduction in σ_c^{2*}/σ for a given change in ρ_{12}.

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 forecasts does not add much to the performance of the initial combination. 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 forecasts does not improve much the performance of the initial combination.

This finding is illustrated in Table 3, which reports on the performance of more forecast combinations. The table shows that the performance of the simple mean combination remains almost the same when: (1) we replace *dy* by *dpr* or *epr*, (2) we replace *tbl* by *tms*, *lty* or *ltr*; and (3) we add with equal weights to the *dy + tbl* combination any one of the following: *dpr*, *epr*, *tms*, *lty* or *ltr*. In all these cases, the results are similar in the sense that all these models deliver a positive and significant R^2_{oos}, which remains positive and significant in both expansions and recessions.

These results imply that the simple mean 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 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.[8]

It is also important to note that the simple mean combination of *dy* and *tbl* significantly outperforms the standard forecast combinations of all 14 predictors. Since Rapach et al. (2010) forecast combinations of all available predictors have become popular,

---

[8] We also report results on typical combinations of idiosyncratic and systematic predictors. As expected, larger groups of idiosyncratic predictors perform well in expansions, whereas larger groups of systematic predictors perform well in recessions. This further confirms the cyclical performance of the two types of predictors.
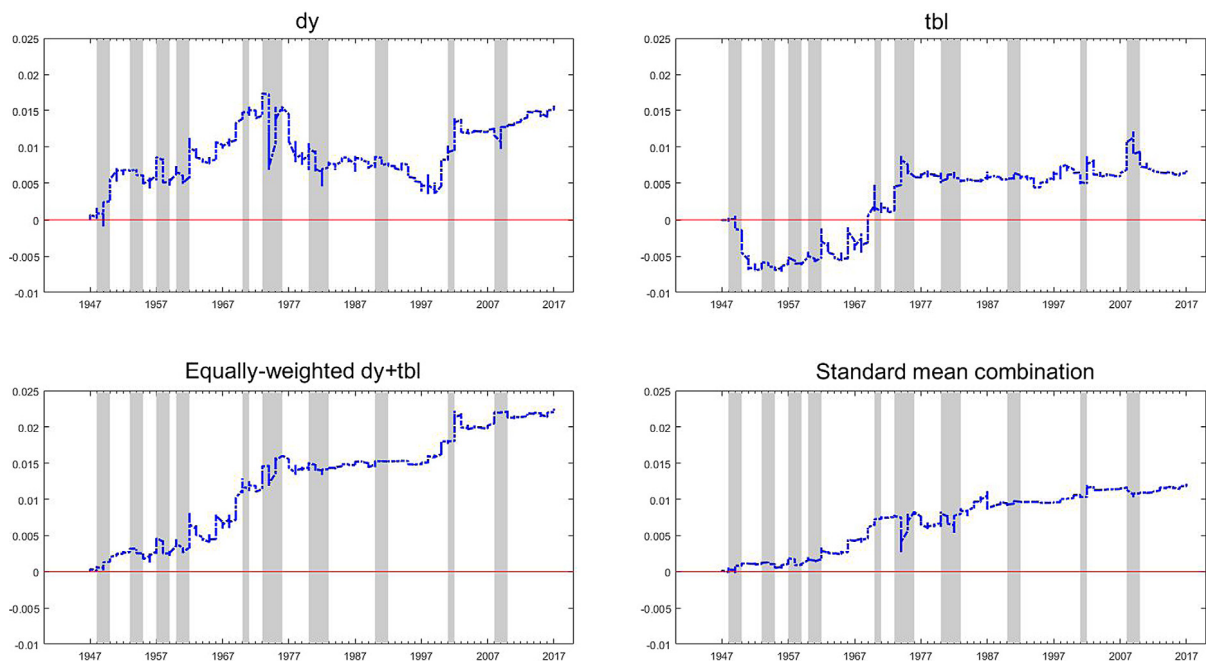
Fig. 1. Out-of-sample equity premium prediction. The figure displays the performance of selected empirical models in out-of-sample equity premium prediction. Each graph shows the difference of the cumulative squared error of the null (historical mean) minus the cumulative squared error of the alternative. 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 2017.

partly due to their simplicity. We show here that simpler combinations can perform much better if we carefully select the forecasts to combine in a way that achieves diversification gains. Indeed, our simple approach not only outperforms the standard combination but also justifies why the standard combination works: by naively combining all predictors, we effectively 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, in addition to the ones that offer no further diversification gains, we can form a simpler, yet superior, combination.

To illustrate these results, Fig. 1 plots 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 the 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 mean combination outperforms the standard mean forecast combination based on all 14 predictors. Finally, the simple mean combination strongly outperforms its two individual components, the dividend yield and the short rate.[9]

4.5.3. The ex-post timing strategy

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. This strategy, however, is informative for the following reason. By design, the ex-post timing strategy can perfectly time which forecasts to use: it uses the dividend yield forecasts in expansions and the short rate forecasts in recessions. Hence there is no averaging in the two sets of forecasts at a given point in time. This generates a profound difference between the simple mean combination and the timing strategy: the former maximizes the diversification gains, whereas the latter maximizes the insurance gains. Therefore, a comparison of the two strategies informs us of how close the simple mean combination comes to achieving the maximum insurance gains.

We find that the R^2_{oos} of the two strategies is very similar: the ex-post timing strategy ($R^2_{oos} = 1.71\%$) is only marginally better than the simple mean combination ($R^2_{oos} = 1.54\%$). The two strategies have almost identical performance in expansions, while the timing strategy performs slightly better in recessions. This is to be expected since there is more predictability in recessions and so perfectly timing the business cycle delivers better performance in recessions. However, the main result here is that the simple mean combination, which is a fully ex-ante strategy, is impressively close to providing the maximum insurance gains in addition to the maximum diversification gains discussed previously.

---

[9] In addition, we have also performed estimation using the adjustment proposed by Kostakis et al. (2015) for persistent regressors. However, this procedure is not designed for forecast combinations based on out-of-sample estimation with economic constraints. Instead this procedure can be used for in-sample estimation using individual predictors without economic constraints. Therefore, we do not report these results but they are available upon request.

Table 4

Two-predictor forecast combinations. The table displays the percent out-of-sample R^2_{oos} and the model rank for two-predictor forecast combinations. The formation of forecasts is as described in Table 2 using the initial in-sample period of January 1927 to December 1946. The training period is the initial out-of-sample period over which models are evaluated. All training periods begin with the first out-of-sample forecast in January 1947. The forecasting period is the final out-of-sample period over which models are implemented. All forecasting periods end with the last out-of-sample forecast in December 2017. Panel A reports the percent out-of-sample R^2_{oos}. Panel B reports the rank of each model based on its R^2_{oos} relative to all other models. The average rank is the rank across all training periods or all forecasting periods or both. The table suppresses evidence of statistical significance to conserve space.

Panel A: R^2_{oos} (%)													
		dy +	dy +	dy +	dy +	dpr +	dpr +	dpr +	dpr +	epr +	epr +	epr +	epr +	ALL 14
Training period														
Begin	End	tbl	tms	lty	ltr	tbl	tms	lty	ltr	tbl	tms	lty	ltr	predictors
194701	196612	2.54	2.44	2.09	2.60	1.93	1.95	1.60	2.34	1.91	1.61	0.91	1.59	1.45
194701	197612	2.85	2.93	2.49	2.96	2.31	2.53	2.02	2.66	1.57	1.78	0.92	1.57	1.46
194701	198612	1.89	2.05	1.60	2.14	1.53	1.81	1.28	1.91	0.91	1.41	0.41	1.22	1.27
194701	199612	1.49	1.47	1.14	1.57	1.19	1.28	0.88	1.48	0.74	1.00	0.23	0.99	0.96
194701	200612	1.62	1.61	1.35	1.61	1.38	1.49	1.14	1.47	1.00	1.26	0.59	1.07	0.93
Forecasting period														
Begin	End													
196701	201712	1.28	1.31	1.22	1.17	1.13	1.22	1.08	1.07	0.78	1.13	0.75	0.90	0.67
197701	201712	0.72	0.68	0.72	0.53	0.66	0.64	0.66	0.50	0.67	0.89	0.70	0.72	0.44
198701	201712	1.14	0.96	1.18	0.78	1.04	0.86	1.08	0.67	1.14	1.02	1.22	0.84	0.33
199701	201712	1.64	1.69	1.95	1.16	1.52	1.57	1.84	1.03	1.62	1.72	1.94	1.17	0.57
200701	201712	1.14	1.04	1.70	0.90	0.87	0.77	1.45	0.62	1.12	1.07	1.77	0.91	0.32

Panel B: Rank of each model based on R^2_{oos} relative to all other models													
		dy +	dy +	dy +	dy +	dpr +	dpr +	dpr +	dpr +	epr +	epr +	epr +	epr +	ALL 14
Training period														
Begin	End	tbl	tms	lty	ltr	tbl	tms	lty	ltr	tbl	tms	lty	ltr	predictors
194701	196612	2	3	5	1	7	6	10	4	8	9	13	11	12
194701	197612	3	2	6	1	7	5	8	4	10	9	13	11	12
194701	198612	4	2	6	1	7	5	9	3	12	8	13	11	10
194701	199612	2	4	7	1	6	5	11	3	12	8	13	9	10
194701	200612	1	3	7	2	6	4	9	5	11	8	13	10	12
Average rank		2.4	2.8	6.2	1.2	6.6	5	9.4	3.8	10.6	8.4	13	10.4	11.2
Rank of average rank		2	3	6	1	7	5	9	4	11	8	13	10	12
Forecasting period														
Begin	End													
196701	201712	2	1	3	5	6	4	8	9	11	7	12	10	13
197701	201712	2	6	3	11	9	10	8	12	7	1	5	4	13
198701	201712	4	8	2	11	6	9	5	12	3	7	1	10	13
199701	201712	6	5	1	11	9	8	3	12	7	4	2	10	13
200701	201712	4	7	2	9	10	11	3	12	5	6	1	8	13
Average rank		3.6	5.4	2.2	9.4	8	8.4	5.4	11.4	6.6	5	4.2	8.4	13
Rank of average rank		2	5	1	11	8	9	5	12	7	4	3	9	13
All Training and Forecasting periods														
Average rank		3	4.1	4.2	5.3	7.3	6.7	7.4	7.6	8.6	6.7	8.6	9.4	12.1
Rank of average rank		**1**	2	3	4	7	5	8	9	10	5	10	12	**13**

5. Ex-ante selection of two-predictor forecast combinations

The main thesis of this paper is that a two-predictor forecast combination is a simple yet powerful tool in out-of-sample equity premium prediction. This forecast combination works well as long as one predictor performs well in expansions and the other one performs well in recessions. In other words, this forecast combination can be thought of as a portfolio of one cyclical and one countercyclical forecast so that the forecasts generated by the two predictors are negatively correlated with each other. Furthermore, the forecasts generated by each of the two predictors are highly positively correlated with the forecasts of all other predictors that perform well in the same state of the world. Accordingly, the two-predictor forecast combination achieves diversification, which implies that the combined forecast performs better than any individual forecast, and insurance, which implies that the combined forecast performs well in every state of the world.

There are several two-predictor combinations that achieve these results. Our analysis focuses on the dividend yield and the short rate for the following reasons: these are popular predictors in the literature, they are theoretically motivated by the present value framework of Campbell and Shiller (1988), one captures systematic risk while the other one idiosyncratic risk, and they perform well in different states of the economy. Therefore, this two-predictor combination is a natural choice to be the focus of our analysis.

Having said that, however, there are other two-predictor combinations that also perform well. In this section, we provide an empirical analysis that informs the ex-ante selection of two-predictor forecast combinations. We consider a set of two-predictor combinations, where the dividend yield (dy) is replaced by the dividend–price ratio (dpr) or the earnings–price ratio (epr); and the short rate (tbl) is replaced by the term spread (tms) or the long-term yield (lty) or the long-term rate of return (ltr). This leads to 12 different forecast combinations: $dy + tbl$; $dy + tms$; $dy + lty$; $dy + ltr$; $dpr + tbl$; $dpr + tms$; $dpr + lty$; $dpr + ltr$; $epr + tbl$; $epr + tms$; $epr + lty$; $epr + ltr$. These are the same two-predictor combinations shown in Table 3.

We assess the ex-ante performance of these models from the point of view of an econometrician who, at a given point in time, must decide which model to use for forecasting based on the out-of-sample performance of each model up to that point. To do so, we divide the out-of-sample period into different training and forecasting periods. The training period defines the information set of the econometrician and always begins with the first out-of-sample forecast in January 1947. The training period is specified to include 20, 30, 40, 50 or 60 years of initial out-of-sample forecasts. The forecasting period begins right after the end of the training period and ends with the last forecast for December 2017. For example, an econometrician making a decision in January 1967 would have knowledge of the out-of-sample performance of the first training period ending in December 1966. Therefore, the econometrician would implement a forecasting period from January 1967 till December 2017.

The findings are reported in Table 4. Panel A reports the R^2_{oos} for different training and forecasting periods, whereas Panel B reports the rank of each model relative to all other models based on the R^2_{oos}. We find that, across all training periods, $dy + tbl$ is ranked second (i.e., its rank of the average rank across all training periods is 2). It is ranked first for the last and longest training period of 1947–2006. Across all training periods, the first-ranked model is $dy + ltr$. Across all forecasting periods, $dy + tbl$ is again ranked second but now the first-ranked model is different: $dy + lty$. Finally, across all training and forecasting periods, $dy + tbl$ is ranked first. In other words, $dy + tbl$ performs consistently well: it is always in the top two models, it is ranked first for the longest training period and it is ranked first across all training and forecasting periods. In short, therefore, there is strong empirical evidence that $dy + tbl$ is a good choice for a two-predictor forecast combination.

The empirical evidence reported in Table 4 fixes the length of training periods to be a multiple of 10 years beginning with 20 years of initial out-of-sample evidence. The evidence might be different for alternative definitions of the training period. A more comprehensive way of reporting the evidence for all possible training periods is by plotting the out-of-sample performance of each two-predictor model against $dy + tbl$. This is illustrated in Figs. 2–4. The figures demonstrate the information (in terms of cumulative out-of-sample performance) available to an econometrician at all points in time. Then, the econometrician considers this information at a given point in time in order to choose one the models. The figures confirm that the $dy + tbl$ combination is not uniquely powerful but it is a combination that is hard to beat at any point in time. Consistent with Table 4, the figures indicate that models using the dividend yield perform better than models using the dividend–price ratio or the earnings–price ratio. In addition, the dividend yield plus short rate combination graphically appears to be as good as or better than any other forecast combination.

In conclusion, we find that there are several two-predictor combinations that perform well in out-of-sample equity premium prediction. The dividend yield and short rate combination is consistently one of the top combinations, and often the top combination, at every point in time during the out-of-sample period. Therefore, an econometrician who considers the ex-ante evidence up to a given date is likely to use this combination for forecasting. For the reasons discussed in this section, the remainder of our analysis focuses on the dividend yield and short rate combination.

6. Predictability and asset allocation

6.1. 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 investor with a one-month ahead horizon, who determines the optimal weights by implementing a maximum expected utility rule as follows:

$$
\begin{aligned}
\max_{w_t} \quad & E_t\left[U\left(r_{p,t+1}\right)\right] & = r_{p,t+1|t} - \frac{\gamma}{2}\sigma^2_{p,t+1|t} \\
\text{s.t.} \quad & r_{p,t+1|t} & = w_t r_{t+1|t} + \left(1 - w_t\right) r_f, \\
& \sigma^2_{p,t+1|t} & = w_t^2 \sigma^2_{t+1|t},
\end{aligned}
\tag{17}
$$

where $r_{p,t+1|t}$ is the $t+1$ forecast of the portfolio return conditional on time t information, γ is the investor's degree of relative risk aversion, $\sigma^2_{p,t+1|t}$ 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 $\sigma^2_{t+1|t}$ is the $t+1$ forecast of the variance to the S&P 500 Index return made at time t. Note that we forecast $\sigma^2_{t+1|t}$ using a five-year rolling average of the variance of past monthly returns as in Campbell and Thompson (2008). Following Della Corte et al. (2009), we consider two degrees of relative risk aversion: $\gamma = 2$ and $\gamma = 6$.

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

$$
w_t = \frac{1}{\gamma} \frac{r_{t+1|t} - r_f}{\sigma^2_{t+1|t}}.
\tag{18}
$$

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, short-selling is not allowed and leverage is limited to no more than 50%.
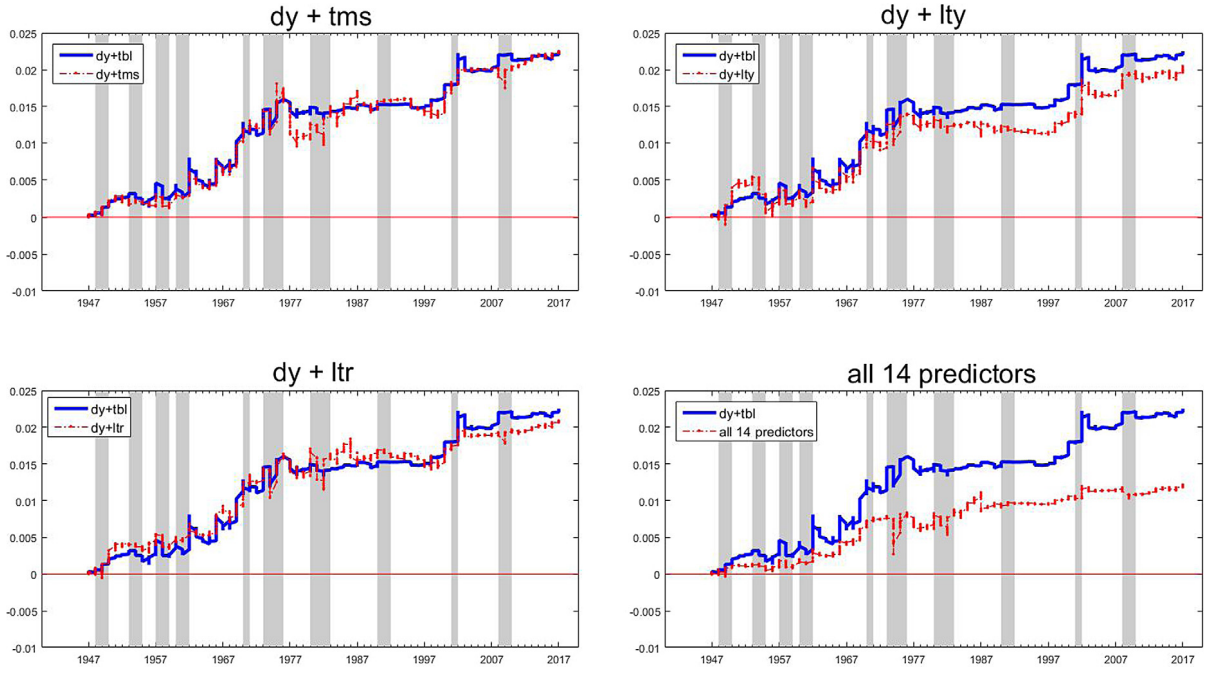
Fig. 2. Two-predictor forecast combinations based on the dividend yield. The figure displays the out-of-sample performance over time in equity premium prediction of two-predictor forecast combinations based on the dividend yield. In each graph, the solid blue line is the benchmark two-predictor model that combines the forecasts of the dividend yield (dy) and the short rate (tbl). The red dashed line is for the alternative two-predictor model. In the bottom right graph, the alternative model is the forecast combination of all 14 predictors. Each line represents the difference of the cumulative squared error of the null (historical mean) minus the cumulative squared error of the selected model. Hence a positive value indicates that the model performs better than the historical mean. 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 2017. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

We evaluate the performance of portfolios generated by a given set of equity premium forecasts using the Sharpe ratio (SR), the Economic Performance Measure (EPM) proposed by Homm and Pigorsch (2012), and the certainty equivalent return (CER). The Sharpe ratio 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. The statistical significance of the SR is assessed using the Ledoit and Wolf (2008) bootstrap two-sided test of whether the SR of the alternative model is different from the benchmark.

The Economic Performance Measure is designed to generalize the Sharpe ratio with respect to non-normal return distributions. Based on the Aumann and Serrano (2008) economic index of riskiness, the EPM is defined as follows:

$$EPM = \frac{18\bar{r}_p}{3\kappa_p \bar{r}_p - 4\bar{r}_p \chi_p^2 - 6\chi_p \bar{\sigma}_p + 9\bar{\sigma}_p^2/\bar{r}_p}, \tag{19}$$

where $\bar{r}_p$ is the mean portfolio return, $\bar{\sigma}_p^2$ is the portfolio variance over the forecast evaluation period, χ_p is the skewness of portfolio returns and κ_p is the excess kurtosis of portfolio returns. For normally distributed excess returns, $EPM = \frac{2\bar{r}_p^2}{\bar{\sigma}_p^2} = 2SR^2$. For this reason, we report $\sqrt{EPM/2}$ so that the EPM is on the same scale as the Sharpe ratio.

The certainty equivalent return is defined as:

$$CER = \left(\bar{r}_p - \frac{\gamma}{2} \bar{\sigma}_p^2 \right). \tag{20}$$

The CER can be interpreted as the performance fee a risk-averse investor is willing to pay for switching from the riskless asset to the risky portfolio. We focus on the difference in CER (ΔCER), 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. ΔCER measures the performance fee a 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. Following Marquering and Verbeek (2004), we compute ΔCER using two values for c, which is the proportional transaction cost per transaction: $c = 50$ bps reflecting medium transaction costs, and $c = 100$ bps reflecting high transaction costs (see also Balduzzi and Lynch, 1999; Neely et al., 2014).
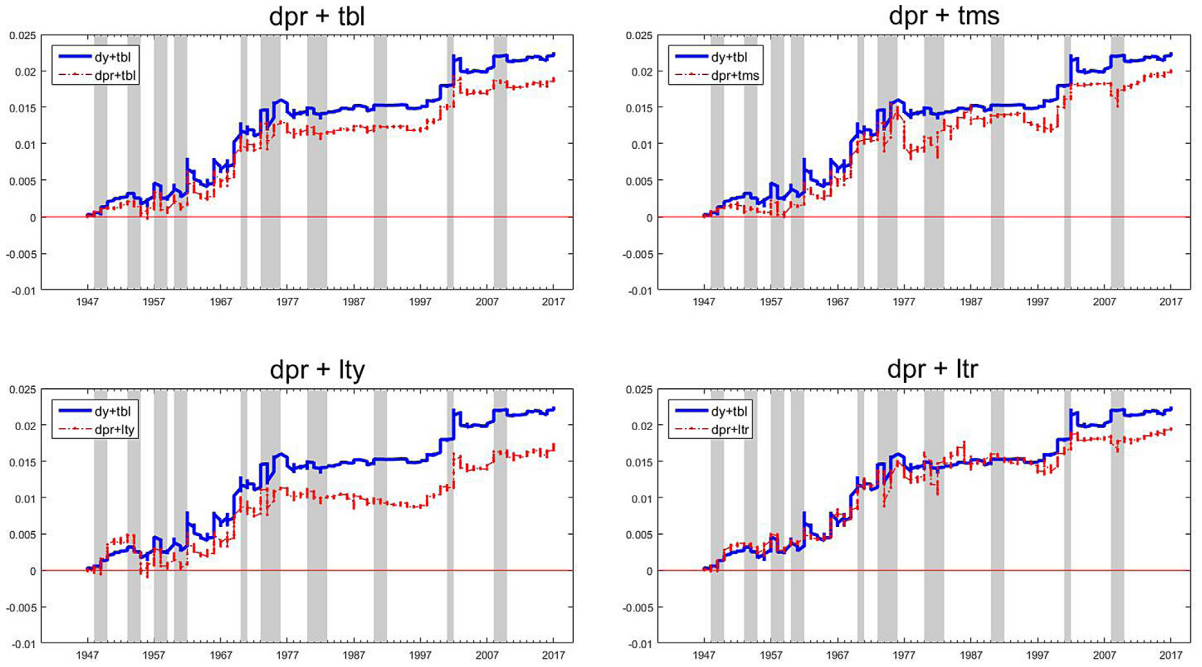
Fig. 3. Two-predictor forecast combinations based on the dividend–price ratio. The figure displays the out-of-sample performance over time in equity premium prediction of two-predictor forecast combinations based on the dividend–price ratio. In each graph, the solid blue line is the benchmark two-predictor model that combines the forecasts of the dividend yield (dy) and the short rate (tbl). The red dashed line is for the alternative two-predictor model. Each line represents the difference of the cumulative squared error of the null (historical mean) minus the cumulative squared error of the selected model. Hence a positive value indicates that the model performs better than the historical mean. 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 2017. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

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

$$TO = \frac{1}{T-1} \sum_{t=1}^{T-1} \left(|w_{t+1} - w_{t+1}^-| \right), \tag{21}$$

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.

6.2. Portfolio performance

We assess the performance of dynamically rebalanced portfolios generated by the monthly forecasts of the predictive models. Table 5 reports the empirical findings for two degrees of relative risk aversion, $\gamma = 2$ (Panel A) and $\gamma = 6$ (Panel B). Our discussion focuses on the case of $\gamma = 6$, which is the standard case considered in the literature. The results for $\gamma = 2$ are similar. The first finding to note is that the historical mean benchmark delivers a $CER = 5.34\%$ per year relative to riskless investing. The CER rises to 6.76% in expansions but falls to −3.08% in recessions. Clearly, using the historical mean is a poor predictor of the equity premium during recessions.

Similar to the statistical findings, our main result here is that the simple mean combination of the dividend yield and the short rate performs well: better than the historical mean benchmark, better than the two individual predictors, and better than the standard forecast combination of all 14 predictors. The ΔCER of the simple mean combination relative to the benchmark is 2.10% per year. More importantly, the ΔCER remains positive for both expansions (1.84%) and recessions (3.63%). Net of transaction costs, it still retains a positive ΔCER of 1.93% per year for medium transaction costs and 1.75% per year for high transaction costs. Moreover, the annualized SR is 0.64, which is significantly higher than the 0.46 of the benchmark strategy according to the Ledoit and Wolf (2008) bootstrap two-sided test. The Economic Performance Measure provides similar values and the same ranking as the Sharpe ratio.

Overall, our evidence shows that, in the context of a dynamic mean–variance strategy, a simple forecast combination of the dividend yield and the short rate has high economic value in predicting the equity premium. The economic gains of this approach
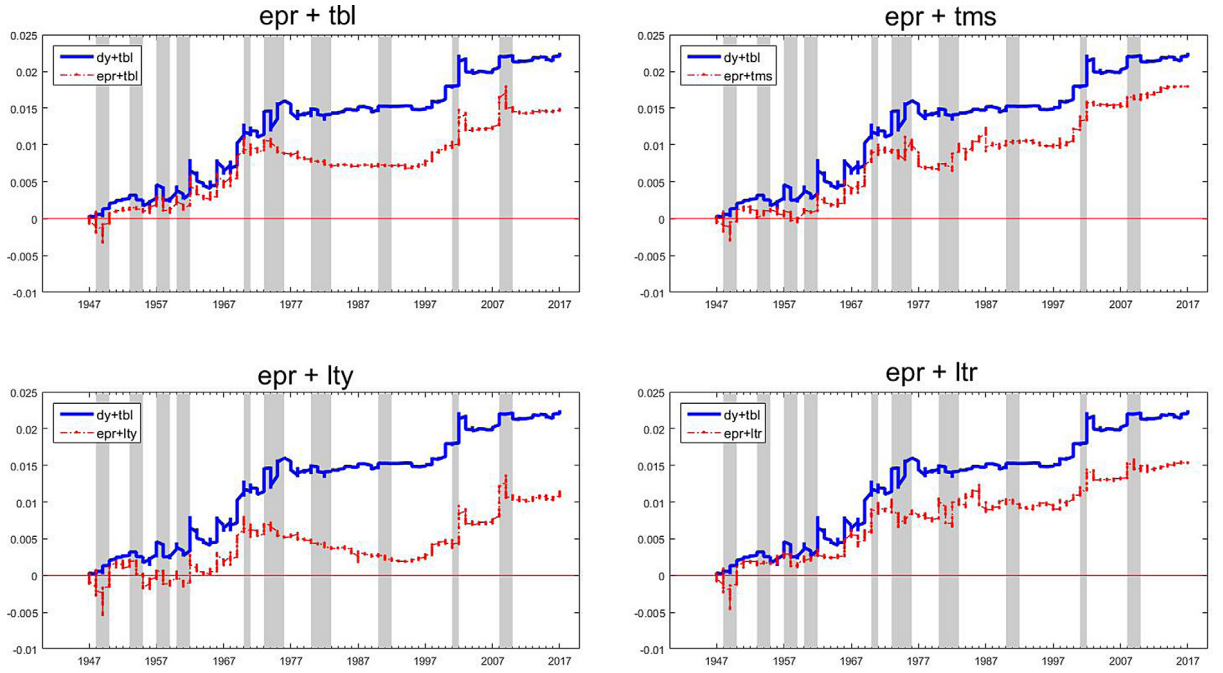
Fig. 4. Two-predictor forecast combinations based on the earnings–price ratio. The figure displays the out-of-sample performance over time in equity premium prediction of two-predictor forecast combinations based on the earnings–price ratio. In each graph, the solid blue line is the benchmark two-predictor model that combines the forecasts of the dividend yield (dy) and the short rate (tbl). The red dashed line is for the alternative two-predictor model. Each line represents the difference of the cumulative squared error of the null (historical mean) minus the cumulative squared error of the selected model. Hence a positive value indicates that the model performs better than the historical mean. 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 2017. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

can be summarized into a performance fee of approximately 2.1% per year overall, 1.8% per year in expansions, 3.6% per year in recessions, together with a statistically significant increase in the SR from 0.46 to 0.64.

Finally, the ex-post timing strategy using the true recession dummy performs even better by generating a performance fee of about 3% per year and a Sharpe ratio of 0.74. This is further evidence that there is high economic value in timing recessions when predictability is high.[10]

6.3. 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.[11]

For example, recall that the equally weighted forecast combination of dy and tbl delivers an R_{oos}^2 of 1.54%. Over the same forecasting period (1947–2017), the risky asset has a squared monthly SR of $0.1560^2 = 0.0243$. Then, the proportional increase in the expected return is $0.0154/0.0243 = 0.6337$. In other words, the simple mean combination will increase the average monthly portfolio return by a factor of 63.37%. Campbell and Thompson (2008) show that this corresponds to an actual increase in the expected return of $\left(\frac{1}{\gamma}\right)\left(\frac{R_{oos}^2}{1-R_{oos}^2}\right)\left(1+SR^2\right) = 0.27\%$ per month or 3.20% per year (for $\gamma = 6$). In conclusion, modest predictive ability for the equity premium can plausibly generate large economic gains.

---

[10]   It is interesting to note that for lower risk aversion ($\gamma = 2$), the performance of the simple mean combination is slightly lower than for higher risk aversion ($\gamma = 6$). This is driven by the fact that the lower the risk aversion the higher the weight on the risky asset. For higher weight on the risky asset, and given that this asset is procyclical, the performance of the historical mean in expansions improves. Similarly, given that the predictive ability of the short rate is countercyclical, the performance of the short rate in recessions also improves. The net effect of lower risk aversion is slightly lower performance for the simple mean combination. In contrast, the ex-post timing strategy performs well for both levels of risk aversion.

[11]   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.

Table 5

Portfolio performance. The table reports 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 2 (Panel A) or 6 (Panel B). ΔCER 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. *EPM* is the Economic Performance Measure. *TO* 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 the benchmark. The last two columns are the annual percent %ΔCER assuming a proportional transaction cost of either 50 or 100 basis points. The formation of forecasts is as described in Table 2. The sample period ranges from January 1927 to December 2017.

Panel A: RRA = 2 ($\gamma = 2$)

| | ΔCER (%) | | | | | | ΔCER (%) | |
Predictor	Full	Exp.	Rec.	SR	EPM	TO	$c = 50$	$c = 100$
Historical mean	9.47	11.89	−4.88	0.48	0.48	0.06	9.36	9.24
Two individual predictors								
dy	1.78	2.09	0.63	0.61*	0.60	2.12	1.78	1.37
tbl	0.18	−1.60	10.71	0.53	0.52	1.42	−0.06	−0.30
Simple mean combination with two predictors								
dy + tbl	1.16	1.06	1.71	0.55*	0.54	1.64	0.86	0.56
Ex-post timing using the true recession dummy								
dy + tbl	3.33	2.09	10.71	0.74***	0.73	2.33	2.81	2.27
Standard forecast combination								
all 14 predictors	0.93	0.87	1.22	0.52	0.51	1.30	0.86	0.79

Panel B: RRA = 6 ($\gamma = 6$)

Historical mean	5.34	6.76	−3.08	0.46	0.46	0.03	5.18	5.01
Two individual predictors								
dy	0.88	2.02	−5.92	0.53	0.51	2.52	0.56	0.24
tbl	1.61	0.38	8.87	0.59	0.56	1.42	1.50	1.40
Simple mean combination with two predictors								
dy + tbl	2.10	1.84	3.63	0.64***	0.62	1.75	1.93	1.75
Ex-post timing using the true recession dummy								
dy + tbl	3.01	2.02	8.87	0.74***	0.73	2.66	2.65	2.29
Standard forecast combination								
all 14 predictors	0.52	0.89	−1.73	0.51	0.50	1.62	0.37	0.23

7. 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. For this reason, instead of using the actual NBER dates, we now turn to implementing two distinct formal rules for ex-ante timing of the business cycle.

7.1. Smooth recession probabilities

The first rule is based on the Markov-switching dynamic factor model (DFMS) of Chauvet (1998), which produces a probability that the economy is in 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, (3) real manufacturing and trade sales, and (4) real personal income excluding transfer payments. This information is used by the model to generate the smoothed recession probabilities, which are obtained from the FRED database of the Federal Reserve Bank of St. Louis.

Our approach for ex-ante timing of the business cycle is as follows. Following Chauvet and Piger (2008), we identify the first month of a recession as the first month for which the smooth recession 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.[12] 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," which are turning points that were established in real time but did not correspond to an actual NBER turning point date. The first available probability in the sample is for June 1967, which is used to forecast the business cycle for the following month.[13]

---

[12] Note that our results are not highly sensitive to the choice of the 50% probability threshold. The performance of the predictive models remains qualitatively the same when using a similar threshold above or below 50%.

[13] 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. Real-time data does not extend back to the 1960s and hence we have no choice but to use the smooth probabilities.

Table 6

Timing the business cycle. The table displays the out-of-sample R^2_{oos} in percent for predictive models of the monthly equity premium against the null of the historical mean. The formation of forecasts is as described in Table 2. The sample period ranges from June 1947 to December 2017 so that the first forecast is for June 1967. 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.

| | R^2_{oos} (%) | | |
Predictor	Full Sample	Expansion	Recession
Two individual predictors			
dy	0.46**	1.09***	−1.04
tbl	0.87**	0.19*	2.51**
Simple mean combination with two predictors			
dy + tbl	1.38***	1.39***	1.34*
Ex-ante timing using the forecasted recession dummy			
dy + tbl	0.98***	0.34**	3.47**
Ex-ante timing using the HP output gap			
dy + tbl	0.47**	−0.55	4.45***
Ex-ante timing using the HP Δ (output gap)			
dy + tbl	2.06***	0.43**	8.50***
Ex-ante timing using the HP 3-month Δ (output gap)			
dy + tbl	1.63***	0.70**	5.29***

7.2. Output gap

The second rule for timing the business cycle is based on the Hodrick and Prescott (HP) (1980, 1997) filter. The HP filter is perhaps the most popular way for estimating the output gap, defined as the difference between a country's actual and potential level of output. The HP filter decomposes output into its trend and cyclical components. The trend component can be viewed as the economy's natural or potential growth path from which growth cyclically deviates. Then, the cyclical component is our empirical proxy for the output gap, defined as the difference between the log of actual output and the log of the trend component. Output is proxied by monthly US real industrial production obtained from the FRED database of the Federal Reserve Bank of St. Louis.

Following Colacito et al. (2019), we use the HP filter to forecast the relative strength of the economy by implementing three HP rules:

1. Consider $t + 1$ to be an expansion if the output gap at time t is positive; and a recession if otherwise.
2. Consider $t + 1$ to be an expansion if the *change* in the output gap at time t is positive (i.e., the output gap is increasing); and a recession if otherwise.
3. Consider $t + 1$ to be an expansion if the output gap at time t is higher than its most recent 3-month average; and a recession if otherwise.

7.3. Empirical results

The ex-ante timing strategies are designed to maximize the ex-ante insurance gains: they predict the state of the economy out of sample and then choose the dividend yield forecasts in expansions or the short rate forecasts in recessions. The statistical performance of these strategies is presented in Table 6. For completion, we also report the performance of *dy*, *tbl* and the simple mean combination of these two predictors. In all cases, the sample period ranges from June 1947 to December 2017 so that the first forecast is for June 1967.

In general, all ex-ante timing strategies perform well. For example, the ex-ante strategy using the forecasted recession dummy based on the smoothed recession probabilities displays an $R^2_{oos} = 0.98\%$, which is highly significant. The R^2_{oos} remains positive and significant in both expansions and recessions. The best HP strategy is the one based on the change in the output gap. This strategy delivers an $R^2_{oos} = 2.06\%$, which is better than the simple mean combination ($R^2_{oos} = 1.38\%$). This HP strategy also delivers a positive and significant R^2_{oos} in both expansions and recessions. Overall, the results in Table 6 indicate that timing the business cycle provides strong performance in out-of-sample equity premium prediction, especially in recessions. As mentioned earlier, this is to be expected since there is more predictability in recessions and thus timing the business cycle delivers better performance in recessions.

7.4. Portfolio performance

The results are similar for the portfolio performance of the ex-ante timing strategies, which are presented in Table 7. Again, the table reports results for $\gamma = 2$ (Panel A) and $\gamma = 6$ (Panel B) but our discussion focuses on the case of $\gamma = 6$. For the full sample, all ex-ante timing strategies perform well and typically better than the simple mean combination. The latter delivers a performance fee of 2.16% per year compared to the ex-ante timing strategies that deliver a performance fee ranging from 2.11% to 2.79%. In terms of Sharpe ratios, the results remain strong: the better HP strategy doubles the Sharpe ratio of the benchmark (0.24 for the benchmark versus 0.50 for the better HP strategy). This difference is statistically significant.

Table 7

Portfolio performance: Timing the business cycle. The table reports the out-of-sample portfolio performance for a mean–variance investor with a degree of relative risk aversion equal to 2 (Panel A) or 6 (Panel B). The formation of forecasts is as described in Table 2. The portfolio formation is as described in Table 4. The sample period ranges from June 1947 to December 2017 so that the first forecast is for June 1967.

Panel A: RRA=2 ($\gamma = 2$)

Predictor	ΔCER (%)			SR	EP	TO	ΔCER (%)	
	Full	Exp.	Rec.				$c = 50$	$c = 100$
Historical mean	6.78	9.67	−8.40	0.29	0.29	0.05	6.63	6.48
Two individual predictors								
dy	2.54	1.79	6.44	0.44	0.43	2.37	2.13	1.73
tbl	2.10	0.02	12.97	0.43	0.42	1.37	1.92	1.74
Simple mean combination with two predictors								
dy + tbl	1.62	0.55	7.17	0.38*	0.38	1.82	1.30	0.98
Ex-ante timing using the forecasted recession dummy								
dy + tbl	3.85	2.39	11.47	0.56**	0.55	2.88	3.22	2.59
Ex-ante timing using the HP output gap								
dy + tbl	2.42	0.42	12.89	0.45	0.45	2.90	1.75	1.08
Ex-ante timing using the HP Δ (output gap)								
dy + tbl	3.26	0.58	17.34	0.49*	0.48	7.91	0.97	−1.34
Ex-ante timing using the HP 3-month Δ (output gap)								
dy + tbl	3.60	1.43	14.96	0.53*	0.53	5.28	2.20	0.78

Panel B: RRA=6 ($\gamma = 6$)

Predictor								
Historical mean	4.09	5.28	−2.18	0.24	0.24	0.03	3.94	3.79
Two individual predictors								
dy	0.07	0.42	−1.86	0.28	0.28	3.02	−0.25	−0.58
tbl	2.53	1.52	7.86	0.47*	0.45	1.35	2.48	2.42
Simple mean combination with two predictors								
dy + tbl	2.16	1.72	4.39	0.43***	0.42	2.01	1.99	1.83
Ex-ante timing using the forecasted recession dummy								
dy + tbl	2.46	2.06	4.53	0.46*	0.42	3.51	2.06	1.66
Ex-ante timing using the HP output gap								
dy + tbl	2.11	1.07	7.50	0.42	0.40	3.23	1.74	1.37
Ex-ante timing using the HP Δ (output gap)								
dy + tbl	2.63	1.22	9.97	0.49**	0.46	9.73	1.25	−0.15
Ex-ante timing using the HP 3-month Δ (output gap)								
dy + tbl	2.79	1.69	8.57	0.50*	0.48	6.59	1.91	1.01

In short, therefore, the evidence indicates that the simple mean combination delivers about 2% per year over and above the historical mean benchmark compared to 2.8% per year for the better ex-ante timing strategies. The latter exhibit twice the Sharpe ratio of the benchmark. Overall, these are important results that further establish the following findings: (1) the equity premium is predictable, (2) this predictability has high economic value, and (3) this predictability rises with the ability to predict the state of the economy. In conclusion, the ex-ante timing strategies predict business cycles well and deliver superior performance especially in recessions.

8. The financial cycle

In this section, we evaluate the cyclical variation of equity premium predictions around a different cycle: the financial cycle. Following Claessens et al. (2012), we identify the phases of the financial cycle based on contractions and expansions of aggregate credit. Credit is a natural aggregate we can use to analyse the financial cycle because it constitutes the most important link between savings and investment. Our measure of credit is aggregate credit in domestic currency offered by domestic banks to the private non-financial sector. Quarterly data are obtained from the Bank for International Settlements.[14] The sample period begins in the first quarter of 1960 and ends in the last quarter of 2017. As a result of data availability, the equity premium sample in this section begins in 1940 so that, given the 20-year estimation window, the first forecast is for January 1960.

We identify the turning points in the log of aggregate credit using the algorithm introduced by Harding and Pagan (2002). This is a well-established and reproducible methodology for dating different phases of a cycle. The algorithm requires a complete cycle to last at least five quarters and each phase to last at least two quarters. Specifically, a peak in the quarterly log-credit series c_t occurs at time t, if:

$$\left\{ \left[(c_t - c_{t-2}) > 0, (c_t - c_{t-1}) > 0 \right] \text{ and } \left[(c_{t+2} - c_t) < 0, (c_{t+1} - c_t) < 0 \right] \right\}. \tag{22}$$

Similarly, a cyclical trough occurs at time t, if:

$$\left\{ \left[(c_t - c_{t-2}) < 0, (c_t - c_{t-1}) < 0 \right] \text{ and } \left[(c_{t+2} - c_t) > 0, (c_{t+1} - c_t) > 0 \right] \right\}. \tag{23}$$

---

[14] We use data from the Bank for International Settlements quarterly series Q:US:C:A:M:USD:A.

Table 8

The financial cycle. The table displays the out-of-sample R^2_{oos} in percent for predictive models of the monthly equity premium against the null of the historical mean. The formation of forecasts is as described in Table 2. The sample period ranges from January 1940 to December 2017 so that the first forecast is for January 1960. Expansions and recessions are according to the NBER. Financial cycle upturns and downturns are defined as in Claessens et al. (2012). *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively, using the Clark and West (2006, 2007) one-sided t-statistic.

	R^2_{oos} (%)		
Predictor	Full Sample	Expansion/ Upturn	Recession/ Downturn
Business Cycle			
dy	0.83**	1.64***	−1.30
tbl	1.02***	0.59**	2.14**
Mean: dy + tbl	1.58***	1.81***	1.00*
Financial Cycle			
dy	0.83**	1.29***	−0.28
tbl	1.02***	0.23**	2.90***
Mean: dy + tbl	1.58***	1.40***	2.02**
Business and Financial Cycle			
dy	0.83**	0.38**	1.35
tbl	1.02***	0.61**	3.30**
Mean: dy + tbl	1.58***	1.20***	3.08**

Using the terminology of Claessens et al. (2012), the recovery phase of the financial cycle (from trough to peak) is called the "upturn," whereas the contraction phase (from peak to trough) is called the "downturn".[15]

In Table 8, we report results for the business cycle, financial cycle and a combination of the two defined as the union of recessions and downturns versus their complement. Our main finding here is that the results for the business cycle also hold for the financial cycle as well as for the combination of the two. For example, the dividend yield performs well in both expansions and upturns, whereas the short rate performs well in both recessions and downturns. More importantly, the simple mean combination of *dy* and *tbl* performs well in expansions as well as upturns and in recessions as well as downturns. In other words, the simple mean combination has a remarkably stable performance across business and financial cycles always delivering a positive and significant R^2_{oos}. In short, therefore, our main finding is robust to determining the state of the economy around either the business cycle or the financial cycle.

9. International evidence

In this section, we report international evidence for four additional countries: Canada, UK, Germany and Japan. All international data used in our analysis are obtained from the Thomson Reuters Eikon database. Specifically, for Canada, the data on the total return index and the dividend yield is for the S&P/TSX composite index. For the UK, Germany and Japan, the data are for that country's FTSE stock index. For all countries, *tbl* is the 3-month treasury bill rate.

It is important to note that the sample period for the international stock indices is substantially shorter than for the Welch and Goyal (2008) US data. The longest sample is for the UK, which begins in December 1964. The shorter sample is for Germany, which begins in December 1972. All samples end on the same month as the US data, which is December 2017. To facilitate comparison, we also report the US results for the longest international sample period of December 1964 to December 2017.

Our empirical setup is the same as before. We report the out-of-sample R^2_{oos} generated by imposing the two economic constraints on slope coefficients and forecasts. We report results for the full sample, expansions and recessions for five models: the individual *dy* predictor; the individual *tbl* predictor; the simple mean combination of *dy* and *tbl*; the ex-post timing strategy of *dy* and *tbl*; and the first ex-ante timing strategy based on the HP filter, which identifies an expansion to be associated with a positive output gap, and a recession otherwise. The results are reported in Table 9.

For the international evidence, the main finding is that there is considerable predictability in three of the four countries: Canada, UK and Japan. The exception is Germany for which there appears to be little or no predictability. More importantly, the three countries behave in a similar way to the US: the dividend yield is a powerful predictor in expansions, whereas the short rate is a powerful predictor in recessions. In all three countries, however, there is a noticeable difference with the US: although *dy* and *tbl* perform very well in one state of the economy, they perform very poorly in other state of the economy. For this reason, the performance of the simple mean combination is positive but weaker than for the US. In contrast, the timing strategies perform very well. Both the ex-post timing strategy and the ex-ante timing strategy based on the HP filter exhibit strong performance in Canada, UK and Japan.

To illustrate these results, consider the case of Canada. Table 9 indicates that *dy* delivers an R^2_{oos} of 2.67% in expansions and *tbl* an R^2_{oos} of 1.97% in recessions. In the other state of world, however, performance is poor: *dy* delivers an R^2_{oos} of −3.14% in recessions

---

[15] Note that we identify the quarterly troughs and peaks to define the monthly financial cycle: the first month following a peak is the beginning of a downturn, and the first month following a trough is the beginning of an upturn.

Table 9

International evidence. The table displays the out-of-sample R^2_{oos} in percent for four additional countries plus the US. The R^2_{oos} 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. 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. Ex-post timing uses the true recession dummy, whereas ex-ante timing uses the HP output gap.

Panel A: Canada, Dec 1969–Dec 2017			
Predictor	*Full Sample*	*Expansion*	*Recession*
dy	−0.64	2.67**	−3.14
tbl	−1.26	−5.55	1.97*
Mean: dy + tbl	0.15	0.31	0.02
Ex-post: dy + tbl	2.27***	2.67**	1.97*
Ex-ante: dy + tbl	3.11***	3.95***	2.48**
Panel B: UK, Dec 1964–Dec 2017			
dy	−0.79	0.28*	−2.11
tbl	−1.89	−3.78	0.44*
Mean: dy + tbl	0.05	0.22	−0.17
Ex-post: dy + tbl	0.35*	0.28*	0.44*
Ex-ante: dy + tbl	0.43*	−0.87	2.02**
Panel C: Germany, Dec 1972–Dec 2017			
dy	0.39*	−2.82	0.46**
tbl	−0.41	−3.04	−0.35
Mean: dy + tbl	0.00	−2.91	0.05
Ex-post: dy + tbl	−0.40	−2.82	−0.35
Ex-ante: dy + tbl	−0.40	−2.74	−0.35
Panel D: Japan, Jan 1986–Dec 2017			
dy	−0.48	3.94*	−0.97
tbl	0.23	−11.13	1.49*
Mean: dy + tbl	0.26	−1.55	0.46*
Ex-post: dy + tbl	1.73*	3.94*	1.49*
Ex-ante: dy + tbl	1.55*	3.72*	1.31*
Panel E: USA, Jan 1964–Dec 2017			
dy	1.14***	1.25***	0.66
tbl	0.16	−0.53	2.90**
Mean: dy + tbl	1.12***	0.90**	2.01***
Ex-post: dy + tbl	1.58***	1.25***	2.90**
Ex-ante: dy + tbl	0.70*	0.15	2.90**

and *tbl* an R^2_{oos} of −5.55% in expansions. As a result, the simple mean combination performance is positive but weaker than for the US as it delivers an R^2_{oos} of 0.15% for the full sample. In contrast, the timing strategies work very well: the ex-post timing strategy produces an R^2_{oos} of 2.27% for the full sample, whereas the ex-ante timing strategy performs even better with an R^2_{oos} of 3.11% for the full sample. In fact, both timing strategies perform well in both states of the economy. The results are similar for the UK and Japan. We conclude, therefore, that there is strong predictability in timing strategies designed to combine the forecasts from *dy* and *tbl* for several countries around the world.[16]

10. Conclusion

We propose a simple way to predict the US 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: diversification and insurance. Diversification in forecasting is the direct analogue to portfolio diversification: by combining two negatively correlated forecasts, which individually display a positive R^2_{oos}, we can create a new set of forecasts that have a substantially higher R^2_{oos} 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. For example, the simple mean combination delivers a certainty equivalent return of about 2% per year. Moreover, strategies designed for ex-ante timing of the business cycle can provide additional economic value as the certainty equivalent return rises to about 3% per year.

These findings are not surprising. The dividend yield and the short rate are popular predictors with solid theoretical foundations. In this paper, we uncover the cyclical variation in their predictive information and take advantage of it by forming a forecast combination. The result is consistent with a recent theory on investor attention allocation that links information choices to the

---

[16] For studies that consider out-of-sample tests of international stock return predictability, see Solnik (1993), Rapach et al. (2005), Cooper and Priestly (2009), Della Corte et al. (2010), Hjalmarsson (2010), Kellard et al. (2010), Henkel et al. (2011), and Rapach et al. (2013).

business cycle. According to this theory, in recessions investors care more about aggregate information, such as interest rates, and in expansions they care more about stock-specific information, such as dividends. Overall, our analysis highlights the cyclical variation in the predictive information of widely-used economic fundamentals and demonstrates that there is a simple and powerful way to predict the US equity premium out of sample.

CRediT authorship contribution statement

Ilias Tsiakas: Conceptualization, Methodology, Formal analysis, Writing - original draft, Writing - review & editing, Project administration, Funding acquisition. **Jiahan Li:** Conceptualization, Methodology, Software, Validation, Formal analysis. **Haibin Zhang:** Methodology, Software, Validation, Formal analysis.

References

Aumann, R., Serrano, R., 2008. An economic index of riskiness. J. Political Econ. 116, 810–836.

Balduzzi, P., Lynch, A.W., 1999. Transaction costs and predictability. J. Financ. Econ. 52, 47–78.

Bekaert, G., Engstrom, E., Xing, Y., 2009. Risk, uncertainty and asset prices. J. Financ. Econ. 91, 59–82.

Campbell, J.Y., Cochrane, J., 1999. Force of habit: A consumption-based explanation of aggregate stock market behavior. J. Political Econ. 107, 205–251.

Campbell, J.Y., Shiller, R.J., 1988. The dividend-price ratio and expectations of future dividends and discount factors. Rev. Financ. Stud. 1, 195–227.

Campbell, J.Y., Thompson, S.B., 2008. Predicting excess stock returns out of sample: Can anything beat the historical average? Rev. Financ. Stud. 21, 1509–1531.

Chauvet, M., 1998. An econometric characterization of business cycle dynamics with factor structure and regime switching. Internat. Econom. Rev. 39, 969–996.

Chauvet, M., Piger, J., 2008. A comparison of the real-time performance of business cycle dating methods. J. Bus. Econom. Statist. 26, 42–49.

Claessens, S., Kose, M.A., Terrones, M.E., 2012. How do business and financial cycles interact? J. Int. Econ. 87, 178–190.

Clark, T.E., West, K.D., 2006. Using out-of-sample mean squared prediction errors to test the martingale difference hypothesis. J. Econometrics 135, 155–186.

Clark, T.E., West, K.D., 2007. Approximately normal tests for equal predictive accuracy. J. Econometrics 138, 291–311.

Cochrane, J., 2008. The dog that did not bark: A defense of return predictability. Rev. Financ. Stud. 21, 1533–1575.

Cochrane, J., 2011. Presidential address: Discount rates. J. Finance 66, 1047–1108.

Colacito, R., Riddiough, S.J., Sarno, L., 2019. Business cycles and currency returns. J. Financ. Econ. http://dx.doi.org/10.1016/j.jfineco.2020.04.005, (forthcoming).

Cooper, I., Priestly, R., 2009. Time-varying risk premiums and the output gap. Rev. Financ. Stud. 22, 2801–2833.

Cujean, J., Hasler, M., 2017. Why does return predictability concentrate in bad times? J. Finance 72, 2717–2758.

Dangl, T., Halling, M., 2012. Predictive regressions with time-varying coefficients. J. Financ. Econ. 106, 157–181.

Della Corte, P., Sarno, L., Tsiakas, I., 2009. An economic evaluation of empirical exchange rate models. Rev. Financ. Stud. 22, 3491–3530.

Della Corte, P., Sarno, L., Valente, G., 2010. A century of equity premium predictability and the consumption wealth ratio: An international perspective. J. Empir. Financ. 17, 313–331.

Fama, E.F., French, K.R., 1989. Business conditions and expected returns on stocks and bonds. J. Financ. Econ. 25, 23–49.

Ferreira, M.I., Santa-Clara, P., 2011. Forecasting stock market returns: The sum of the parts is more than the whole. J. Financ. Econ. 100, 514–537.

Guidolin, M., Timmermann, A., 2007. Asset allocation under multivariate Regime switching. J. Econom. Dynam. Control 31, 3503–3544.

Harding, D., Pagan, A., 2002. Dissecting the cycle: A methodological investigation. J. Monetary Econ. 49, 365–381.

Henkel, S.J., Martin, J.S., Nardari, F., 2011. Time-varying short-horizon predictability. J. Financ. Econ. 99, 560–580.

Hjalmarsson, E., 2010. Predicting global stock returns. Journal of Financial and Quantitative Analysis 45, 49–80.

Hodrick, R.J., Prescott, E.C., 1980. Postwar U.S. Business Cycles: An Empirical Investigation Working Paper. Carnegie-Mellon University.

Hodrick, R.J., Prescott, E.C., 1997. Postwar U.S. business cycles: An empirical investigation. J. Money Credit Bank. 29, 1–16.

Homm, U., Pigorsch, C., 2012. Beyond the sharpe ratio: An application of the Aumann–Serrano index to performance measurement. J. Bank. Financ. 36, 2274–2284.

Kacperczyk, M., Van Nieuwerburgh, S., Veldkamp, L., 2016. A rational theory of mutual funds' attention allocation. Econometrica 84, 571–626.

Kellard, N.M., Nankervis, J.C., Papadimitriou, F.I., 2010. Predicting the equity premium with dividend ratios: Reconciling the evidence. J. Empir. Financ. 17, 539–551.

Kelly, B., Pruitt, S., 2013. Market expectations in the cross-section of present values. J. Finance 68, 1721–1756.

Kostakis, A., Magdalinos, T., Stamatogiannis, M.P., 2015. Robust econometric inference for stock return predictability. Rev. Financ. Stud. 28, 1506–1553.

Ledoit, O., Wolf, M., 2008. Robust performance hypothesis testing with the sharpe ratio. J. Empir. Financ. 15, 850–859.

Lettau, M., Van Nieuwerburgh, S., 2008. Reconciling the return predictability evidence. Rev. Financ. Stud. 21, 1607–1652.

Li, J., Tsiakas, I., 2017. Equity premium prediction: The role of economic and statistical constraints. J. Financial Mark. 36, 56–75.

Ludvigson, S.C., Ng, S., 2007. The empirical risk-return relation: A factor analysis approach. J. Financ. Econ. 83, 171–222.

Maio, P., 2013. The 'Fed Model' and the predictability of stock returns. Rev. Finance 17, 1489–1533.

Marquering, W., Verbeek, M., 2004. The economic value of predicting stock index returns and volatility. J. Financ. Quant. Anal. 39, 407–429.

McCracken, M.W., 2007. Asymptotics for out of sample tests of Granger causality. J. Econometrics 140, 719–752.

Menzly, L., Santos, T., Veronesi, P., 2004. Understanding predictability. J. Political Econ. 112, 1–47.

Neely, C.J., Rapach, D.E., Zhou, G., 2014. Forecasting the equity risk premium: The role of technical indicators. Manage. Sci. 60, 1772–1791.

Paye, B.S., Timmermann, A., 2006. Instability of return prediction models. J. Empir. Financ. 13, 274–315.

Pettenuzzo, D., Timmermann, A., Valkanov, R., 2014. Forecasting stock returns under economic constraints. J. Financ. Econ. 114, 517–553.

Rapach, D.E., Strauss, J.K., Zhou, G., 2010. Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. Rev. Financ. Stud. 23, 821–862.

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

Rapach, D.E., Wohar, M.E., Rangvid, J., 2005. Macro variables and international stock return predictability. Int. J. Forecast. 21, 137–166.

Rapach, D., Zhou, G., 2013. Forecasting stock returns. In: Elliott, G., Timmermann, A. (Eds.), Handbook of Economic Forecasting, Volume 2A. Elsevier., Amsterdam.

Solnik, B., 1993. The Performance of International Asset Allocation Strategies using Conditioning Information. Journal of Empirical Finance 1 33–55.

Timmermann, A., 2006. Forecast combinations. In: Elliott, G., Granger, C.W.J., Timmermann, A. (Eds.), Handbook of Economic Forecasting, Volume 1. North-Holland, Amsterdam.

Welch, I., Goyal, A., 2008. A comprehensive look at the empirical performance of equity premium prediction. Rev. Financ. Stud. 21, 1455–1508.