Out-of-sample equity premium prediction: a scenario analysis approach *

Xiaoxiao Tang[†] Feifang Hu[‡] Peiming Wang[§]

Abstract

We propose two methods of equity premium prediction with a single and multiple predictors respectively and evaluate their out-of-sample performance using US stock data with 15 popular predictors for equity premium prediction. The first method defines three scenarios in terms of the expected returns under the scenarios and assumes a Markov chain governing the occurrence of the scenarios over time. It employs predictive quantile regressions of excess return on a predictor for three quantiles to estimate the occurrence of the scenarios over an in-sample period and the transition probabilities of the Markov chain, predicts the expected returns under the scenarios, and generates an equity premium forecast by combining the predicted returns under three scenarios with the estimated transition probabilities. The second method generates an equity premium forecast by combining the individual forecasts from the first method across all predictors. For most of predictors, the first method outperforms the benchmark method of historical average and the traditional predictive linear regression with a single predictor both statistically and economically, and the second method consistently performs better than several competing methods used in the literature. The performance of our methods is further examined under different scenarios and economic conditions, and is robust for two different out-of-sample periods and specifications of the scenarios.

JEL Classification: C53, C22, G11

Keywords: Combination forecast; equity premium; Markov chain; quantile regression; scenario analysis.

*We are grateful to two anonymous referees for their helpful comments.

[†]Olin Business School, Washington University in St. Louis, MO, United States

[‡]Department of Statistics, George Washington University, Washington, D.C., United States

[§]Corresponding Author. Department of Finance, Business School, Auckland University of Technology, Auckland 1142, Private Bag 92006, New Zealand. Phone: ++64 9 9219999 extn 5393. Email: peiming.wang@aut.ac.nz.

1 Introduction

Equity premium prediction is important in finance because it is related to not only important investment decisions on asset allocation made by practitioners, but also some important issues in finance such as market efficiency and asset pricing models in which academics in finance are interested. While many studies report evidence of U.S. equity risk premium predictability based on predictors such as valuation ratios and interest rates, there are mixed results from the research of equity premium predictability, particularly for out-of-sample forecast.[1] For example, Goyal and Welch (2008) examine the out-of-sample predictability of the traditional predictive linear regression of return on either a single or multiple predictors for a list of popular predictors from the literature with respect to the benchmark method of historical average. They find that the benchmark method of historical average has better out-of-sample performance than the traditional predictive linear regressions. On the other hand, Campbell and Thompson (2008) show that the forecasts based on the predictive linear regression with some restrictions on regression coefficients and predicted returns perform better out-of-sample than the historical average forecast. In addition, using the same list of predictors as investigated in Goyal and Welch (2008), Rapach et al. (2010) provide empirical evidence that the combination forecast of the individual forecasts from the traditional predictive linear regression with a single predictor has better out-of-sample performance than the historical average forecast.

There are also other recent studies on alternatives to the traditional predictive linear regression for equity premium prediction in the literature, which deliver better out-of-sample performance than the historical average forecast. For instance, Markov regime-switching models for equity premium prediction outperform the historical average forecast by modelling regime shifts in the state of an economy according to a finite-state Markov chain and different linear relationships between return and predictors for different regimes (e.g., Guidolin and Timmermann, 2009; Henkel et al. 2011). The linear dynamic models proposed by Dangl and Hailling (2012) improve out-of-sample predictive performance relative to the benchmark method of historical average by allowing the coefficients of the traditional predictive linear regression to be random and time-varying. In addition, Meligkotsidou et al. (2014) propose the quantile regression method of equity premium prediction that first generates individual

[1] See Cochrane (2011) and Rapach and Zhou (2013) for recent surveys.

forecasts from predictive quantile linear regression of return on a single predictor, and then produces an equity premium forecast by combining these individual forecasts across all predictors. They find that their method has better out-of-sample performance than the benchmark method of historical average.

In general, asset returns may depend on many different factors including market and economic conditions. That is why scenario analysis is widely used in practice to predict the expected return on an investment based on the returns expected under a number of possible future scenarios for the investment, such as the bad, normal, and good case scenarios for an investment. It is a what-if analysis in which the likelihood of various scenarios and their investment outcomes are evaluated, and the predicted expected return on an investment is derived from the predicted expected returns under different scenarios by combining them with the probability that they will occur.

In this paper, we follow the principle of scenario analysis to propose a method of equity premium prediction with a single predictor. Specifically, we define three possible scenarios for a stock investment in terms of the expected excess returns achieved under each of the scenarios and assume a first-order three-state Markov chain governing the occurrence of the scenarios over time. In the specification of the three scenarios, the expected excess return under Scenario 1 (3) is below (above) the lower (upper) quartile of excess returns, and the expected excess return under Scenario 2 is equal to the median excess return. We use predictive quantile linear regressions of excess return on a predictor for the three quartiles to estimate the occurrence of the three scenarios over an in-sample period and the transition probabilities of the Markov chain with in-sample data. Then we predict the one-step ahead out-of-sample excess returns expected under each of the three scenarios as follows. The predicted expected excess return under Scenario 1 (3) is the average of the predicted lower (upper) quartile from the predictive quantile linear regression and the excess returns under that scenario during the in-sample period, and the predicted expected excess return under Scenario 2 is the predicted median return from the predictive quantile linear regression. Finally, we generate a one-step ahead out-of-sample forecast of equity premium by combining the predicted expected excess returns under the three scenarios with the estimated transition probabilities, based on the scenario that occurred in the end of the in-sample period.

Methodologically, our prediction method with a single predictor is related to the work of Markov regime-switching models for stock returns (e.g., Guidolin and Timmermann, 2009; Henkel et al. 2011) because we employ a Markov chain to model the switching between scenarios. However, it differs from the previous studies because of the differences in the specification of regimes, the estimation of parameters and regimes, and the determination of the expected return conditional on regimes. Our predition method with a single predictor is also related the studies on the financial applications of quantile regression (e.g., Engle and Manganelli, 1999; Meligkotsidou et al. 2014) because of the use of quantile regression in the analysis, but differs from these studies as quantile regression is used differently in equity premium prediction.

We also propose a method of equity premium prediction with multiple predictors, which combines the individual forecasts from our method with a single predictor across all predictors to form a single forecast of equity premium. While studies show that many predictors such as valuation ratios and term spreads can be used to detect changes in factors associated with asset prices (e.g., Fama and French, 1989), different predictors could capture different factors that have different relationships with asset prices. This implies that individual forecasts based on different predictors could have a weak correlation between each other. As pointed out by Rapach et al. (2010), the combination forecast of individual forecasts like the average of individual forecasts based on different predictors should be more accurate because it reduces the uncertainty or instability risk associated with individual forecasts.

To evaluate the out-of-sample performance of our prediction methods, we not only measure the accuracy of out-of-sample prediction relative to a competing forecast method in terms of mean squared forecast errors (MSFE), but also conduct two encompassing tests to compare the information on equity premium prediction in each of our methods with that in a competing forecast method. In addition, we assess the out-of-sample performance in terms of economic profits by calculating the utility difference between two competing forecasts from the perspective of a risk-averse investor. For our method with a single predictor, we further look into the predictability for the forecasts generated under each of the three scenarios and the changes in the transition probabilities as well as the switching between scenarios over an out-of-sample period. Moreover, we examine the link between the out-of-sample predictive performance of our methods and the real economy by evaluating the predictability for the forecasts during low, average, and high GDP growth periods and during recession and expansion periods separately. Finally, for robustness check, we evaluate the out-of-sample performance of our methods for two different out-of-sample periods and two different specifications of the three scenarios.

In our empirical analysis, we employ the quarterly US stock data with the 15 popular predictors used in Goyal and Welch (2008), and our empirical findings show that our prediction methods perform better out-of-sample than several competing prediction methods in the economics and finance literature. First, our method with a single predictor is superior to the prevailing benchmark method of historical average and the traditional predictive linear regression with a single predictor for most of predictors in terms of the accuracy of out-of-sample prediction and the useful forecast information in a prediction method as well as economic profits. In addition, the improvement in out-of-sample prediction for our method with a single predictor is similar across the three scenarios. Second, based on all the 15 predictors, our prediction method with multiple predictors consistently outperforms the prevailing benchmark method of historical average and the forecast combination method of the individual forecasts from the traditional predictive linear regression with a single predictor in terms of both the accuracy of out-of-sample prediction and the useful prediction information in a prediction method. The forecasting gains in prediction accuracy also lead to economic profits for an investor who uses our equity premium forecast instead of the two competing forecasts. Moreover, these findings hold for three different combination methods and two different out-of-sample periods. Thus, our findings provide evidence that equity premium predictability based on these predictors is likely stronger collectively than individually. Third, using the same data as in Meligkotsidou et al. (2014), we find that our prediction methods generate forecasting gains in the accuracy of out-of-sample prediction relative to their methods.

Furthermore, our findings show that our prediction method with a single predictor performs similarly across the three GDP growth periods as well as the recession and expansion periods, suggesting that the out-of-sample predictive performance for this method is unlikely associated with different economic conditions. On the other hand, our method with multiple predictors performs better during the low and average GDP growth periods than during the high GDP growth period, and better during the recession period than during the expansion period. This implies that the out-of-sample predictive performance of our method with multiple predictors is likely associated with business cycle.

Finally, our findings show that the out-of-sample performance of our prediction methods is robust for the two different out-of-sample periods and two different specifications of the three scenarios. Our findings provide evidence that our prediction methods are useful alternatives to the methods of equity premium prediction commonly used the literature.

The remainder of the paper is organized as follows. In section 2, we describe our prediction methods with a single predictor and multiple predictors respectively, and discuss the evaluation of out-of-sample performance. In section 3, we describe our data and two alternative forecasts, and present empirical results. Section 4 concludes the paper.

2 The scenario analysis method

In this section, we first propose a method with a single predictor for prediction of equity premium, referred to as the scenario analysis method (SAM). We then describe how we combine different single-predictor forecasts to form a single forecast of equity premium. Finally, we discuss the statistical and economic evaluation criteria for out-of-sample performance of equity premium prediction.

2.1 The prediction method with a single predictor

Following the principle of scenario analysis, we define three possible scenarios for a stock investment in terms of the expected excess return on the investment over a risk-free rate for a period as follows. Under scenario one, the excess return on the investment unlikely exceeds the lower quartile of all possible excess returns, and the expected excess return is lower than this lower quartile. Under scenario three, the excess return on the investment is unlikely below the upper quartile of all possible excess returns, and the expected excess return is higher than this upper quartile. Under scenario two, the excess return on the investment is likely in the region between the lower and upper quartiles of all possible excess returns, and the expected excess return is equal to the median excess return for the investment.[2] Thus, we can regard the three scenarios as bad, normal, and good scenarios for a stock investment respectively.

Furthermore, we assume that only one of the three scenarios can occur during an investment period, and a first-order three-state Markov chain governs the occurrence of the three scenarios over time. Specifically, we use a discrete state variable S_t to indicate the scenario that occurs during period t with $S_t = 1, 2, 3$ for the bad, normal, and good scenarios respectively. The transition probability from Scenario j for period $t-1$ to Scenario k for period t is given by

$$P(S_t = k \mid S_{t-1} = j) = p_{kj}, \qquad \text{for } j, k = 1, 2, 3, \tag{1}$$

and $\sum_{k=1}^{3} p_{kj} = 1$, for $j = 1, 2, 3$. Thus, knowing $S_{t-1} = j$ for period $t-1$, we can determine the equity premium for period t, denoted as $E(y_t)$, by

$$E(y_t) = \sum_{k=1}^{3} p_{kj} E(y_t | S_t = k), \tag{2}$$

where y_t is the excess return on a stock investment for period t, and $E(y_t|S_t = k)$ is the expected excess return under Scenarios k for period t. Equation (2) shows that equity premium is determined by a probability-weighted average of the expected excess returns under the three possible scenarios.

For our prediction method with a single predictor, we first use in-sample data up to time $t-1$ to obtain the one-step ahead predicted values of the lower, median, and upper quartiles of excess returns for each of periods up to time t by predictive quantile linear regressions of excess return on a single predictor for the three quartiles. Then we use these predicted quartiles to estimate the values of the state variable for the in-sample period up to time $t-1$ as well as the transition probabilities, and to predict the one-step ahead expected excess returns conditional on each of the three scenarios for period t. Finally, based on the estimated value of the state variable for period $t-1$, we apply Equation (2) to obtain a one-step ahead out-of-sample forecast of equity premium for period t by replacing the transition probabilities with their estimated values and the expected excess returns under the scenarios

[2]Note that we choose the lower and upper quartiles in the specification of the scenarios because boxplot for investment analysis is widely used in practice, in which these two quartiles divide investment returns into three regions of low, average, and high returns.

with their predicted values respectively.

More specifically, with the in-sample data of observed excess returns and observed values of a predictor up to time $t-1$,$\{(y_i, x_i); i = 1, ..., t-1\}$, we first run the predictive quantile regressions of y_{i+1} on x_i for three quartiles to obtain the predicted lower, median, and upper quartiles of excess returns for each of periods up to time t, denoted by

$$Q_{y_{i+1}|x_i}(\tau_j) = \hat{\beta}_{0,j} + \hat{\beta}_{1,j} x_i, i = 1, \ldots, (t-1), j = 1, 2, 3, \tag{3}$$

where $Q_{y_{i+1}|x_i}(\tau_j)$ is the one-step ahead predicted τ_jth quantile of excess returns for period $i+1$ based on the observed value of a predictor x_i, $\tau_1 = 0.25$, $\tau_2 = 0.5$, $\tau_3 = 0.75$, and $\hat{\beta}_{0,j}$ and $\hat{\beta}_{1,j}$ are the regression parameter estimates of the τ_jth quantile linear regression of y_{i+1} on x_i with the in-sample data.[3]

According to the specification of the three scenarios, if the observed excess return for period i is less (larger) than the predicted lower (upper) quartile for period i, the bad (good) scenario most likely occurred during period i. On the other hand, if the observed excess return for period i is between the predicted lower and upper quartiles for period i, the normal scenario most likely occurred during period i. Therefore, we use the predicted quartiles to estimate the values of the state variable for the in-sample period as follows:

$$\hat{S}_i = \begin{cases} 1, \text{ if } y_i < Q_{y_i|x_{i-1}}(\tau_1); \\ 3, \text{ if } y_i > Q_{y_i|x_{i-1}}(\tau_3); \\ 2, \text{ otherwise}, \end{cases} \tag{4}$$

for $i = 2, \ldots, t-1$.

For a first-order finite state Markov chain, researchers usually use information over each period on movements of the chain from one state of the chain to another to estimate transition probabilities. Here the maximum likelihood estimators of the transition probabilities p_{kj} of the transition from state j to state k are derived as the ratios of the numbers of times the chain moving from state j

[3]Note that we obtain $\hat{\beta}_{0,j}$ and $\hat{\beta}_{1,j}$ using the method proposed by Koenker and Bassett (1978) with the in-sample data.

to state k to the numbers of times originally in state j. Many researchers have examined in detail the statistical properties of these estimators and provided evidence that these estimators are excellent rules for calculating estimates of the true transition probabilities (e.g., Anderson and Goodman,1957; Lee et al., 1969). Thus, we employ the maximum likelihood estimators with the estimated values of the state variable, $\{\hat{S}_i; i = 2, \ldots, t-1\}$, to obtain the estimates of the transition probabilities. Particularly, we calculate the numbers of changes from Scenario j to Scenario k over two consecutive periods (n_{kj}) as well as the numbers of times that Scenario j occurred during the in-sample period (n_j), and determine the estimates of p_{kj} by

$$\hat{p}_{kj} = \frac{n_{kj}}{n_j}, \text{for } k, j = 1, 2, 3, \tag{5}$$

where $n_j = \sum_{i=2}^{t-1} I_i(j)$ with $I_i(j) = 1$ if $\hat{S}_i = j$, and 0 otherwise, and $n_{kj} = \sum_{i=2}^{t-2} I_i(k,j)$ with $I_i(k,j) = 1$ if $\hat{S}_i = k$ and $\hat{S}_{i-1} = j$, and 0 otherwise.

Furthermore, we use the in-sample data up to time $t-1$ together with the three predicted quartiles for period t to predict the expected excess return under Scenario j for period t, denoted as $\hat{y}_{t,j}(x_{t-1})$, in such a way that

$$\hat{y}_{t,j}(x_{t-1}) = \begin{cases} \frac{1}{n_j+1}\left(\sum_{i=2}^{t-1} y_i I_i(j) + Q_{y_t|x_{t-1}}(\tau_j)\right), & \text{if } j = 1, 3; \\ Q_{y_t|x_{t-1}}(\tau_2), & \text{if } j = 2. \end{cases} \tag{6}$$

Note that the predicted expected excess return under the bad (good) scenario for period t in Equation (6) is the average of the excess returns below (above) the corresponding predicted lower (upper) quartile for the in-sample period up to time $t-1$ and the predicted lower (upper) quartile for time t. We predict the expected excess returns under different scenarios differently with the information relevant to particular scenarios because the excess returns under different scenarios could be distributed differently. In particular, as the predicted lower (upper) quartile of excess returns for period t is the most optimistic (pessimistic) outcome under the bad (good) scenario, averaging it with the past returns under that scenario likely results in a more credible prediction of the expected excess return for that scenario. Besides, the predicted median excess return for period t is used for the predicted expected excess return under the normal scenario because of the specification of this scenario in our study.

Finally, given $\hat{S}_{t-1} = j$, we employ Equation (2) to construct the one-step ahead out-of-sample forecast of the equity premium for period t, denoted as $\hat{y}_t$, by

$$\hat{y}_t = \sum_{k=1}^{3} \hat{p}_{kj} \hat{y}_{t,k}(x_{t-1}), \tag{7}$$

where $\hat{p}_{kj}$ and $\hat{y}_{t,k}(x_{t-1})$ are the estimates of p_{kj} and the predicted values of $E(y_t|S_t = k)$ respectively.

In summary, our prediction method with a single predictor is implemented with the following iterative procedure:

1. Set $\tau_1 = 0.25$, $\tau_2 = 0.5$, $\tau_3 = 0.75$, $t = m$, and $T = m + m_o$ where m and m_o are the in-sample and out-of-sample sizes respectively.
2. Obtain the three predicted quartiles of excess returns $Q_{y_{i+1}|x_i}(\tau_j)$ by fitting the data $\{(y_i, x_i); i = 1, ..., t\}$ to the τ_jth quantile linear regression of excess return y_{i+1} on a predictor x_i for $i = 1, \ldots, t$ and $j = 1, 2, 3$.
3. Estimate the state variable S_i by $\hat{S}_i$ with Equation (4) for $i = 2, \ldots, t$.
4. Estimate the transition probabilities p_{kj} by $\hat{p}_{kj}$ with Equation (5) for $k, j = 1, 2, 3$.
5. Forecast the expected excess returns under each of the three scenarios for period $t+1$,$\hat{y}_{t+1,j}(x_t)$,with Equation (6) for $j = 1, 2, 3$.
6. Forecast the equity premium for period $t + 1$ based on $\hat{S}_t$, $\hat{p}_{kj}$ and $\hat{y}_{t+1,j}(x_t)$ with Equation (7).
7. If $t + 1 < T$, let $t = t + 1$ and go to Step 2; otherwise, the iterative procedure ends.

2.2 The prediction method with multiple predictors

Given a set of predictors, we propose a prediction method that combines the individual forecasts obtained by using the method discussed in Section 2.1 across all predictors to form a single forecast of equity premium. Such a forecast combination allows us to make use of information across the individual forecasts and improve forecast accuracy. Since the seminal work of Bates and Granger (1969), many studies have shown that forecast combinations can improve the forecasting performance over that offered by individual forecasts.[4] For example, for inflation prediction, Stock and Watson (1999)

[4]Timmermann (2006) provides a comprehensive survey of forecast combinations.

and Wright (2004), among others, show that the combination forecasts of individual forecasts using real activity and financial indicators are usually more accurate than individual forecasts. For equity premium prediction, Rapach et al. (2010) provide empirical evidence that the combination forecasts based on the individual forecasts of predictive linear regression with a single predictor significantly increase out-of-sample forecasting ability of multiple predictors and consistently outperform both the historical average forecast and the forecast based on predictive linear regression with multiple predictors.

We consider forecast combinations of individual forecasts in the form of

$$\hat{y}_t = \sum_{j=1}^{K} w_j \hat{y}_{j,t} \tag{8}$$

where $\hat{y}_t$ is the forecast of equity premium for period t that is combined with K individual forecasts of risk premium $\hat{y}_{j,t}$ for period t, and w_j is the weight on the jth individual forecast in the combination with $\sum_{j=1}^{K} w_j = 1$. In our study, we use three forecast combinations, namely, mean, trimmed mean, and median. In the mean combination, all the weights are equal to $1/K$. In the trimmed mean combination, the two weights corresponding to the largest and smallest individual forecasts are zero and the other weights are equal to $1/(K-2)$. In the median combination, the weights are determined in such a way that the combined forecast is the median of K individual forecasts.

2.3 Out-of-sample forecast evaluation

To evaluate the out-of-sample performance of our prediction methods with respect to a competing prediction method, we consider both statistical and economic evaluation criteria. First, for two competing forecast methods, we measure the accuracy of out-of-sample prediction by computing the ratio of the MSFE for forecast method A to MSFE for forecast method B:

$$MSFE_{A|B} = \frac{MSFE_A}{MSFE_B} = \frac{\frac{1}{m_o}\sum_{s=1}^{m_o}(y_{m+s} - \hat{y}_{A,m+s})^2}{\frac{1}{m_o}\sum_{s=1}^{m_o}(y_{m+s} - \hat{y}_{B,m+s})^2}, \tag{9}$$

where m and m_o are in-sample and out-of-sample sizes respectively, and $\hat{y}_{A,t}$ and $\hat{y}_{B,t}$ are the two predicted values of the observed value y_t for time t obtained by using forecast methods A and B, respectively, with data up to time $t-1$. When $MSFE_{A|B} < 1$, it indicates that forecast method A is superior to forecast method B.[5]

Second, to assess the useful information of equity premium forecast in our prediction methods with respect to that in a competing prediction method, we apply the forecast encompassing test, referred to as the $ENC-T$ test, which is developed by Harvey et al. (1998) for examining information differences in forecasting between two competing forecast methods. This test can provide evidence of whether for two competing forecast methods one embodies any useful information absent in the other in terms of out-of-sample prediction. Specifically, it examines the optimal composite of two competing forecasts in the form of

$$\hat{y}_{o,t} = \lambda \hat{y}_{A,t} + (1-\lambda)\hat{y}_{B,t}, 0 \leq \lambda \leq 1, \tag{10}$$

where $\hat{y}_{o,t}$ is the optimal combination of two competing forecasts $\hat{y}_{A,t}$ and $\hat{y}_{B,t}$ with the optimal weight λ of the first competing forecast. It tests the null hypothesis of $\lambda = 0$ against the one-sided alternative hypothesis of $\lambda > 0$. If the test leads to a conclusion of rejecting the null hypothesis, this indicates that the competing forecast A contains useful information of equity premium prediction absent in the competing forecast B. On the other hand, if the test fails to reject the null hypothesis, this means that the competing forecast B is superior to the competing forecast A because the former encompasses all the useful information of equity premium prediction that the latter has. In our study, we use the following test statistic proposed by Harvey et al. (1998):

$$ENC = \frac{\bar{D}}{SE(\bar{D})}, \tag{11}$$

where $\bar{D}$ is the sample mean of the sequence $D_t = (\hat{y}_{A,t} - \hat{y}_{B,t})(y_t - \hat{y}_{B,t})$ over out-of-sample, and $SE(\bar{D})$ is the estimated standard error of $\bar{D}$. Asymptotically, the $ENC-T$ statistic follows the standard normal distribution under the null hypothesis.

Furthermore, by switching the position of the two forecasts in the formulas of $\hat{y}_{o,t}$ and D_t, we can test whether the competing forecast B contains any useful information absent in the competing forecast A. In this case, the null hypothesis of the $ENC-T$ test is equivalent to $\lambda = 1$ under the specification

[5] Out-of-sample R^2 is equal to $1 - MSFE_{A|B}$ when B is the historical average forecast. Hence, $MSFE_{A|B} < 1$ is equivalent to $R^2 > 0$ when B is the historical average forecast.

of Equation (10). Hence, failing to reject the null hypothesis indicates that the forecast A is superior to the forecast B.

Finally, we employ a measure of relative utility gain to assess the out-of-sample performance of our prediction methods with respect to a competing prediction method, which provides an economic evaluation for two competing forecasts of equity premium. The measure of relative utility gain is based on the following presumption: at a given point in time, one equity premium forecast is better than another if investment decisions based on the former lead to higher expected utility. Similarly, a forecast method of equity premium is preferred if, on average, over many periods, it leads to higher expected utility than an alternative forecast method. As West et al. (1993) show that a utility-based measure is fundamentally different from statistical ones based on mean squared forecast error, the relative utility gain provides economic evaluation of competing forecast methods for the performance of out-of-sample prediction.

Particularly, we consider a mean-variance investor who composes his/her portfolio from a risky asset and a risk-free asset and maximizes a mean-variance utility function for period $t+1$ as follows:

$$\max_{w_{p,t+1}} \{E(r_{p,t+1}) - \frac{1}{2}\gamma Var(r_{p,t+1})\}, \tag{12}$$

where $w_{p,t+1}$ is the proportion of the investor's portfolio allocated to the risky asset at the beginning of period $t+1$, $r_{p,t+1}$ is the return of the investor's portfolio over period $t+1$, and γ is the coefficient representing the investor's degree of risk aversion.[6] The optimal portfolio weight of the risk asset for the investor is given by

$$w^*_{p,t+1} = \frac{E(r_{e,t+1}) - r_{f,t+1}}{\gamma Var(r_{e,t+1})}, \tag{13}$$

where $r_{e,t+1}$ and $r_{f,t+1}$ are the return of the risk asset and the return of the risk-free asset for period $t+1$ respectively. Thus, if the investor uses a forecast of equity risk premium for period $t+1$, denoted as $\hat{y}_{A,t+1}$, the optimal portfolio weight of equity is estimated by

$$\hat{w}^*_{p,t+1} = \frac{\hat{y}_{A,t+1}}{\gamma \hat{\sigma}^2_{t+1}}, \tag{14}$$

[6]We choose $\gamma = 3$ in our empirical analysis as in many other studies in the literature.

where $\hat{\sigma}^2_{t+1}$ is the rolling-window estimate of the variance of equity returns.[7] Consequently, for a unit of initial wealth, we calculate the average utility over an out-of-sample period for the investor with forecast method A by

$$\bar{U}^A = (\bar{r}^A - \frac{1}{2}\gamma\bar{\sigma}^2), \tag{15}$$

where $\bar{r}^A$ and $\bar{\sigma}^2$ are the sample means of $r^A_{p,t+1}$ and $\hat{\sigma}^2_{t+1}$ over the out-of-sample period respectively. The relative utility gain for two competing forecast methods of equity risk premium is defined as

$$\Delta\bar{U}_{A|B} = 400(\bar{U}^A - \bar{U}^B). \tag{16}$$

A positive value of $\Delta\bar{U}_{A|B}$ indicates that forecast method A is better than forecast method B in terms of utility gain. Note that we use a scale of 400 to express the utility gain as average annualized percentage return because quarterly data are used in our empirical analysis.

3 Empirical Results

3.1 Data

In our empirical analysis, we forecast the market equity risk premium for the US stock market using the updated quarterly data of Goyal and Welch (2008) with a sample period from the first quarter of 1947 to the third quarter of 2014.[8] The data set includes the continuously compounded $S\&P500$ quarterly return together with the 15 variables defined below.

- *Dividend-price ratio (log), DP,* is the log of 12-month moving sum of dividends paid on the $S\&P500$ index minus the log of stock prices (i.e., the $S\&P500$ index value).
- *Dividend yield (log), DY,* is the log of a 12-month moving sum of dividends minus the log of 1 month-lagged prices of the stocks included in the $S\&P500$ index.
- *Earning-price ratio (log), EP,* is the log of a 12-month moving sum of earnings on the S&P 500 index minus the log of stock prices.

[7] As in Campbell and Thompson (2008) and Rapach et al. (2010), we use a ten-year rolling window of quarterly returns up to period t to calculate the sample variance as the value of $\hat{\sigma}^2_{t+1}$

[8] The data can be retrieved from the website: http://www.hec.unil.ch/agoyal/ and a detailed description can be found in Goyal and Welch (2008).

- *Dividend-payout ratio (log), DE,* is the log of a 12-month moving sum of dividends minus the log of a 12-month moving sum of earnings.
- *Book-to-market ratio, BM,* of the Dow Jones Industrial Average.
- *Treasury bill rate, TBL,* is the interest rate on a 3-month Treasury bill from secondary market.
- *Default yield spread, DFY,* is the difference in yield between Moody's BAA- and AAA-rated corporate bonds.
- *Long-term yield, LTY,* on long-term government bonds.
- *Term spread, TMS,* is the difference between the long-term government bond yield and the Treasury bill rate.
- *Net equity expansion, NTIS,* is the ratio of a 12-month moving sum of net equity issues by NYSE-listed stocks to the total end-of-year market capitalization of NYSE stocks.
- *Inflation, INFL* , is calculated from the Consumer Price Index (CPI) for all urban consumers.
- *Long-term return, LTR,* on long-term government bonds.
- *Default return spread, DFR,* is the difference in return between long-term corporate and long-term government bonds.
- *Stock variance, SVAR,* is a sum of squared daily returns on the $S\&P500$ index.
- *Investment-to-capital ratio, IK* , is the ratio of aggregate private nonresidential fixed investment to aggregate capital for the entire economy.

In our study, we use the excess return on the $S\&P500$ index over the three-month Treasury bill rate as the response variable (y_t) and the above 15 variables as individual predictors (x_t). We divide the full sample into in- and out-of-samples with the two different out-of-sample periods starting from either the first quarter of 1965 or the first quarter of 1976 to the third quarter of 2014. Note that the starting times of the two out-of-sample periods are the same as in Goyal and Welch (2008) and Rapach et al. (2010).

3.2 Alternative prediction methods

To evaluate the out-of-sample performance of our prediction methods with respect to a competing prediction method, we consider two alternative methods of equity premium prediction from the literature, namely, historical average (HA), and predictive linear regression (OLS). The first method is often served as the prevailing benchmark method in many studies on equity premium prediction in the literature, while the second method is the traditional approach of equity premium prediction with either a single predictor or multiple predictors, which is used in many previous studies (e.g., Goyal and Welch 2008; Rapach et al. 2010). The HA forecast for period t is defined by $\hat{y}_{t,HA} = \bar{y}_{t-1}$, where $\bar{y}_{t-1}$ is the average of the observed excess returns over an in-sample period up to $t-1$. The OLS forecast with a single predictor for period t is given by

$$\hat{y}_{t,OLS} = \hat{\beta}_{0,OLS} + \hat{\beta}_{1,OLS} x_{t-1}, \tag{17}$$

where $\hat{\beta}_{0,OLS}$ and $\hat{\beta}_{1,OLS}$ are the ordinary least squares estimates of the predictive linear regression of excess return y_i on a predictor x_{i-1} with a data set up to time $t-1$, $\{(y_i, x_i); i = 1, ..., t-1\}$.

For combination prediction with all the 15 predictors, we also consider the OLS combination prediction as an alternative to our method with multiple predictors. The OLS combination prediction generates a single equity premium forecast by combining individual single-predictor OLS forecasts across all the predictors with mean, median, and trimmed combination methods separately.

3.3 Results with a single predictor

Panel A of Table 1 reports the results of the out-of-sample performance for our prediction method with a single predictor (SAM) for the out-of-sample period from the first quarter of 1965 to the third quarter of 2014. First, we compare the SAM with the benchmark method of historical average (HA) in terms of the accuracy of out-of-sample prediction. As shown in the second column of Panel A, the ratio of the MSFE for the SAM to the MSFE for the HA is less than 1 for each predictor. This indicates that for all the 15 predictors, our prediction method outperforms the benchmark method of historical average with more accurate forecasts than the benchmark forecast.

[Insert Table 1 here]

Second, we compare the SAM with the HA in terms of useful information on equity premium forecasting in a prediction method. The fourth and sixth columns of Panel A report the p-values of the two encompassing tests for comparison between the SAM and HA forecasts. For either of the two tests, the null hypothesis $H_o : \lambda = 0$ is against the alternative hypothesis $H_a : \lambda > 0$, where λ is the parameter in the optimal composite of two competing forecasts (A and B), defined by $\hat{y}_{o,t} = \lambda\hat{y}_{A,t} + (1 - \lambda)\hat{y}_{B,t}$. In the fourth column of panel A, the p-values of the first test with A=SAM and B=HA are less than 5% for 11 of the 15 predictors, except DP, DY, EP and BM. Hence, for these 11 predictors, we reject the null hypothesis at significance level of 5%, and conclude that for each of the 11 predictors, the SAM contains useful information on equity risk premium forecasting which the HA method does not encompass. In the sixth column of Panel A, the p-values of the second test with A=HA and B=SAM are larger than 5% for all the predictors. Thus, for each of the predictors, we fail to reject the null hypothesis in the second test at significance level of 5%, and conclude that the SAM contains all the information on equity risk premium forecasting that the HA method contains. The results of the two encompassing tests lead to the following two conclusions. (1) For the four predictors (DP, DY, EP and BM), the SAM contains the same useful information as the HA method has; and (2) for the other 11 predictors, the SAM has more useful information than the HA method. On the whole, the findings suggest that the SAM likely has more useful information on equity risk premium prediction than the HA method.

Third, we evaluate the out-of-sample performance of the SAM in terms of economic gains relative to the HA. The differences in utility between the SAM and HA ($\Delta\bar{U}_{A|B}$) in the eighth column of Panel A are positive for all the predictors but DP and IK. This suggests that in general, our prediction method with a single predictor likely generates more economic profits than the benchmark method.

Likewise, we compare the SAM with the traditional method of predictive linear regression with a single predictor (OLS) in the same way as for the comparison between the SAM and HA, and report the results in the four columns (3, 5, 7 and 9) of Panel A. First, as the ratio of the MSFE for the SAM to the MSFE for the OLS in the third column is less than 1 for each of the 15 predictors, it indicates that the SAM forecasts are more accurate than the OLS forecasts. Second, the p-values of the first encompassing test in the fifth column are all less than 5% for the all the predictors but one (SVAR). Thus, we reject the null hypothesis at significance level of 5%, and conclude that for each of the 14 predictors, the SAM contains some useful information on equity premium prediction that the OLS does not have. On the other hand, since the p-values of the second encompassing test in the seventh column exceed 5% for the all the predictors, we fail to reject the null hypothesis at significance level of 5%, and conclude that for each of the predictors, the SAM encompasses all the information that the OLS contains. The results of the two tests suggest that the SAM likely contains more useful information than the OLS. Third, the differences in utility between the SAM and OLS in the ninth column are positive for all the predictors but one (TMS). This provides evidence that the SAM forecasts likely generate more economic profits than the OLS forecasts.

Furthermore, we also assess the out-of-sample performance of the SAM for an alternative out-of-sample period from the first quarter of 1976 to the third quarter of 2014, and present the results in Panel B of Table 1 in the same way as in Panel A. According to the MSFE ratio, the SAM forecasts are more accurate than the HA forecast for 9 out of the 15 predictors and the OLS forecasts for all the predictors. As indicated by the two encompassing tests at significance level of 5%, the information on equity premium forecasting in the SAM is more than that in the HA for six predictors, less than that in the HA for one predictor (BM), and the same as that in the HA for the rest of the predictors. Moreover, with respect to the OLS, the SAM has more information for 12 predictors and the same amount of information for the rest of the predictors. In terms of the difference in utility between two forecasts, the SAM forecasts generate more economic profit than the HA forecast for all the predictors, and the OLS forecasts for all the predictors but one (TMS). These results in Panel B are quite similar to those in Panel A.

For our prediction method with a single predictor, the forecast of equity premium for period t is determined based on the information that a particular scenario occurred for period $t-1$, as indicated in Equation (7). Thus, we further examine the out-of-sample predictive performance of the SAM for each of the three scenarios by measuring the accuracy of the SAM prediction made under each of the three scenarios. Specifically, for each of the predictors, we first divide the out-of-sample SAM forecasts into three groups in such a way that the out-of-sample SAM forecast for period t is in group j if $\hat{S}_{t-1} = j$ $(j = 1, 2, 3)$, and also classify either the HA forecasts or the OLS forecasts into the three groups accordingly. Then we evaluate the accuracy of the SAM prediction for each of the three groups relative to either the HA or the OLS prediction by calculating the ratio of the MSFE for the SAM to the MSFE for either the HA or the OLS and conducting the two encompassing tests. The results for the out-of-sample period from the first quarter of 1965 to the third quarter of 2014 are reported in Table 2 .

[Insert Table 2 here]

In terms of the accuracy of out-of-sample prediction with respect to the HA, the SAM prediction is more accurate for 6 predictors under Scenario 1 (the bad scenario), for 10 predictors under Scenario 2 (the normal scenario), and for 12 predictors under Scenario 3 (the good scenario) respectively.[9] Besides, with respect to the OLS prediction, the SAM prediction is more accurate for all the 15 predictors under the bad scenario, and for 10 predictors under either the normal or the good scenario respectively.

In addition, at significance level of 5%, the two encompassing tests lead to the following two conclusions regarding the useful information on equity premium prediction in the SAM relative to either the HA or the OLS. First, the information in the SAM relative to the HA is more for 1 predictor, less for 2 predictors, and the same for the other predictors under Scenario 1, more for 6 predictors and the same for the other predictors under Scenario 2, and more for 8 predictors and the same for the other predictors under Scenario 3. Second, the information in the SAM relative to the OLS is more for 4 predictors and the same for the other predictors under Scenario 1, more for 3 predictors and the same for the other predictors under Scenario 2, and more for 8 predictors, less for 3 predictors, and the same for the other predictors under Scenario 3.

Moreover, as the out-of-sample SAM prediction with a single predictor is related to the transition probabilities estimated with in-sample data as shown in Equation (7), we look into how the transition probabilities evolve over time for each of the predictors by plotting it over the out-of-sample period from the first quarter of 1965 to the third quarter of 2014. We find that these probabilities vary over

[9] Note that the predictability of a scenario is evaluated over the whole out-of-sample period including both the recession and expansion periods as defined by the US recession indicator, and 14% of time during the whole out-of-sample period is the recession periods and 86% of time the expansion periods.

time considerably for each predictor. To save space, we only show the plots for the predictor DP in Figure 1 as an example.

[Insert Figure 1 here]

The three plots in the first column of Figure 1 are the probabilities of the transition from Scenario 1 to Scenarios 1, 2 and 3 respectively. For the transition probability p_{11}, it rises to a high level near 0.45 during 1970s, and then drops by more than 30% to a level below 0.3 in the early 1980s. For the transition probability p_{21}, it is at a low level around 0.35 in the early 1970s, and increases by more than 28% to a high level above 0.45 in 1980s. For the transition probability p_{31}, it has a sharp change of about 75% from about 0.16 to near 0.28 over a short period from the late 1990s to the early 2000s. These plots in Figure 1 show a considerable time-variation in these probabilities with different patterns of variation.

Finally, we examine the occurrence of the three scenarios over time by plotting the estimated values of the state variable over the out-of-sample period for each predictor. In general, it appears that in these plots, the occurrence of the three scenarios resembles a realization of a three-state Markov chain in which no particular scenario dominates the whole period, and the bad (good) scenario likely occurs during periods when the US economic conditions are bad (good) for stock investments. We show the plot of $\hat{S}_{t-1}$ over the out-of-sample period for the predictor DP in Figure 2 as an example.

[Insert Figure 2 here]

Figure 2 indicates that the bad scenario, for example, occurred in the late 2000s, which coincides with the global financial crisis of 2007-2008.

In summary, the empirical findings indicate that our prediction method with a single predictor could improve the accuracy of out-of-sample prediction relative to the two alternative methods, and likely contains more information on equity premium forecasting and generates more economic profits than the two alternative methods. Moreover, the findings also suggest that all three scenarios rather than a particular scenario contribute the improvement in the out-of-sample prediction for our method. Finally, the level of time-variation in the probabilities used in our forecasts is sizable; and the occurrence of a scenario likely corresponds to the state of the economy in a country or region, which we will further discuss in Section 3.5.3.

3.4 Results with multiple predictors

Table 3 reports the results of the out-of-sample performance for our prediction method with multiple predictors (SAM) for the same out-of-sample periods as in the analysis of our method with a single predictor. These results lead to the following findings for each of the three combination methods: mean, median, and trimmed mean. First, the SAM forecasts are more accurate than the two alternative forecasts: the HA and the OLS combination forecasts, because the ratios of the MSFE for the SAM to the MSFE for either of the HA or the OLS are all less than 1 as shown in the second and third columns of Panels A and B. Second, the SAM contains more useful information on equity premium prediction than either of the HA or the OLS because at significance level of 5%, the null hypothesis is rejected for the first encompassing test, but not rejected for the second encompassing test. [10] Third, the SAM forecasts generate more economic gains than the HA and OLS forecasts as the differences in utility between the SAM and HA or between the SAM and the OLS ($\Delta \bar{U}_{A|B}$) are all positive as shown in the last two columns of Panels A and B. These empirical findings show that our prediction method with multiple predictors consistently improves the performance of out-of-sample prediction of equity premium relative to the HA and the OLS combination methods.

[Insert Table 3 here]

3.5 Predictability under different economic conditions

In this section, following the previous studies on the link between equity returns and real economy (e.g., Liew and Vassalou, 2000; Rapach et al., 2010), we examine the out-of-sample predictive performance of our prediction methods under different economic conditions that are characterized in two different ways. We also analyze the association between the three scenarios and the state of an economy by examining the occurrence of a scenario and the transition probabilities of remaining in the same scenario over two consecutive periods under different economic conditions.

[10] See the p-values of the tests in the four columns: 4, 5, 6 and 7 of Table 3.

3.5.1 Predictive performance during the three different GDP growth periods

First, we define three different economic conditions in terms of the real US GDP growth rate and split the out-of-sample period from the first quarter of 1965 to the third quarter of 2014 into three different GDP growth periods: low, average, and high GDP growth periods. Particularly, we use either the 20th and 80th percentiles or the 25th and 75th percentiles of the real GDP growth rates to divide the out-of-sample period into the three GDP growth periods,[11] and report the results of the out-of-sample predictive performance in Table 4.

[Insert Table 4 here]

Panel A of Table 4 shows the MSFE ratios and the p-values of the two encompassing tests regarding information about equity premium forecast in a prediction method for the SAM with a single predictor and the three different GDP growth periods specified by the 20th and 80th percentiles of the real US GDP growth rates. Based on the MSFE ratios in Panel A, out of the 15 predictors, the SAM forecasts are more accurate than the HA forecast for 12, 9, and 13 predictors during the low, average, and high GDP growth periods respectively. Similarly, the SAM forecasts are more accurate than the single-predictor OLS forecasts for 11, 14, and 10 predictors during the low, average, and high GDP growth periods respectively. The findings provide evidence that for most of the predictors, the SAM with a single predictor could improve out-of-sample equity premium prediction relative to either the HA or the OLS during each of the three GDP growth periods.

Based on the p-values of the two encompassing tests in Panel A, the two tests lead to the following conclusions at significance level of 5%. Compared with the HA, the information in the SAM with a single predictor is more for 5, 7, and 8 predictors, and the same for 10, 7, and 7 predictors during the low, average, and high GDP growth periods respectively, and less for one predictor only during the average GDP growth period. Compared with the OLS, the information in the SAM with a single predictor is more for 4, 10, and 6 predictors, the same for 11, 5, and 8 predictors during the low, average, and high GDP growth periods respectively, and less for one predictor only during the high GDP growth period. The results suggest that the SAM with a single predictor likely has more useful

[11] The quarterly real US GDP growth data in our study can be download from the website of the US. Bureau of Economic Analysis: http://www.bea.gov/national/.

information on equity premium prediction than either the HA or the OLS during each of the three periods.

In addition, we use the 25th and 75th percentiles of the US real GDP growth rates to define an alternative specification of the three GDP growth periods, repeat the same analysis as in Panel A, and present the results in Panel B. The results in Panels A and B are quite similar. For example, based on the MSFE ratios in Panel B, the SAM forecasts are more accurate than the HA forecast for 13, 11, and 10 predictors during the low, average, and high GDP growth periods respectively, and also more accurate than the OLS forecasts for 12, 12, and 11 predictors during the three GDP growth periods respectively.

Panels C and D of Table 4 report the results of the SAM with multiple predictors for the two specifications of the three GDP growth periods respectively. Compared with the HA forecast, the three SAM combination forecasts are more accurate during each of the three GDP growth periods for both the specifications of the three periods because the MSFE ratios relative to the HA are all less than 1 in Panels C and D. Furthermore, the two encompassing tests regarding useful information in a prediction method lead to the following conclusions at significance level of 5%. Each of the three SAM combination prediction methods contains more information than the HA for all the cases evaluated in these two panels except the cases for the high GDP growth period and the second pacification of the three periods, for which both the SAM and HA contain the same information.

Similarly, compared with the OLS combination forecasts, the three SAM combination forecasts are more accurate during either the low or the average GDP growth period for the two specifications and more accurate during the high GDP growth period for the first specification. The two encompassing tests lead to the conclusions at significance level of 5% that the three SAM combinations have more information than the OLS combinations during the average GDP growth period for the second specification and the same information as the OLS combinations for all the other cases evaluated in the two panels.

These findings from Panels C and D suggest that the SAM combination forecasts likely perform better during the low and average GDP growth periods than during the high GDP growth period.

3.5.2 Predictive performance during the recession and expansion periods

Second, we define different economic conditions based on the data of the US recession indicator with 1 for recession and 0 for expansion, produced by the National Bureau of Economic Research (NBER)[12]. We then divide the out-of-sample period into the recession and expansion periods according to the values of the indicator, and report the results of the out-of-sample predictive performance of the SAM during the two periods in Table 5.

[Insert Table 5 here]

Panel A of Table 5 reports the MSFE ratios and the p-values of the two encompassing tests for the SAM with a single predictor during the recession and expansion periods. According to the MSFE ratios, out of the 15 predictors, the SAM forecasts are more accurate than the HA forecast for 14 and 10 predictors during the recession and expansion periods respectively, and more accurate than the OLS forecasts for 12 and 14 predictors during the two periods respectively. Moreover, the two encompassing tests lead to the following conclusions at significance level of 5%. Compared with the information in the HA, the information in the SAM with a single predictor is more for 7 predictors and the same for 8 predictors during the recession period, and more for 7 predictors, the same for 2 predictors, and less for 2 predictors during the expansion period. Likewise, compared with the information in the OLS, the information in the SAM with a single predictor is more for 7 predictors and the same for 8 predictors during the recession period, and more for 12 predictors and the same for 3 predictors during the expansion period. These results indicate that the out-of-sample predictive performance of the SAM with a single predictor is similar during the two periods, and better than that of the HA and OLS for most of predictors during each of the two periods.

Panel B of Table 5 reports the results for the SAM with multiple predictors during the recession and expansion periods respectively. Based on the MSFE ratios, the three SAM combination forecasts are all more accurate than the HA and OLS combination forecasts during the recession period, and

[12]The data were from the website of the Federal Reserve Bank of St. Louis: https://fred.stlouisfed.org/series/USREC, and available up to the third quarter of 2014 at the time when we carried out this study.

more accurate than the HA forecast and less accurate than the OLS combination forecasts during the expansion period. The p-values of the two encompassing tests in the panel lead to the following conclusions at significance level of 5%. The information on equity premium prediction in the SAM with multiple predictors is more than that of the HA and OLS methods for all the three combinations during the recession period, and more than that of the HA and less than that of the OLS method for all the three combinations during the expansion period. The results suggest that relative to the HA and OLS combination methods, the SAM with multiple predictors performs better during the recession period than during the expansion period.

3.5.3 The association between the three scenarios and the state of the economy

Based on the estimated values of the state variable over the out-of-sample period for each predictor, we calculate the percentages of occurrences for each of the three scenarios during the recession and expansion periods and during the three GDP growth periods respectively, and report the results in Table 6.

[Insert Table 6 here]

For the bad scenario, the percentage of its occurrences ranges from 52% to 74% for different predictors with an average of 63% over all the 15 predictors during the recession periods, and from 15% to 26% with an average of 20% during the expansion periods. For the average scenario, the percentage of its occurrences varies from 11% to 30% for different predictors with an average of 20% over all the 15 predictors during the recession periods, and from 48% to 58% with an average of 54% during the expansion periods. For the good scenario, the percentage of its occurrences varies from 7% to 30% for different predictors with an average of 17% over all the 15 predictors during the recession periods, and from 22% to 33% with an average of 26% during the expansion periods. These findings show that the bad scenario is most likely the prevailing scenario during the recession periods, while the average scenario is most likely the prevailing scenario during the expansion periods. They also indicate that the good scenario is more likely to occur during the expansion periods than during the recession periods. In addition, the findings suggest that the association between the occurrence of a scenario and the state of the economy may vary from one predictor to another.

Similarly, the percentages of occurrences for each of the three scenarios vary over the three different GDP growth periods. Specifically, on average over all the 15 predictors, the percentages of occurrences are 55%, 38%, and 7% for the bad, average, and good scenarios, respectively, during the low GDP growth periods, 21%, 51%, and 28% during the average GDP growth periods, and 15%, 56%, and 29% during the high GDP growth periods. This means that the bad scenario is most likely the prevailing scenario during the low GDP growth periods and the average scenario is most likely the prevailing scenario during both the average and high GDP growth periods. Furthermore, the good scenario is more likely to occur during the high GDP growth period than during the other two GDP periods and the least likely to occur during the low GDP growth periods.

We further examine the transition probability of remaining in the same scenario over two consecutive periods (p_{ii}) under different economic conditions.[13] Particularly, for each predictor, we calculate the averages of the estimated values of p_{ii} over the recession, expansion, and the three GDP growth periods respectively, and report them in Table 7.

[Insert Table 7 here]

Table 7 shows that the average of the estimated values of p_{ii} varies for different predictors as well as over different periods. On average over all the 15 predictors, the averages of the estimated p_{ii} are 0.28, 0.03, and 0.14 for the bad, average, and good scenarios respectively during the recession periods, and 0.06, 0.30, and 0.07 respectively during the expansion periods. Similarly, these averages are 0.27, 0.11, and 0.04 for the bad, average, and good scenarios respectively during the low GDP growth periods, 0.06, 0.30, and 0.08 respectively during the average GDP growth periods, and 0.02, 0.30, and 0.11 respectively during the high GDP growth periods. These findings indicate that on average, the bad scenario stays much longer than the other two scenarios during the recession periods, while the average scenario stays much longer than the other two scenarios during the expansion periods. In addition, the bad scenario stays the longest during the low GDP growth periods among the three different GDP growth periods; the average scenario stays much longer during the average and good GDP growth periods than during the low GDP growth period; and the good scenario remains in the same scenario the longest during the high GDP growth periods.

[13]Note that the expected duration of scenario i is equal to $1/(1 - p_{ii})$.

Our findings provide empirical evidence on how the three scenarios are associated with the state of the economy. They suggest that most of the predictability of our single-predictor method comes from the bad scenario during the recession periods and from the average scenario during expansion periods respectively. By contrast, the predictability of either the HA or the single-predictor OLS method is always based on one event characterized by the mean of the distribution of excess returns. Finally, because of the association between the scenarios and the state of the economy, modelling the evolution of the three scenarios over time likely further improves the predictability of our single-predictor method relative to the HA and the single-predictor OLS methods.

3.6 Results with an alternative specification of the three scenarios

Instead of using the lower and upper quartiles of excess returns to define the three scenarios for the SAM as in Section 2.1, we employ the 30th and 70th percentiles of excess returns as an alternative to specify the three scenarios. Specifically, our prediction method with a single predictor is carried out in the same way as in Section 2.1 with $\tau_1 = 0.3$, $\tau_2 = 0.5$, and $\tau_3 = 0.7$. The results for the out-of-sample performance of the SAM for the same out-of-sample period as in both Panel A of Table 1 and Panel A of Table 3 are presented in Table 8.

[Insert Table 8 here]

Both the MSFE ratios and the p-values of the two encompassing tests in Panels A and B of Table 8 are quite similar to the corresponding results in Panel A of Table 1 for the SAM with a single predictor and Panel A of Table 3 for the SAM with multiple predictors respectively. Therefore, these results provide evidence that the out-of-sample performance of our prediction methods is consistent for the two different specifications of the three scenarios.

3.7 Predictive performance relative to the prediction methods of MPVV(2014)

In this section, we examine the accuracy of the out-of-sample prediction for our prediction methods relative to the methods studied in Meligkotsidou et al. (2014), referred to as MPVV(2014). In their paper, they consider two combination methods for equity premium prediction, namely, the robust forecast combination (RFC) and the quantile forecast combination (QFC). For the RFC, they first generate the quantile forecasts for the distribution of the excess returns for the next period by running quantile regression with a single predictor for different quantiles and combine it across all the chosen quantiles to produce point forecasts for each predictor. Then they obtain a final point forecast of equity premium by combining the point forecasts across all predictors with either a fixed or a time-varying weighting scheme. For the QFC, they first generate point forecasts for different quantiles by combining the quantile forecasts for the same quantile across all predictors, and then obtain a final point forecast of equity premium by combining the point forecasts across all the chosen quantiles with either a fixed or a time-varying weighting scheme. They employ the updated quarterly data of Goyal and Welch (2008) for the period from the first quarter of 1947 to the fourth quarterly of 2010 with the out-of-sample period from the first quarter of 1965 to the fourth quarter of 2010, and evaluate the accuracy of the out-of-sample prediction of their methods in terms of MSFE.

Clearly, the study of MPVV(2014) focuses on combination forecasts only, while our study in this paper focuses more on single-predictor forecasts than on combination forecasts. More specifically, our prediction methods differ from that of MPVV(2014) in two aspects. First, our method with a single predictor predicts the expected excess return achieved under each scenario, while the methods of MPVV (2014) focus on quantile forecasts of excess returns for each predictor or a quantile. Second, the probability weights in our single-predictor forecasts are the transition probabilities that are associated with the predicted expected returns under each scenario through a Markov chain governing the scenarios. On the other hand, the weighting schemes used in MPVV(2014) are not necessarily related to point forecasts for each predictor or a quantile. As for combination forecasts, while the study of MPVV(2014) includes many different combination methods with various weighting schemes, our study only considers the three usual combinations of single-predictor forecasts.[14]

Using the same data as in MPVV(2014), we compare our methods with theirs as well as the predictive individual mean regression (IMR) and the mean forecast combination (MFC) methods evaluated in MPVV(2014) in terms of the accuracy of out-of-sample prediction, and report the results in Tables 9 and 10.

[14]Although it is worth exploring other combination methods with various weighting schemes based on the predictive performance of our single-predictor forecasts, this is a research project that we wish to explore in future research.

[Insert Table 9 here]

Table 9 reports the MSFE ratios of either the IMR or the SAM with a single predictor to the prevailing benchmark method for the out-of-sample period from the first quarter of 1965 to the fourth quarter of 2010. For each predictor, the MSFE ratio for the SAM in the second row is less than the corresponding MSFE ratio for the IMR in the first row. This provides evidence that the SAM forecasts likely perform better for out-of-sample equity premium prediction than the IMR forecasts.

[Insert Table 10 here]

Table 10 reports the MSFE ratios of various prediction combination methods to the prevailing benchmark method for the same out-of-sample period as in Table 9. Note that all the MSFE ratios in the table except the three in the last row for the SAM mean, median and trimmed mean combinations respectively are for the prediction combination methods evaluated in MPVV(2014). In this table, the MSFE ratios are 0.9585, 0.9652 and 0.9603 for the SAM mean, median, and trimmed mean combination forecasts respectively, with each less than the corresponding smallest MSFE ratios of 0.9594, 0.9669, and 0.9619 for the best mean, median, and trimmed mean combination forecasts, respectively, studied in MPVV (2014). This shows that for each of these three combination methods, the predictive performance of our combination forecast is at least as good as that of the best combination forecast in MPVV(2014). In addition, 0.9680 is the smallest MSFE ratio for the other combination forecasts reported in the columns 5-11 of the table, which is larger than the largest MSFE ratio of 0.9603 among the three SAM combination forecasts. Thus, these results show that the predictive performance of the best SAM combination forecast is at least as good as that of the best combination forecast in MPVV(2014).

4 Conclusions

In this paper, we propose two methods of equity premium prediction with a single predictor and multiple predictors respectively. The method with a single predictor defines three possible scenarios for a stock investment in terms of the low, normal, and high investment returns expected under each of the three scenarios respectively, and assumes a first-order three-state Markov chain governing the occurrence of the three scenarios over time. It employs predictive quantile regression of excess return on a predictor to estimate the occurrence of the scenarios for an in-sample period as well as the transition probabilities of the Markov chain, and to predict the expected excess returns under each of the three scenarios. It then generates a one- step ahead out-of-sample forecast of equity premium by combining the predicted expected excess returns under the three scenarios with the estimated transition probabilities, based on which scenario is estimated to have occurred in the end of the in-sample period. The method with multiple predictors combines the forecasts of our method with a single predictor across all predictors to generate a single forecast of equity premium.

Using the US stock data, we have evaluated the out-of-sample performance of our two methods relative to either the prevailing benchmark method of historical average or the traditional predictive regression. Our forecasts with a single predictor generate forecasting gains and economic profits relative to the alternative forecasts for most of predictors. Moreover, the improvement in the out-of-sample predictive performance is similar regardless under which scenario we generate our forecasts. In addition, our forecasts with multiple predictors perform better both statistically and economically than the HA and OLS combination forecasts for all the three combination methods. Besides, we have compared our combination methods with the various combination methods studied in MPVV(2014) and shown that the predictive performance of the best forecast for our combination methods is at least as good as that of the best forecast for their combination methods. Overall, the out-of-sample performance of our two methods is consistent for the two different sample periods and the two different specifications of the three scenarios.

Furthermore, we have examined the out-of-sample predictive performance of our prediction methods under different economic conditions in two different ways. First, we define three periods of low, average and high GDP growth in terms of the real US GDP growth rate to characterize different economic conditions. We have found that our forecasts with a single predictor outperform the HA and OLS forecasts for most of predictors across all the three GDP growth periods. On the other hand, our forecasts with multiple predictors consistently outperform the HA and OLS combination forecasts during the low and average GDP growth periods. Second, we define two periods of recession and expansion in terms of the US recession indicator to characterize different economic conditions. In this case, we have found that our forecasts with a single predictor outperform the HA and OLS forecasts for most of predictors across the two periods, and our forecasts with multiple predictors consistently outperform the HA and OLS combination forecasts during the recession period. Our findings imply that the out-of-sample predictive performance of our method with a single predictor is likely independent of economic conditions. Our findings also suggest that the out-of-sample predictive performance of our methods with multiple predictors is likely associated with business cycle.

In addition, we have examined the association between the three scenarios and the state of the economy. Our findings show that the predictability of our single-predictor method comes from the three scenarios and the contribution of a scenario to the predictability varies over time according to the association between the three scenarios and the state of the economy. On the other hand, the predictability of either of the two alternative methods: the HA and the single-predictor OLS prediction methods always comes from one event characterized by the mean of the distribution of excess returns. Our findings also suggest that modelling the stochastic process of the three scenarios by the first-order Markov chain improves the predictability of equity risk premium relative to the two alternative methods. Hence, the difference in the predictive framework between our single-predictor method and the two alternative methods may explain for why our forecasts likely perform better.

Finally, out-of-sample equity premium prediction is a challenging task. While we surely do not claim that our prediction methods can outperform the alternative methods over every possible out-of-sample period, our findings suggest that it is important to predict the expected returns under different scenarios for an investment differently and take the stochastic features of the scenarios over time into account in equity premium prediction. Our findings also provide evidence that our prediction methods are useful alternatives to the several forecasting methods commonly used for out-of-sample equity premium prediction in the economics and finance literature.

References

Anderson, TW, Goodman, LA. 1957. Statistical inference about Markov chains. *the Annals of Mathematical Statistics* **28**: 89-110.

Bates, JM, Granger, CWJ. 1969. The combination of forecasts. *Operational Research Quarterly* **20**: 451-468.

Campbell JY, Thompson SB. 2008. Predicting excess stock returns out of sample: can anything beat the historical average? *Review of Financial Studies* **21**: 1509-1531.

Cochrane, JH. 2011. Presidential address: discount rates. *Journal of Finance* **66**: 1047-1108.

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

Engle, RF, Manganelli, S. 1999. CAViaR: conditional value at risk by quantile regression. *National Bureau of Economic Research Working Paper* No. 7341.

Fama, EE, French, KR. 1989. Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics* **25**: 23-49.

Goyal A, Welch I. 2008. A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies* **21**: 1455-508.

Guidolin, M, Timmermann, A. 2009. Forecasts of US short-term interest rates: a flexible forecast combination approach. *Journal of Econometrics* **150**: 297-311.

Harvey, DI, Leybourne, SJ, Newbold, P. 1998. Tests for forecast encompassing. *Journal of Business and Economic Statistics* **16**: 254-259.

Henkel, SJ, Martin, JS, Nadari, F. 2011. Time-varying short-horizon predictability. *Journal of Financial Economics* **99**: 560-580.

Koenker, R, Bassett, G. 1978. Regression quantiles. *Econometrica* **46**: 33-50.

Lee, TC, Judge, GG, Cain, RL. 1969. A sampling study of the properties of estimators of transition probabilities. *Management Science* **15**: 373-398.

Liew, J, Vassaloum, M. 2000. Can book-to-market, size, and momentum be risk factors that explain economic growth. *Journal of Financial Economics* **57**: 221-245.

Meligkotsidou, L, Panopoulou, E, Vrontos, ID, Vrontos, SD. 2014. A quantile regression approach to equity premium prediction. *Journal of Forecasting* **33**: 558-576.

Rapach D, Strauss J, Zhou G. 2010. Out-of-sample equity premium prediction: combination forecasts and links to the real economy. *Review of Financial Studies* **23**: 821-1862.

Rapach D, Zhou, G. 2013. Forecasting stock returns. In *Handbook of Economic Forecasting*, Vol. 2A, G. Elliott, A. Timmermann (eds). Elsevier: Amsterdam; 328-383.

Stock, JH, Watson, MW. 1999. Forecasting inflation. *Journal of Monetary Economics* **44**: 293-335.

Timmermann A. 2006. Forecast combinations. In *Handbook of Economic Forecasting*, Vol. I, G. Elliott, C. W. J. Granger, A. Timmermann (eds). Elsevier: Amsterdam; 135-196.

West, K, Edison, H, Cho, D. 1993. A utility-based comparison of some models of exchange rate volatility. *Journal of International Economics* **35**: 23-46.

Wright, JH. 2004. Forecasting U.S. inflation by Bayesian model averaging. Working paper, Federal Reserve Board of Governors.

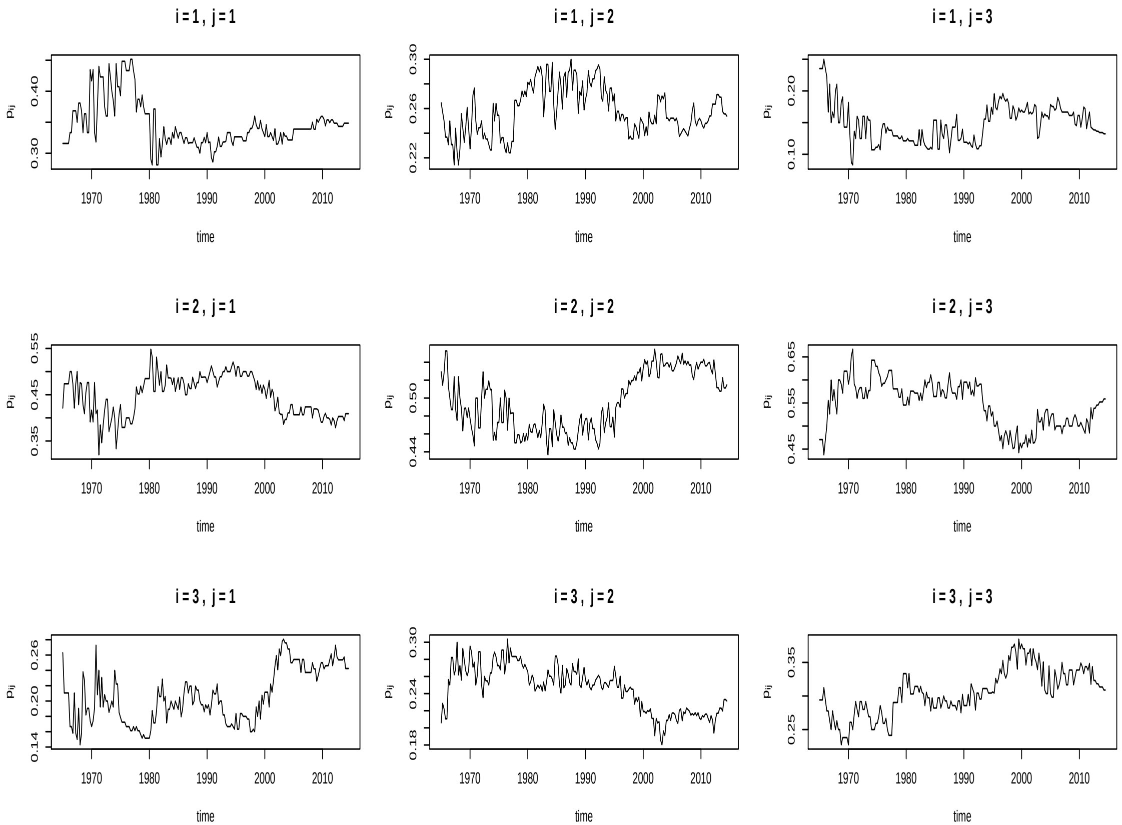

Figure 1: Time series plot of the transition probabilities for predictor DP over the out-of-sample period from Q1, 1965 to Q3, 2014

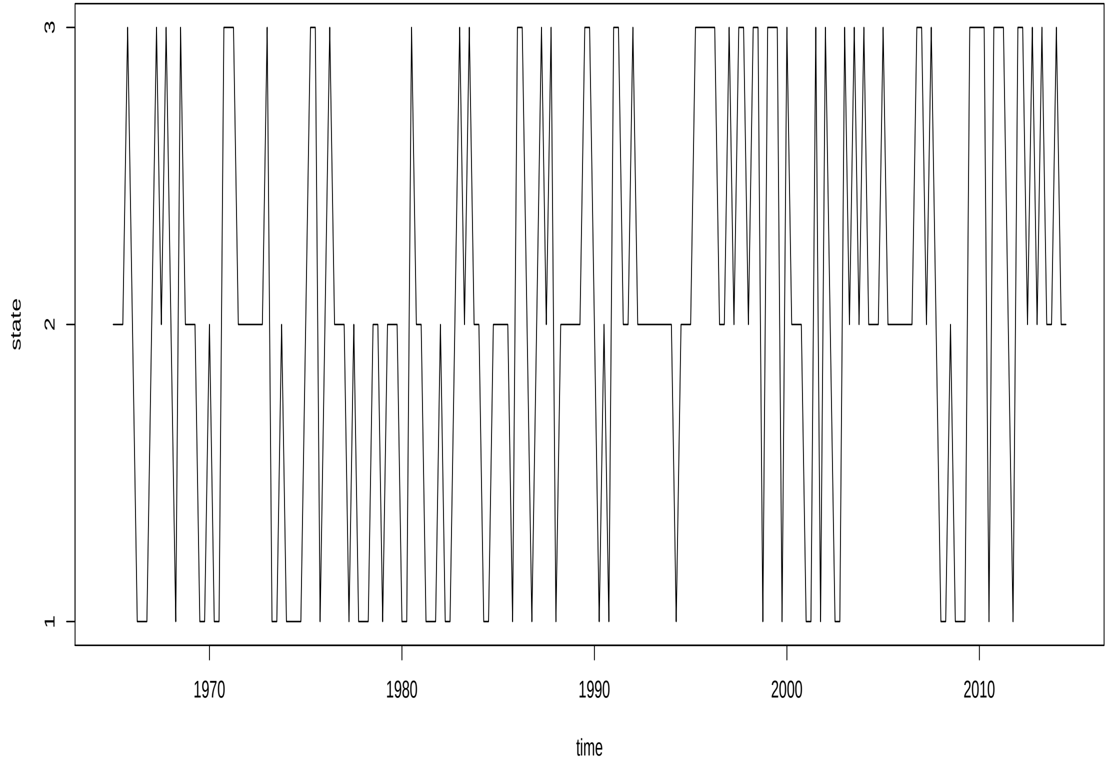

Figure 2: Time series plot of the state variable for predictor DP over the out-of-sample period from Q1, 1965 to Q3, 2014

Table 1: Out-of-sample performance of the SAM with a single predictor

Panel A: Out-of-sample period from Q1,1965 to Q3,2014

	$MSFE_{A\|B}$:		P-value of ENC-T:		P-value of ENC-T:		$\Delta\bar{U}_{A/B}$:	
	A=SAM,		A=SAM vs		B=SAM vs		A=SAM,	
x	B=HA	B=OLS	B=HA	B=OLS	A=HA	A=OLS	B=HA	B=OLS
DP	0.991	0.992	0.114	0.020	0.354	0.067	−0.245	5.231
DY	0.994	0.996	0.141	0.032	0.342	0.064	0.138	5.674
EP	0.991	0.979	0.101	0.027	0.289	0.285	2.107	5.108
DE	0.952	0.935	0.001	0.000	0.601	0.874	2.235	3.270
BM	0.995	0.975	0.117	0.024	0.227	0.302	3.196	3.125
TBL	0.973	0.951	0.004	0.002	0.132	0.271	1.054	3.988
DFY	0.972	0.947	0.012	0.000	0.333	0.859	6.784	1.861
LTY	0.965	0.943	0.003	0.002	0.292	0.489	0.103	5.072
TMS	0.985	0.959	0.011	0.001	0.087	0.538	9.858	−1.200
NTIS	0.985	0.964	0.014	0.002	0.187	0.665	6.349	1.569
INFL	0.957	0.953	0.001	0.008	0.520	0.537	3.057	2.743
ITR	0.979	0.969	0.019	0.004	0.166	0.571	5.442	0.979
DFR	0.970	0.969	0.013	0.014	0.543	0.488	3.937	1.600
SVAR	0.991	0.879	0.010	0.094	0.148	0.751	3.103	235.643
IK	0.965	0.987	0.005	0.033	0.464	0.223	−0.700	1.044

Panel B: Out-of-sample period from Q1,1976 to Q3,2014

	$MSFE_{A\|B}$:		P-value of ENC-T:		P-value of ENC-T:		$\Delta\bar{U}_{A/B}$:	
	A=SAM,		A=SAM vs		B=SAM vs		A=SAM,	
x	B=HA	B=OLS	B=HA	B=OLS	A=HA	A=OLS	B=HA	B=OLS
DP	1.020	0.980	0.490	0.010	0.070	0.150	0.640	4.330
DY	1.010	0.960	0.340	0.000	0.160	0.300	0.870	4.240
EP	1.010	0.980	0.310	0.040	0.110	0.280	3.270	5.180
DE	0.970	0.940	0.000	0.000	0.280	0.740	3.470	4.220
BM	1.030	0.990	0.530	0.070	0.020	0.140	4.050	4.500
TBL	0.990	0.940	0.020	0.000	0.060	0.520	4.500	2.980
DFY	0.990	0.970	0.050	0.000	0.080	0.490	6.830	4.340
ITY	0.980	0.940	0.010	0.000	0.180	0.620	2.690	3.850
TMS	0.990	0.940	0.030	0.000	0.070	0.800	11.540	−2.690
NTIS	0.990	0.970	0.020	0.010	0.130	0.510	7.890	1.260
INFL	0.970	0.950	0.000	0.000	0.270	0.580	5.070	2.560
ITR	0.990	0.970	0.050	0.010	0.090	0.490	5.960	1.220
DFR	0.980	0.980	0.050	0.050	0.280	0.270	4.370	1.980
SVAR	1.020	0.850	0.050	0.110	0.050	0.790	3.280	304.020
IK	1.000	0.990	0.110	0.040	0.150	0.200	2.680	0.340

Note: Panels A and B are the results for the two out-of-sample periods from the first quarter of 1965 to third quarter of 2014 and from the first quarter of 1976 to the third quarter of 2014, respectively. SAM, HA and OLS in the table stand for the SAM method with a single predictor, the prevailing benchmark method of historical average, and traditional predictive linear regression with a single predictor, respectively. $MSFE_{A|B} = \frac{MSFE_A}{MSFE_B}$ is the MSFE ratio of SAM to either HA or OLS. P-value of ENC-T is the p-value of the encompassing test with the null hypothesis $H_o : \lambda = 0$ against the alternative hypothesis $H_a : \lambda > 0$, where λ is the parameter in the optimal composite of two competing forecasts (A and B), defined by $\hat{y}_{o,t} = \lambda \hat{y}_{A,t} + (1-\lambda)\hat{y}_{B,t}$. $\Delta\bar{U}_{A|B} = 400(\bar{U}^A - \bar{U}^B)$ is the difference in utility between two competing forecasts A and B.

Table 2: Out-of-sample predictability under different scenarios for the SAM with a single predictor

	Scenario 1						Scenario 2						Scenario 3					
	$MSFE_{A\|B}$		P-Value of ENC-T		P-Value of ENC-T		$MSFE_{A\|B}$		P-Value of ENC-T		P-Value of ENC-T		$MSFE_{A\|B}$		P-Value of ENC-T		P-Value of ENC-T	
	A=SAM		A= SAM vs		B= SAM vs		A=SAM		A= SAM vs		B= SAM vs		A=SAM		A= SAM vs		B= SAM vs	
	B=HA	B=OLS	B=HA	B=OLS	A=HA	A=OLS	B=HA	B=OLS	B=HA	B=OLS	A=HA	A=OLS	B=HA	B=OLS	B=HA	B=OLS	A=HA	A=OLS
DP	0.981	0.986	0.176	0.144	0.529	0.410	0.996	0.972	0.281	0.069	0.415	0.531	1.004	1.038	0.304	0.117	0.233	0.031
DY	1.001	0.984	0.397	0.160	0.373	0.498	1.021	0.963	0.667	0.044	0.092	0.546	0.950	1.062	0.055	0.230	0.719	0.020
EP	0.976	0.957	0.155	0.074	0.549	0.688	1.000	0.993	0.374	0.191	0.363	0.468	1.004	1.001	0.235	0.131	0.188	0.127
DE	0.975	0.967	0.177	0.034	0.618	0.920	0.963	0.980	0.016	0.083	0.587	0.623	0.860	0.753	0.006	0.002	0.470	0.744
BM	0.967	0.921	0.068	0.019	0.726	0.783	1.012	1.001	0.463	0.298	0.218	0.271	1.028	1.067	0.336	0.337	0.121	0.050
TBL	1.011	0.966	0.511	0.181	0.266	0.472	0.940	1.008	0.008	0.156	0.304	0.099	0.985	0.836	0.069	0.001	0.132	0.541
DFY	0.971	0.940	0.094	0.017	0.707	0.846	1.017	0.988	0.563	0.108	0.071	0.704	0.849	0.848	0.017	0.002	0.396	0.611
LTY	1.031	0.984	0.643	0.265	0.147	0.441	0.910	1.012	0.002	0.371	0.708	0.144	0.984	0.777	0.067	0.000	0.154	0.722
TMS	1.036	0.964	0.784	0.069	0.008	0.644	0.961	0.982	0.032	0.103	0.431	0.366	0.934	0.883	0.036	0.002	0.171	0.591
NTIS	1.012	0.967	0.522	0.063	0.184	0.839	0.991	0.971	0.101	0.049	0.519	0.761	0.874	0.923	0.010	0.034	0.292	0.256
INFL	1.032	0.961	0.725	0.102	0.049	0.786	0.932	1.020	0.003	0.496	0.929	0.131	0.893	0.784	0.007	0.008	0.346	0.677
LTR	1.021	0.957	0.466	0.024	0.118	0.868	1.002	0.993	0.253	0.163	0.229	0.374	0.806	0.935	0.011	0.053	0.488	0.323
DFR	1.000	0.999	0.390	0.322	0.395	0.355	0.980	0.970	0.016	0.017	0.929	0.830	0.870	0.889	0.015	0.044	0.497	0.421
SVAR	0.995	0.750	0.028	0.099	0.266	0.896	0.996	1.015	0.331	0.791	0.451	0.063	0.953	1.061	0.047	0.168	0.139	0.034
IK	1.026	0.995	0.605	0.334	0.095	0.438	0.895	0.983	0.002	0.114	0.896	0.434	0.998	0.978	0.195	0.060	0.222	0.178

Note: The out-of-sample period is from the first quarter of 1965 to the third quarter of 2014. SAM, HA and OLS in the table stand for the SAM method with a single predictor, the prevailing benchmark method of historical average, and traditional predictive linear regression with a single predictor, respectively. $MSFE_{A|B} = \frac{MSFE_A}{MSFE_B}$ is the MSFE ratio of SAM to either HA or OLS. P-value of ENC-T is the p-value of the encompassing test with the null hypothesis $H_o : \lambda = 0$ against the alternative hypothesis $H_a : \lambda > 0$, where λ is the parameter in the optimal composite of two competing forecasts (A and B), defined by $\hat{y}_{o,t} = \lambda \hat{y}_{A,t} + (1 - \lambda)\hat{y}_{B,t}$. The results under Scenario i ($i = 1, 2, 3$) are the results for the out-of-sample forecasts that are generated based on Scenario i.

Table 3: Out-of-sample performance of the SAM with multiple predictors

Combination method	$MSFE_{A\|B}$: A=SAM, B=HA	B=OLS	P-value of ENC-T: A=SAM vs B=HA	B=OLS	P-value of ENC-T: B=SAM vs A=HA	A=OLS	$\Delta\bar{U}_{A\|B}$: A=SAM, B=HA	B=OLS
	Panel A: Out-of-sample period from Q1, 1965 to Q3, 2014							
Mean	0.961	0.989	0.002	0.025	0.812	0.291	1.738	5.137
Median	0.967	0.989	0.002	0.030	0.723	0.284	2.124	4.705
Trim mean	0.962	0.990	0.002	0.031	0.803	0.276	1.774	5.276
	Panel B: Out-of-sample period from Q1, 1976 to Q3, 2014							
Mean	0.978	0.985	0.014	0.013	0.463	0.35	3.348	4.532
Median	0.981	0.989	0.015	0.034	0.379	0.256	3.587	4.214
Trim mean	0.979	0.985	0.014	0.018	0.451	0.345	3.307	4.691

Note: In the table, HA represents the prevailing benchmark method of historical average, and SAM and OLS stand for the SAM with multiple predictors and the combination methods based on the traditional predictive linear regression with a single predictor respectively. $MSFE_{A|B} = \frac{MSFE_A}{MSFE_B}$ is the MSFE ratio of SAM to either HA or OLS. P-value of ENC-T is the p-value of the encompassing test with the null hypothesis $H_o : \lambda = 0$ against the alternative hypothesis $H_a : \lambda > 0$, where λ is the parameter in the optimal composite of two competing forecasts (A and B), defined by $\hat{y}_{o,t} = \lambda \hat{y}_{A,t} + (1 - \lambda)\hat{y}_{B,t}$. $\Delta\bar{U}_{A|B} = 400(\bar{U}^A - \bar{U}^B)$ is the difference in utility between two competing forecasts A and B. Panels A and B include the results for the two out-of-sample periods from the first quarter of 1965 to third quarter of 2014 and from the first quarter of 1976 to the third quarter of 2014 respectively.

Table 4: Out-of-sample predictability during three different GDP growth periods

Panel A: The SAM with a single predictor, the three GDP periods defined by the 20th and 80th percentiles of GDP growth rates

	Low GDP growth						Average GDP growth						High GDP growth					
	$MSFE_{A\|B}$		P-Value of ENC-T		P-Value of ENC-T		$MSFE_{A\|B}$		P-Value of ENC-T		P-Value of ENC-T		$MSFE_{A\|B}$		P-Value of ENC-T		P-Value of ENC-T	
	A=SAM		A= SAM vs		B= SAM vs		A=SAM		A= SAM vs		B= SAM vs		A=SAM		A= SAM vs		B= SAM vs	
	B=HA	B=OLS	B=HA	B=OLS	A=HA	A=OLS	B=HA	B=OLS	B=HA	B=OLS	A=HA	A=OLS	B=HA	B=OLS	B=HA	B=OLS	A=HA	A=OLS
DP	0.940	1.005	0.017	0.325	0.945	0.263	1.018	0.980	0.494	0.020	0.106	0.149	0.994	1.034	0.304	0.313	0.352	0.158
DY	0.940	1.021	0.004	0.511	0.984	0.133	1.018	0.968	0.532	0.012	0.095	0.250	1.014	1.116	0.438	0.644	0.280	0.072
EP	0.920	0.965	0.008	0.157	0.962	0.596	1.022	0.977	0.466	0.048	0.085	0.307	1.030	1.040	0.459	0.424	0.189	0.132
DE	0.945	0.964	0.082	0.054	0.746	0.813	0.971	0.935	0.012	0.004	0.241	0.705	0.861	0.844	0.004	0.022	0.915	0.885
BM	0.938	0.901	0.022	0.044	0.922	0.839	1.028	1.005	0.503	0.144	0.045	0.093	0.973	1.039	0.230	0.364	0.458	0.101
TBL	0.984	0.948	0.110	0.144	0.228	0.556	0.990	0.996	0.037	0.040	0.098	0.058	0.828	0.721	0.003	0.000	0.893	0.944
DFY	1.007	0.955	0.446	0.070	0.298	0.768	0.975	0.939	0.040	0.001	0.292	0.807	0.844	0.968	0.026	0.092	0.852	0.430
LTY	0.931	0.935	0.041	0.099	0.655	0.705	0.984	0.986	0.037	0.063	0.163	0.152	0.961	0.755	0.083	0.004	0.320	0.886
TMS	0.988	0.988	0.171	0.281	0.305	0.492	0.993	0.948	0.050	0.002	0.101	0.512	0.918	0.935	0.075	0.064	0.376	0.633
NTIS	1.008	0.958	0.494	0.073	0.302	0.831	0.989	0.972	0.040	0.022	0.137	0.421	0.888	0.924	0.018	0.047	0.766	0.639
INFL	0.959	0.967	0.092	0.177	0.661	0.615	0.973	0.974	0.013	0.070	0.266	0.288	0.852	0.787	0.007	0.008	0.845	0.934
LTR	1.008	0.936	0.368	0.006	0.259	0.971	0.967	0.990	0.033	0.085	0.284	0.221	0.960	0.953	0.115	0.087	0.247	0.522
DFR	0.940	1.019	0.051	0.560	0.881	0.225	1.003	0.952	0.181	0.017	0.122	0.601	0.866	0.926	0.009	0.010	0.956	0.808
SVAR	0.966	0.897	0.079	0.020	0.837	0.886	1.016	0.848	0.076	0.136	0.074	0.671	0.911	1.084	0.048	0.844	0.694	0.023
IK	0.954	1.008	0.064	0.480	0.714	0.322	0.980	0.978	0.058	0.033	0.280	0.322	0.905	0.977	0.036	0.140	0.617	0.319

Panel B: The SAM with a single predictor, the three GDP periods defined by the 25th and 75th percentiles of GDP growth rates

	Low GDP growth						Average GDP growth						High GDP growth					
	$MSFE_{A\|B}$		P-Value of ENC-T		P-Value of ENC-T		$MSFE_{A\|B}$		P-Value of ENC-T		P-Value of ENC-T		$MSFE_{A\|B}$		P-Value of ENC-T		P-Value of ENC-T	
	A=SAM		A= SAM vs		B= SAM vs		A=SAM		A= SAM vs		B= SAM vs		A=SAM		A= SAM vs		B= SAM vs	
	B=HA	B=OLS	B=HA	B=OLS	A=HA	A=OLS	B=HA	B=OLS	B=HA	B=OLS	A=HA	A=OLS	B=HA	B=OLS	B=HA	B=OLS	A=HA	A=OLS
DP	0.958	1.007	0.044	0.309	0.838	0.228	1.008	0.957	0.375	0.007	0.196	0.346	1.024	1.108	0.470	0.562	0.197	0.027
DY	0.976	1.036	0.110	0.620	0.739	0.089	1.019	0.958	0.555	0.011	0.092	0.345	1.006	1.112	0.405	0.624	0.330	0.017
EP	0.953	0.992	0.068	0.294	0.792	0.391	1.019	0.962	0.421	0.027	0.108	0.420	1.018	1.045	0.412	0.468	0.236	0.084
DE	0.972	0.989	0.146	0.173	0.535	0.405	0.976	0.932	0.027	0.009	0.202	0.736	0.884	0.876	0.012	0.028	0.818	0.702
BM	0.968	0.918	0.101	0.044	0.684	0.788	1.015	1.003	0.387	0.165	0.130	0.133	1.001	1.046	0.309	0.368	0.300	0.035
TBL	0.972	0.956	0.058	0.113	0.238	0.480	0.998	1.001	0.077	0.074	0.091	0.065	0.875	0.770	0.005	0.001	0.811	0.838
DFY	0.986	0.937	0.202	0.016	0.496	0.882	0.964	0.959	0.019	0.009	0.380	0.578	0.970	0.938	0.196	0.020	0.349	0.691
LTY	0.929	0.954	0.022	0.105	0.685	0.567	0.993	0.986	0.085	0.078	0.138	0.186	0.927	0.760	0.015	0.003	0.562	0.864
TMS	0.978	0.974	0.103	0.143	0.346	0.566	0.979	0.952	0.015	0.010	0.157	0.488	0.986	0.910	0.189	0.009	0.239	0.766
NTIS	1.004	0.959	0.440	0.050	0.321	0.839	0.986	0.965	0.041	0.025	0.178	0.530	0.931	0.973	0.072	0.149	0.517	0.353
INFL	0.960	0.978	0.064	0.181	0.643	0.469	0.964	0.957	0.007	0.056	0.375	0.481	0.894	0.845	0.026	0.019	0.669	0.753
LTR	1.013	0.958	0.364	0.027	0.196	0.810	0.945	0.964	0.015	0.026	0.449	0.528	0.958	0.963	0.122	0.119	0.297	0.437
DFR	0.944	0.997	0.036	0.333	0.907	0.402	0.998	0.944	0.147	0.018	0.158	0.639	0.924	0.975	0.072	0.144	0.658	0.401
SVAR	0.976	0.899	0.110	0.013	0.769	0.907	0.992	1.009	0.081	0.152	0.180	0.058	1.007	0.565	0.044	0.168	0.192	0.766
IK	0.961	1.023	0.077	0.580	0.636	0.154	0.964	0.963	0.036	0.014	0.455	0.477	0.948	0.955	0.092	0.046	0.472	0.387

Panel C: The SAM with multiple predictors,

the three GDP periods defined by the 20th and 80th percentiles of GDP growth rates

Combination method	$MSFE_{A\|B}$: A=SAM,		P-value of ENC-T: A=SAM vs		P-value of ENC-T: B=SAM vs	
	B=HA	B=OLS	B=HA	B=OLS	A=HA	A=OLS
	Low GDP growth periods					
Mean	0.948	0.987	0.029	0.157	0.877	0.686
Median	0.960	0.983	0.049	0.159	0.852	0.689
Trimmed mean	0.950	0.985	0.031	0.131	0.881	0.725
	Average GDP growth periods					
Mean	0.979	0.995	0.040	0.087	0.431	0.172
Median	0.981	0.994	0.039	0.081	0.364	0.181
Trimmed mean	0.979	0.997	0.039	0.107	0.422	0.156
	High GDP growth period					
Mean	0.891	0.959	0.015	0.088	0.924	0.622
Median	0.897	0.977	0.018	0.173	0.903	0.488
Trimmed mean	0.894	0.965	0.017	0.109	0.914	0.572

Panel D: The SAM with multiple predictors,

the three GDP periods defined by the 25th and 75th percentiles of GDP growth rates

Combination method	$MSFE_{A\|B}$ A=SAM		P-Value of ENC-T A= SAM vs		P-Value of ENC-T B= SAM vs	
	B=HA	B=OLS	B=HA	B=OLS	A=HA	A=OLS
	Low GDP growth periods					
Mean	0.958	0.994	0.039	0.264	0.837	0.519
Median	0.962	0.983	0.048	0.151	0.843	0.681
Trimmed mean	0.959	0.991	0.040	0.201	0.848	0.601
	Average GDP growth periods					
Mean	0.964	0.983	0.014	0.036	0.640	0.368
Median	0.967	0.981	0.014	0.036	0.567	0.346
Trimmed mean	0.965	0.986	0.014	0.048	0.625	0.316
	High GDP growth period					
Mean	0.952	1.001	0.134	0.275	0.572	0.271
Median	0.967	1.030	0.176	0.513	0.493	0.177
Trimmed mean	0.957	1.007	0.147	0.337	0.547	0.256

Note: The out-of-sample period is from the first quarter of 1965 to the third quarter of 2014. The low, average, and high GDP growth periods correspond to the periods during which the US real GDP growth rate is in the low, average, and high regions, respectively, defined by either 20th and 80th percentiles or the 25th and 75th percentiles of the GDP growth rates. HA represents the prevailing benchmark method of historical average. SAM and OLS stand for the SAM with a single predictor and the traditional predictive linear regression with a single predictor in Panels A and B respectively and for the SAM with multiple predictors and the combination methods based on the traditional predictive linear regression with a single predictor in Panels C and D respectively. $MSFE_{A|B} = \frac{MSFE_A}{MSFE_B}$ is the MSFE ratio of SAM to either HA or OLS. P-value of ENC-T is the p-value of the encompassing test with the null hypothesis $H_o : \lambda = 0$ against the alternative hypothesis $H_a : \lambda > 0$, where λ is the parameter in the optimal composite of two competing forecasts (A and B), defined by $\hat{y}_{o,t} = \lambda\hat{y}_{A,t} + (1-\lambda)\hat{y}_{B,t}$.

Table 5: Out-of-sample predictability during the recession and expansion periods

Panel A: the SAM with a single predictor

	Recession periods						Expansion periods					
	$MSFE_{A\|B}$ A=SAM		P-Value of ENC-T A= SAM vs		P-Value of ENC-T B= SAM vs		$MSFE_{A\|B}$ A=SAM		P-Value of ENC-T A= SAM vs		P-Value of ENC-T B= SAM vs	
	B=HA	B=OLS	B=HA	B=OLS	A=HA	A=OLS	B=HA	B=OLS	B=HA	B=OLS	A=HA	A=OLS
DP	0.939	1.012	0.031	0.436	0.920	0.277	1.017	0.984	0.452	0.011	0.070	0.080
DY	0.952	1.046	0.049	0.774	0.900	0.105	1.014	0.974	0.441	0.008	0.084	0.146
EP	0.942	0.947	0.062	0.116	0.820	0.723	1.015	0.995	0.341	0.063	0.069	0.111
DE	0.908	0.882	0.021	0.012	0.937	0.984	0.974	0.961	0.007	0.003	0.160	0.373
BM	0.928	0.934	0.034	0.118	0.919	0.708	1.027	0.994	0.473	0.048	0.022	0.095
TBL	1.041	0.920	0.459	0.077	0.154	0.681	0.939	0.970	0.000	0.003	0.298	0.070
DFY	0.924	0.925	0.035	0.029	0.882	0.909	0.996	0.957	0.064	0.001	0.110	0.594
LTY	0.949	0.857	0.098	0.014	0.611	0.944	0.974	0.991	0.005	0.024	0.148	0.052
TMS	0.960	0.969	0.104	0.164	0.583	0.678	0.997	0.954	0.027	0.002	0.041	0.424
NTIS	0.989	0.937	0.301	0.021	0.530	0.956	0.983	0.977	0.011	0.017	0.136	0.303
INFL	0.942	0.885	0.052	0.034	0.811	0.907	0.964	0.989	0.003	0.049	0.281	0.130
LTR	0.948	0.901	0.114	0.001	0.588	0.997	0.994	1.004	0.039	0.108	0.069	0.063
DFR	0.927	0.956	0.041	0.124	0.912	0.717	0.992	0.975	0.068	0.030	0.166	0.280
SVAR	0.923	0.892	0.027	0.019	0.961	0.892	1.024	0.874	0.068	0.145	0.034	0.597
IK	0.949	1.005	0.079	0.454	0.823	0.369	0.972	0.978	0.016	0.016	0.255	0.244

Panel B: the SAM with multiple predictors

Combination method	$MSFE_{A\|B}$: A=SAM, B=HA	B=OLS	P-value of ENC-T: A=SAM vs B=HA	B=OLS	P-value of ENC-T: B=SAM vs A=HA	A=OLS
	Recession periods					
Mean	0.939	0.955	0.026	0.011	0.928	0.983
Median	0.943	0.951	0.029	0.014	0.936	0.973
Trim mean	0.939	0.957	0.026	0.011	0.936	0.983
	Expansion periods					
Mean	0.972	1.006	0.009	0.154	0.503	0.042
Median	0.978	1.008	0.013	0.176	0.370	0.040
Trim mean	0.974	1.007	0.010	0.171	0.474	0.043

Note: The out-of-sample period is from the first quarter of 1965 to the third quarter of 2014. The recession and expansion periods correspond to the periods during which the US recession indicator is 1 and 0 respectively. HA represents the prevailing benchmark method of historical average. SAM and OLS stand for the SAM with a single predictor and the traditional predictive linear regression with a single predictor in Panel A respectively and for the SAM with multiple predictors and the combination methods based on the traditional predictive linear regression with a single predictor in Panel B respectively. $MSFE_{A|B} = \frac{MSFE_A}{MSFE_B}$ is the MSFE ratio of SAM to either HA or OLS. P-value of ENC-T is the p-value of the encompassing test with the null hypothesis $H_o : \lambda = 0$ against the alternative hypothesis $H_a : \lambda > 0$, where λ is the parameter in the optimal composite of two competing forecasts (A and B), defined by $\hat{y}_{o,t} = \lambda \hat{y}_{A,t} + (1 - \lambda)\hat{y}_{B,t}$.

Table 6: The percentages of occurrences of a scenario under different economic conditions during the out-of-sample period.

	Recession			Expansion			Low GDP growth			Average GDP growth			High GDP growth		
	S1	S2	S3	S1	S2	S3	S1	S2	S3	S1	S2	S3	S1	S2	S3
DP	67%	19%	15%	18%	51%	31%	58%	33%	8%	19%	48%	33%	10%	53%	37%
DY	63%	15%	22%	18%	49%	33%	58%	33%	8%	18%	46%	36%	10%	53%	37%
EP	63%	19%	19%	20%	53%	27%	56%	39%	6%	21%	50%	29%	10%	57%	33%
DE	59%	26%	15%	20%	56%	24%	50%	44%	6%	20%	53%	27%	20%	53%	27%
BM	70%	19%	11%	22%	48%	30%	61%	33%	6%	23%	44%	32%	13%	53%	33%
TBL	52%	19%	30%	15%	56%	29%	39%	50%	11%	17%	51%	32%	10%	53%	37%
DFY	63%	19%	19%	22%	57%	22%	56%	39%	6%	22%	53%	26%	17%	63%	20%
LTY	52%	30%	19%	15%	56%	29%	39%	50%	11%	17%	54%	29%	10%	50%	40%
TMS	59%	22%	19%	26%	52%	22%	50%	42%	8%	26%	49%	26%	27%	53%	20%
NTIS	67%	19%	15%	24%	54%	22%	64%	31%	6%	23%	53%	23%	17%	53%	30%
INFL	56%	26%	19%	20%	56%	24%	53%	42%	6%	20%	53%	27%	13%	60%	27%
ITR	70%	11%	19%	22%	56%	22%	58%	36%	6%	21%	53%	26%	23%	53%	23%
DFR	67%	19%	15%	23%	54%	23%	61%	31%	8%	23%	53%	25%	20%	57%	23%
SVAR	74%	19%	7%	21%	58%	22%	61%	33%	6%	23%	53%	24%	13%	70%	17%
IK	59%	22%	19%	20%	52%	28%	56%	36%	8%	20%	50%	30%	13%	50%	37%
Average	63%	20%	17%	20%	54%	26%	55%	38%	7%	21%	51%	28%	15%	56%	29%

Note: The out-of-sample period is from the first quarter of 1965 to the third quarter of 2014. The recession and expansion periods correspond to the periods during which the US recession indicator is 1 and 0 respectively. The low, average, and high GDP growth periods correspond to the periods during which the US real GDP growth rate is in the low, average, and high regions, respectively, defined by the 25th and 75th percentiles of the GDP growth rates. S1, S2, and S3 stand for the bad, average, and good scenarios respectively. The percentages of occurrences of a scenarios for each predictor are calculated based on the estimated values of $\hat{S}_{t-1}$ over different periods defined by different economic conditions. The last row presents the average of the percentages of occurrences over all the 15 predictors.

Table 7: The averages of the transition probabilities of remaining in the same scenario over two consecutive periods under different economic conditions during the out-of-sample period

	Recession			Expansion			Low GDP growth			Average GDP growth			High GDP growth		
	p_{11}	p_{22}	p_{33}	p_{11}	p_{22}	p_{33}	p_{11}	p_{22}	p_{33}	p_{11}	p_{22}	p_{33}	p_{11}	p_{22}	p_{33}
DP	0.30	0.00	0.07	0.07	0.26	0.11	0.28	0.06	0.06	0.07	0.26	12%	0.03	0.27	0.10
DY	0.26	0.04	0.15	0.06	0.25	0.12	0.28	0.06	0.06	0.05	0.26	14%	0.00	0.23	0.10
EP	0.26	0.00	0.11	0.06	0.28	0.08	0.25	0.08	0.03	0.07	0.29	9%	0.00	0.27	0.13
DE	0.26	0.00	0.15	0.06	0.33	0.06	0.25	0.17	0.03	0.06	0.32	8%	0.00	0.27	0.10
BM	0.37	0.07	0.07	0.09	0.21	0.09	0.36	0.03	0.03	0.09	0.21	10%	0.03	0.30	0.10
TBL	0.22	0.00	0.22	0.02	0.34	0.08	0.19	0.17	0.06	0.02	0.32	9%	0.00	0.33	0.17
DFY	0.26	0.00	0.11	0.08	0.33	0.05	0.25	0.11	0.03	0.08	0.31	8%	0.03	0.37	0.03
LTY	0.22	0.11	0.19	0.03	0.34	0.08	0.19	0.22	0.06	0.03	0.35	8%	0.00	0.27	0.20
TMS	0.26	0.04	0.15	0.08	0.28	0.05	0.25	0.14	0.06	0.06	0.26	7%	0.13	0.33	0.07
NTIS	0.33	0.04	0.15	0.08	0.28	0.04	0.36	0.08	0.03	0.07	0.29	5%	0.00	0.27	0.10
INFL	0.22	0.07	0.19	0.06	0.35	0.07	0.25	0.14	0.03	0.05	0.35	8%	0.00	0.37	0.17
ITR	0.37	0.00	0.15	0.06	0.32	0.05	0.31	0.11	0.03	0.06	0.32	8%	0.03	0.30	0.07
DFR	0.26	0.00	0.15	0.08	0.31	0.06	0.31	0.08	0.06	0.06	0.32	7%	0.03	0.27	0.10
SVAR	0.41	0.04	0.07	0.08	0.33	0.04	0.36	0.08	0.03	0.09	0.31	5%	0.00	0.43	0.07
IK	0.22	0.04	0.15	0.04	0.28	0.08	0.22	0.11	0.06	0.04	0.28	8%	0.00	0.27	0.17
Average	0.28	0.03	0.14	0.06	0.30	0.07	0.27	0.11	0.04	0.06	0.30	8%	0.02	0.30	0.11

Note: The out-of-sample period is from the first quarter of 1965 to the third quarter of 2014. The recession and expansion periods correspond to the periods during which the US recession indicator is 1 and 0 respectively. The low, average, and high GDP growth periods correspond to the periods during which the US real GDP growth rate is in the low, average, and high regions, respectively, defined by the 25th and 75th percentiles of the GDP growth rates. The averages of p_{ii} $(i = 1, 2, 3)$ for each predictor are the averages of the values of p_{ii} estimated with in-sample data using Equation (5) over different periods defined by different economic conditions. The last row presents the average of the averaged transition probabilities over all the 15 predictors.

Table 8: Out-of-sample performance of the SAM with an alternative specification of the three scenarios

x	$MSFE_{A\|B}$: A=SAM, B=HA	B=OLS	P-value of ENC-T: A=SAM vs B=HA	B=OLS	P-value of ENC-T: B=SAM vs A=HA	A=OLS	$\Delta\bar{U}_{A/B}$: A=SAM, B=HA	B=OLS
	Panel A: the SAM with a single predictor							
DP	0.994	0.995	0.102	0.014	0.233	0.032	1.047	6.523
DY	0.991	0.993	0.110	0.026	0.382	0.067	0.663	6.199
EP	0.990	0.978	0.100	0.026	0.298	0.271	2.135	5.136
DE	0.966	0.949	0.003	0.002	0.705	0.812	1.653	2.688
BM	0.992	0.972	0.104	0.016	0.267	0.297	2.657	2.586
TBL	0.980	0.958	0.020	0.003	0.248	0.252	−0.566	2.369
DFY	0.980	0.954	0.030	0.001	0.336	0.780	4.717	−0.207
ITY	0.967	0.945	0.005	0.003	0.446	0.425	−0.387	4.582
TMS	0.978	0.953	0.013	0.001	0.263	0.536	5.699	−5.359
NTIS	0.976	0.954	0.010	0.002	0.521	0.850	4.891	0.111
INFL	0.963	0.959	0.001	0.017	0.839	0.649	1.203	0.889
ITR	0.987	0.977	0.055	0.047	0.285	0.668	3.553	−0.910
DFR	0.971	0.969	0.007	0.022	0.680	0.513	2.687	0.350
SVAR	0.977	0.868	0.010	0.095	0.313	0.683	5.005	237.544
IK	0.968	0.991	0.004	0.030	0.481	0.104	0.151	1.894
Combination method	Panel B: the SAM with multiple predictors							
Mean	0.967	0.996	0.004	0.083	0.829	0.220	1.118	4.517
Median	0.968	0.990	0.004	0.054	0.814	0.367	1.593	4.174
Trimmed Mean	0.968	0.996	0.004	0.095	0.823	0.227	1.259	4.761

Note: The out-of-sample period is from the first quarter of 1965 to the third quarter of 2014. HA represents the prevailing benchmark method of historical average. SAM and OLS stand for the SAM with a single predictor and the traditional predictive linear regression with a single predictor in Panel A respectively and for the SAM with multiple predictors and the combination methods based on the traditional predictive linear regression with a single predictor in Panel B respectively. $MSFE_{A|B} = \frac{MSFE_A}{MSFE_B}$ is the MSFE ratio of SAM to either HA or OLS. P-value of ENC-T is the p-value of the encompassing test with the null hypothesis $H_o : \lambda = 0$ against the alternative hypothesis $H_a : \lambda > 0$, where λ is the parameter in the optimal composite of two competing forecasts (A and B), defined by $\hat{y}_{o,t} = \lambda\hat{y}_{A,t} + (1-\lambda)\hat{y}_{B,t}$. $\Delta\bar{U}_{A|B} = 400(\bar{U}^A - \bar{U}^B)$ is the difference in utility between two competing forecasts A and B.

Table 9: MSFE ratios of either the predictive individual mean regression or the SAM with a single predictor to the prevailing benchmark method

Forecast	Predictor														
method	D/P	D/Y	E/P	D/E	SVAR	B/M	NTIS	TBL	LTY	LTR	TMS	DFY	DFR	INFL	I/K
IMR	0.9928	0.9900	1.0109	1.016	1.0665	1.0180	1.0210	1.0243	1.0259	1.0115	1.0265	1.0271	0.9909	1.0076	0.9768
SAM	0.9875	0.9898	0.9898	0.943	0.9872	0.9918	0.9853	0.9739	0.9643	0.9779	0.9844	0.9726	0.9661	0.9561	0.9614

Note: IMR and SAM represent the predictive individual mean regression (IMR) and the SAM with a single predictor respectively. The table reports the MSFE ratios of either IMR or SAM to the prevailing benchmark method of historical average for the out-of-sample period from the first quarter of 1965 to the fourth quarter of 2010. The MSFE ratios for the IMR are in MPVV(2014).

Table 10: MSFE ratios of various prediction combination methods to the prevailing benchmark method

Forecast method	Combination method									
	Mean	Median	Trim. Mean	DMSFE(1)	DMSFE(0.9)	Cluster 2	Cluster 3	Prin. Comp.		
MFC	0.9703	0.9781	0.9715	0.9704	0.9702	0.9766	0.9878	1.0169		
RFC-FW1	0.9761	0.9865	0.9778	0.9755	0.9747	0.9726	1.0059	1.0289		
RFC-FW2	0.9768	0.9893	0.9786	0.9763	0.976	0.9778	0.9992	1.0256		
RFC-FW3	0.9741	0.9848	0.9761	0.9737	0.9731	0.9744	1.0017	1.0284		
RFC-FW4	0.9720	0.9794	0.9737	0.9719	0.9716	0.9769	0.9861	1.0287		
RFC-TVW1	0.9635	0.9718	0.9650							
RFC-TVW2	0.9654	0.9760	0.9677							
RFC-TVW3	0.9633	0.9669	0.9667							
				DALFE(1)	DALFE(0.9)	AL Cluster 2	AL Cluster 3	AL Prin. Comp.	AL Lasso	AL Ridge
QFC-FW1	0.9761	0.9886	0.9785	0.9758	0.9752	0.9768	0.9798	1.0079	0.9777	0.9696
QFC-FW2	0.9768	0.9903	0.9791	0.9766	0.976	0.9809	0.9753	1.016	0.9899	0.9719
QFC-FW3	0.9741	0.9866	0.9768	0.9738	0.9731	0.9754	0.9787	1.0062	0.9782	0.9680
QFC-FW4	0.9720	0.9830	0.9746	0.9719	0.9711	0.9733	0.9785	1.0181	0.9866	0.9705
QFC-TVW1	0.9594	0.9669	0.9619							
QFC-TVW2	0.9619	0.9717	0.9648							
QFC-TVW3	0.9677	0.9746	0.9702							
MFC,RFC-TVW1,QFC-TVW1	0.9640	0.9676	0.9746							
MFC,RFC-TVW2,QFC-TVW2	0.9655	0.9683	0.9737							
MFC,RFC-TVW3,QFC-TVW3	0.9661	0.9827	0.9728							
SAM	0.9585	0.9652	0.9603							

Note: This table reports the MSFE ratios of various forecast combination methods to the prevailing benchmark method of historical average for the out-of-sample from the first quarter of 1965 to the fourth quarter of 2010. Different forecast combination methods are classified by a forecast method and a combination method. The forecast methods include mean forecast combination (MRC), robust forecast combination with fixed weights (RFC-FW1-4) and time-varying weights (RFC-TVW1-3), quantile forecast combination with fixed weights (QRC-FW1-4) and time-varying weights (QRC-TVW1-3), the three combinations of different forecast methods ((MFC,RFC-TVW1,QFC-TVW1),(MFC,RFC-TVW2,QFC-TVW2),(MFC,RFC-TVW3,QFC-TVW3)), and the SAM with multiple predictors (SAM). The combination methods include mean, median, trimmed mean combinations as well as the other combination methods studied in MPVV(2014). Except the MSFE ratios for SAM, all the other MSFE ratios are in MPVV(2014).