



Munich Personal RePEc Archive

Stock Return Prediction with Fully Flexible Models and Coefficients

Byrne, Joseph and Fu, Rong

Department of Accountancy, Economics and Finance, Heriot-Watt University

9 November 2016

Online at <https://mpa.ub.uni-muenchen.de/75366/>
MPRA Paper No. 75366, posted 05 Dec 2016 10:12 UTC

Stock Return Prediction with Fully Flexible Models and Coefficients

Joseph P. Byrne ^{*†} Rong Fu ^{*}

November 9, 2016

Abstract

We evaluate stock return predictability using a fully flexible Bayesian framework, which explicitly allows for different degrees of time-variation in coefficients and in forecasting models. We believe that asset return predictability can evolve quickly or slowly, based upon market conditions, and we should account for this. Our approach has superior out-of-sample predictive performance compared to the historical mean, from a statistical and economic perspective. We also find that our model statistically dominates its nested models, including models in which parameters evolve at a constant rate. By decomposing sources of prediction uncertainty into five parts, we find that our fully flexible approach more precisely identifies time-variation in coefficients and in forecasting models, leading to mitigation of estimation risk and forecasting improvements. Finally, we relate predictability to the business cycle.

JEL classification: C11, G11, G12, G17

Keywords: Stock return prediction, Time-varying coefficients and forecasting models, Bayesian econometrics, Forecast combination

^{*}Heriot-Watt University, Department of Accountancy, Economics and Finance, Edinburgh, EH14 4AS, UK.

[†]Corresponding author, E-mail: j.p.byrne@hw.ac.uk.

1 Introduction

Stock return predictability continues to be a core area of research in financial economic. This research is often skeptical about standard models or predictors and their out-of-sample forecasting power (Goyal and Welch, 2003; Cooper and Gulen, 2006; Andrew and Geert, 2007; Campbell and Thompson, 2008; Welch and Goyal, 2008; Joscha and Schüssler, 2014; Turner, 2015). For instance, Welch and Goyal (2008) comprehensively investigate the predictive power of commonly used asset pricing indicators: these predictors only occasionally have a good out-of-sample performance. Predictive power improves for some predictors in specific periods of stress, suggesting the presence of both model uncertainty and instability. In this paper, therefore, we investigate whether flexibly accounting for model uncertainty and parameter instability, can improve stock return prediction.

The literature agrees that ignoring model uncertainty and instability impairs predictability (Stambaugh, 1999; Wachter and Warusawitharana, 2009; Dangl and Halling, 2012; Billio et al., 2013; Johannes et al., 2014; Wachter and Warusawitharana, 2015). Time-varying parameters in predictive regressions have been previously considered in the literature. Bossaerts and Hillion (1999) argue that the poor out-of-sample performances of the prediction models is due to model nonstationarity: in other words, the parameters of the model should be time-varying. Similarly, by splitting the whole sample into different subsample periods, Andrew and Geert (2007) clearly find evidence of time-evolving parameters.

In addition to the problem of parameter instability, a large array of excess stock return predictors has been identified in the literature. Therefore, how to combine from a range of predictors to forecast stock returns should also be considered. This is known as model uncertainty. Stock return determinants and hence forecasting models may also change at each time period, as demonstrated in Avramov (2002), Cremers (2002), Rapach et al. (2010), Dangl and Halling (2012) and Johannes et al. (2014). Specifically, Johannes et al. (2014) argue that incorporating different model features, such as time-evolving expected returns, time-varying volatility and taking account of estimation error, is beneficial for out-of-sample gains.

Although there is agreement that model instability and uncertainty obstruct forecast

accuracy, the exact nature of time-variation in parameters and how to accommodate model uncertainty may matter for prediction. To successfully predict excess stock returns, we tackle these issues in this paper. Our main aim is try to answer the following research questions: Is time-variation in coefficients and in forecasting models rapid or do they evolve more slowly? How important is the problem of model uncertainty and parameter instability for stock return predictability?

Bayesian methods are reasonable to answer these questions, for instance the approaches of Raftery et al. (2010) and Koop and Korobilis (2012), since they take parameter instability and model uncertainty into account. These Dynamic Model Averaging (DMA) methods assume that coefficients and forecasting models vary in the same fashion over time. Harrison and West (1999) suggest that the fixed dynamics for evolution is unappealing and some model specifications can only be appropriate for some periods. For example, investors may rapidly update the relative importance of equity predictors in times of market stress and less rapidly in other times. We consider whether the assumption of Raftery et al. matters when predicting returns. In this sense, we construct Dynamic Mixture Model Averaging (DMMA), which allows for possible degrees of time-variation in coefficients and forecasting models, nesting gradual to abrupt changes in coefficients. In the extreme, our DMMA approach even accommodates constant coefficients and equal model weights. This implies our model is sufficiently flexible to detect and exploit locally appropriate forecasting models over time. In this framework, we can identify which predictors, which degrees of time-variation in coefficients and forecasting models result in better stock market predictability.

Our contribution goes further than constructing DMMA and comparing its out-of-sample performances to alternative models. The added flexibility of our DMMA approach allows us to analyze what leads to forecasting improvements. We can investigate whether forecast improvements are due to different predictors, different degrees of time-variation in coefficients and forecasting models. To that end, we decompose prediction variance into five parts: (i) model uncertainty caused by random fluctuations in the data process (*observational variance*), (ii) uncertainty due to the errors from estimating the coefficients (*estimation risk*), (iii) model uncertainty with respect to predictor selection, (iv) model uncertainty regarding different choices of the degree of time-variation in coefficients, (v) model uncertainty in terms

of different choices of the magnitude of time-variation in forecasting models. To the best of our knowledge, this is the first stock paper that systematically examines the role of model uncertainty in terms of different degrees of time-variation in forecasting models.

Methodologically, the paper most similar to ours is Dangl and Halling (2012), which takes parameter instability as well as model uncertainty into account, by employing time-varying coefficients with Bayesian Model Averaging (BMA) using macroeconomic predictors. However, our paper extends theirs in the following regards. First, the widely used Bayesian Model Averaging (BMA) method (Avramov, 2002; Cremers, 2002; Dangl and Halling, 2012; Turner, 2015) combines different models recursively according to posterior model probabilities. However, BMA does not distinguish the difference between the more recent forecast performance and the more distant past forecast performance. Even though Dangl and Halling (2012) employ recursive forecasting and incorporate some time variation in the forecasting model, posterior model probabilities will only vary slightly as new data is incorporated. Whereas, for DMA, recent forecast performance receives a higher weight. This allows for more rapid changes for different forecasting models' posterior weights. Our DMMA approach is flexible enough to encompass both DMA and BMA. It even allows for models with equal weights. Hence, DMMA is advantageous in embedding the locally proper degree of time-variation in forecasting models at each time. Moreover, we not only consider the macroeconomic predictors emphasized by Dangl and Halling (2012), but also technical predictors used in a recent paper by Neely et al. (2014), providing complementary information about the stock market. Beside, we use stochastic volatility instead of the constant volatility employed in Dangl and Halling (2012), which accommodates the widely identified fat tail property of financial data. Finally, compared with Dangl and Halling (2012), we add one more element of model uncertainty: uncertainty with regard to different choices of time-variation in forecasting models, and we study what leads to forecast improvements according to different sources of uncertainty.

To preview our results, we find that our Dynamic Mixture Model Averaging (DMMA) predominately outperforms the historical mean (HM), with positive out-of-sample R^2 and large utility gains relative to historical mean for a mean-variance investor across different sample periods. Moreover, DMMA dominates alternative model combination methods, in-

cluding equal weights, Bayesian Model Averaging (BMA) and Dynamic Model Averaging (DMA), in terms of point forecast accuracy. We also find that imposing constant coefficient on predictive models worsens out-of-sample performance. The reason why DMMA performs better than its nested models is that by detecting the locally appropriate time-variation in coefficients and in forecasting models, DMMA alleviates the problem of model misspecification, leading to forecast improvements.

Importantly, we further analyze different sources of uncertainty for different predictive model. Regarding DMMA, observational variance, model uncertainty with respect to predictor selection and uncertainty regarding the errors from estimating the coefficients are the top three obstructing factors in forecast performance. In contrast, uncertainty about the degree of time-variation in coefficients is small and uncertainty regarding the degree of time-variation in forecasting models only appears in the initial data points. This suggests that DMMA successfully adapts to stock market instability by embedding the precise level of time-variation in coefficients and forecasting models. Interestingly, when we fix the variability in coefficients and in forecasting models, the percentage weight of estimation error is large. The flexible DMMA, in contrast, has small estimation risk. As a consequence, one possible explanation for DMMA's superior out-of-sample performance is that considering a possible range of time-variation in coefficients increases coefficient variability, and offsets the loss in forecast accuracy caused by the coefficients estimation uncertainty. In addition, by allowing for different degrees of time-variation in forecasting models, we flexibly enhance model adaptability and quickly detect locally appropriate models, therefore, further compensate the losses due to estimation risk.

Linking DMMA's predictability to the business cycle, we find that DMMA has superior out-of-sample performances during recessions than expansions. DMMA also outperforms HM statistically and economically during all the expansions. Importantly, the predicted equity premium increases at the end of recessions and decreases when the recessions start. This is consistent with the asset pricing theory in the literature (Fama and French, 1989; Campbell and Cochrane, 1995; Cochrane, 1999, 2005; Rapach et al., 2010; Dangl and Halling, 2012). Hence, the investor who follows DMMA perfectly times the stock market. The agreement between DMMA's predictions and asset price theory provide more economic insights of equity

premium predictability.

2 Econometrics Framework

In this section we set out our approach to excess return prediction, while accommodating three dimensions of model uncertainty: different predictor selection, different degrees of time-variation in coefficients and different degrees of time-variation in forecasting models. This flexible prediction framework takes different sources of uncertainty into account. Specially, in Section 2.1, we demonstrate a dynamic linear model which allows for time-varying coefficients for a certain choice of predictors. In Section 2.2, we construct the Dynamic Mixture Model Averaging method to attach posterior probabilities to individual models that have alternative predictors and different degrees of time-variation in coefficients. Then combine individual models together using different degrees of time-variation in forecasting models.

2.1 Dynamic Linear Model

Consistent with the majority of research on stock return prediction, we assume a linear prediction model (Avramov, 2002; Cremers, 2002; Andrew and Geert, 2007; Welch and Goyal, 2008; Dangl and Halling, 2012; Joscha and Schüssler, 2014). Moreover, our model allows for time-variation in the coefficients of the prediction model. As Dangl and Halling (2012) acknowledge, by reducing estimation errors and calibrating coefficients to observed data, random-walk coefficients have better out-of-sample performances compared to autocorrelated coefficients.¹ In light of this, we allow the time-varying coefficients to follow a random walk process, signalling that changes in coefficients are unpredictable. Note that our prediction and out-of-sample performance are in real time: we only use information at or before time t if we want to predict the excess stock return at $t + 1$.

Particularly, assume r_t is the expected stock returns at time t , X_{t-1} is the specific predictor for each individual model at time $t - 1$ and time-varying parameter models are allowed. We

¹We further check this hypothesis in an online appendix and find that random-walk coefficients dominate those with autocorrelated coefficients.

perform the return prediction on a monthly basis.

$$r_t = X_t \theta_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, H_t) \quad (\text{observation equation}) \quad (1)$$

$$\theta_t = \theta_{t-1} + u_t, \quad u_t \sim N(0, Q_t) \quad (\text{transition equation}) \quad (2)$$

where the errors ε_t and u_t are normally distributed with zero mean and uncorrelated across all lags with time-varying variances H_t and Q_t , respectively. We refer to H_t as the observational variance in the following; $X_t = [1, x_{t-1}]$, where x_t is the single predictor we choose at time t . Therefore, dynamic linear models differ with respect to the choices of predictors. In the Section 3.1, we discuss the set of predictors in x_t .

Given that we take a Bayesian perspective, denote $D_t = [r_t, r_{t-1}, \dots, X_t, X_{t-1}, \dots, Priors]$ as the information set available at t , which includes all the previous information about excess stock return values, predictor values, as well as the priors for coefficients θ_0 and observational variance H_0 . Essentially, we use a simple Kalman filter, following Raftery et al. (2010) and Koop and Korobilis (2012), to incorporate forgetting factors into the evolution of the parameters to capture the dynamics of the estimated coefficients and reduce the large computational burden. To explain how it works, start from the standard Kalman filter. We can obtain the posterior distribution for θ_{t-1} as u_t follows a normal distribution with mean zero and covariance Q_t :

$$\theta_{t-1} | D_{t-1} \sim N(\hat{\theta}_{t-1}, \Sigma_{t-1|t-1}) \quad (3)$$

Kalman filter predicts θ_t conditional on the information up to time $t - 1$:

$$\theta_t | D_{t-1} \sim N(\hat{\theta}_t, \Sigma_{t|t-1}) \quad (4)$$

$$\Sigma_{t|t-1} = \Sigma_{t-1|t-1} + Q_t \quad (5)$$

Raftery et al. (2010) and Koop and Korobilis (2012) suggest using a form of forgetting to ease computational demands in stead of specifying the matrix Q_t . In particular, equation (5) is replaced by:

$$\Sigma_{t|t-1} = \frac{1}{\lambda} \Sigma_{t-1|t-1} \quad (6)$$

where λ is the forgetting factor and $0 < \lambda \leq 1$. Therefore, according to equation (6), the forgetting factor λ is essential in influencing the degree of time-variation in coefficients. Effectively, the forgetting factor λ can be interpreted as the age-weights for different time point, thus, the effective window size used for estimation is $1/(1-\lambda)$. For example, setting $\lambda = 1$ means that the covariances of coefficients, $\Sigma_{t|t-1}$ are constant over time, thus, the coefficients themselves are constant over time. This also implies that a constant-coefficients predictive model is nested in equation (1). Whereas, setting $\lambda < 1$ implies that the covariances, $\Sigma_{t|t-1}$ increase over time and coefficients are time-varying. Moreover, the lower the value of λ s, the more abrupt the coefficients change. If we assume $\lambda=0.99$ with monthly data, it means that observations for the covariances of coefficients, $\Sigma_{t|t-1}$, last year has approximately 89% as much weight as last month. This is therefore the case when we have gradually changing coefficients. When the coefficients change suddenly, with $\lambda=0.90$, observations one year ago only account for 28% as much weight as last month's. This implies that λ has substantial influence on coefficient stability and different degrees of λ lead to different dynamic linear models.

In the appendix, we provide details on the predictive distribution and time-varying volatility.

2.2 Dynamic Mixture Model Averaging

In Section 2, we argued that different predictors and degrees of time-variation in coefficients may influence our ability to forecast stock returns. Improper model selection can increase total variance of return prediction and affect the accuracy of statistical inference.

However, one specification from the substantial pool is unlikely to dominate all others at each point in time. Since a single model may not be systematically the most successful at prediction, one potential approach is to take model uncertainty into account and compare all the models simultaneously. Dynamic Model Averaging (DMA) allows for time-varying coefficients and weights attached to each model changing over time based upon their past forecasting performances (Raftery et al., 2010; Koop and Korobilis, 2012). However, Raftery et al. (2010) and Koop and Korobilis (2012) assume ex-ante that coefficients in the predictive regressions and forecasting models change in the same fashion over time. Unlike DMA,

we construct Dynamic Mixture Model Averaging (DMMA) to allow for possible degrees of time-variation in coefficients and forecasting models, nesting gradual to abrupt change, and even constant coefficients and forecasting models. Therefore, in addition to the uncertainty about the choice of the predictors, we consider two more model uncertainties compared to Raftery et al. (2010) and Koop and Korobilis (2012): uncertainty regarding different degrees of time-variation in coefficients and in forecasting models. Using a Bayesian approach, the data identifies the precise degree of time-variation in coefficients and forecasting models by attaching posterior weights to possible models. This flexibility allows us to detect locally proper models over time.

In general, DMA suggests that the degrees of time-variation for coefficients and forecasting models are controlled by two forgetting factors. The first forgetting factor is the λ we mentioned above, which controls the degree of time-variation in coefficients. The other forgetting factor is α , which controls the degree of time-variation in forecasting models and will be explained later. Particularly, denote k_i a choice of predictors from K candidates, λ_j a choice of time-variation in coefficients from d candidates and α_z a choice of time-variation in forecasting models from a candidates. Then, there are $d \cdot K$ possible dynamic linear models and a different ways of combining them. Consequently, the predictive density of individual models and final forecasting results depends on the selection of predictors (k_i), degrees of time-variation in coefficients (λ_j) and degrees of time-variation in forecasting models (α_z) as well. Further details on the model combination are available in the appendix.

Note that the prediction equation for different variables can be written in the form of conditional predictive density, thus, the predictive weight attached to each variable k_i is :

$$\begin{aligned}
 P(L_t = k_i \mid \lambda_j, \alpha_z, D_{t-1}) &\propto [P(L_{t-1} = k_i \mid \lambda_j, \alpha_z, D_{t-2})P(r_{t-1} \mid L_{t-1} = k_i, \lambda_j, \alpha_z, D_{t-2})]^{\alpha_z} \\
 &= \prod_{s=1}^{t-1} [P(r_{t-s} \mid L_{t-1} = k_i, \lambda_j, \alpha_z, D_{t-s-1})]^{\alpha_z}
 \end{aligned}
 \tag{7}$$

where L_t represents the model selected at time t .

Therefore, if the predictive density for variable k_i conditional on the degree of time-variation in coefficients (λ_j) and forecasting models (α_z) is high in the past, it will obtain

more weight at time t , which is controlled by the forgetting factor α . If a certain value is allowed for α , it might only be locally suitable and the model can be misspecified. Our task is to make the model averaging process as flexible as possible and allow data to detect appropriate models by considering different degrees of α .

Interestingly, some conventional models are nested in DMMA. When models are equally weighted, $\alpha = 0$, forecasting models are constant over time. Huang and Lee (2010) suggest equal weights dominates other forecasting methods, using equity premium prediction as an example. Similarly, Rapach et al. (2010) demonstrate the superior out-of-sample equity premium prediction of equal weights combination method. When $\alpha = 1$, there is no discounting, and therefore, no role for α , model averaging shrinks to normal BMA. BMA, the approach widely used in the stock return prediction literature (see for example Avramov (2002); Cremers (2002); Dangl and Halling (2012) and Turner (2015)), equally weights all the data from the more distant past and the more recent past. When $\alpha < 1$, model weights are time-varying. For instance, if $\alpha = 0.99$, given monthly data, forecast performance last year receives 89% weight compared to that last month, which is a gradual change. Whereas, if $\alpha = 0.90$, 89% abruptly changes to 28%.

The posterior probability of each $a \cdot d \cdot K$ specification is updated according to Bayes rule. In the appendix, we provide more details about Dynamic Mixture Model Averaging.

3 Empirical study design

3.1 Data description

We examine stock predictability using monthly data of the S&P 500 index excess returns, and in particular the difference between monthly return on the stock market and the risk-free rate. The choice of predictors is guided by the literature. In particular, numerous research relies on macroeconomic predictors to forecast equity premium while paying little attention to technical indicators. Neely et al. (2014), however, address the fact that macroeconomic predictors and technical indicators provide complementary information in terms of improving forecast accuracy and economic gains. Therefore, we consider macroeconomic predictors as well as technical predictors.

The macroeconomic predictors we apply are the most widely used in empirical excess returns models.² Therefore we use the data set from Welch and Goyal (2008).³ Table 1 provides a short description of 12 macroeconomic predictors⁴ for the sake of brevity (see Welch and Goyal (2008) for details).

Table 1: List of Predictors

<i>No.</i>	<i>Predictor ID</i>	<i>Description</i>
Macroeconomic Indicators		
1	dy	Dividend yield
2	ep	Earnings-price ratio
3	de	Dividend-payout ratio
4	svar	Stock variance
5	bm	Book-to-Market ratio
6	ntis	Net Equity Expansion
7	tbl	Treasury bill rate
8	lty	Long-term yield
9	ltr	Long-term return
10	dfy	Default Yield Spread
11	dfr	Default Return Spread
12	infl	Inflation
Technical Indicators		
1	MA(1,9)	Moving average strategy
2	MA(2,9)	
3	MA(3,9)	
4	MA(1,12)	
5	MA(2,12)	
6	MA(3,12)	
7	MOM(9)	Momentum strategy
8	MOM(12)	
9	VOL(1,9)	Volume strategy
10	VOL(2,9)	
11	VOL(3,9)	
12	VOL(1,12)	
13	VOL(2,12)	
14	VOL(3,12)	

Notes: This tables shows the description of macroeconomic and technical predictors. Data is from December 1950 to December 2015.

Following Neely et al. (2014), we also construct 14 technical predictors based on three

²Some powerful macroeconomic predictors are uncovered, including net payout yield (Boudoukh et al., 2007), investor sentiment aligned (Huang et al., 2015) and short interest (Rapach et al., 2016). All of them are suggested to have comparably good out-of-sample performances. However, in our paper, we exclude these predictors due to data availability.

³The data are available from Amit Goyal’s webpage at <http://www.hec.unil.ch/agoyal/>.

⁴Compared with Welch and Goyal (2008), we exclude quarterly observations, also ‘dividend-to-price ratio’ and ‘term spread’ for collinearity reasons. ‘Cross-section premium’ is also excluded as it is only available from May 1937 to December 2002. Whereas, except for ‘cross-section premium’, all the other data can extend to a longer time period: January 1927 to December 2015.

strategies: moving-averages (MA), momentum (MOM) and volume (VOL). Table 1 illustrates the technical predictors we consider.⁵ $MA(s, l)$ generates a buy signal if the stock price in the short(s) MA is larger than that in the long(l) MA ($s = 1, 2, 3$ and $l = 9, 12$). $MOM(m)$ shows a positive momentum effect if the current prices is larger than the price m periods ago ($m = 9, 12$). $VOL(s, l)$ indicates a strong market trend if recent stock market volume as well as the stock price increases, where $s = 1, 2, 3$ ($l = 9, 12$) controls the recent(distant) past. Hence, we employ 26 macro and technical predictors to predict equity premium,⁶ and our whole data set is from December 1950 to December 2015.⁷

3.2 Prior choices

The method described in Section 2 requires appropriate priors and choices of λ and α . First, we suggest the prior of the coefficient θ_0 in the predictive regression is:

$$\theta_0 \sim N(\theta_{0|0}, \Sigma_{0|0}) \quad (8)$$

$\theta_{0|0}$ is the OLS estimate of coefficients in the training period. Similarly, the variance-covariance matrix $\Sigma_{0|0}$ is the corresponding OLS estimate of coefficients' covariance in the training period.⁸

Second, we need to choose possible degrees of λ and α . In experiments, we find that results deteriorate after λ drops to 0.90 and α decreases to 0.90. Models with equal weights ($\alpha=0$) are also included. Thus, we consider $\lambda = [0.90, 0.95; 0.99; 1]$ and $\alpha = [0; 0.90; 0.95; 0.99; 1]$ for time-variation in coefficients and forecasting models, which seems to cover all the likely values given monthly data⁹. $\lambda, \alpha = 0.95$ or 0.99 , are the values considered by Koop and Korobilis (2012) for DMA in an inflation prediction context. The lower bound for λ and α is 0.90 (except $\alpha = 0$), implying sudden changing coefficients. Based on the range of time-variation

⁵See online appendix for details.

⁶If we consider all the models generated by the 26 predictors, there would be 2^{26} model specifications, which would present excessive computational demands. Consequently, in this paper, we consider single predictor in each regression as demonstrated in Section 2.

⁷According, if the training period is ten years, the sample period will start from November 1960. Choosing such a long period and omitting other predictors, our aim is to alleviate worries of sample selection bias for our results.

⁸ We repeat the analysis using noninformative prior $\theta_0 \sim N(0, \Sigma_{0|0})$ and get similar results.

⁹Dangl and Halling (2012) consider a more narrow range, $\lambda = [0.96; 0.98; 1]$.

in coefficients and forecasting models, we then study which value of λ and α is supported by the data.

We initially assign a diffuse conditional prior for different choices of predictor, different degrees of time-variation in coefficients as well as in forecasting models, which means, $P(\alpha_z | D_0) = 1/a = 1/5$, $P(\lambda_j | \alpha_z, D_0) = 1/d = 1/4$ and $P(k_i | \lambda_j, \alpha_z, D_0) = 1/K = 1/26$. Therefore, each predictor and model specification has the sample probability at the beginning.

4 Empirical Results

Our results section begins by examining whether there is out-of-sample predictability for the DMMA model, and whether incorporating possible degrees of time-variation in coefficients and in forecasting models lead to forecasting improvements. Next, we decompose prediction variance and highlight the different sources in forecasting power compared to other models. Finally, we link predictability to the business cycle.

4.1 Out-of-sample predictability

We begin our formal analysis by comparing our Dynamic Mixture Model Averaging (DMMA) with an historical mean (HM) model, in terms of the out-of-sample R^2 (R_{OS}^2), Clark and West (2007) statistics and model predictive log likelihoods. Welch and Goyal (2008) indicate that HM can be a strict out-of-sample benchmark, which most predictors fail to outperform. Specifically, HM excludes predictors and only includes a constant term in the regressions. Thus, HM is nested in our set of predictive regressions. Here we assume that the coefficient and volatility of HM is constant following previous studies (Welch and Goyal, 2008; Campbell and Thompson, 2008; Dangl and Halling, 2012).

Moreover, we consider some representative models for comparison and the model set is summarized in Table 2:

Table 2: Model Sets

<i>Model specification</i>	<i>Predictors</i>	<i>Coefficients</i>	<i>Forecasting models</i>
DMMA	-	-	-
EW	-	-	$\alpha = 0$
CC-EW	-	$\lambda = 1$	$\alpha = 0$
BMA	-	-	$\alpha = 1$
CC-BMA	-	$\lambda = 1$	$\alpha = 1$
DMA	-	$0 < \lambda < 1$	$0 < \alpha < 1$
HM	$k = 0$	$\lambda = 1$	-

Notes: The table shows the model set with imposed restrictions on choices of predictors, time-variation in coefficients and time-variation in forecasting models. k refers to number of predictors, λ specifies the degree of time-variation in coefficient and α indicates the degree of time-variation in forecasting model. (-) implies that no restrictions are imposed.

- DMMA: Forecasts using Dynamic Mixture Model Averaging. Specially, possible degrees of time-variation in coefficients $\lambda = [0.9, 0.95; 0.99; 1]$ and possible degrees of time-variation in forecasting models $\alpha = [0; 0.90; 0.95; 0.99; 1]$.
- EW: Forecasts using equal weighted models ($\alpha = 0$).
- CC-EW: Forecasts using constant coefficients and equal weights ($\lambda = 1$ and $\alpha = 0$).
- BMA: Forecasts using Bayesian Model Averaging ($\alpha = 1$).
- CC-BMA: Forecasts using constant coefficients and Bayesian Model Averaging ($\lambda = 1$ and $\alpha = 1$).
- DMA: Forecasts using Dynamic Model Averaging with time-varying coefficients and forecasting models ($0 < \lambda < 1$ and $0 < \alpha < 1$).
- HM: Forecasts using historical mean model without any predictors while keeping the coefficients and volatility constant ($k = 0$ and $\lambda = 1$).

Although DMMA is flexible enough to nest different simple model specifications, our goal is not to construct the most general model specification. Rather, we aim to incorporate a number of features that may be essential for forecast accuracy and portfolio allocation, including multiple predictors, time-varying volatility, time-evolving coefficients and time-evolving forecasting models.

Note that as out-of-sample predictability can be spurious and driven by some outliers, it would be inaccurate and unreasonable to focus on only one sample period. Hence, in light of the analysis from Dangl and Halling (2012), we study three different sample periods (1960+, 1976+ and 1988+) to confirm our results. In particular, the literature has suggested that out-of-sample stock return predictability is mainly driven by exceptional periods such as oil price shock (1975) and the stock market crash (1987) (Welch and Goyal, 2008; Rapach et al., 2010; Dangl and Halling, 2012). To get rid of the disturbances of distress, two subsamples begin from 1976 and 1988, respectively.

4.1.1 Statistical evaluation

We use out-of-sample R^2 (R_{OS}^2), Clark and West (2007) statistics and model predictive log likelihoods for different subsamples to statistically evaluate our model's out-of-sample predictability. In detail, the first evaluation, R_{OS}^2 , as acknowledged by Campbell and Thompson (2008), is the fractional reduction in mean squared forecast error (MSFE) for the predictive model compared to HM benchmark,

$$R_{OS}^2 = 1 - \frac{\sum_{t=\underline{t}}^{\bar{t}} (r_{t+1} - \hat{r}_{t+1})^2}{\sum_{t=\underline{t}}^{\bar{t}} (r_{t+1} - \bar{r}_{t+1})^2} \quad (9)$$

where \underline{t} is the starting of the evaluation period, \bar{t} is the end of the evaluation period, \hat{r}_{t+1} is the estimated prediction from a regression using the information at time t and \bar{r}_{t+1} denotes the estimated HM at time t . If $R_{OS}^2 > 0$, MSFE of the predictive regression is smaller than that of HM, thus, \hat{r}_{t+1} has more accurate prediction than \bar{r}_{t+1} .

The second measurement we report is the widely used Clark and West (2007) test (CW), which evaluates the statistical differences in forecasts. The advantage of the CW test is that it still follows an asymptotically standard normal distribution when comparing to the forecasting results of nested models, which is exactly our case since HM is nested in our general DMMA framework. CW statistics test the null hypothesis that the MSFE of HM is less than or equal to that of predictive model, against the upper tail alternative hypothesis that MSFE of HM is greater than that of predictive regression.

The third statistical criteria we use to assess our model is the log predictive likelihood

($\text{Log}(PL)$). Involving the entire predictive distribution, log predictive likelihood is a frequently used evaluation method for Bayesian models (Geweke and Amisano, 2011). The larger the log predictive likelihood, the better the forecasts in a Bayesian comparison.

Table 3 presents the first set of core statistical results from our empirical analysis. The overall story is clear: in terms of out-of-sample R^2 , Clark and West (2007) test and log predictive likelihood, DMMA outperforms HM for all subsample periods. Moreover, DMMA has better results than other model combination methods including equal weights, BMA, DMA in terms of point forecast accuracy.

We first examine Dynamic Mixture Model Averaging’s performances in greater detail compared to the no-predictability benchmark in Table 3. DMMA takes all sources of model uncertainties into account, allowing for different choices of predictors, varying degrees of coefficient and forecasting models adaptivity. We find that DMMA significantly and consistently outperforms HM, with R_{OS}^2 larger than zero and substantially larger log likelihoods than HM across three different sample periods. The conclusion that DMMA has statistically lower forecast errors than HM is confirmed by the CW test. This test rejects the null hypothesis that the MSFE of HM is less than or equal to that of predictive model for all time periods. It is worth noting that DMMA’s out-of-sample statistical performances slightly worsen during period 1976+ and period 1988+, which is consistent with the finding that prediction accuracy for forecasting excess stock returns may be driven by oil price shock in 1973-1975 and the stock price crash in 1987 (Campbell and Thompson, 2008; Dangl and Halling, 2012; Joscha and Schüssler, 2014). However, DMMA still outperforms HM for sample periods 1976+ and 1988+, signalling that our framework is robust to many sample periods.

Next, we study the models that are nested by in DMMA and examine what features lead to forecasting improvements. In particular, we present the results for equal weighted models, Bayesian Model Averaging (BMA), and Dynamic Model Averaging (DMA) we built on, in Panel B, C and D of Table 3 respectively. Looking at the results for equal weights (EW) in Panel B, we find that the simple combination method, has reasonable performances. It is not a surprise as a model with equal weights is a tough benchmark in the forecasting combination literature (Rapach et al., 2010; Huang and Lee, 2010; Geweke and Amisano, 2012). We further study how time-varying coefficient influences the results. R_{OS}^2 for constant coefficient with

Table 3: Statistical Evaluation

Panel A: Dynamic Mixture Model Averaging (DMMA)						
	Period:1960+		Period:1976+		Period:1988+	
	$R^2_{OS}(\%)$	$Log(PL)$	$R^2_{OS}(\%)$	$Log(PL)$	$R^2_{OS}(\%)$	$Log(PL)$
DMMA	1.72**	1141.70	0.91*	802.66	0.88*	573.13
Panel B: Equal Weights (EW)						
EW ($\alpha=0$)	1.44**	1142.80	0.88*	803.05	0.67	573.31
CC-EW ($\lambda=1, \alpha=0$)	0.92***	1141.00	0.75**	802.78	0.62*	573.52
Panel C: Bayesian Model Averaging (BMA)						
BMA ($\alpha=1$)	-1.26*	1133.30	-2.23	795.27	-0.15	571.08
CC-BMA ($\lambda=1, \alpha=1$)	-1.01*	1128.20	-2.35	793.53	-0.24	571.12
Panel D: Dynamic Model Averaging (DMA)						
$\lambda=0.90, \alpha=0.90$	-10.07	1092.70	-8.83	768.47	-10.34	547.14
$\lambda=0.95, \alpha=0.90$	-10.60	1091.10	-9.41	767.47	-11.15	546.34
$\lambda=0.99, \alpha=0.90$	-11.71	1091.00	-10.00	768.31	-11.48	547.33
$\lambda=0.90, \alpha=0.95$	-6.18	1115.40	-5.67	784.89	-6.92	559.32
$\lambda=0.95, \alpha=0.95$	-6.52	1114.10	-6.12	784.14	-7.57	558.79
$\lambda=0.99, \alpha=0.95$	-6.32	1113.40	-5.89	784.28	-6.94	559.29
$\lambda=0.90, \alpha=0.99$	-0.39	1136.50	-1.03	798.44	-1.23	569.43
$\lambda=0.95, \alpha=0.99$	-0.53	1135.60	-1.16	798.11	-1.46	569.17
$\lambda=0.99, \alpha=0.99$	-0.54	1133.80	-1.13	797.59	-1.47	569.04
HM	0	128.56	0	293.27	0	304.43

Notes: Statistical predictability for different models using out-of-sample R^2 ($R^2_{OS}(\%)$), Clark and West test (*, ** and *** show that the null hypothesis that the MSFE of HM is less than or equal to that of predictive model, is rejected at the 10%, 5% and 1% significance level, respectively) and predictive log likelihoods ($Log(PL)$). We consider different model combination method, including equal weights (EW) in Panel B, Bayesian Model Averaging (BMA) in Panel C and also all the possible Dynamic Model Averaging (DMA) models in Panel D. Bold font suggests the statistics of that predictive model is larger than the corresponding one of HM. Following Dangl and Halling (2012), we study three different evaluation periods: 1960+, 1976+ and 1988+.

equal model weights (CC-EW) deteriorates compared to time-varying coefficient with equal weights (EW), confirming the finding in the literature that parameter instability matters for return predictability.

Turning to Table 3 Panel C, we investigate Bayesian Model Averaging (BMA). Interestingly, we find BMA cannot outperform HM for our dataset, with negative R_{OS}^2 . BMA is a commonly used technique to tackle model uncertainty and to combine models together according to their posterior probabilities (Avramov, 2002; Cremers, 2002; Dangi and Halling, 2012; Turner, 2015). However, the biggest problem of equal weights and BMA, as acknowledged by Geweke and Amisano (2012), is that “they both condition on one of the models under consideration being true.” Moreover, equal weights and BMA procedures assume their method is appropriate across the whole sample period, which seems inappropriate especially for the sporadically volatile stock market. DMMA, however, detects the locally appropriate time-variation in coefficient and in forecasting model at each time. Therefore, it is not surprising to find that DMMA improves upon models with equal weights (EW) and models with BMA with regard to R_{OS}^2 . These results suggest that allowing different degrees of time-variation in forecasting model is important for improving forecast accuracy, as it flexibly accommodates accommodating changes in data.

Last but not least, we employ possible specifications of Dynamic Model Averaging (DMA) in Panel D of Table 3, in which we assume that coefficients and forecasting models change in the same fashion over time. All DMAs do not forecast as well as HM and none of them have positive R_{OS}^2 and p values less than 10% for CW test. Moreover, data prefers gradual changes in forecasting models, as the DMAs with $\alpha = 0.9$ is much worse than the ones with $\alpha = 0.99$, indicating the importance of choosing the exact time-variation in forecasting models. In contrast, DMMA, the method based on DMA, has superior out-of-sample improvement compared to DMA. This further confirms that even if we take time-varying coefficients and forecasting models into account, predictability can still disappear if we ignore the importance of evolving degrees of coefficients and models adaptivity.

All in all, DMMA substantially outperforms HM, equal weights, time-varying coefficients with BMA as well as DMA across different subsample periods. By detecting locally appropriate degree of time-variation in coefficients and forecasting models, DMMA can predict

with misspecified models and quickly adapt the dynamics in the data generating process.¹⁰ Consequently, DMMA improves upon all the nested models in terms of out-of-sample R^2 .

4.1.2 Economic evaluation

In the previous section we confirm that large out-of-sample R_{OS}^2 can be obtained using DMMA to predict stock returns, but is this meaningful for investors and traders?

Campbell and Thompson (2008) propose that if we can observe the predictors, the percentage increase for the expected return is :

$$\left(\frac{R_{OS}^2}{1 - R_{OS}^2}\right)\left(\frac{1 + SR^2}{SR^2}\right) \quad (10)$$

where SR^2 is the square of the Sharpe ratio. Equation (10) is larger than R_{OS}^2/SR^2 , and converges to R_{OS}^2/SR^2 if R_{OS}^2 and SR^2 are both small.

In light of this, the appropriate way to evaluate R_{OS}^2 is to calculate R_{OS}^2/SR^2 . If the ratio is larger than 1, indicating that R_{OS}^2 is larger than SR^2 , investors can obtain higher portfolio return by using the results in the predictive model. For instance, from Table 3 and 4 period 1960+ for DMMA model, we know $R_{OS}^2 = 1.72\%$ and $SR^2 = 4.9\%$. A mean-variance investor thus can use DMMA to find out the optimal portfolio returns and increase his average monthly portfolio returns substantially by $1.72\%/4.9\% = 35.1\%$.

Note that investors need to take risks to get the high portfolio return mentioned above. To exclude risk especially for a mean-variance risk-averse investor who allocates his wealth between equities and risk-free assets using forecasting results from our variance models, we report certainty equivalent return (CER). The expected utility for the mean-variance investor is:

$$U(R_p) = E(R_p) - \frac{1}{2}\gamma Var(R_p) \quad (11)$$

where R_p is the investors' portfolio return, $E(R_p)$ is the expected value of the return, $Var(R_p)$ is the variance of the return and $\gamma = 3$.¹¹ At the end of t , the investor optimally allocates a

¹⁰Raftery et al. (2010) show that DMA rapidly accommodates changes in coefficients and changes in the entire forecasting models, by employing a simulation study. DMMA offers greater flexibility than DMA, and is capable to detect changes in the stock market.

¹¹We repeat same analysis for $\gamma = 5$. Similar results are obtained.

portfolio weight in the risky asset

$$w_t = \frac{1}{\gamma} \frac{\hat{r}_{t+1}}{\hat{\sigma}_{t+1}^2}$$

where \hat{r}_{t+1} is the forecast of excess stock return and $\hat{\sigma}_{t+1}^2$ is the forecast of its variance. Following Campbell and Thompson (2008) and Neely et al. (2014), we limit the percentage invested in equities to be between 0% and 150% and assume that a five-year moving window of past returns is used to estimate the variance forecasts.

The CER for the portfolio is

$$CER_p = \hat{\mu}_p - \frac{1}{2} \gamma \hat{\sigma}_p^2 \tag{12}$$

where $\hat{\mu}_p$ and $\hat{\sigma}_p^2$ are the mean and variance for the investor's entire portfolio over the sample period. Monthly CER are annualized by multiplying by 1200. Moreover, we consider the effect of transaction costs in CER following Balduzzi and Lynch (1999) and Neely et al. (2014), where the costs are measured using the percentage change of wealth traded each month and assuming a proportional transactions cost equal to 50 basis points per transaction.

Table 4 shows certainty equivalent return (*CER*) and Sharp Ratio (*SR*) for different models over different sample periods. The conclusion that DMMA has superior forecasts than HM is confirmed and strengthened using this economic criteria. In particular, DMMA has *CER* at most 741 basis points and its *SR* is larger than HM's across all sample periods.

Compared with other predictive models, DMMA has good economic performance, and in no case much worse than the best alternatives. A model with equal weights and Dynamic Model Averaging (DMA) perform occasionally better than DMMA. Although DMMA consistently has the best point forecast accuracy, there is slight disagreement between statistical evaluation and economic evaluation. This is in consonance with the finding of Cenesizoglu and Timmermann (2012) who suggest that there is a weak link between point forecast accuracy (e.g. out-of-sample R^2) and economic value. Interestingly, imposing constant coefficient restriction negatively affect model's out-of-sample performance, indicating the importance of time-varying coefficients for economic gains.

Table 4: Economic Evaluation

Panel A: Dynamic Mixture Model Averaging (DMMA)						
	Period:1960+		Period:1976+		Period:1988+	
	<i>CER</i>	<i>SR</i>	<i>CER</i>	<i>SR</i>	<i>CER</i>	<i>SR</i>
DMMA	5.15	0.08	6.24	0.10	7.41	0.15
Panel B: Equal Weights (EW)						
EW ($\alpha=0$)	5.44	0.09	6.34	0.11	7.26	0.14
CC-EW ($\lambda=1, \alpha=0$)	4.15	0.07	5.54	0.10	7.09	0.14
Panel C: Bayesian Model Averaging (BMA)						
BMA ($\alpha=1$)	5.04	0.07	5.56	0.09	6.30	0.13
CC-BMA ($\lambda=1, \alpha=1$)	2.49	0.03	4.13	0.07	5.74	0.12
Panel D: Dynamic Model Averaging (DMA)						
$\lambda=0.90, \alpha=0.90$	3.27	0.04	4.90	0.08	7.04	0.14
$\lambda=0.95, \alpha=0.90$	3.39	0.04	4.48	0.07	6.62	0.13
$\lambda=0.99, \alpha=0.90$	3.23	0.04	4.53	0.07	6.52	0.13
$\lambda=0.90, \alpha=0.95$	4.12	0.05	5.05	0.08	7.01	0.14
$\lambda=0.95, \alpha=0.95$	4.23	0.06	5.55	0.09	7.53	0.15
$\lambda=0.99, \alpha=0.95$	4.89	0.07	5.85	0.09	7.56	0.15
$\lambda=0.90, \alpha=0.99$	4.90	0.07	4.97	0.08	6.00	0.12
$\lambda=0.95, \alpha=0.99$	4.72	0.07	4.81	0.08	6.05	0.12
$\lambda=0.99, \alpha=0.99$	4.92	0.07	6.22	0.10	6.50	0.13
HM	3.39	0.07	3.93	0.08	5.25	0.12

Notes: Economic predictability for different predictive models using certainty equivalent return gain (*CER*) and sharp ratio (*SR*) compared with historical mean (HM) for a mean-variance investor who allocates wealth between equities and risk-free assets using different forecasts from different models. We consider different model combination method, including equal weights (EW) in Panel B, Bayesian Model Averaging (BMA) in Panel C and also all the possible Dynamic Model Averaging (DMA) models in Panel D. The risk aversion coefficient for a mean-variance investor is 3 and we assume a 50 basis points percentage transactions cost per transaction when calculating *CER*. Bold font suggests that the statistics of that predictive model is larger than the corresponding one of HM. Following Dangl and Halling (2012), we study three different evaluation periods: 1960+, 1976+ and 1988+.

4.2 Sources of Prediction Uncertainty

Our study is innovative since our DMMA approach can outperform the historical mean statistically and economically, but also because we delineate forecast errors beyond the standard approach. This means that the relative importance for predictors, time-varying coefficients, and the individual model weights can be tracked over time. In this framework, the prediction variance of the excess stock return can be decomposed. By doing so, we understand our model's underlying features and the source of forecasting power. This constitutes one of the critical contributions of this paper.

We begin with the decomposition with regard to different degrees of time-variation in models α_z , based on the *Law of Total Variance*, prediction variance can be written as:

$$\text{Var}(r_t) = \mathbb{E}_{\alpha_z}(\text{Var}(r_t | \alpha_z)) + \text{Var}_{\alpha_z}(\mathbb{E}(r_t | \alpha_z)) \quad (13)$$

where \mathbb{E}_{α_z} and Var_{α_z} are the expectation and prediction variance with regard to α_z . We can further decompose the term $\text{Var}(r_t | \alpha_z)$ in equation (13) with respect to different degrees of time-variation in coefficients λ_j into:

$$\text{Var}(r_t | \alpha_z) = \mathbb{E}_{\lambda_j}(\text{Var}(r_t | \lambda_j, \alpha_z)) + \text{Var}_{\lambda_j}(\mathbb{E}(r_t | \lambda_j, \alpha_z)) \quad (14)$$

Similarly, $\text{Var}(r_t | \lambda_j, \alpha_z)$ in equation (14) can be written conditional on different choices of predictors k_i :

$$\text{Var}(r_t | \lambda_j, \alpha_z) = \mathbb{E}_{k_i}(\text{Var}(r_t | k_i, \lambda_j, \alpha_z)) + \text{Var}_{k_i}(\mathbb{E}(r_t | k_i, \lambda_j, \alpha_z)) \quad (15)$$

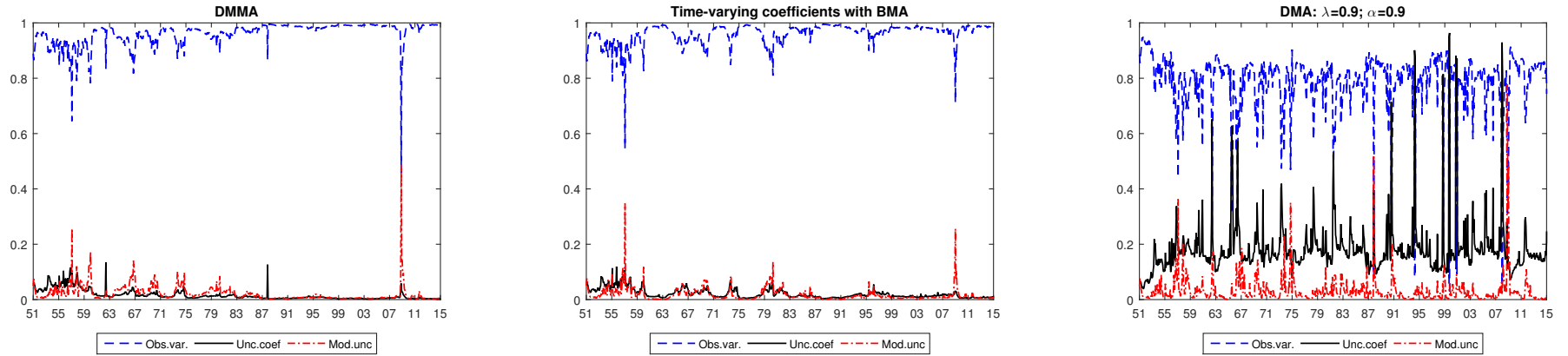
Finally, substitute equation (15) and (14) into (13), we obtain:

$$\begin{aligned}
Var(r_t) &= \mathbb{E}_{k_i, \lambda_j, \alpha_z} [Var(r_t | k_i, \lambda_j, \alpha_z)] + \mathbb{E}_{\lambda_j, \alpha_z} \{Var_{k_i}[\mathbb{E}(r_t | k_i, \lambda_j, \alpha_z)]\} \\
&\quad + \mathbb{E}_{\alpha_z} \{Var_{\lambda_j}[\mathbb{E}(r_t | \lambda_j, \alpha_z)]\} + Var_{\alpha_z}[\mathbb{E}(r_t | \alpha_z)] \\
&= \sum_z \left\{ \sum_j [(H_t | k_i, \lambda_j, \alpha_z, D_t)] P(k_i | \lambda_j, \alpha_z, D_t) P(\lambda_j | \alpha_z, D_t) \right\} P(\alpha_z | D_t) \\
&\quad + \sum_z \left\{ \sum_j [(X_t \Sigma_{t+1|t} X_t' | k_i, \lambda_j, \alpha_z, D_t)] P(k_i | \lambda_j, \alpha_z, D_t) P(\lambda_j | \alpha_z, D_t) \right\} P(\alpha_z | D_t) \\
&\quad + \sum_z \left\{ \sum_j [(\hat{r}_{t+1,i}^{j,z} - \hat{r}_{t+1}^{j,z})^2 P(k_i | \lambda_j, \alpha_z, D_t)] P(\lambda_j | \alpha_z, D_t) \right\} P(\alpha_z | D_t) \\
&\quad + \sum_z \left[\sum_j (\hat{r}_{t+1}^{j,z} - \hat{r}_{t+1}^z)^2 P(\lambda_j | \alpha_z, D_t) \right] P(\alpha_z | D_t) \\
&\quad + \sum_z (\hat{r}_{t+1}^z - \hat{r}_{t+1})^2 P(\alpha_z | D_t)
\end{aligned} \tag{16}$$

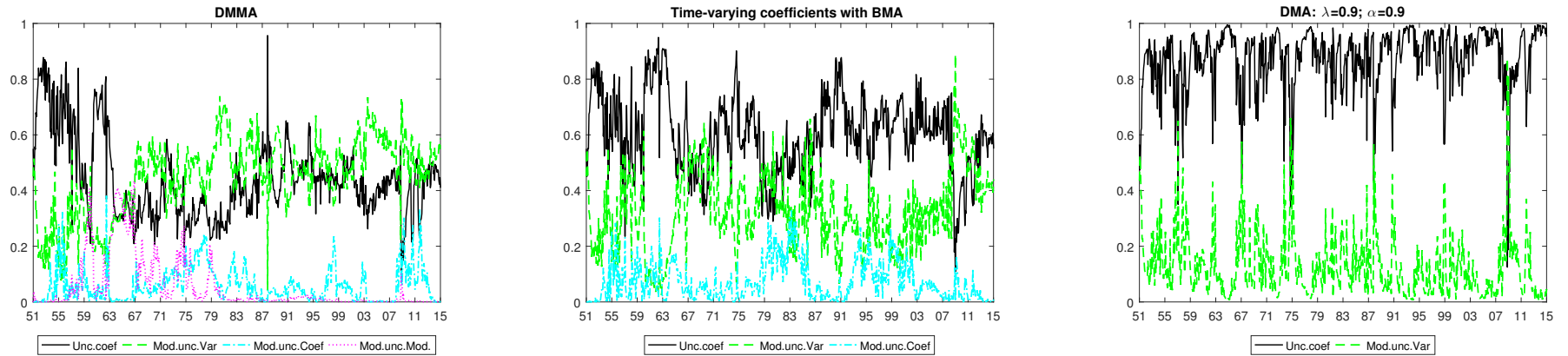
Hence, equation (16) sheds light on the sources of uncertainty of return prediction. Intuitively, the first term captures the expected variance of the innovation term in the measurement equation, conditional on the choices of predictors k_i , degree of time-variation in coefficients λ_j and model α_z . We call it observational variance. The second term indicates the expected variance of errors in coefficients, which can be classified as estimation uncertainty in coefficients. Whereas, the remaining components of equation (16) are referred to as model uncertainty. In particular, the third term characterizes model uncertainty with regard to predictor selection. The fourth term measures model uncertainty in terms of degree of time-variation in the coefficients. Finally, the fifth term states model uncertainty regarding the degree of time-variation in forecasting models. Dangi and Halling (2012) consider the first four sources of forecast errors. To the best of our knowledge, ours is the first paper in the return predictability literature that investigates model uncertainty with respect to different choices of time-variation in forecasting models.

Figure 1: Sources of Prediction Variance for DMMA, BMA and DMA

Panel A: All sources of prediction variance



Panel B: Prediction variance exc. Obs.var



Notes: Decomposition of the prediction variance for DMMA, possible degrees of time-variation in coefficients with BMA and DMA (take $\lambda = 0.9, \alpha = 0.9$ as an example). Panel A of the Figure plots the relative weights of observational variance (Obs.var.), expected variance from errors in the estimation of coefficients (Unc.coef.) and variance caused by the model uncertainty (Mod.unc.). Panel B excludes observational variance and investigates expected variance from errors in the estimation of coefficients (Unc.coef.), variance caused by the uncertainty regarding the variable selection (Mod.unc.Var), variance caused by the uncertainty regarding different degrees of time-variation in coefficients (Mod.unc.Coef) and in forecasting models (Mod.unc.Mod). Particularly, in Panel B, for time-varying coefficients with BMA, Mod.unc.Mod is neglected as $\alpha = 1$. Similarly, for DMA, Mod.unc.Mod and Mod.unc.Coef are excluded as λ and α are invariant.

Figure 1 depicts different sources of prediction variance for three approaches: (i) DMMA, (ii) possible degrees of time-variation in coefficients with BMA ($\alpha = 1$) and (iii) DMA model (take $\lambda = 0.9, \alpha = 0.9$ as an example).¹² Panel A of the Figure shows the relative weights of observational variance (Obs.var.), uncertainty about estimating coefficients (Unc.coef.) and model uncertainty (Mod.unc.).¹³ For all three approaches observational variance dominates. Dengl and Halling (2012), claim that this is conventional for stock return prediction because random fluctuations are expected to cause considerable volatility, especially for the one month forecast horizons we consider. Specifically, for DMMA and BMA, estimation risk and model uncertainty are small except for the initial data-points of the out-of-sample predictive process¹⁴ and the peak of model uncertainty during the financial crisis in 2008. Whereas, for DMA, estimation uncertainty accounts for around 20% of the total prediction variance and model uncertainty is nonnegligible.

There are also notable differences between our three models. With respect to estimation uncertainty and variance caused by predictor selection in Panel B of Figure 1, there are notable differences among DMMA, BMA and DMA. Importantly, for DMMA, uncertainty regarding coefficient estimation accounts for the largest proportion in the initial data points, however, it is of less importance after 1965. In the meantime, uncertainty with regard to predictor selection becomes crucial. Whereas, when we fix time-variation in the forecasting model using BMA, estimation uncertainty in coefficients dominates the remaining variance for most periods, only with occasional switches to model uncertainty with respect to predictor selection. If the coefficient and forecasting models change in the same fashion over time (DMA), estimation risk is prominent at the start of the sample. These first imply that observational uncertainty, estimation uncertainty and uncertainty with respect to predictor selection are the top three sources of prediction variance that hinder forecasting performance for different models. Second, DMMA has the smallest estimation error among different

¹²We consider $\lambda = 0.9, \alpha = 0.9$ because they represent extremely rapidly changing coefficient and forecasting models, which seems unreasonable in the stock market. Our aim is to demonstrate different sources of uncertainty when we choose inappropriate parameters.

¹³We present absolute values of different variances in the online appendix. DMA has the largest prediction variance. DMMA, in contrast, has the smallest variance.

¹⁴As priors influence results in a Bayesian framework, Geweke and Amisano (2010) argue that it is reasonable for the prediction variance to be sensitive for the initial forecasting data-points. However, results will be invariant to the prior distribution after data has been accumulated.

alternatives, whereas, when we select an inappropriate degree of time-variation in coefficients and in forecasting models, the estimation risk can be large.

Besides, turning to the rest of the model uncertainty in DMMA, we uncover that uncertainty regarding different choices of time-variation is negligible after the initial 30 years, implying that learning the dynamics in the time-varying forecasting models can take some time. Uncertainty with regard to different choices of time-variation in coefficients, however, is low except for the fluctuations around recessions (e.g., the post-Korean War recession between 1953-1954, the oil shock around 1973 and the financial crisis from 2007).

Importantly, uncertainty with respect to predictor selection cannot be neglected and it seems to be critical in leading forecasting improvements. As we will demonstrate in section 4.4 when we study the role of different predictors, all the predictors tend to be important in forecasting equity premium. This implies that the ensemble of features of DMMA is necessary, including combining multiple predictors, allowing varying degrees of coefficients adaptivity and different degrees of time-variation in forecasting models.

In spite of the fact that estimation uncertainty in coefficients is one of the key factors obstructing forecasting performance, our fully flexible DMMA model outperforms alternatives because it makes use of all the information in the dataset. In other words, our model effectively adapts the pattern in the unstable stock market by embedding the precise level of time-variation in coefficients and forecasting models, thus, the variances due to uncertainty about the choice of degree of time-variation in coefficients and forecasting models are small. When comparing the differences of variance decomposition between BMA and DMA, one possible explanation for DMMA's superior out-of-sample performance is that considering possible range of time-variation in coefficients increase coefficient variability, thus, offset the loss in forecast accuracy caused by the second largest source of prediction variance: coefficients estimation uncertainty. Moreover, by allowing for different degrees of time-variation in forecasting models, DMMA enhances model adaptability and quickly detects locally appropriate models, therefore, improves upon time-varying coefficients with BMA and further compensates the losses due to the estimation risk.

4.3 Link predictability to the business cycle

Previously, we provided evidence of economic and statistical predictability of our DMMA model relative to others over different sample periods. We also analyze what leads to forecasting improvements by decomposing variance. In this section, we link DMMA’s predictability to the business cycle.

Theoretically, excess stock returns predictability is closely related to the business cycle (Fama and French, 1989; Campbell and Cochrane, 1995; Cochrane, 1999, 2005; Rapach et al., 2010; Dangl and Halling, 2012). In general, investors are more risk-averse during recessions, who, in turn, ask for much higher excess stock returns for risk compensation. As a consequence, the equity premium tends to decrease during expansions and increase during recessions. Moreover, local maxima of the equity risk premia often appears to be near business cycles troughs, whereas, local minima occurs near business cycles peaks (Fama and French, 1989; Campbell and Cochrane, 1995; Cochrane, 1999). In this framework, DMMA’s predictability would rise if it could capture the business cycle (Rapach et al., 2010; Henkel et al., 2011; Dangl and Halling, 2012).

Rapach et al. (2010) systematically and empirically study the link between prediction improvements and business cycle. Similarly, they predict equity premium by combining individual predictive regression models together and find that the combination method has superior out-of-sample prediction of excess stock returns which, in turn, better links to the business cycle when comparing to individual forecasts and the historical mean model. Rapach et al. (2010) argue that the reason why the historical mean model cannot capture business-cycle fluctuations is because it always produces a very smooth prediction, therefore, fails to incorporate macroeconomic information. With respect to the individual predictive regressions, they may contain false signals and exhibit implausible fluctuations.

We use NBER recessions and expansions data to identify how close the predictability of the DMMA model is linked to the business cycle. Table 5 reports two statistics: out-of-sample R^2 (R_{OS}^2) and CER gains (ΔCER) relative to the no-predictability benchmark during recessions and expansions over different sample periods for various predictive models. When looking at results for DMMA, we find that out-of-sample R^2 is substantially larger during

Table 5: Business Cycle Analysis

Panel A: Dynamic Mixture Model Averaging (DMMA)				
	$R_{OS}^2(\%)$		ΔCER	
	Recession	Expansion	Recession	Expansion
DMMA	5.19	0.23	19.01	0.22
Panel B: Equal Weights (EW)				
EW ($\alpha=0$)	4.22	0.25	17.90	0.18
CC-EW ($\lambda=1, \alpha=0$)	1.93	0.48	5.18	0.63
Panel C: Bayesian Model Averaging (BMA)				
BMA ($\alpha=1$)	2.60	-2.92	20.65	-1.31
CC-BMA ($\lambda=1, \alpha=1$)	2.83	-2.66	17.90	-1.90
Panel D: Dynamic Model Averaging				
$\lambda=0.90, \alpha=0.90$	-10.70	-9.80	18.14	-2.78
$\lambda=0.95, \alpha=0.90$	-11.56	-10.19	18.56	-2.88
$\lambda=0.99, \alpha=0.90$	-12.40	-11.42	19.66	-2.62
$\lambda=0.90, \alpha=0.95$	-7.74	-5.52	22.80	-2.87
$\lambda=0.95, \alpha=0.95$	-8.40	-5.71	21.25	-2.78
$\lambda=0.99, \alpha=0.95$	-6.70	-6.15	23.00	-2.39
$\lambda=0.90, \alpha=0.99$	2.11	-1.46	12.04	-1.87
$\lambda=0.95, \alpha=0.99$	1.79	-1.53	12.32	-1.63
$\lambda=0.99, \alpha=0.99$	2.44	-1.81	17.37	-1.23

Notes: Business cycle analysis using out-of-sample R^2 ($R_{OS}^2(\%)$) and certainty equivalent return gain (ΔCER) compared with historical mean (HM). Bold font suggests that the statistics of that predictive model is larger than the corresponding one of HM. The sample period is from November 1961 to December 2015.

recessions than expansions. In terms of economic evaluation, DMMA shows its unique power to generate considerable returns especially in recessions, with *CER* gains during recessions approximately 41 times larger than that during expansions. The fact that predictability will rise during recessions is in line with the empirical evidence provided by Rapach et al. (2010), Henkel et al. (2011) and Dangel and Halling (2012). Researchers argue that this is because HM model overestimates the equity premium, therefore, suffers from huge losses particularly in recessions. Interestingly, DMMA also has positive out-of-sample R^2 and *CER* gains during expansions, which further confirms the strong predictability of DMMA. This result is consistent with the findings of Dangel and Halling (2012), who claim there is predictability during expansions using their predictors and econometric method.

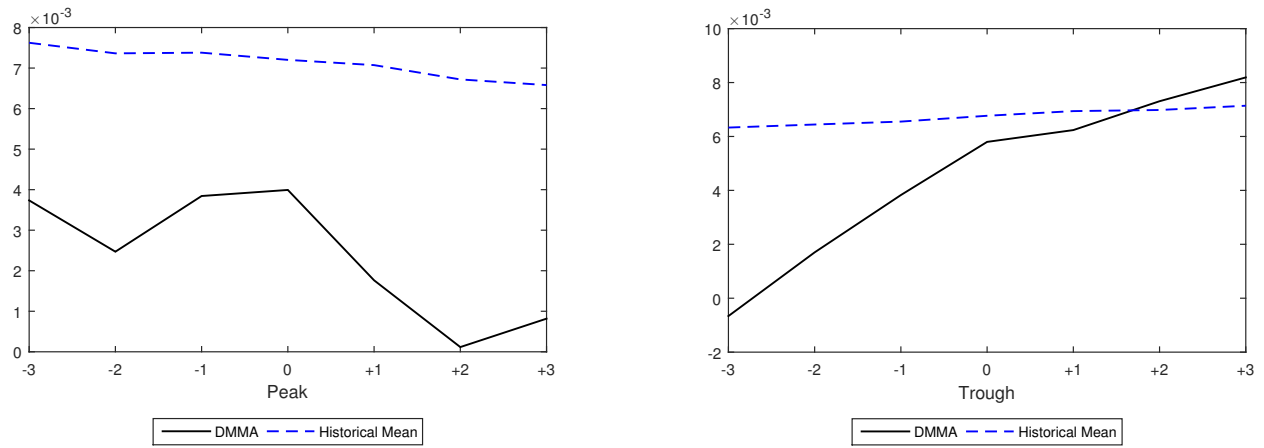
Comparing with other models, DMMA has superior predictive power especially during recessions. Whereas, constant coefficients with equal weights (CC-EW) perform well during expansions, with its out-of-sample R^2 and *CER* gains both slightly larger than DMMA's. We cannot find predictability for DMA and BMA models during expansions, confirming the conclusion in Section 4.4.2 that data is mainly in favor of models with equal weights and constant coefficients. However, the biggest difference between DMMA and CC-EW is that DMMA detects dynamics in the stock market by embedding the exact degree of time-variation in coefficients and in forecasting models at each point in time. In contrast, CC-EW is static and over-optimistic, thus will be unable to adapt to changes during recessions. This leads to DMMA's improvements upon CC-EW especially during downturns.

Next, we closely look at the equity premium predictions and portfolio weights of risky asset around turning points of the business cycle. Figure 2 Panel A shows the predicted equity premiums around peaks and troughs. Predictions from DMMA fit the theoretical pattern acknowledged by Cochrane (1999, 2005): the predicted equity premium increases at the end of recession, signalling greater risk-aversion during recessions. In addition, local minima seems to be around the peak. In contrast, investors who believe in the HM model are over optimistic and predictions from HM are too smooth to capture the fluctuations around business-cycle turning points.

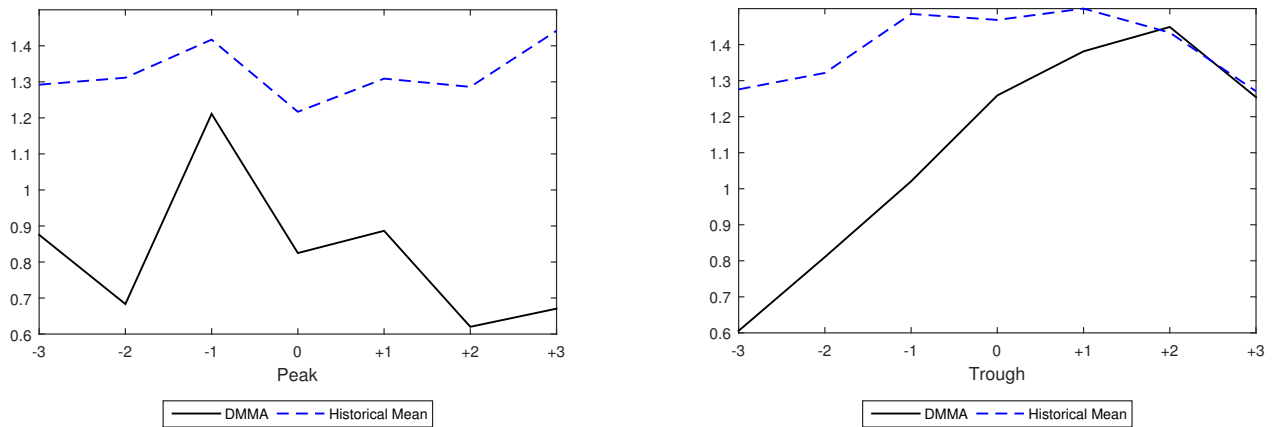
We also uncover that a mean-variance optimizer who relies on DMMA appears to perfectly time the stock market and seize investment opportunities well. In essence, investors employing

Figure 2: Equity Premium Forecasts and Portfolio Weights Around Peaks and Troughs

Panel A: Equity Premium Forecasts



Panel B: Portfolio Weights



Notes: Panel A demonstrate the predicted equity premium using DMMA and historical mean. Panel B presents the portfolio weights of the risky asset for a mean-variance investor, using predictions from DMMA and historical mean. As mentioned in Section 4.1.2, we limit the percentage invested in equities to be between 0% and 150%.

DMMA pull out of the stock market rapidly when the recession starts, and gradually increase equity holdings towards the end of recession. Whereas, HM gives investors false signals, making them fail to withdraw money from the equity market at the beginning of a recession.

We conclude that the predictions from the historical mean cannot capture the abrupt changes in the stock market and are less economically meaningful. The agreement between DMMA's predictions and asset price theory suggested by Cochrane (1999, 2005) provide more economic insights of equity premium predictability.

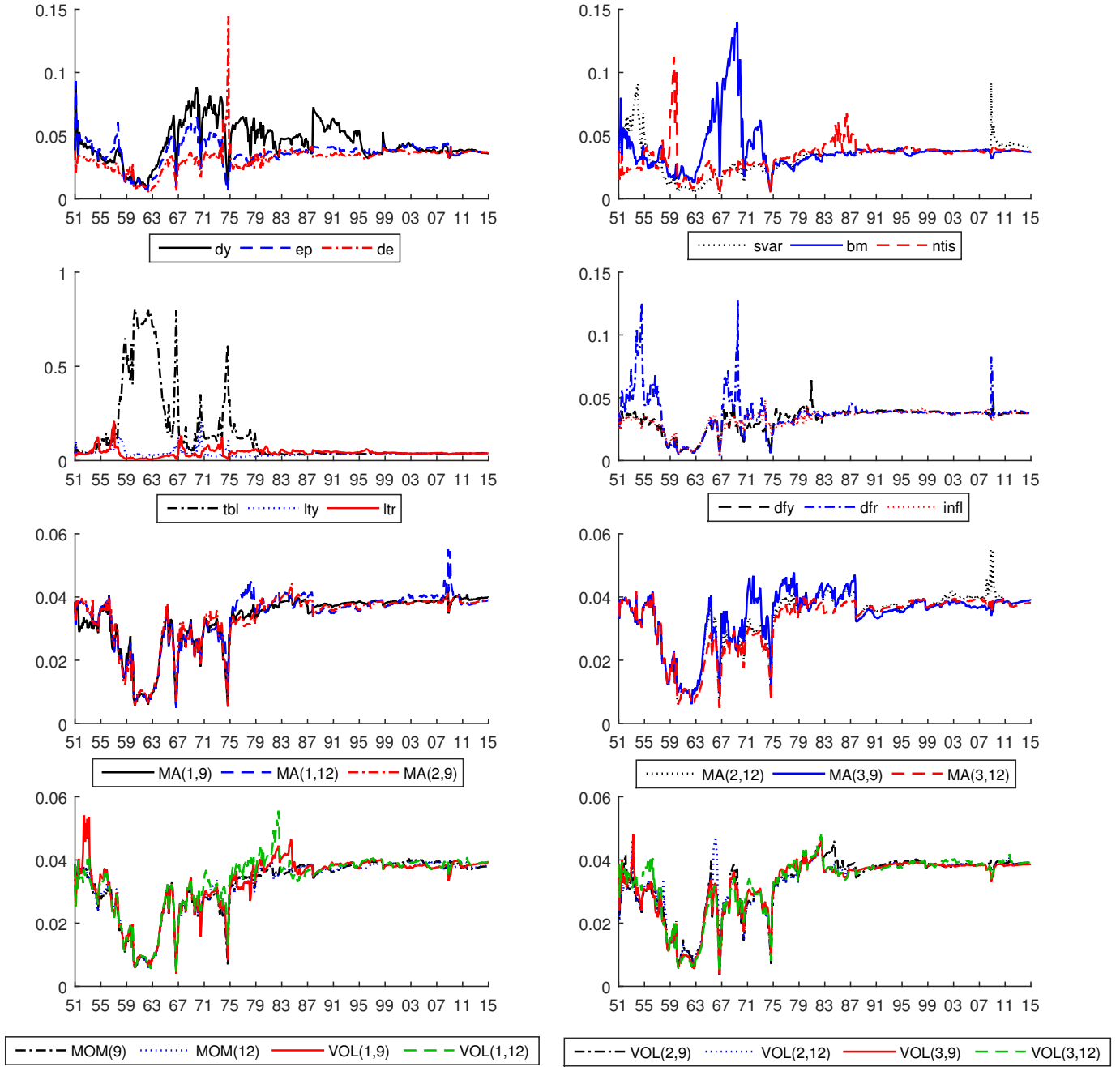
4.4 Model characteristics

4.4.1 Which predictor is important?

Given that we have 12 macroeconomic predictors and 14 technical predictors for excess stock return, it would be interesting to see which one is the most important and how a predictor evolves over time. We measure this by presenting the posterior inclusion probability for each predictor at each time, which can be obtained using DMMA. Following the econometric framework mentioned above, we know that our predictive models are constructed in a way that only a single predictor is included in each model, thus, the posterior inclusion probability for each predictor can be treated as the posterior model probabilities. Therefore, if the posterior model probability for a model or the posterior inclusion probability for a variable is high, that model is likely to be the true model and that variable may play an important role in predicting excess stock returns.

Figure 3 presents the time-varying posterior probabilities for the 26 predictors. From the initial data points until 1975, the shifts between different predictors are occasional and mostly between macroeconomic predictors: treasury bill rate (tbl), default yield spread (dfr), book-to-market ratio (bm), dividend yield (dy) and net equity expansion (ntis). Especially, treasury bill rate (tbl) is highly informative for stock prediction during 1957 to 1967 and 1973 to 1975. After 1975, however, technical indicators become essential and have the similar predictive power as macroeconomic indicators, with inclusion probabilities for different predictors all around 0.04, only with several spikes around the financial crisis in 2008 (e.g., see stock variance (svar), default yield spread (dfr), MA(1,12) and MA(2,12) for examples). This hints at the

Figure 3: Time-Varying Inclusion Probabilities for Different Predictors



Notes: The description of the predictors is as follows: dp is dividend-price ratio; dy is dividend yield; ep is earnings-price ratio; de is dividend-payout ratio; svar is stock variance; bm is book-to-market- ratio; ntis is net equity expansion; tbl is treasury bill rate; lty is long-term yield; ltr is long-term return; tms is term spread; dfy is default yield spread; dfr is default return spread; infl is inflation. $MA(s,l)$, $MOM(l)$, $VOL(s,l)$ are technical indicators, based on moving average, momentum and volume strategy respectively ($s=1,2,3$ and $l=9,12$).

view that DMMA attaches approximately equal weights to each predictor from 1975 to the end of the sample period, and is consistent with the finding acknowledged by Neely et al. (2014) that macroeconomic predictors and technical predictors have complementary information. Moreover, we find that none of the predictor’s posterior probabilities consistently exceeds the prior of $1/26$ over time. This confirms the result in Section 4.2 that there is nonnegligible uncertainty about the best predictor. Under this condition, DMMA automatically detects the best predictor while attaching low posterior weight to the ones that perform poorly over time.

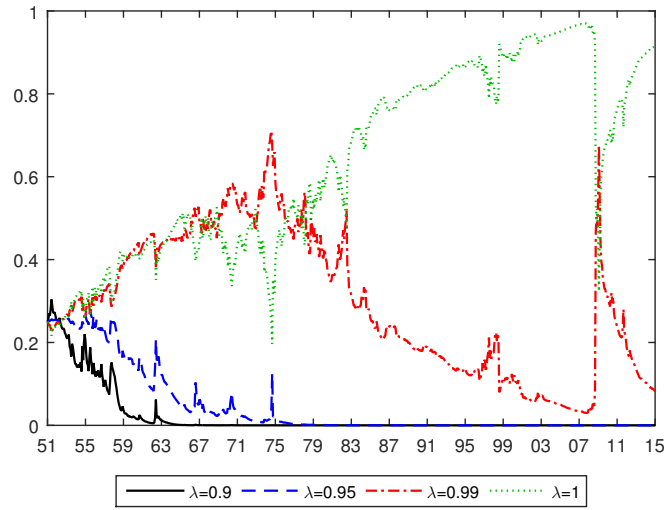
4.4.2 Analysis of different degrees of time-variation in coefficients and in forecasting models

The preceding results show that using DMMA, we can adapt the pattern in data by embedding the exact level of time-variation in coefficients and forecasting models. Next we demonstrate the empirical evidence for that.

We closely look at posterior probabilities for possible degrees of time-variation in coefficients for DMMA in Figure 4. In general, models with constant coefficients and gradually changing coefficients are informative about the movements of equity premium. In contrast, models with sudden changes in coefficients lose data support at the beginning of the sample. Moreover, observations around the oil shock in 1975 and the financial crisis in 2008 enhance the occasional evidence in favor of time-varying coefficients.

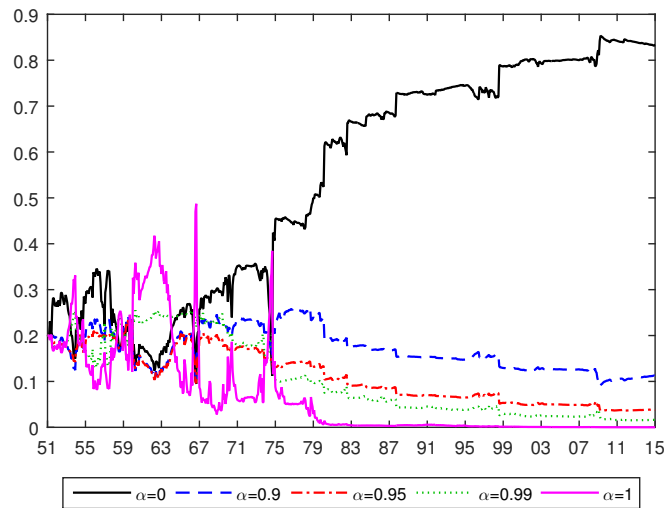
Figure 5 presents the posterior probabilities of different degrees of time-variation in forecasting models. Equal weighted ($\alpha = 0$) models dominate other situations such as BMA ($\alpha = 1$), abruptly changing predictive density combination and a gradually changing predictive density combination after a period of adjustment until 1975. Whereas, BMA is favored by the data at the beginning of out-of-sample period, with its inclusion probability larger than prior 0.2 during 1960 to 1965 and spikes in 1967 and 1975. The dominance of a certain value of α for a prolonged period is reflected in the negligible uncertainty with respect to the degree of time-variation in forecasting models. Moreover, high inclusion probability for equal weighted models from 1975 is in line with the finding in Section 4.4.1: DMMA attaches similar weights to different predictors.

Figure 4: Posterior Probabilities of Degrees of Time-variation in Coefficients (λ)



Notes: Posterior probabilities of models with a specific degree of time-variation of coefficients (λ) for DMMA. Particularly, $\lambda \in [0.90, 0.95, 0.99, 1]$.

Figure 5: Posterior Probabilities of Degrees of Time-variation in Forecasting Models (α)



Notes: Posterior probabilities of models with a specific degree of time-variation of forecasting models (α) for DMMA. Particularly, $\alpha \in [0, 0.90, 0.95, 0.99, 1]$.

5 Conclusion

The literature on stock return forecasting suggests that the out-of-sample predictability is erratic (Cooper and Gulen, 2006; Andrew and Geert, 2007; Campbell and Thompson, 2008; Welch and Goyal, 2008; Joscha and Schüssler, 2014; Turner, 2015). Even though occasionally predictive power is found, it seems to be specific to some predictors in some sample periods, signalling the presence of model instability and uncertainty. Although there have been several attempts to take account of them, there is not a consensus on the exact degree of time-variation in coefficients and the method to combine all the individual models using different predictors. In this paper, we solve these problems by constructing Dynamic Mixture Model Averaging (DMMA), which incorporates possible degrees of time-variation in coefficients and in forecasting models, to detect locally appropriate models. Especially, instead of imposing ex-ante that coefficients and forecasting models vary in the same fashion over time, we encompass moderate to abrupt changes and even no-change in coefficients and forecasting models.

What we uncover is that DMMA model generate more accurate forecasts compared to the historical mean (HM) benchmark across different sample periods. These statistical gains also lead to superior economic profits for a mean-variance investor. Most importantly, in terms of point accuracy, DMMA dominates its nested model combination method including Bayesian Model Averaging (BMA), Dynamic Model Averaging (DMA) and equal weighted models. This implies the importance of accommodating different degrees of time-variation in coefficients and multiple degrees of time-varying forecasting model adaptability.

We further pin down the origins of forecasting improvements by tracking different sources of uncertainty in the predictive regressions. Besides the observational variance, uncertainty regarding the errors from estimating the coefficients and model uncertainty with respect to predictor selection are the key factors hindering forecast accuracy. In contrast, uncertainty about the degree of time-variation in coefficients is small and uncertainty regarding the degree of time-variation in forecasting models is only notable at the initial data points. Essentially, DMMA successfully reduces uncertainty regarding estimation error compared to other predictive models. These all hint at the view that DMMA successfully adapts the pattern in

the unstable stock market by embedding the precise level of time-variation in coefficients and forecasting models, resulting in the mitigation of estimation risk and leading to higher forecast accuracy.

Finally, we find DMMA's predictability is closely linked to the business cycle. Particularly, DMMA's superior performance is mainly driven by recessions, with better forecast results during recessions than during expansions. Interestingly, DMMA also outperforms HM during expansions, indicating DMMA's predictive power is robust to different periods of the business cycle. Consistent with the asset pricing theory acknowledged by Cochrane (1999), our methodology forecasts an increasing equity premium at the end of the recession and the investor who follows DMMA can better time the stock market. This provides more insights about DMMA's superior predictive power.

Overall, DMMA is a powerful approach to predict stock returns. Our findings not only shed light on the roles of different degrees of time-variation in coefficients and in forecasting models, but also have essential implications for monitoring the ups and downs of the business cycle.

References

- Andrew, Ang and Bekaert Geert (2007), “Stock return predictability: Is it there?” *Review of Financial Studies*, 20, 651–707.
- Avramov, Doron (2002), “Stock return predictability and model uncertainty.” *Journal of Financial Economics*, 64, 423–458.
- Balduzzi, Pierluigi and Anthony W. Lynch (1999), “Transaction costs and predictability: Some utility cost calculations.” *Journal of Financial Economics*, 52, 47–78.
- Billio, Monica, Roberto Casarin, Francesco Ravazzolo, and Herman K. Van Dijk (2013), “Time-varying combinations of predictive densities using nonlinear filtering.” *Journal of Econometrics*, 177, 213–232.
- Bossaerts, Peter and Pierre Hillion (1999), “Implementing statistical criteria to select return forecasting models: what do we learn?” *Review of Financial Studies*, 12, 405–428.
- Boudoukh, Jacob, Roni Michaely, Matthew Richardson, and Michael R. Roberts (2007), “On the importance of measuring payout yield: Implications for empirical asset pricing.” *The Journal of Finance*, 62, 877–915.
- Campbell, John Y. and John H. Cochrane (1995), “By force of habit: A consumption-based explanation of aggregate stock market behavior.” Technical report, National Bureau of Economic Research.
- Campbell, John Y. and Samuel B. Thompson (2008), “Predicting excess stock returns out of sample: Can anything beat the historical average?” *Review of Financial Studies*, 21, 1509–1531.
- Cenesizoglu, Tolga and Allan Timmermann (2012), “Do return prediction models add economic value?” *Journal of Banking & Finance*, 36, 2974–2987.
- Clark, Todd E. and Kenneth D. West (2007), “Approximately normal tests for equal predictive accuracy in nested models.” *Journal of Econometrics*, 138, 291–311.
- Cochrane, John (2005), “Financial markets and the real economy.” Technical report, National Bureau of Economic Research.
- Cochrane, John H. (1999), “New facts in finance.” Technical report, National Bureau of Economic Research.
- Cooper, Michael and Huseyin Gulen (2006), “Is time-series-based predictability evident in real time?” *The Journal of Business*, 79, 1263–1292.
- Cremers, KJ. Martijn (2002), “Stock return predictability: A bayesian model selection perspective.” *Review of Financial Studies*, 15, 1223–1249.
- Dangl, Thomas and Michael Halling (2012), “Predictive regressions with time-varying coefficients.” *Journal of Financial Economics*, 106, 157–181.
- Fama, Eugene F. and Kenneth R. French (1989), “Business conditions and expected returns on stocks and bonds.” *Journal of Financial Economics*, 25, 23–49.
- Geweke, John and Gianni Amisano (2010), “Comparing and evaluating bayesian predictive distributions of asset returns.” *International Journal of Forecasting*, 26, 216–230.
- Geweke, John and Gianni Amisano (2011), “Hierarchical markov normal mixture models with applications to financial asset returns.” *Journal of Applied Econometrics*, 26, 1–29.
- Geweke, John and Gianni Amisano (2012), “Prediction with misspecified models.” *The American Economic Review*, 102, 482–486.
- Goyal, Amit and Ivo Welch (2003), “Predicting the equity premium with dividend ratios.” *Management Science*, 49, 639–654.
- Harrison, Jeff and Mike West (1999), *Bayesian Forecasting & Dynamic Models*. Springer.
- Henkel, Sam James, J. Spencer Martin, and Federico Nardari (2011), “Time-varying short-

- horizon predictability.” *Journal of Financial Economics*, 99, 560–580.
- Huang, Dashan, Fuwei Jiang, Jun Tu, and Guofu Zhou (2015), “Investor sentiment aligned: a powerful predictor of stock returns.” *Review of Financial Studies*, 28, 791–837.
- Huang, Huiyu and Tae-Hwy Lee (2010), “To combine forecasts or to combine information?” *Econometric Reviews*, 29, 534–570.
- Johannes, Michael, Arthur Korteweg, and Nicholas Polson (2014), “Sequential learning, predictability, and optimal portfolio returns.” *The Journal of Finance*, 69, 611–644.
- Joscha, Beckmann and Rainer Schüssler (2014), “Forecasting equity premia using bayesian dynamic model averaging.” Technical report, Center for Quantitative Economics (CQE), University of Muenster.
- Koop, Gary and Dimitris Korobilis (2012), “Forecasting inflation using dynamic model averaging.” *International Economic Review*, 53, 867–886.
- Neely, Christopher J., David E. Rapach, Jun Tu, and Guofu Zhou (2014), “Forecasting the equity risk premium: the role of technical indicators.” *Management Science*, 60, 1772–1791.
- Raftery, Adrian E., Miroslav Kárný, and Pavel Ettler (2010), “Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill.” *Technometrics*, 52, 52–66.
- Rapach, David E., Matthew C. Ringgenberg, and Guofu Zhou (2016), “Short interest and aggregate stock returns.” *Journal of Financial Economics*, 121, 46–65.
- Rapach, David E., Jack K. Strauss, and Guofu Zhou (2010), “Out-of-sample equity premium prediction: Combination forecasts and links to the real economy.” *Review of Financial Studies*, 23, 821–862.
- Stambaugh, Robert F. (1999), “Predictive regressions.” *Journal of Financial Economics*, 54, 375–421.
- Turner, James A. (2015), “Casting doubt on the predictability of stock returns in real time: Bayesian model averaging using realistic priors.” *Review of Finance*, 19, 785–821.
- Wachter, Jessica A. and Missaka Warusawitharana (2009), “Predictable returns and asset allocation: Should a skeptical investor time the market?” *Journal of Econometrics*, 148, 162–178.
- Wachter, Jessica A. and Missaka Warusawitharana (2015), “What is the chance that the equity premium varies over time? evidence from regressions on the dividend-price ratio.” *Journal of Econometrics*, 186, 74–93.
- Welch, Ivo and Amit Goyal (2008), “A comprehensive look at the empirical performance of equity premium prediction.” *Review of Financial Studies*, 21, 1455–1508.

Appendices

A Further Details On Dynamic Linear Models

Given the state space model in equation (1) and (2) in Section 2.1, we update the belief about the coefficients and observational variance recursively. Assume coefficient θ_0 follows a normally distributed prior, Kalman filter process is updated by:

$$\theta_t|D_t \sim N(\hat{\theta}_t, \Sigma_{t|t}) \quad (\text{A.1})$$

where ,

$$\hat{\theta}_t = \hat{\theta}_{t-1} + \Sigma_{t|t-1} X'_{t-1} (H_t + X_{t-1} \Sigma_{t|t-1} X'_{t-1})^{-1} (r_t - X_{t-1} \hat{\theta}_{t-1}) \quad (\text{A.2})$$

$$\Sigma_{t|t} = \Sigma_{t|t-1} - \Sigma_{t|t-1} X'_{t-1} (H_t + X_{t-1} \Sigma_{t|t-1} X'_{t-1})^{-1} X_{t-1} \Sigma_{t|t-1} \quad (\text{A.3})$$

Therefore, the predictive distribution is

$$r_t|D_{t-1} \sim N(X_t \theta_{t|t-1}, H_t + X_t \Sigma_{t|t-1} X'_t) \quad (\text{A.4})$$

Note that all the derivations are conditional on H_t , the observational variance. Evidence for time-varying volatility is strong as it generates fat-tailed return distribution for stock market (Johannes et al., 2014). Moreover, numerous studies find the important role time-varying volatility plays in predicting excess stock return (Johannes et al., 2014; Joscha and Schüssler, 2014). As a consequence, following Koop and Korobilis (2012), we use exponentially weighted Moving Average (EWMA) to estimate H_t , which updates at each time and can be approximated by a recursive form:

$$\hat{H}_{t+1|t} = \kappa \hat{H}_{t|t-1} + (1 - \kappa) (r_t - X_{t-1} \hat{\theta}_t)^2 \quad (\text{A.5})$$

where κ is referred to as the delay factor. We set $\kappa = 0.95$ for monthly data. This fits monthly data's property: a relatively rapid delay.

B Further Details On Dynamic Mixture Model Averaging

Following Raftery et al. (2010) and Koop and Korobilis (2012), prediction equation for different choices of predictors k_i conditional on the degree of time-variation in coefficients (λ_j) and forecasting models (α_z) at time t is:

$$P(L_t = k_i | \lambda_j, \alpha_z, D_{t-1}) = \frac{P(L_{t-1} = k_i | \lambda_j, \alpha_z, D_{t-1})^{\alpha_z}}{\sum_k P(L_{t-1} = k_i | \lambda_j, \alpha_z, D_{t-1})^{\alpha_z}} \quad (\text{B.1})$$

where L_t indicates the certain predictor selected at time t and α_z is the other forgetting factor.

Based on Bayes rule, the model updating equation is given by:

$$P(L_t = k_i | \lambda_j, \alpha_z, D_t) = \frac{P(r_t | L_t = k_i, \lambda_j, \alpha_z, D_{t-1})P(L_t = k_i | \lambda_j, \alpha_z, D_{t-1})}{P(r_t | \lambda_j, \alpha_z, D_{t-1})} \quad (\text{B.2})$$

where

$$P(r_t | \lambda_j, \alpha_z, D_{t-1}) = \sum_k P(r_t | L_t = k_i, \lambda_j, \alpha_z, D_{t-1})P(L_t = k_i | \lambda_j, \alpha_z, D_{t-1}) \quad (\text{B.3})$$

and the conditional density given by equation (A.4) in Appendix A is

$$P(r_t | L_t = k_i, \lambda_j, \alpha_z, D_{t-1}) \sim N(\hat{r}_{t,i}^{j,z}, H_t + X_{t-1}\Sigma_{t|t-1}X'_{t-1}) \quad (\text{B.4})$$

The one-step ahead prediction of excess stock returns conditional on these parameters is:

$$\hat{r}_{t,i}^{j,z} = E(r_t | k_i, \lambda_j, \alpha_z, D_{t-1}) = \theta_t x_{t-1} | k_i, \lambda_j, \alpha_z, D_{t-1} \quad (\text{B.5})$$

Given α_z and λ_j , the return prediction over all the choices of different variables at time t is :

$$\hat{r}_t^{j,z} = \sum_k P(L_t = k_i | \lambda_j, \alpha_z, D_{t-1})\hat{r}_{t,i}^{j,z} \quad (\text{B.6})$$

Next, we integrate over the degrees of time-variation in λ_j and α_z . Specially, we calculate the posterior probability for λ_j given a specific choice of α_z :

$$P(\lambda_j | \alpha_z, D_t) = \frac{P(r_t | \lambda_j, \alpha_z, D_{t-1})P(\lambda_j | \alpha_z, D_{t-1})}{P(r_t | \alpha_z, D_{t-1})} \quad (\text{B.7})$$

where

$$P(r_t | \alpha_z, D_{t-1}) = \sum_\lambda P(r_t | \lambda_j, \alpha_z, D_{t-1})P(\lambda_j | \alpha_z, D_{t-1}) \quad (\text{B.8})$$

Then given a certain value of α_z , prediction is:

$$\hat{r}_t^z = \sum_\lambda P(\lambda_j | \alpha_z, D_{t-1})\hat{r}_t^{j,z} \quad (\text{B.9})$$

Finally, the posterior probability for a certain α_z is obtained:

$$P(\alpha_z | D_t) = \frac{P(r_t | \alpha_z, D_{t-1})P(\alpha_z | D_{t-1})}{P(r_t | D_{t-1})} \quad (\text{B.10})$$

where

$$P(r_t | D_{t-1}) = \sum_\alpha P(r_t | \alpha_z, D_{t-1})P(\alpha_z | D_{t-1}) \quad (\text{B.11})$$

The unconditional prediction of all the model specifications is :

$$\hat{r}_t = \sum_{\alpha} P(\alpha_z | D_{t-1}) \hat{r}_t^z \quad (\text{B.12})$$

According to Bayes rule, the total posterior of a specific model (i.e., choice of predictive variables k_i , choice of λ_j and choice of α_z) is given by:

$$\begin{aligned} P(k_i, \lambda_j, \alpha_z | D_t) &= P(k_i, \lambda_j | \alpha_z, D_t) P(\alpha_z | D_t) \\ &= P(k_i | \lambda_j, \alpha_z, D_t) P(\lambda_j | \alpha_z, D_t) P(\alpha_z | D_t) \end{aligned} \quad (\text{B.13})$$

C Technical Predictors

We form 14 technical predictors based on three technical strategies following Neely et al. (2014).

The first strategy is to compare two moving-averages (*MA*):

$$S_{i,t} = \begin{cases} 1 & \text{if } MA_{s,t} \geq MA_{l,t} \\ 0 & \text{if } MA_{s,t} < MA_{l,t} \end{cases}, \quad (\text{C.1})$$

where

$$MA_{j,t} = (1/j) \sum_{i=0}^{j-1} P_{t-i}, \quad j = s, l, \quad s = 1, 2, 3 \quad l = 9, 12 \quad (\text{C.2})$$

We obtain a buy signal when $S_{i,t} = 1$ or a sell signal when $S_{i,t} = 0$. A *MA* indicator with s and l lags can be presented as $MA(s, l)$.

The second strategy is based on momentum (*MOM*)

$$S_{i,t} = \begin{cases} 1 & \text{if } P_t \geq P_{t-m} \\ 0 & \text{if } P_t < P_{t-m} \end{cases}, \quad m = 9, 12 \quad (\text{C.3})$$

When the current stock price is higher than that m period ago, it generates a positive momentum, therefore, a buy signal. The indicator is $MOM(m)$.

The last strategy is based on volume (*VOL*). Define

$$OBV_t = \sum_{k=1}^t VOL_k D_k, \quad D_k = \begin{cases} 1 & \text{if } P_k \geq P_{k-1} \\ -1 & \text{if } P_k < P_{k-1} \end{cases} \quad (\text{C.4})$$

where VOL_k is the trading volume of stocks during period k . We then construct a trading

signal based on OBV_t :

$$S_{i,t} = \begin{cases} 1 & \text{if } MA_{s,t}^{OBV} \geq MA_{l,t}^{OBV} \\ 0 & \text{if } MA_{s,t}^{OBV} < MA_{l,t}^{OBV} \end{cases}, \quad (\text{C.5})$$

where

$$MA_{j,t}^{OBV} = (1/j) \sum_{i=0}^{j-1} OBV_{t-i}, \quad j = s, l, \quad s = 1, 2, 3 \quad l = 9, 12 \quad (\text{C.6})$$

If volume and prices are both high recently, this imply a positive trend, and thus, generate a buy signal. Denote the indicator as $VOL(s, l)$.

D Absolute Values of Different Sources of Prediction Variance

Table D.1: Absolute Values of Different Sources of Prediction Variance

	DMMA	BMA	DMA
Obs.var	1.3498	1.4162	1.4938
Unc.coef	0.0193	0.0260	0.6317
Mod.unc.Var	0.0221	0.0203	0.1299
Mod.unc.Coef	0.0037	0.0045	-
Mod.unc.Mod	0.0043	-	-
Total	1.3992	1.4671	2.2554

Notes: See notes for Figure 1. (-) implies no corresponding values.

E Results for DMMS

Table E.1: Evaluation for DMMS

	Period:1960+		Period:1976+		Period:1988+	
	Statistical Evaluation					
	R_{OS}^2	$Log(PL)$	R_{OS}^2	$Log(PL)$	R_{OS}^2	$Log(PL)$
DMMS	-0.79	1141.20	-4.74	796.58	-5.22	568.71
	Economic Evaluation					
	CER	SR	CER	SR	CER	SR
DMMS	5.11	0.06	4.91	0.05	0.22	0.01

Notes: Statistical predictability for DMMS. See details in Table 3.

DMMS, different from DMMA, selects the individual model with the highest posterior probability at each time, among choices of predictors, degrees of time-variation in coefficients and forecasting models. Table E.1 presents statistical and economical results for DMMS. Compared with the results in Table 3 and 4, we find that although DMMS occasionally

outperforms HM in terms of log likelihoods and certainty equivalent return, it is worse than DMMA. We conclude that this is because DMMS switches more rapidly than DMMA and cannot make use of all the information data provides.

F Empirical Robustness Checks

In this section, we relax the assumption that the time-varying coefficients in the state space model follows a random walk process (see equation (2) in Section 2.1). As asset pricing theory suggests that expected returns are nonstationary, we check whether the results are sensitive to random walk assumption and address the stationary issue using autoregressive process for the transition equation. Specially, following Dangl and Halling (2012), we rewrite our transition equation by introducing autoregression in the following form:

$$\theta_t = G\mathbb{I}\theta_{t-1} + u_t \quad (\text{F.1})$$

with \mathbb{I} denoting an identity matrix and $0 < G \leq 1$ a scalar. Hence, if $G = 1$, equation (F.1) is the same as equation (2). We also consider several alternatives of G , including $G = 0.95, 0.90, 0.80$ to generate nonstationary process. Our goal is to compare the results of random walk coefficients and any other parameters choices of G .

From Table F.1, in the race between random walk coefficients and autoregressive coefficients, model with random walk coefficients works better than any parameter choice of G less than one. DMMA model with random walk coefficients consistently has the highest out-of-sample R_{OS}^2 and predictive log likelihoods over different sample periods. Regarding economic evaluation, in Table F.2, model with random walk coefficients is the only one that could consistently outperform HM according to CER and SR . Furthermore, except for the CER in the period of 1960+, DMMA model with random walk coefficients dominates other autoregression coefficients. All these demonstrate the advantage of applying random walk coefficients.

Table F.1: Statistical Evaluation- Robustness Check

	Period:1960+		Period:1976+		Period:1988+	
	R_{OS}^2	$Log(PL)$	R_{OS}^2	$Log(PL)$	R_{OS}^2	$Log(PL)$
$G = 1.00$	1.72**	1141.70	0.91*	802.66	0.88*	573.13
$G = 0.95$	-0.63	1140.50	-0.57	802.00	-0.67	571.56
$G = 0.90$	-0.76	1139.80	-0.39	801.88	-0.35	571.77
$G = 0.80$	-0.37	1141.40	-0.24	801.85	-0.20	571.60
HM	0	128.56	0	293.27	0	304.43

Notes: Statistical predictability for different models using various autoregression coefficient (G). See details in Table 3.

Table F.2: Economic Evaluation- Robustness Check

	Period:1960+		Period:1976+		Period:1988+	
	<i>CER</i>	<i>SR</i>	<i>CER</i>	<i>SR</i>	<i>CER</i>	<i>SR</i>
$G = 1.00$	5.15	0.08	6.24	0.10	7.41	0.15
$G = 0.95$	5.34	0.06	3.17	0.02	6.37	0.13
$G = 0.90$	5.18	0.06	2.94	0.02	6.50	0.13
$G = 0.80$	5.62	0.07	1.78	0.00	5.37	0.11
HM	3.39	0.07	3.93	0.08	5.25	0.12

Notes : Economic predictability for different predictive models various autoregression coefficient (G). See details in Table 4.