International Stock Return Predictability under Model Uncertainty*

Andreas Schrimpf

Centre for European Economic Research (ZEW), Mannheim, Germany

January 15, 2008

Abstract

This paper examines return predictability when the investor is uncertain about the right state variables. A new feature of the "frequentist" model averaging approach used in this paper is to account for finite-sample bias of the coefficients in the predictive regressions. The effects of model uncertainty in predictive regressions are studied in an international data set with five countries and nine financial and macroeconomic state variables. Our results indicate that interest-rate related variables are usually among the most prominent predictive variables, whereas valuation ratios perform rather poorly in our setup. There is also some evidence that risk premia vary with the output gap. Yet, predictability of market excess returns clearly weakens, once model uncertainty is accounted for. We find (weak) evidence for out-of-sample predictability by model averaging methods in France and the United States but not the remaining stock markets.

JEL Classification: G11, G12

Keywords: Stock return predictability, Empirical asset pricing, Bayesian model averaging, Model uncertainty, International Stock Markets

---

*Correspondence: Andreas Schrimpf, Centre for European Economic Research (ZEW) Mannheim, P.O. Box 10 34 43, 68034 Mannheim, Germany, email: schrimpf@zew.de, phone: +49 621 1235160, fax: +49 621 1235223. I am grateful to Todd E. Clark, Joachim Grammig, Francois Laisney, Maik Schmeling, Peter Schmidt, Qingwei Wang and seminar participants at University of Tübingen, ZEW Mannheim and the German Finance Association PhD seminar (2007, Dresden) for helpful comments and suggestions. I am also grateful to Amit Goyal and Ivo Welch for providing data on their webpages. Nataliya Matosova and Jörg Breddermann provided able research assistance.

1 Introduction

Empirical studies have found a plethora of variables to be informative about future excess returns in regressions of the form:

$$r_t = \alpha + \beta' x_{t-1} + u_t, \tag{1}$$

where r_t denotes the return of the aggregate stock market portfolio in excess of the risk-free rate and x_{t-1} is a vector of predictive variables, such as the dividend yield, a term spread or certain macroeconomic variables.[2] Statistically significant β coefficients in (1) are interpreted as evidence for predictability and as evidence that risk premia are time-varying.[3]

Given the large number of variables proposed in the literature, a typical investor is confronted by a high degree of uncertainty on what the "right" state variables are. Moreover, the fact that so many variables have found to be valuable predictors of returns naturally raises the concern that the apparent predictability in the literature may well arise due to data-snooping rather than genuine variation of economic risk premia.[4] The aim of this paper is to explore the robustness of several predictive variables in international stock markets in the context of model uncertainty. We follow the spirit of the seminal work by Cremers (2002) and Avramov (2002) and use Bayesian model averaging in order to account for model uncertainty. Unlike the classical framework, the Bayesian approach does not assume the existence of a "true" model. By contrast, a-posteriori model probabilities can be derived for the different candidate models, which are then used to weight the coefficients accordingly in a composite model. In this way, model uncertainty can be accounted for in a coherent way.

[2]See e.g. Fama and French (1988), Fama and French (1989), Campbell and Shiller (1988a), Campbell and Shiller (1988b), Lettau and Ludvigson (2001) etc.

[3]Based on the evidence provided by the articles mentioned in the text documenting stock return predictability, by the late 1990s the consensus among financial economists considered expected excess returns to be time-varying. In particular, predictability of market excess returns has been labeled as one of the "new facts in finance" Cochrane (1999).

[4]See e.g. Bossaerts and Hillion (1999), Ferson et al. (2003) for critical views. Most notably, after a comprehensive out-of-sample forecast evaluation, Goyal and Welch (2006) come to the conclusion that knowledge of different state variables is of little use for a real-time investor. They interpret their findings as a strong case against stock return predictability.

A new feature of our approach is to account for finite-sample bias of the coefficients in the predictive regressions in a "frequentist" model averaging framework. A pure Bayesian model averaging framework as in Cremers (2002) and Avramov (2002) requires prior elicitation for the relevant parameters conditional on the different models. This can be a problematic task when the set of models becomes very large.[5] Therefore, in order to reduce the impact of subjective prior information, we base our empirical analysis on Bayesian averaging of classical estimates (BACE) as in Sala-i-Martin et al. (2004). BACE can be seen as a limiting case of the Bayesian approach as the prior information becomes dominated by the data (See Leamer 1978). Another less-attractive feature of the pure Bayesian model averaging approach as used by Cremers (2002) and Avramov (2002) is that it treats the predictive variables as exogenous, an assumption which is clearly violated in the context of predictive regressions. How to conduct reliable inference in predictive regressions taking the time-series properties of the predictive variables (such as the dividend yield) into account has been the subject of a great amount of recent research (See for instance Stambaugh 1999, Campbell and Yogo 2003, Lewellen 2004, Amihud and Hurvich 2004, Torous et al. 2004, and Moon et al. 2006). In order to account for problems due to the persistence of the predictive variables, we estimate the models by classical OLS, where the coefficients are adjusted for finite-sample bias using the approach put forth in Amihud and Hurvich (2004). The bias-corrected coefficients in the particular models are then weighted by their posterior model probabilities which are derived according to the BACE approach of Sala-i-Martin et al. (2004).

This paper also contributes to the literature by conducting a comprehensive analysis of stock return predictability in international stock markets. It is fair to say that the profession's view on stock return predictability has been shaped for the most part by empirical studies on the US stock market. However, looking in greater depth at other important capital markets may provide important additional insights, especially in a controversial field such as return predictability. Moreover, investigation of international markets also provides a way of guarding against data-snooping concerns. We thus examine the predictive performance of nine variables in a total of five international stock markets (France, Germany, Japan, United

[5] Avramov (2002) responds to this problem using an empirical Bayes approach which uses sample data for prior elicitation. In the Bayesian tradition, Cremers (2002) specifies subjective prior distributions based on different skeptical or optimistic beliefs about predictability.

Kingdom, United States). Other important recent papers which provide evidence on international stock markets with up-to-date methods include Hjalmarsson (2004), Rapach et al. (2005), Ang and Bekaert (2005) or Giot and Petitjean (2006).[6] To our knowledge evidence on return predictability in international stock markets under model uncertainty has been lacking so far.

There is a long list of variables which has been proposed in the literature on stock return predictability. In particular, valuation ratios such as dividend yields or earnings yields (e.g. Fama and French 1988, Campbell and Shiller 1988a, Lewellen 2004), interest rate related variables such as short-term interest rates (e.g. Fama and Schwert 1977, Hodrick 1992, Ang and Bekaert 2005) or default and term spreads (e.g. Campbell 1987, Fama and French 1989) have featured prominently in predictive regressions. Lamont (1998) has proposed the dividend-payout ratio as a predictive variable. The predictive power of stock market volatility has been studied by French et al. (1987). Pure macroeconomic variables used in predictive regressions include for instance the inflation rate (e.g. Fama 1981), consumption-wealth ratio (Lettau and Ludvigson 2001), price-GDP ratio (Rangvid 2006), industrial production growth (e.g. Fama 1990 or Avramov 2002), and more recently the output gap (Cooper and Priestley 2006). Variables motivated from a behavioral point of view (such as stock market sentiment as in Brown and Cliff 2005) have also been shown to predict returns.

The reading of the literature in the previous paragraph suggests that there is not much consensus on what the important variables are, or put differently that there is tremendous model uncertainty in predictive regressions. In particular, some variables may appear significant in one specification and be insignificant in others, as researchers may only report their preferred specifications. As time elapses more variables are sure to be added to the list of predictors.

While in-sample predictability is a debated topic, the question whether stock returns may be predictable out-of-sample (OOS) has been even more controversial. Empirical results on OOS predictability are mixed. Recently, several authors – most notably Goyal and Welch (2003; 2006) – argue against stock return predictability or time-varying risk premia based on the

[6]Hjalmarsson considers only four financial variables. Rapach et al. (2005) focus merely on macroeconomic variables and do not consider valuation ratios. Giot and Petitjean (2006) consider finite-sample bias but do not address the issue of model uncertainty. Their set of predictive variables is limited to financial variables.

lacking evidence for out-of-sample predictability.[7] Campbell and Thompson (2006), however, find that once sensible restrictions are imposed on the predictive regression coefficients, the OOS forecast performance can be improved. It has also been argued that averaging forecasts of various models enhances out-of sample forecast performance substantially. Avramov (2002) finds that the out-of-sample performance of the weighted model is superior to the performance of models selected by information criteria and better than a naive benchmark. Another aim of the paper is therefore to look closer at the out-of-sample forecast performance of model averaging, in particular the variation of OOS performance in the spirit of Goyal and Welch (2006).

Our results can be summarized as follows. Several differences with regard to return predictability are found across countries. We find that interest rate related variables are usually among the most prominent predictive variables, whereas valuation ratios perform rather poorly. There is also some evidence that risk premia vary with the output gap. The earnings yield often appears to be a more robust predictor than the dividend yield. Yet, predictability of market excess returns clearly weakens, once model uncertainty is accounted for. We find (weak) evidence for out-of-sample predictability by model averaging methods in France and the United States but not the remaining stock markets.

The remainder of the paper is structured as follows. Section II discusses the econometric framework of predictive regressions and how model uncertainty can be accounted for in a model averaging framework. Section III briefly discusses our data set. Empirical findings are discussed in Section IV. Section V concludes.

2 Methodology

In this paper we assess predictive ability in the conventional framework of predictive regressions. When there are multiple predictive variables (depending on the particular model $\mathcal{M}_j$),

[7] Cochrane (2006) defends predictability based on the argument that even though predictability from the dividend-price ratio may be weak on statistical grounds, the fact that dividend growth is not predictable at all, may be interpreted as evidence that the variation of the dividend-price ratio is informative about future expected returns.

the predictive equation for the future stock return is given by

$$r_t = \alpha + \beta_j' x_{j;t-1} + u_{j,t}, \tag{2}$$

where r_t denotes the (log)-return on the market portfolio in excess of the (log) risk-free rate and $x_{j;t-1}$ is a k_j-dimensional vector of predictive variables, whose dimension and composition depends on the particular model $\mathcal{M}_j$. In total, we have κ different predictive variables which results in 2^κ different subsets, i.e. vectors $x_{j;t-1}$ $(j =, 1, \cdots, 2^\kappa)$. β_j is a k_j-dimensional vector of regression coefficients on the predictive variables. As is common in the extant literature, the vector of predictive variables is assumed to follow a first-order VAR:

$$x_{j;t} = \Theta_j + \Phi_j x_{j;t-1} + \nu_{j;t}. \tag{3}$$

Θ_j is a k_j-dimensional intercept and Φ_j is a $k_j \times k_j$ matrix with all eigenvalues smaller than one in absolute value to ensure stationarity of the process. The errors $(u_{j;t}, \nu_{j;t}')'$ are i.i.d. multivariate normal with mean zero.

2.1 Accounting for Model Uncertainty

We want to put ourselves in the position of an investor who is confronted by the voluminous literature on evidence for stock return predictability, yet is uncertain about which variables are actually important. In such a context, a Bayesian framework is attractive, since model uncertainty can be considered coherently. In a classical framework, however, the search for the "true model" usually implies running a series of model specification tests. Moreover, a classical approach is less appealing, because once a single model is determined, information in the remaining $2^\kappa - 1$ models is neglected. The approach taken in this paper is therefore to combine the Bayesian feature of model averaging with coefficients estimated by classical OLS (adjusted for finite-sample bias) which reduces dependence on prior distributions (See

Sala-i-Martin et al. 2004).[8]

We explore the usefulness of $\kappa = 9$ candidate predictive variables in total, which implies that $2^\kappa = 512$ different model combinations are assessed. In a Bayesian framework, posterior probabilities $p(\mathcal{M}_j|y)$ for each model $j = 1, \ldots, 2^\kappa$ can be derived. These posterior model probabilities are used in the Bayesian model averaging framework as weights of the composite model:

$$E[\beta|y] = \sum_{j=1}^{2^\kappa} p(\mathcal{M}_j|y)\beta_j|y, \tag{4}$$

where $\beta_j|y$ denotes the posterior mean of the predictive coefficients in the jth model. In the same way, the posterior standard deviation in the composite model is obtained from the corresponding diagonal element of the matrix

$$Var(\beta|y) = \sum_{j=1}^{2^\kappa} p(\mathcal{M}_j|y)[Var(\beta_j|y) + (\beta_j - E[\beta|y])(\beta_j - E[\beta|y])']. \tag{5}$$

Note that the posterior variance of the composite model in (5) contains essentially two components: the first term in the brackets accounts for estimation risk, whereas the second measures the variation of the predictive coefficients across the different models and thus accounts for model uncertainty.[9]

For determining the weights, the marginal likelihood for the different models $\mathcal{M}_j$ must be computed.[10] In the pure BMA framework, analytical solutions can be found only for certain prior distribution families.[11] In the "frequentist" model averaging framework of Sala-i-Martin

[8] Bayesian and classical results are numerically identical when diffuse priors are specified.

[9] Following Avramov (2002), we report posterior standard deviations with and without adjustment for model uncertainty in order to demonstrate the effects of accounting for model uncertainty in the inference.

[10] Mathematically, the marginal likelihoods can be obtained by integrating out the parameters from the combination of the likelihood and the prior conditional on the model.

[11] Avramov (2002), for instance, uses an "empirical Bayes" approach for prior elicitation, which uses data-information from the sample in order to determine the prior specification. Yet, such an approach can be criticized for using information of the dependent variable, which violates the rules of probability necessary for conditioning (Fernández et al., 2001).

et al. (2004), however, the marginal likelihood of a particular model is approximated as $exp(-0.5BIC_j)$. The posterior model probability for $\mathcal{M}_j$ can then be derived as

$$p(\mathcal{M}_j|y) = \frac{p(\mathcal{M}_j)exp(-0.5BIC_j)}{\sum_{i=1}^{2^\kappa} p(\mathcal{M}_i)exp(-0.5BIC_i)}. \tag{6}$$

As discussed in Sala-i-Martin et al. (2004) this formula can be derived in a standard g-prior framework when one takes the limit as the information in the data gets large relative to the prior information. Thus, using posterior model probabilities as in Equation (6) essentially implies using a "prior" that becomes dominated by the data.

2.2 Finite-sample Bias in Predictive Regressions

In the following we outline how we correct for finite sample bias in the BACE framework. In order to provide some intuition on the econometric problems arising from predictive variables which are not exogenous but rather predetermined, we first briefly review the single predictor case by Stambaugh (1999).

$$r_t = \alpha + \beta x_{t-1} + \epsilon_t, \tag{7}$$

where r_t denotes the (log)-return on the market portfolio in excess of the (log) risk-free rate and x_{t-1} is a predictive variable such as the dividend yield. The predictive variable itself is modeled as a first-order autoregressive process

$$x_t = \gamma + \rho x_{t-1} + \xi_t. \tag{8}$$

The errors in Equations (7) and (8) are assumed to be i.i.d. jointly normally distributed. Stambaugh (1999) then derives an analytical formula for the finite-sample bias of the predictive coefficient:

$$E(\hat{\beta}-\beta) \approx \gamma E(\hat{\rho}-\rho), \tag{9}$$

where $\gamma = \frac{\sigma_{\epsilon\xi}}{\sigma_{\xi}^2}$ is the ratio of the covariance of the errors in both equations $\sigma_{\epsilon\xi}$ and the variance σ_{ξ}^2 of the error term ξ_t . As Equation (9) shows, the bias of the predictive coefficients arises from the (downward) bias of the autoregressive parameter for the predictive variable $\hat{\rho}$ in combination with the correlation of the errors v_t and the error term u_t in the predictive equation. The latter effect can be particularly severe in the case of valuation ratios (where the covariance between the shocks $\sigma_{\epsilon\xi}$ is typically negative, which results in an upward bias of $\hat{\beta}$). A bias-corrected estimator $\hat{\beta}^s = \hat{\beta} + \hat{\gamma}(1+3\hat{\rho})/n$, where n denotes the sample size and $\hat{\gamma}$ is a sample estimate of γ, has been used e.g. by Giot and Petitjean (2006) in the single predictor case.

Since this paper is concerned about the issue of model uncertainty involving a multiplicity of variables, we work with the generalized case of multiple predictors as in Equations (2) and (3). A bias-corrected estimator for the vector of predictive coefficients β_j in Equation (2) has recently been derived by Amihud and Hurvich (2004). The approach by Amihud and Hurvich (2004) amounts to running an augmented regression

$$r_t = \alpha + \beta_j' x_{j;t-1} + \phi_j' \nu_{j,t}^c + e_{j,t}, \tag{10}$$

Equation (10) is equivalent to running the predictive regression in (2) augmented by a corrected $k_j \times 1$ residual series $\nu_{j,t}^c$. As shown by Amihud and Hurvich (2004), this procedure yields a reduced-bias estimator $\hat{\beta}_j^c$ for the vector of predictive coefficients. The residual series $\nu_{j,t}^c = x_{j;t} - (\hat{\Theta}_j^c + \hat{\Phi}_j^c x_{j;t-1})$ is based on a reduced-bias estimator for the autoregressive parameters $\hat{\Phi}_j$ in the multivariate AR(1) model in Equation (3). Our estimate of $\hat{\Phi}_j^c$ follows the approach put forth by Amihud and Hurvich (2004) for the case when Φ_j is constrained to be diagonal.[12] Hence, the different series $x_{j,t}^i$ $(i = 1, \cdots, k_j)$ are considered separately.

[12] Allowing for a non-diagonal structure raises the need to estimate a multiplicity of parameters, in particular as k_j gets large. This may result in a degradation of performance (See Amihud and Hurvich (2004)). We therefore impose a diagonal structure.

The individual error series are constructed as $\nu^i_{j,t} = x^i_{j,t} - \hat{\theta}^i_j - \hat{\rho}^c_i x^i_{j,t-1}$. The autoregressive parameters are adjusted according to finite-sample bias by $\hat{\rho}^c_i = \hat{\rho} + (1+3\hat{\rho})/n + 3(1+3\hat{\rho})/n^2$ The reduced bias-estimator β^c_j is then obtained by regressing stock excess returns on the set of k_j lagged predictive and the corrected error proxies ($\nu^i_{j,t}$ $(i = 1, \cdots, k_j)$). Standard errors for β^c_j are adjusted for the two-step procedure as proposed in Amihud and Hurvich (2004).

3 Empirical Results

3.1 Data

Our dataset comprises monthly and quarterly data for five international stock markets: France, Germany, Japan, the United States and the United Kingdom. The dependent variables are log (returns) on broad stock indexes in excess of the (log) short-term interest rate. Monthly summary statistics on the dependent variables and the predictive variables can be found in Table 1.

– Insert Table 1 about here –

We assemble a data set of nine financial and macroeconomic predictive variables for the different international stock markets. The following variables comprise our set of predictors:

Interest rate variables: Difference between the yield on long-term government bonds and the three-month interest rate (Term spread, TRM), short term interest rate relative to its 12-month backward-looking moving average (RTB), long-term government bond yield relative to its 12-month backward-looking moving average (RBR).

Valuation Ratios and other Financial Variables: Dividends paid over the past 12 months in relation to the current price (dividend yield, LDY) and earnings over the past 12 months in relation to the current price (earnings yield, LEY), both in logs. (Log) realized stock market volatility (LRV).

Macro Variables: Annual inflation rate (INF) based upon the Consumer Price Index, annual industrial production growth (IPG), estimate of the output gap obtained by the HP-filter (GAP).

Due to data availability, the different sample periods differ across markets. For most countries, the sample periods start in the early 1970s and end in mid 2000. The US sample starts already in the late 1950s. The selection of variables is guided mainly by the previous US literature, as well as data availability.[13] Unfortunately, a default yield spread based on the yield difference of BAA and AAA rated corporate bonds (as used e.g. by Avramov 2002 or Cremers 2002) does not exist in the different international stock markets outside the US in a reasonable quality. For further information on data sources and construction the reader is referred to Appendix I.

Table 1 provides monthly summary statistics on the mean, standard deviation and first-order autocorrelation of the particular state variables. The AC(1) coefficients in the table shows that some series, in particular for valuation ratios and the inflation rate, exhibit a fairly strong degree of persistence. For this reason, taking into account the time series properties and finite-sample biases seems to be warranted.

3.2 In-sample Results: Return Predictability in International Stock Markets

We first discuss the results of the in-sample analysis of return predictability in international stock markets. The only subjective element of the BACE approach is the choice of the a-priori expected model size $\bar{k}$, i.e. the researcher's belief of how many variables are a-priori likely to be included in the predictive model. We choose a rather moderate specification of this hyperparameter, consistent with the principle of parsimony prevailing in econometrics.

[13] (Subsets) of these variables are used for instance in Fama and Schwert (1977), Fama and French (1988), Campbell and Shiller (1988a), Fama and French (1989)) Fama (1981), Fama (1990), Hodrick (1992),Avramov (2002), Cremers (2002), Lewellen (2004), Ang and Bekaert (2005), Rapach et al. (2005), Cooper and Priestley (2006), Pastor and Stambaugh (2006).

We therefore set the a-priori expected model size to $\bar{k} = 2$ variables.[14] This implies a prior probability of inclusion of $\pi = 2/\kappa = 0.\bar{2}$ for each variable. The choice of the expected model size is linked to the a-priori model probability $p(\mathcal{M}_j)$ which is given as $p(\mathcal{M}_j) = \pi^{k_j}(1-\pi)^{\kappa-k_j}$.[15] Note that a prior probability of inclusion smaller than 0.5 amounts to an a-priori down weighting of larger model specifications. This implies an additional penalty for highly parameterized models beside the penalty implied by the degree of freedom adjustment of the BIC.

The tables for the different stock markets, which will be discussed in the following, are all organized in the same way. Panel A and C are based on monthly data while Panel B and D present results for quarterly data. Panel A and B report results for the composite model with bias-corrected slope coefficients. $\pi|y$ denotes the posterior probability of inclusion for each variable. The posterior probability of inclusion is defined as the total sum of the posterior probabilities of all models, in which the particular variable is included: $\mathcal{C}'\mathcal{P}$, where $\mathcal{C}$ is a $2^\kappa \times \kappa$ matrix denoting inclusion (exclusion) of a particular variable in model j by 1 (0), and $\mathcal{P}$ is a $2^\kappa \times 1$ vector containing the posterior model probabilities $p(\mathcal{M}_j|y)$. Posterior means of the predictive coefficients in the weighted model are reported in the second column of Panels A/B. The third and fourth column report posterior Bayesian t-ratios. Following Avramov (2002), we report both t-ratios based on posterior standard deviations which ignore model uncertainty and t-ratios adjusted for model uncertainty (see discussion in section 2).

We also assess the robustness of the different state variables according to two other criteria. In Panels A/B we report the proportion of cases when the coefficient on a particular variable (every time it is included in one of the $j = 1, \cdots, 2^\kappa$ models) has the same sign as the posterior mean in the composite model (denoted as sgn in the tables). Moreover, we also report the fraction of cases across the different models, which include the particular variable, when a classical t-statistic is greater than two in absolute value. This statistic serves as another indicator of the robustness or fragility of a particular predictive variable. In Panels C and D, the five top-performing models are displayed (highest posterior probability of all models).

[14]The main results reported in this paper are not sensitive to this choice.

[15]In principle, one could also specify different prior probabilities of inclusion for the different variables based on economic considerations.

The models are defined by inclusion (1) or exclusion (0) of the specific variable. Moreover, the corresponding posterior model probabilities and the adjusted R^2 are reported.

3.2.1 France

Estimation results for the French stock market are provided in Table 2. As Panel A (monthly predictive regressions) shows, the only variable for which the posterior probability of inclusion $\pi|y$ rises compared to the prior probability of inclusion is the relative bond-rate RBR. In the case of the other variables, inspection of the data leads us to reduce us our prior opinion about their usefulness. Panel C reports monthly results for the five best-performing model specifications. After having seen the data, the model which includes RBR as a single predictive variable receives a posterior model probability of more than 50%, which is greatly higher than the one of the next best model specifications. The relative bond rate together with the output gap is also significant according to a posterior t-ratio.

Robustness of a particular variable can also be assessed by the sign certainty probability. This is the fraction the coefficient on the variable (when included in one of the 2^{κ} Models) has the same sign as its coefficient in the weighted model. According to this criterion, the relative bond rate is rather successful. The relative bond rate (RBR), the term spread (TRM), industrial production growth (IPG) and the output gap (GAP) all have sign certainty probabilities exceeding 90%, whereas several other popular predictors such as the dividend yield perform clearly worse. However, Table 2 also makes clear that none of the variables remains significant when the additional variability of estimates across models is accounted for.[16]

– Insert Table 2 about here –

[16]This is a general result which holds for almost all predictive variables and almost all stock markets considered. In this way, we provide evidence consistent with Avramov (2002) that predictive regressions in finance are subject to a great deal of model uncertainty. Avramov also finds that almost all variables which appear to be significant, lose their significance once model uncertainty is considered.

Panels B and D show that the evidence for predictability in the French stock market is stronger in the quarterly case. Both, the relative bond rate and the dividend yield have higher posterior probabilities of inclusion than $0.\bar{2}$. Yet, the dividend yield does not appear to have significant slope coefficient in the composite model, which may be due to finite-sample bias. It is also worth noting that the earnings yield performs relatively well in-terms of sign certainty in the quarterly case.

3.2.2 Germany

Table 3 provides estimation results for the German stock market. As can be seen in Panel A and C of Table 3, predictability of monthly stock returns is rather weak on statistical grounds. The case for predictability is clearly less pronounced than in the French stock market discussed the previous subsection. The model receiving the highest posterior probability is the one which does not include any lagged state variables (iid case). None of the variables in the monthly model receives a higher posterior inclusion probability compared to the prior inclusion probability of $0.\bar{2}$. Among the variables considered only the relative bond rate (5%) and the output gap (10%) can be considered as significant according to a Bayesian t-ratio, but this does not hold true when the dispersion of coefficients across models is considered.

– Insert Table 3 about here –

Similar to the French case, the relative bond rate is rather important in the quarterly regressions where the probability of inclusion rises after having seen the data. Evidence for predictability with quarterly data is somewhat stronger than for monthly data. This can be seen from the result in Panel D that the most likely quarterly model is now the one which includes the relative bond rate. This model achieves an adjusted R^2 of about 5% in the quarterly regressions, which is quite high for the stock return predictability literature. Several variables appear rather robust with regard to sign certainty: The term spread (TRM), the relative bond rate (RBR), industrial production growth (IPG), and the dividend yield (LDY) have the same sign as the posterior mean in the composite model in more than 90% of all models in which they are included.

3.2.3 Japan

Results for the Japanese stock market are given in Table 4. As for Germany, there is no compelling evidence that monthly stock returns in Japan are predictable: The model with clearly the highest posterior probability in Panel C is the model with no explanatory variables (iid-model). The dividend yield and the output gap are somewhat marginally important. Note also that the term spread (TRM), industrial production growth (IPG) and the earnings yield (LEY) are rather robust in terms of sign probability.

– Insert Table 4 about here –

A greater amount of predictability can be detected in the quarterly case. In particular, the dividend yield receives a higher posterior probability of inclusion than expected a-priori (Panel D of Table 4). The relative short-term interest rate (RTB) and the earnings yield (LEY) are the most robust variables with regard to sign certainty. It is also worth noting that Japan is the only stock market where the relative bond rate has virtually no predictive power for stock market excess returns.

3.2.4 United Kingdom

Table 5 shows, that both for monthly and quarterly predictive regressions, the case for return predictability in the United Kingdom is rather weak. Panel C shows, that the largest posterior probability in the monthly regressions is assigned to the iid model (as for the monthly regressions for Germany and Japan). Among the economically motivated predictors, the output gap

(GAP) is the most relevant variable, in that our prior opinion on its importance is updated upwards after having seen the data. According to the posterior t-ratio, also the relative bond rate (RBR) can be considered as significant in the monthly composite model. Yet, as before, accounting for model uncertainty greatly reduces the evidence for predictability.

– Insert Table 5 about here –

The inflation rate (INF) and industrial production growth (IPG) can be regarded as rather robust with respect to sign certainty. The output gap – which performs rather well according to other criteria – appears to be a rather fragile variable in that regard. Also notice that industrial production growth and the earnings yield generally appear often as significant variables according to a classical t-test, but since the models which include these variables receive a low posterior probability, they are not ranked among the most important variables in the composite model.

3.2.5 United States

Table 6 reports results on in-sample return predictability for the US stock market. As shown by the table, evidence for predictability is clearly stronger in the US than for the stock markets previously discussed. Variables which appear important after having seen the data are the relative bond rate (RBR) and, most importantly, the output gap (GAP). The output gap is the only variable which can also be considered as significant once model uncertainty is accounted for. It receives a posterior probability of inclusion of more than 80%, which is a substantial upward revision of the prior probability of inclusion. Thus, our results corroborate the recent findings by Cooper and Priestley (2006) that risk-premia may be vary with the output-gap. Good economic conditions as measured by the output gap are associated with low risk premia. However, these results must be taken with a grain of salt: the output gap is usually not the most robust variable with regard to sign certainty, a result which also holds true for most of the other stock markets considered. In this way, this paper contributes additional evidence lacking in Cooper and Priestley (2006) on the role of the output gap in predictive regressions. Several other variables are important when model uncertainty is ignored (the relative bond rate (RBR), inflation rate (INF), and industrial production growth (IPG)), but lose their significance once model uncertainty is considered.

– Insert Table 6 about here –

In the quarterly model, the output gap (GAP), earnings yield (LEY), realized volatility (LRV) and inflation (INF) appear as important variables a-posteriori. Contrary to the monthly case, the output gap, however, does not survive the model uncertainty adjustment anymore. Also note that the relative bond-rate is less important in the quarterly regressions. Panels A and B further show that the earnings yield appears to be very robust with regard to sign certainty, which holds both in the monthly and the quarterly models.

3.3 Out-of-Sample Analysis of Return Predictability

The question whether predictability of stock returns exists out-of-sample (OOS) has been a much debated topic. Results in the literature are mixed.[17] There are several theoretical reasons why OOS performance of stock return prediction models may be poor. Cochrane (2006), for instance shows by simulations that even in a world where stock returns are truly predictable, the results of Goyal and Welch (2006) will happen frequently. Inoue and Kilian (2004) argue that in-sample predictability tests are more powerful than out-of sample tests and are therefore more trustworthy when assessing the existence of a predictive relationship.

It is not the purpose of this paper to discuss the entire debate in the literature or to take a particular side. Rather, we are interested in a thorough investigation of the performance of

[17] The recent predictability debate has been spurred by the question whether the documented (limited) in-sample predictability is of any use for an investor in real-time. See the different conclusions obtained by e.g. Goyal and Welch (2006) and Campbell and Thompson (2006).

model averaging in the context of OOS predictability of excess returns. Avramov (2002) argues that averaging the forecasts of the different competing models in a Bayesian model averaging framework can substantially improve the out-of-sample forecast performance. Therefore, the main motivation of our analysis in this subsection is to reassess the findings by Avramov (2002) in our set of international stock markets.

For the purpose of evaluating OOS forecast performance, we estimate the 2^{κ} models using a recursive scheme. The first ten years are used as initialization period. Afterwards, the models are estimated recursively. We compare the performance of several (conditional) models to the results of an unconditional (or naive) benchmark model which takes the prevailing historical mean as the forecast of the dependent variable. The model-based forecasts include Bayesian averaging of OLS coefficients adjusted for finite-sample bias (BACE-adj), a pure Bayesian model averaging approach (BMA) with g-prior specification[18], the individual model which receives the highest posterior model probability according to BMA (denoted as TOP), and an all-inclusive specification (ALL). Following Bossaerts and Hillion (1999), we also assess the performance of individual models selected by the conventional model selection criteria: Akaike criterion (AIC), Schwarz criterion (BIC), as well as the adjusted R^2. The corresponding (pseudo-) OOS forecasts are then evaluated according to several criteria for assessing forecast accuracy.

– Insert Table 7 about here –

Tables 7 and 8 report the results of the evaluation of OOS performance for our international set of stock markets. The evaluation of forecast accuracy uses standard criteria. MPE denotes the mean prediction error. Testing the significance of the MPE amounts to testing the unbiasedness of the forecasts. U_1 is the ratio of the mean absolute prediction error (MAPE)

[18] The approach is similar to Cremers (2002). However, rather than motivating the g hyperparameter from economic reasoning, we follow recommended practice and set this parameter to $g = \max\{n, \kappa^2\}^{-1}$, where n denotes the sample size (See Fernández et al. 2001; Koop 2003).

of the particular model-based forecast to the one of the naive benchmark model. U_2 is the ratio of the mean square prediction error (MSPE) of the particular model-based forecast to the one of the naive benchmark model.[19] Contrary to the previous literature we also provide an evaluation of directional accuracy of forecasts obtained by model averaging. We therefore report the fraction of times the direction of the dependent variable is correctly predicted by the model (denoted as Hit in the table). PT denotes the test-statistic for directional accuracy proposed by Pesaran and Timmermann (1992). Net-SSE plots in the spirit of Goyal and Welch (2006) are presented in Figures 1 and 2. These graphs display the cumulated sum of the squared forecast errors of the benchmark model minus the squared forecast errors of the model of interest. These plots show how the OOS performance of the predictive model evolves over time. Periods where the line in the graph is upward sloping are times when the conditional model outperforms the naive model in terms of squared forecast errors.

– Insert Figure 1 about here –

As the evaluation of the monthly forecasts in Table 7 shows, OOS stock return predictability is very limited. Moreover, notable differences of OOS return predictability can be detected across countries. It is worth noting that for those countries for which the in-sample evidence for predictability was stronger also the evidence for return predictability out-of-sample appears to be stronger.

The results for the French stock market, presented in Panel A of Table 7, show some evidence for out-of-sample predictability. This is consistent with the in-sample results for the composite model, where also the evidence was stronger compared to other capital markets (such as the UK or Germany). Panel A also shows that the model averaging approaches (BACE-adj, BMA) outperform the naive model and model selection criteria in terms of MSPE. All model-

[19] Note that U_1 and U_2 are merely descriptive criteria. In the case of nested models, the mean square prediction error MSPE of the smaller nested model is expected to be smaller than the MSPE under the null of equal predictive power, a point raised by Clark and West (2007). This is due to the fact that the larger model needs to estimate parameters which are zero in population, which introduces noise in the forecasts.

based forecasts generally appear to be unbiased for the French case. The Net-SSE plot (a) in Figure 1 shows the relative OOS performance of the forecasts produced by the BACE-adj model over time.[20] As shown by the graph, the model has produced lower squared forecast errors relative to the benchmark up to about 1999. Afterwards, the relative performance has somewhat decreased.

In the case of Germany (Panel B of Table 7), BACE-adj and BMA generally do a better job compared to other model specifications, but are not able to outperform the iid model in terms of MSPE or MAPE. This is consistent with the modest results for in-sample predictability, where the most likely model a-posteriori with monthly data was the iid model. The Net-SSE plot (b) in Figure 1 shows that OOS predictability has been stronger in the 1990s, where lagged state variables contributed to lower squared prediction errors. Since about 1999, OOS predictability has declined.

Also in the Japanese stock market the case for OOS predictability is fairly weak, as Panel C of Table 7 makes clear: forecasts of the naive model generally produce the lower MAPE or MSPE than conditional models. This is confirmed by the Net-SSE plot (c) of Figure 1. The plot shows a steady decline of OOS forecast performance of the weighted model forecast from the late 1980s onwards. Since the early 1990s the OOS performance has remained rather constant.

Panel D of Table 7 presents results of the OOS evaluation for the United Kingdom. Analogously to Germany and Japan, OOS predictability is very poor. Moreover, the United Kingdom is an exception in that it is the only stock market where the conditional models produce forecasts with a substantial bias. Also note that the model averaging methods BACE-adj and BMA outperform the other selection criteria but fail to outperform the naive model in terms of MAPE and MSPE.

OOS evaluation results for the US stock market are given in Panel E of Table 7. As in the case of France, consistent with the stronger evidence for in-sample predictability, there is some evidence for OOS predictability. The forecasts generally appear to be unbiased. When

[20]Net-SSE plots based on the BMA approach are generally quite similar.

judged according to MSPE and MAPE, the forecasts produced by BACE-adj show a better performance compared to other models. Furthermore, forecasts based on Bayesian model averaging techniques (BACE-adj, BMA) are the only ones which achieve a lower MAPE and MSPE than the naive model. There is also some (modest) statistical evidence that these approaches have managed to get the direction of the stock market right (BACE: significant at the 10% significance level in a one-sided test of "no market timing" against the alternative of "market timing", BMA: significant at the 5% level). The Net-SSE plot for the United States in (e) of Figure 1 shows a decline of predictability since the early 1990s, which has slightly reversed since 2003.

– Insert Table 8 about here –

The results for OOS predictability of quarterly market excess returns (reported in Table 8) are quite similar to the monthly case. Again, there is more evidence for OOS predictability in stock markets where evidence for in-sample predictability was more pronounced (i.e. France and the United States). There is no evidence however, that OOS predictability increases with the horizon of the forecast. Quite to the contrary, OOS predictability is somewhat smaller than the (already modest) OOS predictability in the monthly case.

– Insert Figure 2 about here –

In the quarterly case for the French stock market (Panel A of Table 8), we find again some evidence that BACE and BMA outperform the iid model in terms of MAPE and MSPE. This also holds for models selected by the AIC (in contrast to the monthly case). Some evidence against unbiasedness is found for the all-inclusive specification. The quarterly Net-SSE plot for France (a) in Figure 2 shows a decline of OOS performance of the composite model during the period 1997-2003; since then OOS predictability has picked up somewhat toward the end of the sample.

Results for the German stock market (Panel B of Table 8) are quite similar to the monthly case. However, modest evidence of market timing possibility can be found for quarterly models. This happens in particular for models which tend to be highly parameterized (i.e. ALL, AIC, R^2_{adj}), with PT-statistics significant at the 10% level. Quarterly results for Japan (Panel C) and UK (Panel D) are very similar to the monthly case.

For the US stock market (Panel E) evidence for OOS predictability is less pronounced compared to the monthly case. The forecasts by BMA are the only ones which have a (slightly) lower MSPE compared to the naive model. There exists no evidence for directional predictability in the case of the quarterly models. According to the Net-SSE plot for the US in Figure 2, OOS performance has been poor over most of the 1990s; however, there is some evidence for a slight increase in recent years.

4 Conclusion

This paper explores stock return predictability in international stock markets in the context of model uncertainty. A Bayesian averaging of classical estimates (BACE) approach is used to account for the tremendous uncertainty of a typical investor to know what the important predictive variables are. This approach is combined with a finite-sample bias correction which accounts for the persistence of the usually employed state variables. Using a comprehensive dataset for international stock markets allows us to gain fresh insights into the empirical evidence for return predictability, which has to the date been mainly based on results for the US stock market.

We find interesting differences across countries in terms of return predictability. Evidence for in-sample predictability is stronger for France and the United States compared to the other countries. For these two stock markets, also a (modest amount) of out-of-sample predictability can be detected. Out-of-sample predictability by model averaging methods appears to be stronger for monthly than for quarterly data. Consistent with Avramov (2002), we find that model averaging often produces better OOS forecasts than individual models based on selection criteria. Nevertheless, we also document a substantial amount of time-variation of OOS forecast performance by averaged forecasts.

Two variables appear to be rather robust predictors across countries: the relative bond rate and the output gap. The latter is the only variable which also remains a significant predictor of market excess returns in the US, once model uncertainty is accounted for. The earnings yield often appears to be a more robust predictive variable than the dividend yield. In general, however, our results show that evidence for in-sample predictability for the excess returns in international equity markets is substantially weakened once model uncertainty is accounted for.

The empirical work in this paper can still be extended along the following lines. It would be interesting to link the evidence for time-series predictability with predictability in the cross-section. An international analysis with size and book-to-market sorted portfolios may provide interesting insights whether there are differences in the state variables which drive risk-premia in the cross-section. Moreover, a major factor for poor out-of-sample performance may be potential model instability. Our out-of-sample forecasting exercise, for instance, shows that top-performing models often change over time. To study the time-varying performance of the predictive models in greater depth may therefore provide important insights on why out-of-sample performance is rather poor. These issues are on our agenda for further research.

References

Amihud, Yakov, and Clifford Hurvich, 2004, Predictive regressions: a reduced-bias estimation method, *Journal of Financial and Quantitative Analysis* 39, 813–842.

Ang, Andrew, and Geert Bekaert, 2005, Stock return predictability: Is it there?, forthcoming Review of Financial Studies.

Avramov, D., 2002, Stock return predictability and model uncertainty, *Journal of Financial Economics* 64(3), 423–458.

Bossaerts, Peter, and Pierre Hillion, 1999, Implementing statistical criteria to select return forecasting models: What do we learn?, *The Review of Financial Studies* 12, 405–428.

Brown, Gregory W., and Michael T. Cliff, 2005, Investor sentiment and asset valuation, *Journal of Business* 78, 405–440.

Campbell, J. Y., and S. Thompson, 2006, Predicting the equity premium out of sample: Can anything beat the historical average?, NBER, Working Paper No. 11468, forthcoming Review of Financial Studies.

Campbell, J. Y., and M. Yogo, 2003, Efficient tests of stock return predictability, forthcoming Journal of Financial Economics.

Campbell, John Y., 1987, Stock returns and the term structure, *Journal of Financial Economics* 18, 373–399.

Campbell, John Y., and Robert J. Shiller, 1988a, The dividend-price ratio and expectations of future dividends and discount factors, *Review of Financial Studies* 1, 195–228.

Campbell, John Y., and Robert J. Shiller, 1988b, Stock prices, earnings, and expected dividends, *Journal of Finance* 43, 661–676.

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, J. H., 2006, The dog that did not bark: A defense of return predictability, Working Paper, University of Chicago, forthcoming Review of Financial Studies.

Cochrane, John H., 1999, New facts in finance, *Federal Reserve Bank of Chicago* 11, 3–31.

Cooper, Ilan, and Richard Priestley, 2006, Time-varying risk premia and the output gap, Working Paper, Norwegian School of Management.

Cremers, K. J. Martijn, 2002, Stock return predictability: A Bayesian model selection perspective, *The Review of Financial Studies* 15, 1223–1249.

Fama, E., and G. W. Schwert, 1977, Asset returns and inflation, *Journal of Financial Economics* 5, 115–146.

Fama, Eugene F., 1981, Stock returns, real activity, inflation, and money, *The American Economic Review* 71, 545–565.

Fama, Eugene F., 1990, Stock returns, expected returns, and real activity, *Journal of Finance* 45, 1089–1108.

Fama, Eugene F., and Kenneth R. French, 1988, Dividend yields and expected stock returns, *Journal of Financial Economics* 22, 3–25.

Fama, Eugene F., and Kenneth R. French, 1989, Business conditions and expected returns on stocks and bonds, *Journal of Financial Economics* 25, 23–49.

Fernández, C., E. Ley, and M. F. J. Steel, 2001, Benchmark priors for Bayesian model averaging, *Journal of Econometrics* 100, 381–427.

Ferson, Wayne E., Sergei Sarkissian, and Timothy T. Simin, 2003, Spurious regressions in financial economics, *Journal of Finance* 58, 1393–1413.

French, K., G. W. Schwert, and R. F. Stambaugh, 1987, Expected stock returns and volatility, *Journal of Financial Economics* 19, 3–29.

Giot, Pierre, and Mikael Petitjean, 2006, International stock return predictability: Statistical evidence and economic significance, CORE Discussion paper 2006/88.

Goyal, Amit, and Ivo Welch, 2003, Predicting the equity premium with dividend ratios, *Management Science* 49, 639–654.

Goyal, Amit, and Ivo Welch, 2006, A comprehensive look at the empirical performance of equity premium prediction, forthcoming Review of Financial Studies.

Hjalmarsson, Erik, 2004, On the predictability of global stock returns, working paper.

Hodrick, Robert J., 1992, Dividend yields and expected stock returns: Alternative procedures for inference and measurement, *Review of Financial Studies* 5, 357–386.

Hodrick, Robert J., and Edward Prescott, 1997, Postwar U.S. business cycles: an empirical investigation, *Journal of Money, Credit and Banking* 29, 1–16.

Inoue, Atsushi, and Lutz Kilian, 2004, In-sample or out of sample tests of predictability: Which one should we use?, *Econometric Reviews* 23, 371–402.

Koop, Gary, 2003, *Bayesian Econometrics* (Wiley, Chichester).

Lamont, Owen, 1998, Earnings and expected returns, *Journal of Finance* 53, 1563–1587.

Leamer, E. E., 1978, *Specification Searches: Ad Hoc inference with Non-experimental data* (Wiley).

Lettau, Martin, and Sydney Ludvigson, 2001, Consumption, aggregate wealth, and expected stock returns, *Journal of Finance* 56, 815–849.

Lewellen, Jonathan, 2004, Predicting returns with financial ratios, *Journal of Financial Economics* 74, 209–235.

Moon, R., A. Rubia, and R. Valkanov, 2006, Long-horizon regressions when the predictor is slowly varying, Working Paper.

Pastor, Lubos, and Robert Stambaugh, 2006, Predictive systems: Living with imperfect predictors, Working Paper, GSB Chicago.

Pesaran, M.H., and A. Timmermann, 1992, A simple non-parametric test of predictive performance, *Journal of Business and Economic Statistics* 10, 461–465.

Rangvid, Jesper, 2006, Output and expected returns, *Journal of Financial Economics* 81, 595–624.

Rapach, David E., Mark E-. Wohar, and Jesper Rangvid, 2005, Macro variables and international stock return predictability, *International Journal of Forecasting* 21, 137–166.

Sala-i-Martin, Xavier, Gernot Doppelhofer, and Ronald I. Miller, 2004, Determinants of long-term growth: A Bayesian averaging of classical estimates (BACE) approach, *American Economic Review* 94, 813–835.

Stambaugh, R. F., 1999, Predictive regressions, *Journal of Financial Economics* 54, 375–421.

Torous, W., R. Valkanov, and S. Yan, 2004, On predicting returns with nearly integrated explanatory variables, *Journal of Business* 77, 937–966.

A Appendix I: Data Description

This section of the appendix provides a more detailed analysis of the stock returns as well as the information variables used in our analysis. The original data are monthly but we also report estimation results using quarterly data. Information on the sample periods for the international stock markets can be found in table 1.

Dependent variables The dependent variables for the international stock markets are taken from various sources. In the case of Germany, the return on the DAFOX is used, which is a broad stock index published for research purposes by Karlsruher Kapitalmarktdatenbank. It comprises all German stocks traded in the top segment (Amtlicher Handel) of the Frankfurt stock exchange. For the US, the value-weighted return on the CRSP market portfolio is used.[21] For the other stock markets broad stock market indexes provided by Datastream are used. Excess returns are constructed by subtracting a risk-free rate proxy. When available a 3-month T-Bill is used as the risk-free rate proxy. Otherwise a three-month money market rate is used. Data are provided by Reuters-Ecowin. In the case of Germany, the money market rate for three-month deposits obtained from the time series database of Deutsche Bundesbank is used as our proxy for the risk-free rate.

Interest rate related variables The term spread (TRM) is defined as the difference of the yield on long-term government bonds and the short-term interest rate (3-month). The necessary yield curve and interest rate data were obtained from the time series database of Deutsche Bundesbank (Germany), St-Louis Fed (USA), Econstats (France, United Kingdom and Japan). Following much of the extant literature, the relative short-term interest rate (RTB) is calculated as the short-term interest rate minus its 12-month backward looking moving average. The relative long-term bond rate (RBR) is calculated as the long-term government bond yield minus its 12-month backward looking moving average.

[21] We thank Amit Goyal and Ivo Welch for providing these data on their webpages.

Valuation ratios and other financial variables The time series of dividend yields (LDY) and earnings yield (LEY) are defined as dividends (earnings) over the past 12 months in relation to the current price. Both series are used in logs, which improves their time-series properties as noted by Lewellen (2004). The US data are taken from Amit Goyal's webpage, while the rest of the valuation ratios refer to the broad stock market indexes provided by Datastream. Realized stock market volatility (LRV) is computed as the sum of the squared daily stock returns and is also used in logs.

Macroeconomic variables The annual inflation rate (INF) is calculated from the seasonally--adjusted Consumer Price Index (CPI). Another macroeconomic variable is the annual growth rate of industrial production (IPG). The time series of the CPI as well as industrial production for the calculation of industrial production growth (IPG) and the output gap (OPG) measure are from the IMF/IFS database and were obtained from Reuters-Ecowin. Following Cooper and Priestley (2006), we construct the output gap measure by applying the filter by Hodrick and Prescott (1997) (HP-filter) to the logarithmic series of industrial production. As in Following Cooper and Priestley (2006) the smoothing parameter to 128800 for the monthly data and 1600 for the quarterly data. The cyclical component of the series is taken as the output gap.

Table 1: Summary Statistics, Monthly

	EXRET	TRM	RTB	RBR	INF	IPG	LRV	LDY	LEY	GAP
France: 1973:02-2005:10										
Mean	0.0044	1.0938	-0.0054	-0.0054	5.1294	0.9889	-6.1797	-3.3346	-2.5178	0.0005
Std.	0.0621	1.2517	0.1099	0.0710	4.0892	4.4328	0.7752	0.3515	0.3275	0.0287
AC(1)	0.0798	0.9207	0.9183	0.9237	0.9966	0.8737	0.5835	0.9782	0.9673	0.8598
Germany: 1972:02-2004:12										
	EXRET	TRM	RTB	RBR	INF	IPG	LRV	LDY	LEY	GAP
Mean	0.0031	1.3726	-0.1960	-0.0943	2.8285	1.2246	-6.5848	-3.7179	-2.7080	-0.0019
Std.	0.0513	1.6839	1.1858	0.6146	1.8475	4.0470	0.9695	0.3530	0.2514	0.0289
AC(1)	0.0872	0.9723	0.9566	0.9054	0.9777	0.8178	0.7488	0.9824	0.9568	0.8354
Japan: 1973:02-2005:11										
	EXRET	TRM	RTB	RBR	INF	IPG	LRV	LDY	LEY	GAP
Mean	0.0016	0.6874	-0.0750	-0.0009	3.0833	2.1889	-6.5518	-4.5379	-3.5609	-0.0004
Std.	0.0522	1.1971	1.1642	0.0066	4.6170	6.2448	1.0202	0.5050	0.4687	0.0416
AC(1)	0.0838	0.9518	0.9611	0.9066	0.9890	0.9426	0.7242	0.9930	0.9905	0.9402
United Kingdom: 1973:01-2005:11										
	EXRET	TRM	RTB	RBR	INF	IPG	LRV	LDY	LEY	GAP
Mean	0.0037	0.7989	-0.0020	-0.0062	6.6242	1.1718	-6.4382	-3.1629	-2.5142	0.0009
Std.	0.0566	2.1353	0.1168	0.0671	5.2741	4.0886	0.8087	0.2748	0.3977	0.0270
AC(1)	0.1092	0.9774	0.9271	0.9092	0.9930	0.8562	0.6812	0.9747	0.9856	0.8691
United States: 1958:01-2005:12										
	EXRET	TRM	RTB	RBR	INF	IPG	LRV	LDY	LEY	GAP
Mean	0.0044	1.6348	0.0000	0.0010	4.0387	3.0665	-6.6934	-3.5071	-2.7798	-0.0011
Std.	0.0423	1.4360	0.0860	0.0480	2.7614	4.8664	0.8540	0.3958	0.3923	0.0309
AC(1)	0.0282	0.9493	0.9034	0.8765	0.9936	0.9609	0.8188	0.9918	0.9926	0.9637

Note: The table reports summary statistics of (log) market excess returns (EXRET) and predictive variables in five international stock markets. The set of predictors comprises the term spread (TRM), the short-term interest rate relative to its 12-month moving average (RTB), a long-term government bond yield relative to its 12-month moving average (RBR), annual inflation rate (INF), annual growth of industrial production (IPG), (log) realized volatility (LRV), (log) dividend yield (LDY), (log) earnings yield (LEY), output gap (GAP).

Table 2: Estimation Results, In-Sample: France

Panel A: Composite Model, Monthly

	$\pi\|y$	post. mean	t-ratio	t-ratio (adj)	sgn	fraction $\|t\| > 2$
TRM	0.034	0.0001	1.196	0.555	0.945	0.023
RTB	0.043	-0.0019	-1.456	-0.555	0.473	0.020
RBR	0.745	-0.1068	-3.373	-1.551	1.000	0.340
INF	0.042	0.0000	0.508	0.273	0.820	0.117
IPG	0.019	0.0000	0.907	0.438	0.945	0.250
LRV	0.050	0.0002	1.210	0.549	0.441	0.000
LDY	0.063	-0.0001	-0.100	-0.073	0.566	0.000
LEY	0.038	0.0001	0.202	0.163	0.742	0.000
GAP	0.113	-0.0284	-2.263	-0.752	0.949	0.172

Panel B: Composite Model, Quarterly

	$\pi\|y$	post. mean	t-ratio	t-ratio (adj)	sgn prob.	fraction $\|t\| > 2$
TRM	0.038	0.0002	0.988	0.506	0.902	0.004
RTB	0.063	-0.0066	-1.249	-0.557	0.625	0.027
RBR	0.666	-0.2088	-2.635	-1.048	1.000	0.414
INF	0.165	0.0007	1.689	0.683	0.891	0.453
IPG	0.030	0.0000	0.162	0.093	0.961	0.242
LRV	0.038	0.0001	0.267	0.166	0.250	0.000
LDY	0.264	0.0006	0.066	0.041	0.500	0.000
LEY	0.156	0.0048	0.960	0.528	0.910	0.000
GAP	0.105	-0.0667	-1.780	-0.675	0.816	0.027

Panel C: Top 5 Models, Monthly

TRM	0	0	0	0	0
RTB	0	0	0	0	0
RBR	1	1	1	1	1
INF	0	0	0	0	0
IPG	0	0	0	0	0
LRV	0	1	0	0	0
LDY	0	0	1	0	0
LEY	0	0	0	1	0
GAP	0	0	0	0	1
$p(\mathcal{M}_j\|y)$	0.567	0.033	0.028	0.026	0.022
$\bar{R}^2$	0.027	0.032	0.031	0.031	0.030

Panel D: Top 5 Models, Quarterly

TRM	0	0	0	0	0
RTB	0	0	0	0	0
RBR	1	1	1	0	1
INF	0	0	0	1	1
IPG	0	0	0	0	0
LRV	0	0	0	0	0
LDY	0	0	1	1	1
LEY	0	1	0	0	0
GAP	0	0	0	0	0
$p(\mathcal{M}_j\|y)$	0.314	0.105	0.088	0.070	0.035
$\bar{R}^2$	0.065	0.094	0.092	0.089	0.123

Note: Panel A (monthly) and B (quarterly) report estimation results for the composite model. The coefficients in the weighted model are the coefficients in individual models weighted by the posterior model probabilities. Posterior probabilities of inclusion are calculated as the sum of the posterior probabilities of the models which include the respective variable. Bayesian t-ratios are reported, without and with adjustment for model uncertainty (adj). Panel C (monthly) and D (quarterly) display the five best-performing model specifications (highest posterior model probability), where 0 indicates exclusion and 1 inclusion of the respective predictive variable. Also the adjusted R^2 and the posterior model probabilities of the models which are receive the highest posterior model probability are reported. The set of predictors comprises the term spread (TRM), the short-term interest rate relative to its 12-month moving average (RTB), a long-term government bond yield relative to its 12-month moving average (RBR), annual inflation rate (INF), annual growth of industrial production (IPG), (log) realized volatility (LRV), (log) dividend yield (LDY), (log) earnings yield (LEY), output gap (GAP).

Table 3: Estimation Results, In-Sample: Germany

Panel A: Composite Model, Monthly

	$\pi\|y$	post. mean	t-ratio	t-ratio (adj)	sgn	fraction $\|t\| > 2$
TRM	0.0279	0.0000	1.113	0.535	0.934	0.492
RTB	0.0247	-0.0001	-0.995	-0.503	0.750	0.082
RBR	0.1827	-0.0019	-2.385	-0.809	0.996	0.262
INF	0.0152	0.0000	-0.129	-0.111	0.402	0.027
IPG	0.0174	0.0000	-0.197	-0.127	0.047	0.285
LRV	0.0174	0.0000	0.339	0.235	0.133	0.520
LDY	0.0155	-0.0001	-0.950	-0.491	1.000	0.102
LEY	0.0159	-0.0001	-0.313	-0.239	0.496	0.000
GAP	0.0714	-0.0121	-1.807	-0.675	0.809	0.117

Panel B: Composite Model, Quarterly

	$\pi\|y$	post. mean	t-ratio	t-ratio (adj)	sgn prob.	fraction $\|t\| > 2$
TRM	0.038	0.0002	0.934	0.491	0.914	0.445
RTB	0.042	-0.0003	-0.792	-0.432	0.758	0.082
RBR	0.477	-0.0186	-2.645	-1.092	1.000	0.527
INF	0.027	0.0000	-0.094	-0.083	0.422	0.000
IPG	0.032	0.0000	0.298	0.177	0.965	0.324
LRV	0.043	0.0003	0.604	0.352	0.262	0.000
LDY	0.033	-0.0004	-0.453	-0.315	0.977	0.000
LEY	0.040	0.0004	0.252	0.191	0.688	0.000
GAP	0.128	-0.0748	-1.825	-0.698	0.715	0.074

Panel C: Top 5 Models, Monthly

TRM	0	0	0	1	0
RTB	0	0	0	0	1
RBR	0	1	0	0	0
INF	0	0	0	0	0
IPG	0	0	0	0	0
LRV	0	0	0	0	0
LDY	0	0	0	0	0
LEY	0	0	0	0	0
GAP	0	0	1	0	0
$p(\mathcal{M}_j\|y)$	0.653	0.157	0.058	0.020	0.019
$\bar{R}^2$	0.000	0.013	0.007	0.001	0.001

Panel D: Top 5 Models, Quarterly

TRM	0	0	0	0	0
RTB	0	0	0	0	0
RBR	1	0	0	1	0
INF	0	0	0	0	0
IPG	0	0	0	0	0
LRV	0	0	0	0	0
LDY	0	0	0	0	0
LEY	0	0	0	0	0
GAP	0	0	1	1	1
$p(\mathcal{M}_j\|y)$	0.353	0.324	0.076	0.022	0.016
$\bar{R}^2$	0.055	0.000	0.029	0.060	0.055

Note: Panel A (monthly) and B (quarterly) report estimation results for the composite model. The coefficients in the weighted model are the coefficients in individual models weighted by the posterior model probabilities. Posterior probabilities of inclusion are calculated as the sum of the posterior probabilities of the models which include the respective variable. Bayesian t-ratios are reported, without and with adjustment for model uncertainty (adj). Panel C (monthly) and D (quarterly) display the five best-performing model specifications (highest posterior model probability), where 0 indicates exclusion and 1 inclusion of the respective predictive variable. Also the adjusted R^2 and the posterior model probabilities of the models which are receive the highest posterior model probability are reported. The set of predictors comprises the term spread (TRM), the short-term interest rate relative to its 12-month moving average (RTB), a long-term government bond yield relative to its 12-month moving average (RBR), annual inflation rate (INF), annual growth of industrial production (IPG), (log) realized volatility (LRV), (log) dividend yield (LDY), (log) earnings yield (LEY), output gap (GAP).

Table 4: Estimation Results, In-Sample: Japan

Panel A: Composite Model, Monthly

	$\pi\|y$	post. mean	t-ratio	t-ratio (adj)	sgn	fraction $\|t\| > 2$
TRM	0.018	0.0000	0.706	0.416	0.930	0.137
RTB	0.017	0.0000	0.005	0.003	0.887	0.129
RBR	0.043	-0.0012	-1.047	-0.485	0.820	0.152
INF	0.032	0.0000	0.393	0.256	0.664	0.000
IPG	0.026	0.0000	1.198	0.551	0.914	0.477
LRV	0.029	0.0001	1.075	0.514	0.481	0.000
LDY	0.257	0.0044	1.510	0.703	0.840	0.000
LEY	0.066	0.0004	0.793	0.456	0.984	0.289
GAP	0.119	-0.0193	-1.751	-0.572	0.344	0.387

Panel B: Composite Model, Quarterly

	$\pi\|y$	post. mean	t-ratio	t-ratio (adj)	sgn prob.	fraction $\|t\| > 2$
TRM	0.032	0.0001	0.599	0.330	0.797	0.094
RTB	0.034	0.0007	0.509	0.226	0.918	0.246
RBR	0.062	-0.0014	-0.402	-0.248	0.809	0.059
INF	0.097	0.0000	0.552	0.343	0.824	0.012
IPG	0.032	0.0000	0.938	0.408	0.867	0.348
LRV	0.030	-0.0001	-0.569	-0.360	0.856	0.106
LDY	0.585	0.0388	1.840	1.005	0.973	0.211
LEY	0.121	0.0034	1.259	0.598	1.000	0.246
GAP	0.148	-0.0236	-0.765	-0.232	0.352	0.313

Panel C: Top 5 Models, Monthly

TRM	0	0	0	0	0
RTB	0	0	0	0	0
RBR	0	0	0	1	0
INF	0	0	0	0	0
IPG	0	0	0	0	0
LRV	0	0	0	0	0
LDY	0	1	0	1	1
LEY	0	0	0	0	0
GAP	0	0	1	0	1
$p(\mathcal{M}_j\|y)$	0.522	0.181	0.074	0.016	0.015
$\bar{R}^2$	0.000	0.015	0.010	0.022	0.022

Panel D: Top 5 Models, Quarterly

TRM	0	0	0	0	0
RTB	0	0	0	0	0
RBR	0	0	0	1	0
INF	0	1	0	0	1
IPG	0	0	0	0	0
LRV	0	0	0	0	0
LDY	1	1	1	1	1
LEY	0	0	0	0	0
GAP	0	0	1	0	1
$p(\mathcal{M}_j\|y)$	0.343	0.051	0.044	0.032	0.009
$\bar{R}^2$	0.060	0.080	0.078	0.072	0.102

Note: Panel A (monthly) and B (quarterly) report estimation results for the composite model. The coefficients in the weighted model are the coefficients in individual models weighted by the posterior model probabilities. Posterior probabilities of inclusion are calculated as the sum of the posterior probabilities of the models which include the respective variable. Bayesian t-ratios are reported, without and with adjustment for model uncertainty (adj). Panel C (monthly) and D (quarterly) display the five best-performing model specifications (highest posterior model probability), where 0 indicates exclusion and 1 inclusion of the respective predictive variable. Also the adjusted R^2 and the posterior model probabilities of the models which are receive the highest posterior model probability are reported. The set of predictors comprises the term spread (TRM), the short-term interest rate relative to its 12-month moving average (RTB), a long-term government bond yield relative to its 12-month moving average (RBR), annual inflation rate (INF), annual growth of industrial production (IPG), (log) realized volatility (LRV), (log) dividend yield (LDY), (log) earnings yield (LEY), output gap (GAP).

Table 5: Estimation Results, In-Sample: United Kingdom

Panel A: Composite Model, Monthly

	$\pi \vert y$	post. mean	t-ratio	t-ratio (adj)	sgn	fraction $\vert t\vert > 2$
TRM	0.018	0.0000	0.534	0.342	0.809	0.309
RTB	0.054	-0.0002	-1.645	-0.633	0.574	0.106
RBR	0.189	-0.1781	-2.421	-0.803	0.824	0.051
INF	0.071	-0.0001	-1.899	-0.687	1.000	0.371
IPG	0.024	0.0000	1.037	0.452	0.910	0.336
LRV	0.022	-0.0001	-1.000	-0.500	0.375	0.000
LDY	0.061	-0.0001	-0.234	-0.174	0.793	0.231
LEY	0.028	0.0000	-0.067	-0.040	0.238	0.481
GAP	0.251	-0.0394	-2.508	-0.829	0.492	0.219

Panel B: Composite Model, Quarterly

	$\pi \vert y$	post. mean	t-ratio	t-ratio (adj)	sgn prob.	fraction $\vert t\vert > 2$
TRM	0.028	0.0000	-0.076	-0.059	0.434	0.066
RTB	0.037	-0.0002	-0.628	-0.341	0.582	0.102
RBR	0.050	-0.0716	-1.159	-0.539	0.664	0.000
INF	0.057	-0.0001	-0.735	-0.434	0.883	0.031
IPG	0.044	0.0001	1.121	0.457	0.895	0.348
LRV	0.027	-0.0001	-0.586	-0.373	0.703	0.027
LDY	0.105	0.0002	0.103	0.086	0.418	0.027
LEY	0.069	0.0006	0.447	0.309	0.840	0.500
GAP	0.314	-0.1530	-2.217	-0.768	0.645	0.195

Panel C: Top 5 Models, Monthly

TRM	0	0	0	0	0
RTB	0	0	0	0	0
RBR	0	0	1	0	1
INF	0	0	0	1	0
IPG	0	0	0	0	0
LRV	0	0	0	0	0
LDY	0	0	0	1	0
LEY	0	0	0	0	0
GAP	0	1	0	0	1
$p(\mathcal{M}_j \vert y)$	0.428	0.195	0.150	0.028	0.012
$\bar{R}^2$	0.000	0.015	0.014	0.024	0.020

Panel D: Top 5 Models, Quarterly

TRM	0	0	0	0	0
RTB	0	0	0	0	0
RBR	0	0	0	0	0
INF	0	0	0	1	0
IPG	0	0	1	0	0
LRV	0	0	0	0	0
LDY	0	0	0	1	0
LEY	0	0	0	0	1
GAP	0	1	1	0	1
$p(\mathcal{M}_j \vert y)$	0.454	0.216	0.018	0.017	0.016
$\bar{R}^2$	0.000	0.037	0.047	0.046	0.046

Note: Panel A (monthly) and B (quarterly) report estimation results for the composite model. The coefficients in the weighted model are the coefficients in individual models weighted by the posterior model probabilities. Posterior probabilities of inclusion are calculated as the sum of the posterior probabilities of the models which include the respective variable. Bayesian t-ratios are reported, without and with adjustment for model uncertainty (adj). Panel C (monthly) and D (quarterly) display the five best-performing model specifications (highest posterior model probability), where 0 indicates exclusion and 1 inclusion of the respective predictive variable. Also the adjusted R^2 and the posterior model probabilities of the models which are receive the highest posterior model probability are reported. The set of predictors comprises the term spread (TRM), the short-term interest rate relative to its 12-month moving average (RTB), a long-term government bond yield relative to its 12-month moving average (RBR), annual inflation rate (INF), annual growth of industrial production (IPG), (log) realized volatility (LRV), (log) dividend yield (LDY), (log) earnings yield (LEY), output gap (GAP).

Table 6: Estimation Results, In-Sample: United States

Panel A: Composite Model, Monthly

	$\pi \vert y$	post. mean	t-ratio	t-ratio (adj)	sgn	fraction $\vert t\vert > 2$
TRM	0.014	0.0000	0.394	0.257	0.930	0.457
RTB	0.025	-0.0008	-1.403	-0.564	0.473	0.254
RBR	0.299	-0.0339	-3.077	-0.952	0.949	0.418
INF	0.103	-0.0003	-3.707	-0.853	0.820	0.402
IPG	0.063	-0.0001	-2.443	-0.695	0.414	0.477
LRV	0.015	0.0000	0.076	0.060	0.078	0.313
LDY	0.031	0.0000	-0.293	-0.226	0.898	0.008
LEY	0.131	0.0025	3.267	0.851	1.000	0.766
GAP	0.803	-0.1758	-3.868	-1.998	0.734	0.363

Panel B: Composite Model, Quarterly

	$\pi \vert y$	post. mean	t-ratio	t-ratio (adj)	sgn prob.	fraction $\vert t\vert > 2$
TRM	0.025	0.0001	1.270	0.509	0.871	0.555
RTB	0.026	0.0022	1.398	0.488	0.918	0.570
RBR	0.062	-0.0030	-0.553	-0.218	0.746	0.168
INF	0.372	-0.0012	-1.795	-0.804	0.606	0.215
IPG	0.056	0.0001	1.860	0.581	0.820	0.523
LRV	0.353	0.0014	0.898	0.526	0.434	0.172
LDY	0.065	0.0000	-0.041	-0.024	0.500	0.000
LEY	0.482	0.0138	1.753	0.913	1.000	0.531
GAP	0.932	-0.3629	-2.690	-0.883	0.633	0.441

Panel C: Top 5 Models, Monthly

TRM	0	0	0	0	0
RTB	0	0	0	0	0
RBR	0	1	0	0	1
INF	0	0	1	1	0
IPG	0	0	1	0	0
LRV	0	0	0	0	0
LDY	0	0	0	0	0
LEY	0	0	1	1	1
GAP	1	1	0	1	1
$p(\mathcal{M}_j \vert y)$	0.531	0.132	0.037	0.032	0.012
$\bar{R}^2$	0.029	0.037	0.046	0.046	0.043

Panel D: Top 5 Models, Quarterly

TRM	0	0	1	0	0
RTB	0	0	0	0	1
RBR	0	0	0	1	0
INF	1	1	1	1	1
IPG	0	0	0	0	0
LRV	1	1	1	1	1
LDY	0	1	0	0	0
LEY	1	1	1	1	1
GAP	1	1	1	1	1
$p(\mathcal{M}_j \vert y)$	0.230	0.009	0.008	0.007	0.005
$\bar{R}^2$	0.159	0.160	0.159	0.158	0.155

Note: Panel A (monthly) and B (quarterly) report estimation results for the composite model. The coefficients in the weighted model are the coefficients in individual models weighted by the posterior model probabilities. Posterior probabilities of inclusion are calculated as the sum of the posterior probabilities of the models which include the respective variable. Bayesian t-ratios are reported, without and with adjustment for model uncertainty (adj). Panel C (monthly) and D (quarterly) display the five best-performing model specifications (highest posterior model probability), where 0 indicates exclusion and 1 inclusion of the respective predictive variable. Also the adjusted R^2 and the posterior model probabilities of the models which are receive the highest posterior model probability are reported. The set of predictors comprises the term spread (TRM), the short-term interest rate relative to its 12-month moving average (RTB), a long-term government bond yield relative to its 12-month moving average (RBR), annual inflation rate (INF), annual growth of industrial production (IPG), (log) realized volatility (LRV), (log) dividend yield (LDY), (log) earnings yield (LEY), output gap (GAP).

Table 7: Estimation Results: OOS evaluation, Monthly

Panel A:	France						
	BACE-adj	BMA	TOP	ALL	AIC	BIC	R^2_{adj}
MPE	0.001	0.001	0.002	-0.004	-0.001	0.002	-0.003
	(0.308)	(0.163)	(0.419)	(-1.133)	(-0.402)	(0.431)	(-0.897)
U_1	0.993	0.987	0.996	0.965	0.980	0.996	0.976
U_2	0.995	0.995	0.998	0.997	1.000	0.998	0.998
Hit	0.572	0.572	0.565	0.572	0.557	0.572	0.579
PT	0.204	0.326	0.386	0.588	0.381	0.623	0.863
Panel B:	Germany						
	BACE-adj	BMA	TOP	ALL	AIC	BIC	R^2_{adj}
MPE	-0.001	-0.001	-0.001	-0.003	-0.002	-0.001	-0.002
	(-0.336)	(-0.273)	(-0.245)	(-0.886)	(-0.500)	(-0.251)	(-0.617)
U_1	1.007	1.009	1.012	1.009	1.023	1.009	1.014
U_2	1.003	1.003	1.009	1.021	1.021	1.009	1.022
Hit	0.546	0.530	0.542	0.574	0.506	0.542	0.550
PT	-0.002	-0.387	0.011	1.338	-0.824	0.011	0.646
Panel C:	Japan						
	BACE-adj	BMA	TOP	ALL	AIC	BIC	R^2_{adj}
MPE	-0.003	-0.002	-0.002	0.002	-0.003	-0.002	0.000
	(-0.944)	(-0.571)	(-0.603)	(0.589)	(-0.812)	(-0.608)	(0.027)
U_1	1.009	1.007	1.014	1.016	1.005	1.014	1.011
U_2	1.007	1.003	1.003	1.009	1.005	1.003	1.009
Hit	0.533	0.552	0.515	0.518	0.563	0.515	0.522
PT	0.044	0.680	-0.217	0.698	1.363	-0.217	0.506
Panel D:	UK						
	BACE-adj	BMA	TOP	ALL	AIC	BIC	R^2_{adj}
MPE	0.006	0.008	0.008	0.013	0.013	0.008	0.013
	(1.900)	(2.901)	(2.873)	(4.624)	(4.609)	(2.901)	(4.458)
U_1	1.029	1.049	1.058	1.099	1.093	1.058	1.090
U_2	1.008	1.017	1.030	1.042	1.038	1.030	1.040
Hit	0.484	0.418	0.432	0.484	0.443	0.432	0.473
PT	0.196	-0.674	-0.973	0.908	-0.091	-0.973	0.571
Panel E:	US						
	BACE-adj	BMA	TOP	ALL	AIC	BIC	R^2_{adj}
MPE	0.000	0.001	0.001	0.000	0.001	0.001	0.000
	(0.141)	(0.248)	(0.339)	(-0.191)	(0.464)	(0.404)	(0.135)
U_1	0.991	0.994	1.000	1.027	1.018	1.003	1.021
U_2	0.995	0.998	1.005	1.011	1.002	1.005	1.007
Hit	0.575	0.564	0.515	0.522	0.518	0.513	0.515
PT	1.377	1.700	0.116	0.315	0.325	0.077	0.222

Note: The table reports evaluation results of out-of-sample performance of different predictive models (monthly data). After 10 years of initialization, the models are estimated recursively. $BACE_{adj}$ uses the forecasts of the weighted model whose coefficients are adjusted for finite-sample bias. BMA is based on a pure Bayesian model averaging framework with a g-prior specification. TOP denotes the forecast by the model specification which receives the highest posterior model probability according to BMA. ALL is the all-inclusive specification. AIC, BIC, R^2_{adj} are based on the best models selected by the Akaike, Schwarz criterion or adjusted R^2, respectively. MPE denotes the mean prediction error. U_1 is the ratio of the mean absolute error of the particular model-based forecast to the one of the naive benchmark model. U_2 is the ratio of the root mean square error of the particular model-based forecast to the one of the naive benchmark model. Hit denotes the fraction of times the direction of the dependent variable is correctly predicted by the model. PT denotes the test-statistic for directional accuracy by Pesaran and Timmermann (1992).

Table 8: Estimation Results: Out-of-sample, Quarterly

Panel A:	France						
	BACE-adj	BMA	TOP	All	AIC	BIC	R^2_{adj}
MPE	0.006	0.001	0.005	-0.021	-0.009	0.003	-0.015
	(0.496)	(0.105)	(0.432)	(-1.702)	(-0.740)	(0.227)	(-1.237)
U_1	0.981	0.976	1.027	1.005	0.983	1.023	1.000
U_2	0.998	0.988	1.012	1.008	0.997	1.018	1.005
Hit	0.730	0.697	0.652	0.730	0.730	0.663	0.708
PT	0.393	0.263	0.183	1.031	1.198	0.284	0.840

Panel B:	Germany						
	BACE-adj	BMA	TOP	All	AIC	BIC	R^2_{adj}
MPE	-0.003	0.000	0.003	-0.009	-0.005	0.005	-0.006
	(-0.224)	(-0.031)	(0.217)	(-0.637)	(-0.336)	(0.335)	(-0.464)
U_1	1.019	1.023	1.050	1.038	1.042	1.053	1.026
U_2	1.007	1.009	1.029	1.022	1.044	1.030	1.010
Hit	0.568	0.568	0.531	0.617	0.605	0.506	0.617
PT	0.165	0.283	0.199	1.362	1.346	0.016	1.492

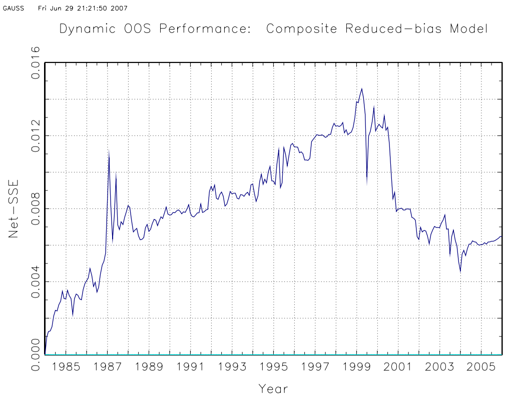
Panel C:	Japan						
	BACE-adj	BMA	TOP	All	AIC	BIC	R^2_{adj}
MPE	-0.015	-0.009	-0.008	0.008	-0.005	-0.008	0.005
	(-1.150)	(-0.682)	(-0.625)	(0.621)	(-0.375)	(-0.639)	(0.391)
U_1	1.024	1.012	1.047	1.050	1.058	1.062	1.026
U_2	1.019	1.004	1.026	1.027	1.030	1.039	1.008
Hit	0.607	0.640	0.640	0.573	0.528	0.629	0.562
PT	0.000	0.000	0.000	0.001	0.000	-0.001	0.001

Panel D:	UK						
	BACE-adj	BMA	TOP	All	AIC	BIC	R^2_{adj}
MPE	0.018	0.025	0.030	0.033	0.032	0.031	0.032
	(1.962)	(2.664)	(3.120)	(3.342)	(3.280)	(3.195)	(3.326)
U_1	1.072	1.130	1.143	1.257	1.207	1.170	1.207
U_2	1.014	1.043	1.084	1.113	1.097	1.110	1.095
Hit	0.629	0.528	0.506	0.472	0.438	0.472	0.506
PT	0.471	1.035	0.815	0.487	-0.343	0.487	0.815

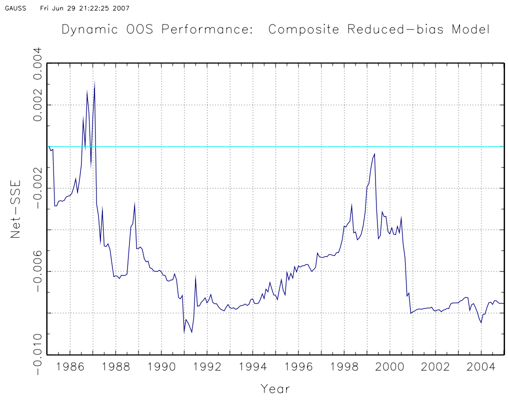
Panel E:	US						
	BACE-adj	BMA	TOP	ALL	AIC	BIC	R^2_{adj}
MPE	0.002	0.000	0.001	-0.002	0.001	0.000	-0.001
	(0.222)	(0.045)	(0.100)	(-0.279)	(0.084)	(0.057)	(-0.088)
U_1	1.036	1.010	1.029	1.053	1.021	1.026	1.034
U_2	1.022	0.999	1.013	1.022	1.011	1.014	1.010
Hit	0.640	0.620	0.593	0.580	0.587	0.580	0.587
PT	-0.096	0.544	0.524	0.231	0.612	0.323	0.331

Note: The table reports evaluation results of out-of-sample performance of different predictive models (quarterly data). After 10 years of initialization, the models are estimated recursively. $BACE_{adj}$ uses the forecasts of the weighted model whose coefficients are adjusted for finite-sample bias. BMA is based on a pure Bayesian model averaging framework with a g-prior specification. TOP denotes the forecast by the model specification which receives the highest posterior model probability according to BMA. ALL is the all-inclusive specification. AIC, BIC, R^2_{adj} are based on the best models selected by the Akaike, Schwarz criterion or adjusted R^2, respectively. MPE denotes the mean prediction error. U_1 is the ratio of the mean absolute error of the particular model-based forecast to the one of the naive benchmark model. U_2 is the ratio of the root mean square error of the particular model-based forecast to the one of the naive benchmark model. Hit denotes the fraction of times the direction of the dependent variable is correctly predicted by the model. PT denotes the test-statistic for directional accuracy by Pesaran and Timmermann (1992).

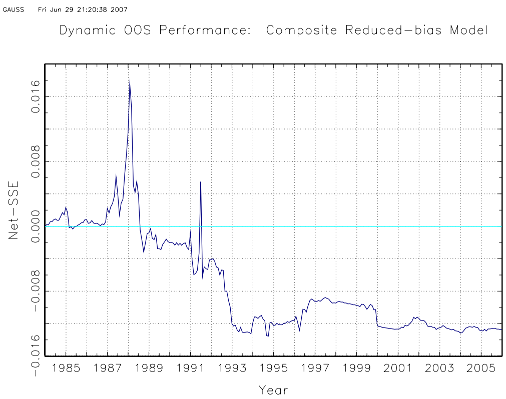
Figure 1: Net-SSE plots, Monthly



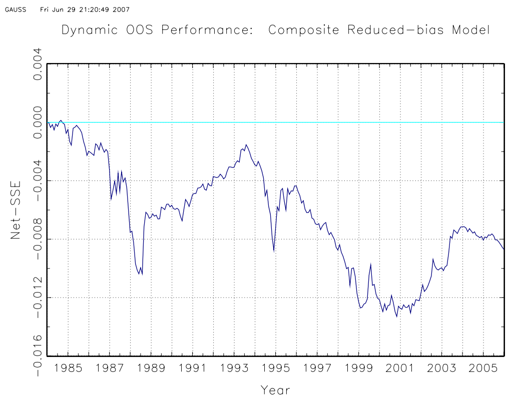
(a) France



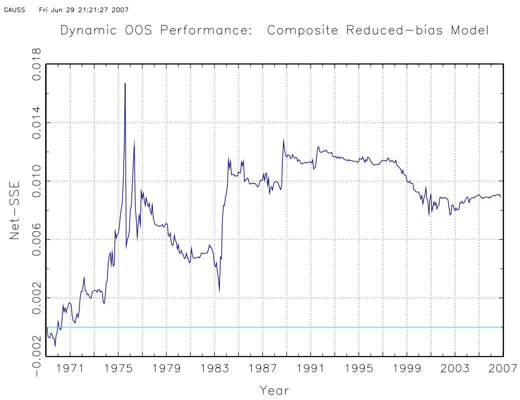
(b) Germany



(c) Japan



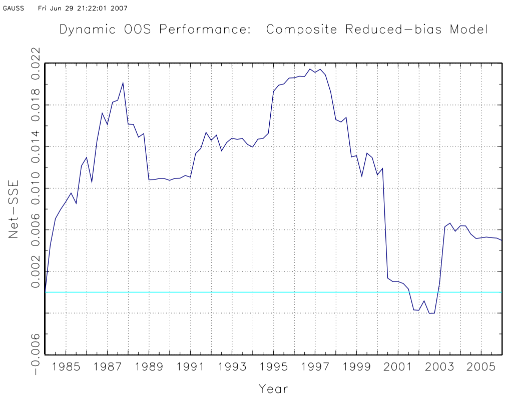
(d) United Kingdom



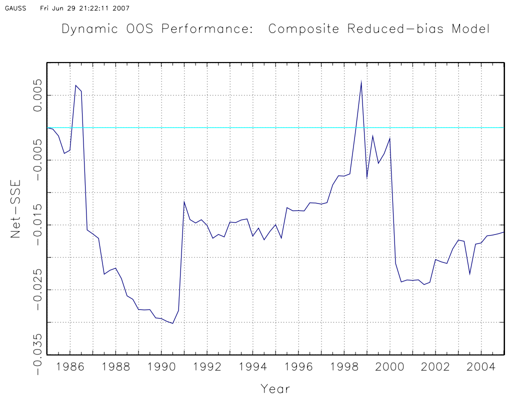
(e) United States

Note: The figure shows Net-SSE plots for the aggregate stock market following Goyal and Welch (2003). Net-SSE is the cumulated difference of squared forecast errors of the unconditional benchmark model (iid) and the conditional model (all, weighted, top): Net-SSE(τ) $= \sum_{t=1}^{\tau}(e_{uc,t}^2 - e_{c,t}^2)$, where $e_{uc,t}$ is the forecast error of the unconditional benchmark, and $e_{c,t}$ is the error of the conditional model. A decrease of the slope represents a better forecast performance of the unconditional model at the particular point in time.

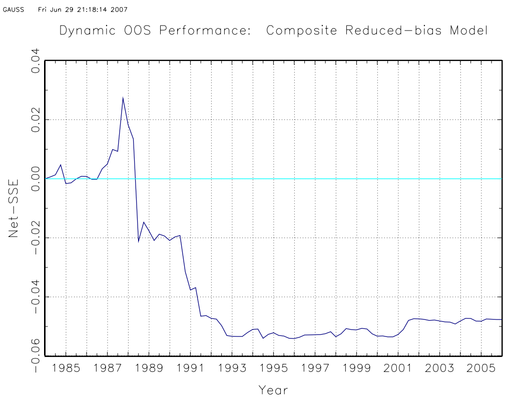
Figure 2: Net-SSE plots, Quarterly



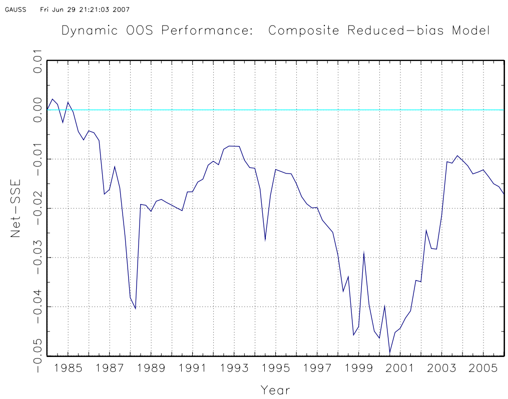
(a) France



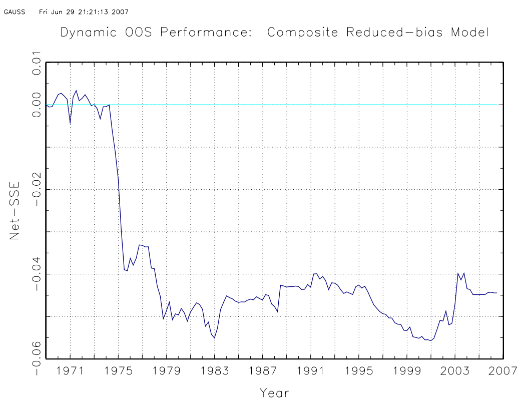
(b) Germany



(c) Japan



(d) United Kingdom



(e) United States

Note: The figure shows Net-SSE plots for the aggregate stock market following Goyal and Welch (2003). Net-SSE is the cumulated difference of squared forecast errors of the unconditional benchmark model (iid) and the conditional model (all, weighted, top): Net-SSE(τ) $= \sum_{t=1}^{\tau}(e_{uc,t}^2 - e_{c,t}^2)$, where $e_{uc,t}$ is the forecast error of the unconditional benchmark, and $e_{c,t}$ is the error of the conditional model. A decrease of the slope represents a better forecast performance of the unconditional model at the particular point in time.