

Discussion Paper No. 08-048

**International Stock Return Predictability
Under Model Uncertainty**

Andreas Schrimpf

ZEW

Zentrum für Europäische
Wirtschaftsforschung GmbH

Centre for European
Economic Research

Discussion Paper No. 08-048

International Stock Return Predictability Under Model Uncertainty

Andreas Schrimpf

Download this ZEW Discussion Paper from our ftp server:

<ftp://ftp.zew.de/pub/zew-docs/dp/dp08048.pdf>

Die Discussion Papers dienen einer möglichst schnellen Verbreitung von neueren Forschungsarbeiten des ZEW. Die Beiträge liegen in alleiniger Verantwortung der Autoren und stellen nicht notwendigerweise die Meinung des ZEW dar.

Discussion Papers are intended to make results of ZEW research promptly available to other economists in order to encourage discussion and suggestions for revisions. The authors are solely responsible for the contents which do not necessarily represent the opinion of the ZEW.

Non-technical Summary

Empirical studies have found a plethora of variables to be informative about future excess returns in predictive regressions. In particular, valuation ratios (e.g. dividend yields) and interest rate related variables (e.g. short-term interest rates as well as default and term spreads) have featured prominently in predictive regressions, but also macroeconomic variables – e.g. the consumption-wealth ratio by Lettau and Ludvigson (2001) or more recently the output gap by Cooper and Priestley (2006) – have been used to predict returns.

Given the large number of variables proposed in the literature, a high amount of uncertainty exists regarding the right choice of state variables. Moreover, the fact that so many variables have found to be valuable predictors of returns naturally raises the concern that the apparent predictability documented in the extant literature may well arise due to data-snooping rather than genuine variation of economic risk premia.

The aim of this paper is therefore 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. A novel feature of the model averaging approach used in this paper is to account for potential finite-sample bias of the coefficients in the predictive regressions. This issue has not been addressed before in work on return predictability using model averaging methods.

Drawing on an extensive international dataset covering major international stock markets, we find that interest-rate related variables are usually among the most prominent predictive variables, whereas valuation ratios generally 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 also document notable differences in the degree of in-sample and out-of-sample predictability across different international stock markets. This finding suggests that return predictability is not a uniform and a universal feature across international capital markets.

Nicht-technische Zusammenfassung (Summary in German)

In empirischen Studien wurde eine Vielzahl von Variablen identifiziert, die eine signifikante Prognosekraft für künftige Überschussrenditen auf Aktienmärkten aufweisen. Besonders häufig wurde in der Literatur die Prognosekraft von Bewertungsmaßen (z.B. Dividendenrenditen) und Zins-basierten Variablen (z.B. Default Spread und Term Spread) untersucht. Ferner wurden in vergangenen Studien auch originär makroökonomische Variablen für die Prognose von Aktienrenditen vorgeschlagen – wie beispielsweise das Verhältnis von Konsum und Vermögen von Lettau and Ludvigson (2001) oder auch der Output Gap von Cooper and Priestley (2006).

Angesichts der großen Anzahl an Variablen, die in der Literatur betrachtet wurden, besteht typischerweise eine große Unsicherheit darüber, welches die entscheidenden Variablen für Prognosezwecke sind. Ausserdem gibt die Tatsache, dass derartig viele Variablen als signifikante Prognosevariablen identifiziert wurden, Anlass zur Sorge, dass die empirisch dokumentierte Prognostizierbarkeit möglicherweise nicht auf einer tatsächlichen ökonomischen Variation von Risikoprämien, sondern im Gegensatz dazu eventuell auf “Data-Snooping” beruhen könnte.

Das Ziel der vorliegenden Studie liegt nunmehr darin, die Frage, inwieweit Aktienüberschussrenditen prognostizierbar sind, einer Robustheitsüberprüfung mit internationalen Kapitalmarktdaten zu unterziehen. Aufbauend auf Arbeiten von Avramov (2002) und Cremers (2002) wird im Rahmen der vorliegenden Studie ein Bayesianischer “Model-Averaging” (BMA) Ansatz gewählt, um der oben genannten Modellunsicherheit Rechnung zu tragen. Ein neuartiges Merkmal des BMA-Ansatzes, welcher in diesem Papier Verwendung findet, besteht darin, dass wir potentiellen Verzerrungen der Koeffizienten in kleinen Stichproben Rechnung tragen.

Unter Verwendung eines umfassenden internationalen Datensatzes zeigen die Ergebnisse, dass Zins-basierte Variablen typischerweise zu den prominentesten Prognosevariablen zu zählen sind. Dagegen schneiden Bewertungsmaße, wie zum Beispiel Dividendenrenditen, recht schlecht ab. Ferner finden wir gewisse Indizien dafür, dass Risikoprämien mit dem Output Gap variieren. Allerdings ist auch festzustellen, dass sich die Evidenz für Prognostizierbarkeit deutlich abschwächt, wenn Modellunsicherheit Rechnung getragen wird. Bemerkenswerte Unterschiede der betrachteten Aktienmärkte in Bezug auf den Grad der Prognostizierbarkeit (in-sample und out-of-sample) von Aktienrenditen werden dokumentiert. Dieses Resultat deutet darauf hin, dass empirische Renditeprognostizierbarkeit kein uniformes bzw. universelles Merkmal internationaler Kapitalmärkte ist.

International Stock Return Predictability under Model Uncertainty*

Andreas Schrimpf^{a,†}

^a *Centre for European Economic Research (ZEW, Mannheim),
P.O. Box 103443, 68034 Mannheim, Germany*

July 8, 2008

Abstract

This paper examines return predictability when the investor is uncertain about the right state variables. A novel feature of the model averaging approach used in this paper is to account for finite-sample bias of the coefficients in the predictive regressions. Drawing on an extensive international dataset, we find that interest-rate related variables are usually among the most prominent predictive variables, whereas valuation ratios perform rather poorly. Yet, predictability of market excess returns weakens substantially, once model uncertainty is accounted for. We document notable differences in the degree of in-sample and out-of-sample predictability across different stock markets. Overall, these findings suggest that return predictability is not a uniform and a universal feature across international capital markets.

JEL Classification: G12; G14; G15; E44

Keywords: Stock Return Predictability; Bayesian Model Averaging; Model Uncertainty; International Stock Markets

*This paper benefited from valuable comments and suggestions by Todd E. Clark, Joachim Grammig, Olaf Korn, Francois Laisney, Maik Schmeling, Peter Schmidt, Qingwei Wang and seminar participants at the annual conference of the EFMA (2008, Athens), Royal Economic Association (2008, Warwick), the German Finance Association doctoral seminar (2007, Dresden), University of Tübingen and ZEW Mannheim. I am also grateful to Amit Goyal and Ivo Welch for providing data on their webpages. Financial support by the Centre for European Research (ZEW, Mannheim) through the project “medium-term predictability of stock returns” is gratefully acknowledged. Nataliya Matosova and Jörg Breddermann provided able research assistance. The usual disclaimer applies.

†Corresponding Author. Tel.: +49 621 1235160; fax: +49 621 1235223.
E-mail address: schrimpf@zew.de.

1 Introduction

Empirical studies have asserted that a plethora of variables contain information about future excess returns in regressions of the form:

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

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.¹ Statistically significant β coefficients in Eq. (1) are interpreted as evidence for predictability and as evidence that risk premia are time-varying.²

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 may well arise due to data-snooping rather than genuine variation of economic risk premia.³ The aim of this paper is, therefore, to explore the robustness of several predictive variables in international stock markets in the context of model uncertainty. One of the major results of the paper is that few of the predictive variables put forth in the literature are truly robust predictors of returns. Second, substantial differences in the degree of in-sample and out-of-sample predictability can be observed across different stock markets.

In this paper, we follow the spirit of the seminal work by Cremers (2002) and Avramov

¹See e.g. Fama and French (1988), Fama and French (1989), Campbell and Shiller (1988a), Campbell and Shiller (1988b), Lettau and Ludvigson (2001) etc.

²Based on the evidence for return predictability provided by the aforementioned articles, 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).

³See 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 strong counterevidence against stock return predictability.

(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.

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. The specification of prior beliefs can be a problematic task when the set of models becomes very large.⁴ Therefore, in order to reduce the impact of subjective prior information, we base our empirical study 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 invalid in the context of predictive regressions. How to conduct reliable inference in predictive regressions while 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, 2006; 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 existing literature by conducting a comprehensive analysis of stock return predictability in major international stock markets. It is fair to

⁴Avramov (2002) addresses 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.

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, examining other important capital markets more closely may provide important additional insights, especially in a controversial field such as return predictability. Moreover, investigation of international markets also provides another 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 Kingdom, United States). Other important recent papers which provide evidence on international stock markets include Neeley and Weller (2000), Hjalmarsson (2004), Rapach et al. (2005), Paye and Timmermann (2006), Giot and Petitjean (2006) or Ang and Bekaert (2007).⁵ To the best of our knowledge, however, evidence on the effects of model uncertainty for return predictability in major international stock markets 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 the dividend yield or the earnings yield (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, 2007) 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 brief review of the literature in the previous paragraph suggests that there is not

⁵Hjalmarsson (2004) and Paye and Timmermann (2006) consider only four financial variables. Rapach et al. (2005) focus merely on macroeconomic variables and do not consider financial 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 five financial variables.

much consensus on what the important variables are, or, put differently, that there is a 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 (2006) – argue against stock return predictability or time-varying risk premia based on the lacking evidence for out-of-sample predictability.⁶ Campbell and Thompson (2007), 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 therefore is to look closer at the out-of-sample forecast performance of model averaging, in particular the time-variation of OOS performance in the spirit of Goyal and Welch (2006).

Our main results can be summarized as follows. Several notable differences with regard to return predictability are found across countries. We find that interest rate related variables are usually among the most robust predictive variables in international stock markets, which corroborates recent results by Rapach et al. (2005) and Ang and Bekaert (2007). Valuation ratios such as the dividend yield, however, perform rather poorly. There is also some evidence across countries that the output gap is related to expected returns and thus that risk premia vary with the state of the economy as pointed out recently by Cooper and Priestley (2006). The earnings yield often appears to be a more robust predictor than the dividend yield. Yet, predictability of market excess returns

⁶Cochrane (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.

clearly weakens, once model uncertainty is accounted for. We only find some evidence for out-of-sample predictability by model averaging methods in the case of France but not for the remaining stock markets. Overall, our international analysis reveals that return predictability is not a uniform and a universal feature across international capital 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), the predictive equation for future stock returns is given by

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

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 utilize κ different predictive variables which results in 2^κ different subsets, i.e. vectors of predictive variables $x_{j;t-1}$ ($j = 1, \dots, 2^\kappa$). 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:

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

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 of importance. 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 to combine the Bayesian feature of model averaging with coefficients estimated by classical OLS (BACE approach put forth by Sala-i-Martin et al. 2004).⁷ The major advantage is that the BACE approach allows us correct for finite-sample bias of predictive slope coefficients, which is an issue previously neglected in the Bayesian model averaging literature as noted for instance by Stock and Watson 2004, p.34. Moreover, the approach largely avoids the drawback of the dependence on prior distributions (See Sala-i-Martin et al. 2004).

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 $j = 1, \dots, 2^\kappa$ can be derived. These posterior model probabilities are used in the Bayesian model averaging framework as weights of the composite model:

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

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

⁷Bayesian and classical results are numerically identical when diffuse priors are specified.

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

Note that the posterior variance of the composite model in Eq. (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.⁸

For determining the weights, the marginal likelihood for the different models \mathcal{M}_j must be computed.⁹ In the pure BMA framework, analytical solutions can be found only for certain prior distribution families.¹⁰ In the “frequentist” model averaging framework of Sala-i-Martin et al. (2004), however, the marginal likelihood of a particular model is approximated using the Schwarz criterion as $\exp(-0.5BIC_j)$. The posterior model probability for \mathcal{M}_j can then be derived as

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

where $p(\mathcal{M}_j)$ denotes the probability assigned to model j a-priori. As discussed in Sala-i-Martin et al. (2004), this formula can be derived in a standard g-prior framework taking the limit as the data-information increases relative to the prior information. Thus, using posterior model probabilities as in Eq. (6) essentially implies using a prior that becomes dominated by the data.

⁸Following 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.

⁹Mathematically, the marginal likelihoods can be obtained by integrating out the parameters from the combination of the likelihood and the prior conditional on the model.

¹⁰Avramov (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).

2.2 Finite-sample Bias in Predictive Regressions

In the following we outline our approach to 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)

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

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

$$x_t = \theta + \rho x_{t-1} + \xi_t. \quad (8)$$

The errors in Eq. (7) and Eq. (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

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

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 Eq. (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 innovations in the predictive variable ξ_t and the error term ϵ_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 strongly 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 Eq. (2) and Eq. (3). In order to obtain a bias-corrected estimator for the vector of predictive coefficients β_j in Eq. (2), we use the method recently put forth by Amihud and Hurvich (2004). Their approach amounts to running an augmented regression

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

which is equivalent to running the predictive regression in Eq. (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 an unbiased 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 autoregressive model in Eq. (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.¹¹ Hence, the different series $x_{j,t}^i$ ($i = 1, \dots, k_j$) are considered separately. The individual error series are computed as $\nu_{j,t}^{c,i} = x_{j,t}^i - \hat{\theta}_i^c - \hat{\rho}_i^c x_{j,t-1}^i$. The autoregressive parameters are adjusted according to finite-sample bias by $\hat{\rho}_i^c = \hat{\rho}_i + (1+3\hat{\rho}_i)/n + 3(1+3\hat{\rho}_i)/n^2$. The reduced bias-estimator $\hat{\beta}_j^c$ is then obtained by regressing stock excess returns on the set of k_j lagged predictive and the corrected error proxies $\nu_{j,t}^i$ ($i = 1, \dots, k_j$). Standard errors for $\hat{\beta}_j^c$ 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, United Kingdom and the United States. The dependent variables are (log) returns on broad stock indices in excess of the (log) short-term interest

¹¹Allowing for a non-diagonal structure raises the need to estimate a multiplicity of parameters, in particular as k_j increases. This may result in a degradation of performance (See Amihud and Hurvich (2004)). We therefore impose a diagonal structure.

rate. Monthly summary statistics on the dependent variables and the predictive variables can be found in Table 1.

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).

The selection of variables is guided mainly by the previous US literature, as well as data availability. The main economic motivation for the different variables is that they are considered to be informative about future expected aggregate cash-flows in the economy or the discount rate applied to these cash-flows.¹² Hence, these variables have typically also featured prominently as state variables in empirical tests of intertemporal asset pricing models, e.g. Campbell (1996) or Campbell and Vuolteenaho (2004).

¹²Subsets of these variables are used for instance in Fama and Schwert (1977), Fama (1981), Fama and French (1988), Campbell and Shiller (1988a), Fama and French (1989), Fama (1990), Hodrick (1992), Avramov (2002), Cremers (2002), Lewellen (2004), Rapach et al. (2005), Cooper and Priestley (2006), Pastor and Stambaugh (2006), Ang and Bekaert (2007).

Table 1: Summary Statistics, Monthly

France: 1973:02-2005:10										
	EXRET	TRM	RTB	RBR	INF	IPG	LRV	LDY	LEY	GAP
MEAN	0.0044	1.0938	-0.0677	-0.0654	5.1294	0.9889	-6.1797	-3.3346	-2.5178	0.0456
STD	0.0621	1.2517	1.4400	0.8517	4.0892	4.4328	0.7752	0.3515	0.3275	2.8689
AC(1)	0.0798	0.9207	0.9171	0.9237	0.9966	0.8737	0.5835	0.9782	0.9673	0.8598
Germany: 1974:02-2004:12										
	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	-0.1855
STD	0.0513	1.6839	1.1858	0.6146	1.8475	4.0470	0.9695	0.3530	0.2514	2.8891
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	-0.0850	3.0833	2.1889	-6.5518	-4.5379	-3.5609	-0.0424
STD	0.0522	1.1971	1.1642	0.6554	4.6170	6.2448	1.0202	0.5050	0.4687	4.1581
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	-0.0258	-0.0796	6.6242	1.1718	-6.4382	-3.1629	-2.5142	0.0887
STD	0.0566	2.1353	1.5307	0.8933	5.2741	4.0886	0.8087	0.2748	0.3977	2.7035
AC(1)	0.1092	0.9774	0.9268	0.9089	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	0.0126	4.0387	3.0665	-6.6934	-3.5011	-2.7798	-0.1089
STD	0.0423	1.4360	1.1069	0.6257	2.7614	4.8664	0.8540	0.3962	0.3923	3.0940
AC(1)	0.0282	0.9493	0.9000	0.8764	0.9936	0.9609	0.8188	0.9920	0.9926	0.9637

The table reports summary statistics of (log) market excess returns (EXRET) and predictive variables in five international stock markets. MEAN, STD, AC(1) denote the mean, standard deviation and first-order autocorrelation coefficient respectively. 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).

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 already starts in the late 1950s. Unfortunately, a default 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 markets outside the US in a reasonable quality. For further detailed information on data sources and construction the reader is referred to Appendix A.

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

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

First, we 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. We therefore set the a-priori expected model size to $\bar{k} = 2$ variables.¹³ 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 $p(\mathcal{M}_j) = \pi^{k_j}(1 - \pi)^{\kappa - k_j}$.¹⁴ It is important to 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

¹³We discuss the sensitivity of the results to this choice of hyperparameter in section 3.3.

¹⁴In principle, one could also specify different prior probabilities of inclusion for the different variables based on economic considerations.

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; it is computed as $\mathcal{C}'\mathcal{P}$, 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 based on Eq. (4) 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 predictive 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 $j = 1, \dots, 2^\kappa$ models) has the same sign as the posterior mean in the composite model (denoted as *sgn prob.* in the tables). Furthermore, we also report the fraction of cases across the different models when a classical t-statistic for the particular variable is greater than two in absolute value. This statistic serves as another indicator of the robustness or fragility of a particular predictive variable (Sala-i-Martin et al., 2004). Panels C and D, presents the five top-performing model specifications which receive the highest posterior probability of all models. 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 of the five top models are also 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

retract 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. A negative relation of the realtive bond rate and expected excess returns is reasonable from an economic point of view, given that higher yields on long-term bonds are typically reflected in a higher level of corporate loan rates and thus may have a negative impact on subsequent real activity. 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 which measures the fraction of cases where the coefficient on the particular variable (when included in one of the 2^k Models) has the same sign as its coefficient in the weighted model. According to this criterion, the relative bond rate is again 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.¹⁵

Panels B and D show that the evidence for predictability in the French stock market is somewhat weaker in the quarterly case. Again, only the relative bond rate receives a posterior probability of inclusion larger than $0.\bar{2}$. It is also worth noting that the earnings yield performs relatively well in-terms of sign certainty in the quarterly case.

¹⁵This 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 many variables which appear to be significant, lose their significance once model uncertainty is considered.

Table 2: Estimation Results, In-Sample: France

Panel A: Composite Model, Monthly				Panel B: Composite Model, Quarterly							
	πy	post. mean	t-ratio (adj)	sgn prob.	fraction $ t > 2$		πy	post. mean	t-ratio (adj)	sgn prob.	fraction $ t > 2$
TRM	0.033	0.010	1.205	0.557	0.969	0.027	TRM	0.036	1.008	0.511	0.004
RTB	0.042	-0.016	-1.535	-0.572	0.473	0.016	RTB	0.064	-1.462	-0.601	0.023
RBR	0.743	-0.958	-3.494	-1.837	1.000	0.340	RBR	0.641	-2.400	-1.461	0.414
INF	0.017	-0.001	-0.404	-0.272	0.180	0.117	INF	0.021	0.002	0.179	0.453
IPG	0.017	0.001	0.880	0.424	0.945	0.246	IPG	0.030	-0.002	-0.127	0.238
LRV	0.042	0.022	1.311	0.581	0.441	0.000	LRV	0.031	0.024	0.598	0.000
LDY	0.003	0.001	0.245	0.197	0.434	0.000	LDY	0.015	0.034	0.772	0.000
LEY	0.003	0.001	0.190	0.159	0.742	0.000	LEY	0.018	0.051	0.890	0.000
GAP	0.116	-0.031	-2.347	-0.764	0.949	0.172	GAP	0.129	-0.104	-0.732	0.027

Panel C: Top 5 Models, Monthly				Panel D: Top 5 Models, Quarterly							
	πy	post. mean	t-ratio (adj)	sgn prob.	fraction $ t > 2$		πy	post. mean	t-ratio (adj)	sgn prob.	fraction $ t > 2$
TRM	0	0	0	0	0	TRM	0	0	0	0	0
RTB	0	0	0	0	1	RTB	0	0	0	1	0
RBR	1	0	0	1	0	RBR	1	0	0	0	1
INF	0	0	0	0	0	INF	0	0	0	0	0
IPG	0	0	0	0	0	IPG	0	0	0	0	0
LRV	0	0	0	1	0	LRV	0	0	0	0	0
LDY	0	0	0	0	0	LDY	0	0	0	0	0
LEY	0	0	0	0	0	LEY	0	0	0	0	0
GAP	0	0	1	0	0	GAP	0	0	1	0	1
Prob	0.635	0.116	0.079	0.031	0.028	Prob	0.527	0.162	0.085	0.040	0.025
\bar{R}^2	0.027	0.000	0.017	0.032	0.012	\bar{R}^2	0.065	0.000	0.038	0.027	0.068

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 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).

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 fairly weak on statistical grounds. The case for predictability is clearly less pronounced than in the French stock market discussed in the previous subsection. The model receiving the highest posterior probability is the one without any lagged state variables (i.i.d. case). None of the variables in the monthly model receives a higher posterior inclusion probability compared to the prior inclusion probability of $\pi = 0.2$. Among the variables considered only the relative bond rate (RBR) and the output gap (GAP) may 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.

Similar to the French case, the relative bond rate is rather important in the quarterly regressions (Panel B of Table 3) 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 quite robust with regard to sign certainty: The term spread (TRM), the relative bond rate (RBR), industrial production growth (IPG), and the two valuation ratios (LDY, LEY) have the same sign as the posterior mean in the composite model in more than 90% of all models in which they are included.

Table 3: Estimation Results, In-Sample: Germany.

Panel A: Composite Model, Monthly				Panel B: Composite Model, Quarterly								
	πy	post. mean	t-ratio (adj)	sgn prob.	fraction $ t > 2$	πy	post. mean	t-ratio (adj)	sgn prob.	fraction $ t > 2$		
TRM	0.029	0.006	1.198	0.554	0.934	0.469	0.038	0.021	0.980	0.504	0.922	0.418
RTB	0.025	-0.006	-1.064	-0.520	0.766	0.039	0.043	-0.032	-0.905	-0.473	0.777	0.066
RBR	0.188	-0.196	-2.418	-0.816	1.000	0.285	0.467	-1.896	-2.683	-1.103	1.000	0.527
INF	0.015	-0.001	-0.342	-0.254	0.488	0.000	0.026	-0.004	-0.272	-0.212	0.504	0.000
IPG	0.017	-0.000	-0.241	-0.152	0.066	0.273	0.031	0.001	0.172	0.105	0.949	0.297
LRV	0.021	0.004	0.761	0.435	0.160	0.430	0.039	0.033	0.836	0.462	0.250	0.000
LDY	0.005	-0.003	-0.855	-0.462	1.000	0.004	0.007	-0.006	-0.292	-0.226	0.938	0.000
LEY	0.013	0.000	0.032	0.031	0.609	0.000	0.027	0.058	0.576	0.369	0.965	0.000
GAP	0.089	-0.016	-1.981	-0.707	0.816	0.133	0.144	-0.093	-1.995	-0.737	0.676	0.078

Panel C: Top 5 Models, Monthly				Panel D: Top 5 Models, Quarterly							
	πy	post. mean	t-ratio (adj)	sgn prob.	fraction $ t > 2$		πy	post. mean	t-ratio (adj)	sgn prob.	fraction $ t > 2$
TRM	0	0	0	1	0	TRM	0	0	0	0	0
RTB	0	0	0	0	1	RTB	0	0	0	0	1
RBR	0	1	0	0	0	RBR	1	0	0	1	0
INF	0	0	0	0	0	INF	0	0	0	0	0
IPG	0	0	0	0	0	IPG	0	0	0	0	0
LRV	0	0	0	0	0	LRV	0	0	0	0	0
LDY	0	0	0	0	0	LDY	0	0	0	0	0
LEY	0	0	0	0	0	LEY	0	0	0	0	0
GAP	0	0	1	0	0	GAP	0	0	1	1	0
Prob	0.635	0.163	0.075	0.022	0.020	Prob	0.364	0.322	0.093	0.026	0.024
\bar{R}^2	0.000	0.013	0.008	0.002	0.001	\bar{R}^2	0.052	0.000	0.031	0.060	0.009

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 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).

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 (i.i.d.-model). The output gap (GAP) and the relative bond rate (RBR) are somewhat marginally important, but their explanatory power is fairly low. Note also that industrial production growth (IPG) and inflation (INF) are quite robust in terms of sign certainty probability.

With quarterly data, the evidence for predictability is even more modest. Again the model which does not include any predictors receives the highest probability a-posteriori. Only the output gap receives a higher posterior probability of inclusion than expected a-priori (Panel D of Table 4). However, model uncertainty again plays a substantial role as evinced by the adjusted Bayesian t-ratios. It is also worth noting that according to the sign certainty measure, the output gap must be considered as a rather fragile predictor.

3.2.4 *United Kingdom*

Table 5 reveals, that both for monthly and quarterly predictive regressions, the case for return predictability in the United Kingdom is quite weak. Panel C shows, that the largest posterior probability in the monthly regressions is assigned to the i.i.d.-model (as in the case of monthly regressions for Germany and Japan). Contrary to the countries discussed so far, interest rate variables do not show up among the most prominent predictors, which confirms the recent findings by Giot and Petitjean (2006) based on univariate return prediction models. By contrast, the dividend yield (LDY) has some predictive content for future stock returns in the UK. Yet, as before, accounting for model uncertainty greatly reduces the evidence for predictability and explanatory power of return prediction models in the UK is rather low.

Table 4: Estimation Results, In-Sample: Japan

Panel A: Composite Model, Monthly				Panel B: Composite Model, Quarterly								
	πy	post. mean	t-ratio (adj)	sgn prob.	fraction $ t > 2$		πy	post. mean	t-ratio (adj)	sgn prob.	fraction $ t > 2$	
TRM	0.015	0.002	0.450	0.303	0.809	0.309	TRM	0.023	-0.003	-0.166	0.434	0.066
RTB	0.055	-0.022	-1.694	-0.648	0.574	0.105	RTB	0.035	-0.021	-0.696	0.582	0.102
RBR	0.192	-0.194	-2.479	-0.826	0.824	0.051	RBR	0.048	-0.083	-1.255	0.664	0.000
INF	0.030	-0.002	-1.391	-0.592	1.000	0.371	INF	0.026	-0.003	-0.521	0.883	0.031
IPG	0.023	0.001	0.944	0.424	0.910	0.336	IPG	0.041	0.007	0.941	0.405	0.348
LRV	0.022	-0.006	-1.017	-0.510	0.375	0.000	LRV	0.025	-0.013	-0.527	0.703	0.027
LDY	0.003	-0.001	-0.584	-0.365	0.793	0.230	LDY	0.017	0.003	0.080	0.071	0.418
LEY	0.005	-0.001	-0.296	-0.185	0.238	0.480	LEY	0.020	0.012	0.307	0.235	0.500
GAP	0.256	-0.043	-2.603	-0.887	0.492	0.219	GAP	0.309	-0.182	-2.448	0.645	0.195

Panel C: Top 5 Models, Monthly				Panel D: Top 5 Models, Quarterly							
	πy	post. mean	t-ratio (adj)	sgn prob.	fraction $ t > 2$		πy	post. mean	t-ratio (adj)	sgn prob.	fraction $ t > 2$
TRM	0	0	0	0	0	TRM	0	0	0	0	0
RTB	0	0	0	1	0	RTB	0	0	0	1	0
RBR	0	0	1	0	0	RBR	0	0	1	0	0
INF	0	0	0	0	1	INF	0	0	0	0	0
IPG	0	0	0	0	0	IPG	0	0	0	0	1
LRV	0	0	0	0	0	LRV	0	0	0	0	0
LDY	0	0	0	0	0	LDY	0	0	0	0	0
LEY	0	0	0	0	0	LEY	0	0	0	0	0
GAP	0	1	0	0	0	GAP	0	1	0	0	1
Prob	0.467	0.213	0.163	0.045	0.017	Prob	0.531	0.253	0.034	0.024	0.019
\bar{R}^2	0.000	0.015	0.014	0.007	0.002	\bar{R}^2	0.000	0.037	0.007	0.001	0.047

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 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												
	πy	post. mean	t-ratio (adj)	t-ratio (adj)	sgn prob.	fraction $ t > 2$	πy	post. mean	t-ratio (adj)	t-ratio (adj)	sgn prob.	fraction $ t > 2$
TRM	0.017	0.001	0.660	0.399	0.902	0.156	TRM	0.029	0.004	0.558	0.302	0.762
RTB	0.014	-0.000	-0.138	-0.089	0.113	0.152	RTB	0.018	-0.001	-0.095	-0.049	0.098
RBR	0.026	-0.010	-1.345	-0.580	0.781	0.148	RBR	0.026	-0.011	-0.708	-0.405	0.785
INF	0.016	0.000	0.509	0.317	0.594	0.008	INF	0.023	0.002	0.817	0.426	0.805
IPG	0.021	0.002	1.219	0.556	0.926	0.398	IPG	0.023	0.002	0.771	0.387	0.875
LRV	0.021	0.006	0.883	0.451	0.250	0.012	LRV	0.033	-0.019	-1.056	-0.520	0.922
LDY	0.203	0.363	1.725	0.724	0.633	0.020	LDY	0.559	3.974	2.169	1.101	0.992
LEY	0.043	0.030	0.993	0.509	1.000	0.480	LEY	0.083	0.269	1.429	0.619	1.000
GAP	0.073	-0.016	-2.038	-0.701	0.316	0.313	GAP	0.041	-0.023	-1.747	-0.621	0.320
Panel B: Composite Model, Quarterly												
Panel C: Top 5 Models, Monthly												
TRM	0	0	0	0	0		TRM	0	0	0	0	0
RTB	0	0	0	0	0		RTB	0	0	0	0	0
RBR	0	0	0	0	1		RBR	0	0	0	0	0
INF	0	0	0	0	0		INF	0	0	0	0	0
IPG	0	0	0	0	0		IPG	0	0	0	0	0
LRV	0	0	0	0	0		LRV	0	0	0	0	1
LDY	0	1	0	0	0		LDY	1	0	0	0	1
LEY	0	0	0	1	0		LEY	0	0	1	0	0
GAP	0	0	1	0	0		GAP	0	0	0	1	0
Prob	0.604	0.185	0.060	0.039	0.019		Prob	0.481	0.268	0.070	0.032	0.019
\bar{R}^2	0.000	0.015	0.007	0.008	0.002		\bar{R}^2	0.062	0.000	0.036	0.016	0.063
Panel D: Top 5 Models, Quarterly												

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 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).

3.2.5 *United States*

As shown by Table 6, evidence for in-sample return predictability is clearly stronger in the US compared to other international stock markets such as Germany, Japan or the UK. Variables which appear important after having seen the data include the relative bond rate (RBR) and, most notably, the output gap (GAP). The output gap is the only variable which can be considered as a significant predictor 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.¹⁶ The output gap also appears to be a less fragile predictor in the US compared to the other countries. It is also worth noting that the earnings yield (LEY) provides more explanatory power than the dividend yield (LDY). Several other variables – such as the relative bond rate (RBR), inflation (INF), and industrial production growth (IPG) – are important when model uncertainty is ignored, but lose their significance once model uncertainty is considered.

When we consider predictive models at a quarterly horizon, the output gap (GAP) again appears as an important variable a-posteriori and also survives the model uncertainty adjustment. 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.

¹⁶Thus, our results corroborate the results of the recent paper by Cooper and Priestley (2006) who find that risk-premia are varying with the output-gap. Good economic conditions as measured by the output gap are associated with low risk premia.

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

Panel A: Composite Model, Monthly				Panel B: Composite Model, Quarterly			
	πy	post. mean	t-ratio (adj)	sgn prob.	fraction $ t > 2$		
TRM	0.014	0.001	0.387	0.254	0.789	0.059	TRM
RTB	0.025	-0.007	-1.450	-0.578	0.898	0.168	RTB
RBR	0.302	-0.278	-3.127	-0.982	1.000	0.809	RBR
INF	0.086	-0.031	-3.787	-0.846	0.918	0.457	INF
IPG	0.059	-0.006	-2.443	-0.701	0.852	0.254	IPG
LRV	0.010	0.000	0.017	0.014	0.117	0.000	LRV
LDY	0.030	0.011	0.695	0.286	0.527	0.172	LDY
LEY	0.117	0.228	3.211	0.838	0.996	0.594	LEY
GAP	0.805	-0.181	-3.904	-2.179	1.000	0.621	GAP

Panel C: Top 5 Models, Monthly				Panel D: Top 5 Models, Quarterly			
	πy	post. mean	t-ratio (adj)	sgn prob.	fraction $ t > 2$		
TRM	0	0	0	0	0	0	TRM
RTB	0	0	0	0	0	0	RTB
RBR	0	1	1	0	0	0	RBR
INF	0	0	0	1	1	0	INF
IPG	0	0	0	1	0	0	IPG
LRV	0	0	0	0	0	0	LRV
LDY	0	0	0	0	0	0	LDY
LEY	0	0	0	1	1	0	LEY
GAP	1	1	0	0	1	0	GAP
Prob	0.544	0.136	0.103	0.034	0.025		Prob
\bar{R}^2	0.029	0.037	0.023	0.046	0.046		\bar{R}^2

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 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).

3.3 Sensitivity to the Choice of Hyperparameter

The previous discussion of in-sample predictability and differences in the relevance of particular predictors across countries was based on a fairly moderate expected model size of two variables. In this sub-section, we analyze the robustness of our main findings to the specific choice of this hyperparameter \bar{k} which is linked to the prior probability of inclusion π . For this purpose, we check whether our earlier conclusions on the relevance of a particular variable – as measured by a posterior probability of inclusion $\pi|y$ exceeding the prior probability of inclusion π – are affected by the choice of the expected model size. Table 7 reports posterior probabilities of inclusion of the predictor variables for different prior probabilities of inclusion π corresponding to model sizes with $\bar{k} = 2, 4, 6$ and 8 variables.

As shown in Table 7, our main conclusions on the relevance of a specific predictor are largely unaffected by the choice of the expected model size. Panel A for France, for instance shows that the relative bond rate can be considered an important predictor for almost all choices of prior probabilities of inclusion. There is not a single case where another predictive variable becomes relevant for a different choice of π , i.e. that there is an upward revision of the probability of inclusion after having seen the data. The same result holds true for the German (Panel B) and the British (Panel D) stock market. Results on the usefulness of the output gap in Japan are only slightly dependent on the choice of π but no other variable shows an upward revision of the probability of inclusion for different choices of hyperparameters. Results for quarterly predictive regressions for the US stock market (Panel E) are also largely unaffected. In the monthly case, however, the earnings yield and the inflation rate play a more prominent role in larger models, while the relative bond rate only serves as a significant predictor in the case of small expected model sizes.

Table 7: Sensitivity, Hyperparameter Expected Model Size

π	Monthly				Quarterly			
	0.222	0.444	0.667	0.889	0.222	0.444	0.667	0.889
Panel A: France								
TRM	0.033	0.080	0.167	0.376	0.036	0.085	0.167	0.362
RTB	0.042	0.066	0.112	0.282	0.064	0.103	0.170	0.359
RBR	0.743	0.819	0.857	0.908	0.641	0.745	0.805	0.878
INF	0.017	0.043	0.098	0.299	0.021	0.054	0.122	0.348
IPG	0.017	0.049	0.125	0.422	0.030	0.069	0.145	0.383
LRV	0.042	0.113	0.250	0.597	0.031	0.082	0.182	0.475
LDY	0.003	0.004	0.005	0.006	0.015	0.020	0.027	0.055
LEY	0.003	0.004	0.006	0.008	0.018	0.033	0.056	0.107
GAP	0.116	0.174	0.280	0.592	0.129	0.194	0.291	0.538
Panel B: Germany								
TRM	0.029	0.070	0.139	0.324	0.038	0.090	0.175	0.362
RTB	0.025	0.057	0.109	0.283	0.043	0.087	0.164	0.380
RBR	0.188	0.358	0.529	0.763	0.467	0.649	0.770	0.902
INF	0.015	0.039	0.088	0.241	0.026	0.069	0.155	0.419
IPG	0.017	0.047	0.110	0.344	0.031	0.085	0.204	0.552
LRV	0.021	0.054	0.115	0.271	0.039	0.098	0.198	0.418
LDY	0.005	0.010	0.017	0.033	0.007	0.011	0.014	0.024
LEY	0.013	0.033	0.069	0.171	0.027	0.063	0.124	0.275
GAP	0.089	0.182	0.300	0.553	0.144	0.255	0.413	0.731
Panel C: Japan								
TRM	0.015	0.039	0.091	0.287	0.023	0.062	0.142	0.395
RTB	0.055	0.093	0.142	0.308	0.035	0.080	0.157	0.390
RBR	0.192	0.307	0.410	0.616	0.048	0.102	0.189	0.452
INF	0.030	0.071	0.147	0.391	0.026	0.067	0.145	0.372
IPG	0.023	0.072	0.174	0.473	0.041	0.125	0.287	0.651
LRV	0.022	0.062	0.145	0.399	0.025	0.065	0.134	0.295
LDY	0.003	0.004	0.006	0.008	0.017	0.029	0.039	0.071
LEY	0.005	0.008	0.011	0.009	0.020	0.032	0.036	0.044
GAP	0.256	0.419	0.572	0.804	0.309	0.535	0.718	0.890
Panel D: UK								
TRM	0.017	0.043	0.094	0.266	0.029	0.071	0.155	0.460
RTB	0.014	0.035	0.085	0.298	0.018	0.043	0.102	0.402
RBR	0.026	0.061	0.126	0.346	0.026	0.066	0.152	0.476
INF	0.016	0.041	0.094	0.300	0.023	0.052	0.112	0.356
IPG	0.021	0.071	0.184	0.506	0.023	0.061	0.139	0.401
LRV	0.021	0.052	0.111	0.310	0.033	0.091	0.199	0.470
LDY	0.203	0.347	0.438	0.453	0.559	0.704	0.782	0.886
LEY	0.043	0.073	0.090	0.098	0.083	0.109	0.131	0.222
GAP	0.073	0.149	0.255	0.488	0.041	0.059	0.086	0.236
Panel E: US								
TRM	0.014	0.036	0.081	0.238	0.022	0.058	0.132	0.336
RTB	0.025	0.048	0.098	0.260	0.025	0.059	0.129	0.330
RBR	0.302	0.399	0.438	0.567	0.060	0.133	0.264	0.518
INF	0.086	0.349	0.684	0.928	0.011	0.037	0.123	0.429
IPG	0.059	0.208	0.389	0.570	0.028	0.061	0.121	0.277
LRV	0.010	0.028	0.059	0.113	0.067	0.171	0.364	0.750
LDY	0.030	0.088	0.207	0.501	0.001	0.001	0.002	0.011
LEY	0.117	0.415	0.754	0.957	0.006	0.019	0.069	0.230
GAP	0.805	0.734	0.663	0.722	0.938	0.956	0.964	0.982

The table contains detailed results on the sensitivity of estimation results with respect to the choice of the expected model size. For different prior probabilities of inclusion π corresponding to model sizes with 2, 4, 6 and 8 variables the posterior probabilities of inclusion are reported. The predictors include 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).

3.4 *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 and 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 risk premia are truly time-varying, the results of Goyal and Welch (2006) will occur 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. Another reason for poor OOS predictability may be temporal instability of the return prediction models.¹⁸ We address the latter issue by studying the time-variation of OOS forecast errors in international stock markets using Net-SSE plots in the spirit of Goyal and Welch (2006).

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 model averaging in the context of OOS predictability of excess returns. Avramov (2002), for instance, 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 the context of major international stock markets.

For the purpose of evaluating OOS forecast performance, we estimate the 2^k 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 future excess return. The model-based forecasts include Bayesian averaging of OLS coefficients adjusted for finite-sample

¹⁷The 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 (2007).

¹⁸See also the recent papers by Paye and Timmermann (2006), Dangl et al. (2006) and Ravazzolo et al. (2006).

bias (BACE-adj), a conventional Bayesian model averaging approach (BMA) with g-prior specification¹⁹, 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.

Table 8 reports the results of the evaluation of OOS performance for our international set of stock markets. The evaluation of forecast accuracy uses standard criteria. ME denotes the mean prediction error. Testing the significance of the ME amounts to testing the unbiasedness of the forecasts. Theil's U (TU) is the ratio of the mean square prediction error (MSPE) of the particular model-based forecast to the one of the naive benchmark model.²⁰ In order to provide an evaluation of directional accuracy of forecasts obtained by model averaging, we also 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 are depicted in Figure 1. 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. One can use these plots to infer how the OOS performance of the predictive model evolves over time and where major forecast breakdowns occur. Periods where the line in the graph is upward sloping represent times when the conditional model outperforms the naive model in terms of squared forecast errors.

¹⁹The 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 or Koop 2003).

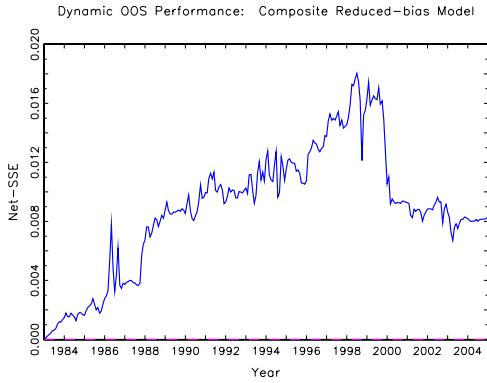
²⁰Note that TU is merely a descriptive criterion. In the case of nested models, the mean square prediction error MSPE of the smaller nested model is expected to be smaller than the MSPE under the null of equal predictive power, a point raised by Clark and West (2007). This is due to the fact that the larger model needs to estimate parameters which are zero in population, which introduces noise in the forecasts.

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

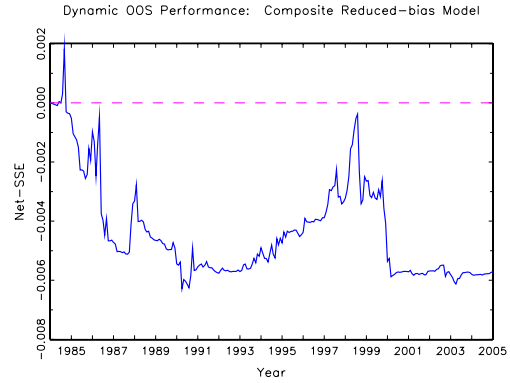
Panel A: France		BACE-adj	BMA	TOP	All	AIC	BIC	\bar{R}^2
ME		0.0010	0.0014	0.0015	-0.0050	-0.0012	0.0015	-0.0035
t-stat		0.2845	0.4114	0.4190	-1.4247	-0.3421	0.4314	-1.0058
TU		0.9947	0.9959	0.9993	1.0025	1.0019	0.9986	1.0035
Hit		0.5978	0.5941	0.5646	0.5830	0.5535	0.5720	0.5720
PT		1.0524	0.9375	0.3858	0.9105	0.3264	0.6230	0.6579
Panel B: Germany		BACE-adj	BMA	TOP	All	AIC	BIC	\bar{R}^2
ME		-0.0003	-0.0003	-0.0003	-0.0040	-0.0017	-0.0003	-0.0039
t-stat		-0.0693	-0.0837	-0.0883	-1.0591	-0.4659	-0.0924	-1.0324
TU		1.0034	1.0038	1.0087	1.0278	1.0221	1.0085	1.0309
Hit		0.5221	0.5181	0.5100	0.5542	0.4940	0.5100	0.5221
PT		-0.6551	-0.7653	-0.6806	0.8108	-1.0818	-0.6806	-0.2041
Panel C: Japan		BACE-adj	BMA	TOP	All	AIC	BIC	\bar{R}^2
ME		-0.0010	-0.0013	-0.0021	-0.0011	-0.0029	-0.0021	-0.0011
t-stat		-0.2827	-0.3769	-0.5894	-0.3138	-0.8317	-0.6030	-0.3255
TU		1.0034	1.0047	1.0034	1.0095	1.0095	1.0038	1.0054
Hit		0.5257	0.5257	0.5037	0.4853	0.5037	0.5037	0.5257
PT		-0.1541	-0.1541	-0.5811	-0.9397	-0.4146	-0.5811	0.4359
Panel D: UK		BACE-adj	BMA	TOP	All	AIC	BIC	\bar{R}^2
ME		0.0023	0.0047	0.0083	0.0112	0.0104	0.0083	0.0114
t-stat		0.8009	1.6394	2.8486	3.8104	3.5598	2.8549	3.8779
TU		1.0032	1.0093	1.0287	1.0517	1.0390	1.0289	1.0495
Hit		0.5678	0.4396	0.4322	0.4542	0.4359	0.4322	0.4322
PT		0.0810	-2.0820	-0.9730	0.0165	-0.5822	-0.9730	-0.5855
Panel E: US		BACE-adj	BMA	TOP	All	AIC	BIC	\bar{R}^2
ME		-0.0005	-0.0005	0.0007	-0.0009	0.0006	0.0007	-0.0003
t-stat		-0.2439	-0.2472	0.3159	-0.4444	0.2806	0.3149	-0.1522
TU		1.0010	1.0009	1.0129	1.0118	1.0065	1.0115	1.0117
Hit		0.5507	0.5485	0.5088	0.5220	0.4934	0.5066	0.5132
PT		0.6817	0.6526	-0.2493	0.0433	-0.8345	-0.3297	-0.1500

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, \bar{R}^2 are based on the best models selected by the Akaike, Schwarz criterion or adjusted R^2 , respectively. ME denotes the mean prediction error (t-statistic reported below). TU 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).

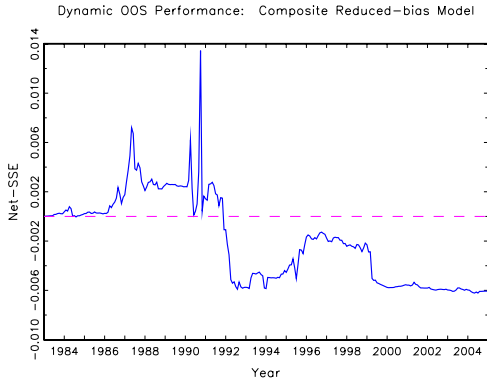
Figure 1: Net-SSE Plots, Monthly



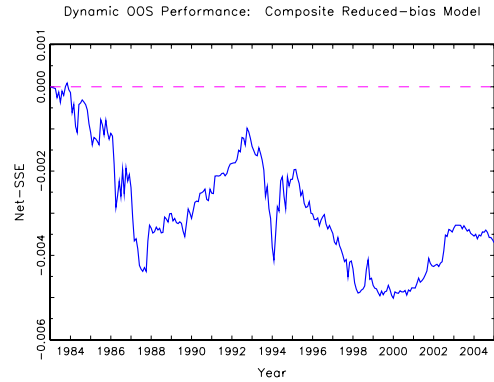
(a) France



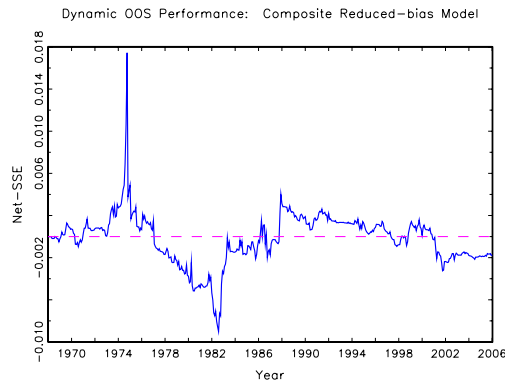
(b) Germany



(c) Japan



(d) United Kingdom



(e) United States

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 (i.i.d. model) and the conditional model (BACE-adj): $\text{Net-SSE}(\tau) = \sum_{t=1}^{\tau} (e_{uc,t}^2 - e_{c,t}^2)$, 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.

As the evaluation of the monthly forecasts in Table 8 shows, out-of-sample predictability of monthly stock returns is generally very limited. Moreover, notable differences of OOS return predictability can be detected across countries. Table 8 also shows that the BACE approach with bias adjustment generally compares rather favorably in terms of forecast accuracy compared to conventional Bayesian model averaging for most stock markets.

The results for the French stock market, presented in Panel A of Table 8, 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 model averaging approaches (BACE-adj, BMA) typically outperform the naive model and model selection criteria in terms of MSPE, i.e. have a Theil's U (TU) smaller than one. All model-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.²¹ As shown by the graph, the model has produced lower squared forecast errors relative to the benchmark up to about 2000. In the aftermath of the climax of the internet boom no outperformance relative to the naive benchmark can be detected anymore.

In the case of Germany (Panel B of Table 8), BACE-adj and BMA generally do a better job compared to other model specifications, but are not able to outperform the i.i.d. model in terms of MSPE. This is consistent with the modest results for in-sample predictability in Table 3, where little evidence for return predictability was detected at a monthly horizon. The Net-SSE plot (b) in Figure 1 shows that OOS predictability has been clearly stronger in the 1990s, where lagged state variables contributed to lower squared prediction errors relative to the benchmark. Also note that, similar to the French case, return prediction models did not provide better forecast accuracy than the benchmark since the height of the new economy boom until the end of the sample.

For the Japanese stock market the case for OOS predictability is also fairly weak, as Panel C of Table 8 reveals: forecasts of the naive model generally produce a lower MSPE than models conditioning on predictive variables. This is confirmed by the Net-SSE plot (c) of Figure 1. The plot shows a decline of OOS forecast performance of the weighted

²¹Net-SSE plots based on the BMA approach are generally quite similar.

model forecast from the early 1990s onwards. Analogously to Germany and Japan, OOS predictability in the United Kingdom (Panel D of Table 8) is very poor. Moreover, the United Kingdom is the only stock market where conditional models produce forecasts with a substantial bias (however less pronounced when model averaging techniques are used). Also note that the model averaging methods (BACE-adj and BMA) again outperform the other selection criteria but fail to outperform the naive model in terms of mean-square prediction error.

Evaluation results for the US stock market are given in Panel E of Table 8. Contrary to the in-sample regressions, out-of-sample predictability of US excess returns is rather poor. Hence, our OOS results are more in line with Goyal and Welch (2006) than Avramov (2002). The Net-SSE plot for the United States in (e) of Figure 1 illustrates the time-variation in the degree of OOS predictability. In particular, a steady decline of predictability since the late 1980s can be recognized. This is consistent with other studies for the US documenting poor return predictability over the 1990s (e.g. Paye and Timmermann, 2006; Ang and Bekaert, 2007).

Results for quarterly market excess returns are quite similar to the monthly case and are therefore provided in the Appendix B. We do not find much evidence that OOS predictability increases with the horizon of the forecast. Quite to the contrary, OOS predictability is somewhat weaker than the OOS predictability in the monthly case (e.g. for the US). Again, France is the only stock market where out-of-sample return predictability by model averaging methods can be observed (Panel A of Table B.1). Results for the German stock market (Panel B of Table B.1) 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 highly parameterized models (i.e. ALL, AIC, \bar{R}^2), with significant PT-statistics 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 with quarterly data is weaker compared to the monthly case. According to the Net-SSE plot for the US in Figure B.1, a substantial forecast breakdown can be detected in the aftermath of the first oil price shock (around 1974). As evinced by Figure B.1, OOS performance of return prediction models in the US has been poor over

most of the 1990s consistent with previous studies mentioned before.

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 in order to find out 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 so far been mainly based on results for the US stock market.

We find substantial 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. In the French case also a (modest amount) of out-of-sample predictability can be detected. Out-of-sample predictability by model averaging methods appears to be more accurate 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 quite 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 model averaging approach accounting for finite-sample bias employed in this paper may be useful beyond the context of return predictability. In the field of macroeconomic forecasting (e.g. inflation or real activity), for instance, also a large amount of model

uncertainty exists and the typical predictors often exhibit a fairly strong degree of persistence (cf. Stock and Watson, 2004). Moreover, another promising subject for future research would be to link the evidence for time-variation in expected returns with the cross-sectional variation of expected returns. An international analysis under model uncertainty with size and book-to-market sorted portfolios may provide additional insights into the particular risks which are relevant to investors.

References

- Amihud, Y., Hurvich, C., 2004. Predictive regressions: a reduced-bias estimation method. *Journal of Financial and Quantitative Analysis* 39, 813–842.
- Ang, A., Bekaert, G., 2007. Stock return predictability: Is it there? *Review of Financial Studies* 20 (3), 651–707.
- Avramov, D., 2002. Stock return predictability and model uncertainty. *Journal of Financial Economics* 64(3), 423–458.
- Bossaerts, P., Hillion, P., Summer 1999. Implementing statistical criteria to select return forecasting models: What do we learn? *The Review of Financial Studies* 12 (2), 405–428.
- Brown, G. W., Cliff, M. T., 2005. Investor sentiment and asset valuation. *Journal of Business* 78, 405–440.
- Campbell, J. Y., 1987. Stock returns and the term structure. *Journal of Financial Economics* 18, 373–399.
- Campbell, J. Y., 1996. Understanding risk and return. *Journal of Political Economy* 104 (2), 298–345.
- Campbell, J. Y., Shiller, R. J., Fall 1988a. The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies* 1 (3), 195–228.
- Campbell, J. Y., Shiller, R. J., June 1988b. Stock prices, earnings, and expected dividends. *Journal of Finance* 43 (3), 661–676.
- Campbell, J. Y., Thompson, S., 2007. 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., Vuolteenaho, T., 2004. Inflation illusion and stock prices. *The American Economic Review (Papers & Proceedings)* 94, 19–23.

- Campbell, J. Y., Yogo, M., 2006. Efficient tests of stock return predictability. *Journal of Financial Economics* 81 (1), 27–60.
- Clark, T. E., West, K. D., 2007. Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics* 138, 291–311.
- Cochrane, J. H., 1999. New facts in finance. *Federal Reserve Bank of Chicago* 11, 3–31.
- 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*.
- Cooper, I., Priestley, R., 2006. Time-varying risk premia and the output gap. *Review of Financial Studies* Forthcoming.
- Cremers, K. J. M., Autumn 2002. Stock return predictability: A Bayesian model selection perspective. *The Review of Financial Studies* 15 (4), 1223–1249.
- Dangl, T., Halling, M., Randl, O., 2006. Equity return prediction: Are coefficients time varying?, Unpublished Working Paper, University of Vienna.
- Fama, E. F., 1981. Stock returns, real activity, inflation, and money. *The American Economic Review* 71, 545–565.
- Fama, E. F., September 1990. Stock returns, expected returns, and real activity. *Journal of Finance* 45 (4), 1089–1108.
- Fama, E. F., French, K. R., October 1988. Dividend yields and expected stock returns. *Journal of Financial Economics* 22 (1), 3–25.
- Fama, E. F., French, K. R., Nov. 1989. Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics* 25 (1), 23–49.
- Fama, E. F., Schwert, G. W., 1977. Asset returns and inflation. *Journal of Financial Economics* 5, 115–146.
- Fernández, C., Ley, E., Steel, M. F. J., 2001. Benchmark priors for Bayesian model averaging. *Journal of Econometrics* 100, 381–427.

- Ferson, W. E., Sarkissian, S., Simin, T. T., 2003. Spurious regressions in financial economics. *Journal of Finance* 58, 1393–1413.
- French, K., Schwert, G. W., Stambaugh, R. F., 1987. Expected stock returns and volatility. *Journal of Financial Economics* 19, 3–29.
- Giot, P., Petitjean, M., 2006. International stock return predictability: Statistical evidence and economic significance, CORE Discussion paper 2006/88.
- Goyal, A., Welch, I., 2003. Predicting the equity premium with dividend ratios. *Management Science* 49, 639–654.
- Goyal, A., Welch, I., 2006. A comprehensive look at the empirical performance of equity premium prediction, forthcoming *Review of Financial Studies*.
- Hjalmarsson, E., 2004. On the predictability of global stock returns, working paper.
- Hodrick, R. J., 1992. Dividend yields and expected stock returns: Alternative procedures for inference and measurement. *Review of Financial Studies* 5 (3), 357–386.
- Hodrick, R. J., Prescott, E., 1997. Postwar U.S. business cycles: an empirical investigation. *Journal of Money, Credit and Banking* 29, 1–16.
- Inoue, A., Kilian, L., 2004. In-sample or out of sample tests of predictability: Which one should we use? *Econometric Reviews* 23 (4), 371–402.
- Koop, G., 2003. *Bayesian Econometrics*. Wiley, Chichester.
- Lamont, O., Oct. 1998. Earnings and expected returns. *Journal of Finance* 53 (5), 1563–1587.
- Leamer, E. E., 1978. *Specification Searches: Ad Hoc inference with Non-experimental data*. Wiley.
- Lettau, M., Ludvigson, S., Jun. 2001. Consumption, aggregate wealth, and expected stock returns. *Journal of Finance* 56 (3), 815–849.
- Lewellen, J., Nov. 2004. Predicting returns with financial ratios. *Journal of Financial Economics* 74 (2), 209–235.

- Moon, R., Rubia, A., Valkanov, R., 2006. Long-horizon regressions when the predictor is slowly varying, Working Paper.
- Neeley, C. J., Weller, P., 2000. Predictability in international asset returns: A reexamination. *Journal of Financial and Quantitative Analysis* 35 (4), 601–620.
- Pastor, L., Stambaugh, R., 2006. Predictive systems: Living with imperfect predictors, Working Paper, GSB Chicago.
- Paye, B., Timmermann, A., 2006. Instability of return prediction models. *Journal of Financial Economics* 13 (1), 274–315.
- Pesaran, M., Timmermann, A., 1992. A simple non-parametric test of predictive performance. *Journal of Business and Economic Statistics* 10, 461–465.
- Rangvid, J., 2006. Output and expected returns. *Journal of Financial Economics* 81, 595–624.
- Rapach, D. E., Wohar, M. E., Rangvid, J., 2005. Macro variables and international stock return predictability. *International Journal of Forecasting* 21, 137–166.
- Ravazzolo, F., Paap, R., van Dijk, D., Franses, P. H., 2006. Bayesian model averaging in the presence of structural breaks, Unpublished Working Paper, Erasmus University Rotterdam.
- Sala-i-Martin, X., Doppelhofer, G., Miller, R. I., 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.
- Stock, J. H., Watson, M. W., 2004. An empirical comparison of methods for forecasting using many predictors, working paper, Harvard University.
- Torous, W., Valkanov, R., Yan, S., 2004. On predicting returns with nearly integrated explanatory variables. *Journal of Business* 77, 937–966.

Appendix A. Data Description

This section of the appendix provides a more detailed description of the stock returns as well as the predictive 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.

Excess returns: 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 employed.²² For the other stock markets, we use broad stock market indexes by Datastream. 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. Interest rates are taken from the Reuters-Ecowin database. 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 databases 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.

²²We would like to 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 taken 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 HP-filter to the logarithmic series of industrial production. As in Cooper and Priestley (2006), the smoothing parameter is set to 128800 for the monthly data and 1600 for the quarterly data. The cyclical component of the series is taken as the output gap.

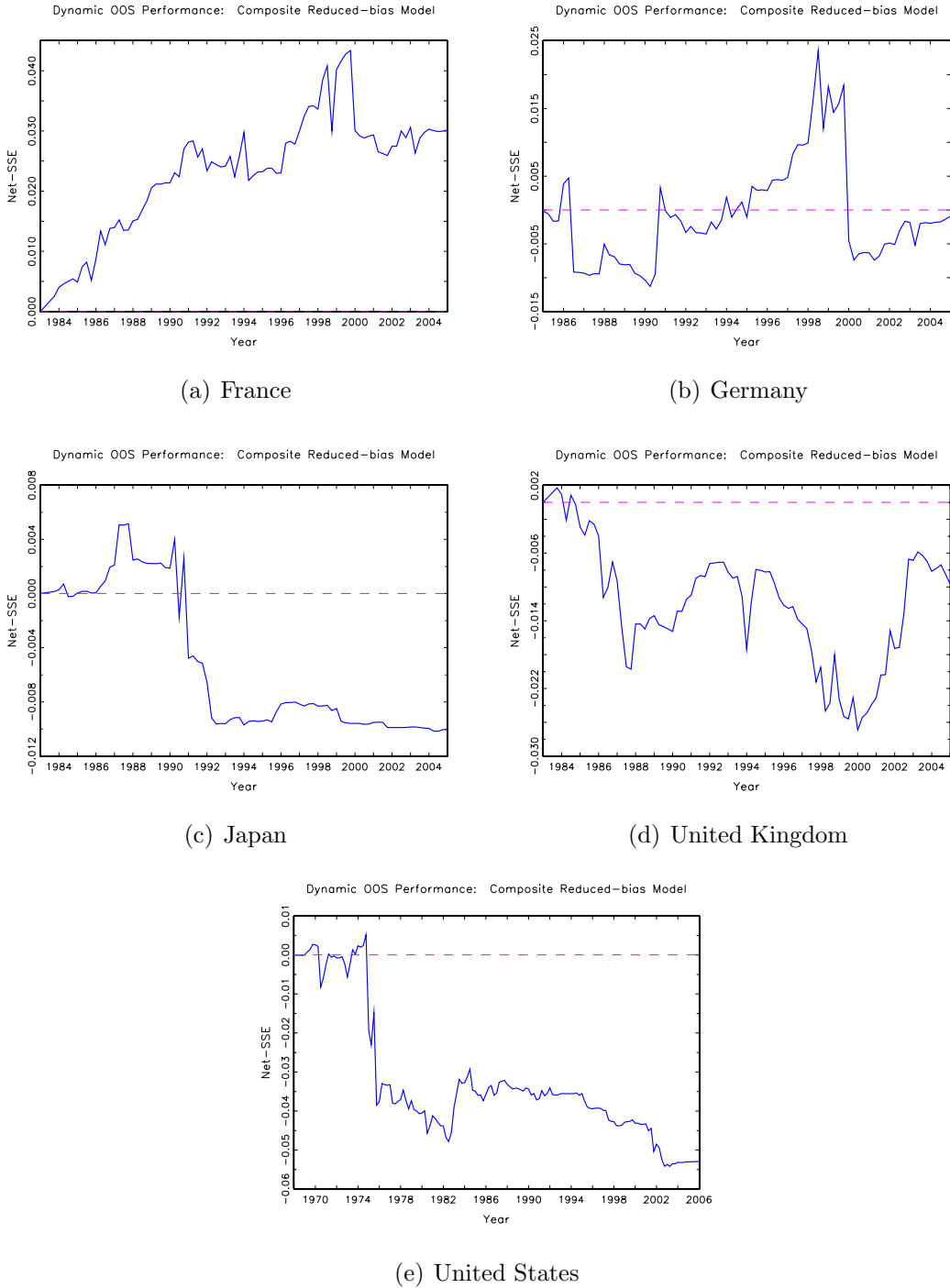
Appendix B. Out-of-Sample Results at the Quarterly Horizon

Table B.1: Estimation Results: Out-of-sample, Quarterly

Panel A: France		BACE-adj	BMA	TOP	All	AIC	BIC	\bar{R}^2
ME		-0.0007	0.0045	0.0055	-0.0198	-0.0083	0.0029	-0.0131
t-stat		-0.0608	0.3652	0.4345	-1.6140	-0.6705	0.2270	-1.0596
TU		0.9864	0.9914	1.0158	1.0020	1.0054	1.0173	1.0060
Hit		0.7416	0.6966	0.6517	0.7079	0.7303	0.6629	0.6966
PT		0.6404	0.1086	0.1834	0.6767	1.1980	0.2841	0.7432
Panel B: Germany		BACE-adj	BMA	TOP	All	AIC	BIC	\bar{R}^2
ME		-0.0021	-0.0016	0.0026	-0.0090	-0.0016	0.0036	-0.0055
t-stat		-0.1628	-0.1235	0.1933	-0.6634	-0.1179	0.2764	-0.4104
TU		1.0004	1.0012	1.0031	1.0343	1.0032	0.9985	1.0192
Hit		0.5926	0.5926	0.5679	0.6420	0.6296	0.5432	0.6173
PT		0.1036	0.1036	0.7742	1.6455	1.8985	0.5957	1.4920
Panel C: Japan		BACE-adj	BMA	TOP	All	AIC	BIC	\bar{R}^2
ME		-0.0077	-0.0078	-0.0071	-0.0104	-0.0087	-0.0068	-0.0071
t-stat		-0.6123	-0.6179	-0.5535	-0.8190	-0.6875	-0.5282	-0.5553
TU		1.0040	1.0053	1.0210	1.0197	1.0106	1.0270	1.0188
Hit		0.5955	0.6067	0.5955	0.4944	0.6180	0.5955	0.5618
PT		-0.2404	-0.1190	-0.2404	-1.1983	0.7746	-0.2404	0.0810
Panel D: UK		BACE-adj	BMA	TOP	All	AIC	BIC	\bar{R}^2
ME		0.0081	0.0171	0.0300	0.0273	0.0271	0.0310	0.0261
t-stat		0.8604	1.8151	3.1203	2.6937	2.7411	3.1558	2.6423
TU		1.0092	1.0260	1.0839	1.1266	1.1010	1.1075	1.0970
Hit		0.7191	0.5955	0.5056	0.5281	0.5056	0.4944	0.5169
PT		0.0000	-0.1719	0.8151	1.0346	0.6477	0.7056	0.7572
Panel E: US		BACE-adj	BMA	TOP	All	AIC	BIC	\bar{R}^2
ME		-0.0003	0.0013	0.0025	-0.0026	0.0035	0.0036	0.0017
t-stat		-0.0430	0.1880	0.3643	-0.3642	0.5132	0.5137	0.2491
TU		1.0252	1.0233	1.0307	1.0453	1.0155	1.0285	1.0287
Hit		0.6333	0.5933	0.5867	0.5600	0.5933	0.5933	0.5867
PT		0.2928	-0.4239	0.2397	-0.5101	1.0035	0.5237	0.4235

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, \bar{R}^2 are based on the best models selected by the Akaike, Schwarz criterion or adjusted R^2 , respectively. ME denotes the mean prediction error (t-statistic reported below). TU 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 B.1: Net-SSE Plots, Quarterly



The figure shows Net-SSE plots for the aggregate stock market following Goyal and Welch (2006). Net-SSE is the cumulated difference of squared forecast errors of the unconditional benchmark model (i.i.d. model) and the conditional model (BACE-adj): $\text{Net-SSE}(\tau) = \sum_{t=1}^{\tau} (e_{uc,t}^2 - e_{c,t}^2)$, 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.