Equity Return Prediction: Are Coefficients Time Varying?*

Thomas Dangl
ISK
and Vienna University of Technology

Michael Halling
University of Vienna

Otto Randl
Anaxo Financial Services
and University of Vienna

March 2006

Abstract

Most papers in equity return prediction rely on the assumption that coefficients are constant over time. We want to question this assumption and evaluate it empirically. Therefore, we develop an econometric framework enabling us to compare models that assume static coefficients and models that allow for time variation in coefficients. We find strong empirical support for the validity of models allowing for time varying regression coefficients and document that the uncertainty about the true level of time-variation in coefficients is of equal magnitude than the uncertainty about the true choice of predictive variables. Furthermore, we find that best-performing static models change more often over time and contain a larger number of predictive variables than best performers among dynamic models. Finally, we address the question of overall predictability. We find in-sample predictability, are unable to unambiguously document out-of-sample predictability, and show that only predictive models with time varying coefficients consistently outperform the non-predictability benchmark model.

JEL Classifications: G12, C11
Keywords: Equity return prediction, Bayesian econometrics, empirical asset pricing.

*Acknowledgment: We thank M. Baltutis, B. Hanke, Z. Rachev, L. Pastor, F. Schausberger, A. Timmermann, and J. Zechner for their helpful comments. Further more we appreciate discussions at the 10th Symposium on Finance, Banking and Insurance.

1 Introduction

The issue of predicting equity returns is one of the most widely discussed topics in financial economics. Yet no consensus exists on the fundamental questions: whether predictability exists and which variables show best predictive performance. Recently, the focus of academic work in this area has been on evaluating the robustness of the current state of knowledge (e.g., Ang and Bekaert (2005), Campbell and Thompson (2005), and Goyal and Welch (2005))

In our opinion, the issue of robustness of prediction results corresponds to the issue of sources of uncertainty in prediction models. Standard sources of uncertainty accounted for in linear regression frameworks include the variance in predictive variables and the uncertainty in coefficients. A third source of uncertainty is frequently labeled model uncertainty and denotes the uncertainty in selecting the variables to be included in the regression model (e.g., given k explanatory variables one can think of 2^k different models). A seminal paper is Pesaran and Timmermann (1995), which accounts for model uncertainty by determining the best predictive model among all possible models using standard diagnostic statistics for linear regressions. Bossaerts and Hillion (1999) evaluate several approaches for the use of statistical criteria for the selection of the best return forecasting model. Contrasting these papers that account for model uncertainty by determining the best model, only recently Avramov (2002) and Cremers (2002) addressed model uncertainty in a Bayesian setup by determining an average model using Bayesian Model Averaging rather than selecting an individual model (see Raftery, Madigan, and Hoeting (1997)).

While all these papers identified in-sample predictability, they failed to or succeeded only partly in documenting out-of-sample predictability. Bossaerts and Hillion (1999) state, for example, that "The poor external validity of the prediction models that formal model selection criteria chose indicates model nonstationarity: the parameters of the best prediction model change over time." Also Cremers (2002) claims in his conclusion that his model is limited by the assumption of parameter stability. In fact, all the previously mentioned papers and, in fact, most papers dealing with equity return prediction (see Paye and Timmermann (2003)) have made an important but so far unverified assumption, namely, that regression coefficients are unknown but constant over time. We doubt that this assumption is valid; in this paper we present an econometric technique to account explicitly for time variation in coefficients.

However, we are not the first to raise doubts about the assumption of constant coefficients. Paye and Timmermann (2003) summarize several potential reasons for time variation in coefficients (i.e., major changes in market sentiments, regime switches in monetary policies, institutional changes, large macroeconomic changes, disappearance of market inefficiencies). Ang and Bekaert (2005) test for time variation in coefficients in an ad-hoc way by splitting their entire sample into different sub-periods. They report several figures that clearly document the time varying pattern of coefficients and find, for example, that the coefficient for the dividend yield is twice as large if estimated from a sample that excludes the 1990s than if estimated on their entire sample.

While Ang and Bekaert (2005) exogenously defined their sub-periods, other papers estimated switching regime models searching for structural breaks in equity returns and forecasting variables. Viceira (1997) is to our knowledge the first paper searching for structural changes in predictive relationships. He, however, does not find evidence for structural breaks between the dividend yield and equity returns. Pastor and Stambaugh (2001) and Kim, Morley, and Nelson (2000) use Bayesian econometrics to identify structural breaks in equity returns. Both papers report that they identify empirical evidence for the existence of structural breaks, which consequently supports our intuition that relationships between forecasting variables and equity returns vary over time.In this strand of literature, the paper most closely related to our work is probably Paye and Timmermann (2003) which identifies several structural breaks in the coefficients of state variables. However, we differ from papers that search for structural breaks in the sense that we do not approximate the time variation in coefficients by a step function, but allow the coefficients to change gradually over time.[1]

Concluding this small literature survey, we think that there is ample empirical support for our belief that the relationship between predictive variables and equity returns varies over time (see also Dimson, Marsh, and Staunton (2002) for more empirical evidence on correlation changes between economic variables). Our paper presents a Bayesian econometric method (following West and Harrison (1997) that allows us to model time varying coeffi-

[1] Note that there is an extensive literature (see, for example, Jostova and Philipov (2005) for a recent paper) that focuses on models with dynamic (i.e., time-varying) Beta which is to some extent related to our work. However, these papers differ considerably in goals and methodology from our approach.

cients that follow a random walk. In fact, the variance/covariance structure of this random walk in the coefficients has to be specified exogenously. This uncertainty about the "amount" of variability of the coefficients can be treated as another dimension of model uncertainty. Consequently, we address this issue in a consistent manner within the Bayesian Model Averaging approach.

The papers most closely related to our paper—from a technical point of view—are Avramov (2002) and Cremers (2002) which both use Bayesian Model Averaging to capture model uncertainty. Cremers (2002) tries to predict returns of the S&P 500 equity index using a standard linear framework with 14 predictive variables that have been used in various previous academic studies. He distinguishes different economically meaningful priors for the expected R-squared, for the variance of the residuals, and for the number of included variables that represent the views of investors who are either skeptical or confident about predictability. His most important results are (i) that the data imply posterior probabilities supporting the existence of stock return predictability and (ii) that the average model calculated as the sum of each models prediction weighted by its posterior probability outperforms the best individual model selected by a standard statistical criterion.

While Cremers (2002) analyzes the monthly excess returns of the S&P 500 index, Avramov (2002) uses Bayesian Model Averaging to account for model uncertainty in regressions of individual stock returns on predictive variables and illustrates the implications for asset allocation decisions. His main results show that ignoring model uncertainty could lead to incorrect inferences about the relevance of predictive variables and that in-sample as well as out-of-sample predictability exists. We show in this paper that, similarly, incorrent inferences about the importance or non-importance of explanatory variables can be made if time variation in coefficients is ignored.

Kandel and Stambaugh (1996), Barberis (2000), and Xia (2001) also present related papers, as they explicitly take into account parameter uncertainty and evaluate the influence of return predictability on portfolio selection using Bayesian methods. MacKinlay and Pastor (2000), Pastor (2000), and Pastor and Stambaugh (2000) model the impact of prior mispricing uncertainty in asset pricing models on portfolio choice. Pettenuzzo and Timmermann (2005) address the issue of model instability (i.e., structural breaks in predictive relationships) and document that it can have a larger impact on optimal asset allocations than other sources of risk such as parameter estimation uncertainty. As far as our paper is concerned, we focus exclusively on the predictive power of our econometric setup and on the question of whether time variation in coefficients is an important source of uncertainty that should be incorporated into predictive models. We do not discuss implications for portfolio selection.

Our paper shows that empirical evidence supports the existence of time variation in regression coefficients. Using S&P 500 returns from the period March 1951 to March 2005, the total posterior probability assigned to models with static regression coefficients is less than 10 percent and this result is robust to changes in the prior. While static models dominate in the beginning of the analyzed horizon, dynamic models successfully compete with static models in the late twentieth century. In the period from 2000 to 2005, models with time varying coefficients are assigned approximately 90 percent of the posterior probability. This indicates that the US equity market became increasingly more dynamic over the five decades we analyzed. Comparing static models with highest posterior probability to best-performers among the class of dynamic models, we find that "good" static models tend to be larger than "good" dynamic models. Thus, static models seem to compensate for the lack of flexibility due to the assumption of constant coefficients by adding additional variables into the models. Therefore, relatively bigger models seem to play an important role. This could explain the frequently reported discrepancy between the apparently strong in-sample predictability and the weak out-of-sample predictability of predictive regressions (see, for example, Goyal and Welch (2005), Ang and Bekaert (2005), and Paye and Timmermann (2003)).

Furthermore, our results are able to shed new light on specific questions pertaining to equity return prediction. In particular, we focus on the importance of the dividend yield as a predictive variable. This question has received a considerable amount of attention among academics. We document—confirming results presented in Ang and Bekaert (2005) and Paye and Timmermann (2003)—that the dividend yield became a less important predictive variable during the 1990s. However, according to our analysis it has regained large parts of its explanatory power since 2000. Another frequently discussed myth in equity return prediction is the January effect. Haugen and Jorion (1996) argue, for example, that the January effect was still present in the 1990s and that its magnitude had not changed. The pattern of the

January dummy in our empirical analysis confirms this. We find that models including the January dummy receive a relatively constant amount of posterior probability of around 20 percent between the 1960s and the 1990s. However, according to our analysis the January dummy has lost predictive importance for the S&P 500 since 2000.

Finally, our approach offers a natural explanation of why the best predictive models identified, for example, in Pesaran and Timmermann (1995) and Bossaerts and Hillion (1999) have been so unstable. As pointed out in Bossaerts and Hillion (1999), there are no intuitive economic reasons why the set of significant variables should erratically change from one month to the next. In fact, our analysis illustrates that this is not necessarily the case. We show that if we consider the fact that coefficients of predictive variables vary over time, we are able to identify top models with respect to their posterior probabilities that are comparatively stable over time. In contrast, if we force coefficients to be constant over time, we observe large fluctuations among top performing models.

The final question that we address in our paper is whether we can identify out-of-sample predictability. Using various measures and empirical analyses, we identify in-sample predictability but fail to unambiguously show the existence or non-existence of out-of-sample predictability. Our most important observation in this context is, however, that only predictive models that explicitly allow for time varying coefficients succeed in consistently outperforming the non-predictability benchmark model containing only the constant across all periods of time. We interpret this observation as another indication that allowing for time variation in coefficients is important in identifying models with predictive power.

The paper is structured in the following way. Section 2 describes the dynamic linear models that we use to account for time variation in coefficients. Section 3 describes the data sources and variables used in the empirical study. Section 4 reports our empirical results and discusses their implications for the answer to our economic questions. Section 5 concludes and summarizes.

2 Prediction Models with Time-Varying Coefficients

2.1 Dynamic Linear Models

Similar to Avramov (2002) and Cremers (2002), we assume a linear relationship between k predictive variables (including the constant) and the dependent variable, i.e., the excess return r of some asset, of the form

$$r_t = F_t'\theta_t + v_t \quad \text{(observation equation)}, \tag{1}$$

where r_t is the return at time t, F_t is the vector of realizations of the explanatory variables at t, θ_t is the vector of unobservable coefficients for each explanatory variable and v_t is the normally distributed error term (with mean 0 and unknown, but constant variance V).

It is important to note that Avramov (2002) and Cremers (2002) (and, with few exceptions, all other prediction models that have been evaluated thus far) assume that the unobservable regression coefficients θ are constant over time. However, we doubt that this is the case. Consequently, we explicitly allow for a time varying nature of θ_t. For this purpose, we develop a dynamic linear model with Bayesian updating (according to West and Harrison (1997)). Given a set of candidate predictive variables, we do not ad hoc specify which of these variables should be included in our model. As in Avramov (2002) and Cremers (2002), we estimate models for all possible combinations and assign posterior probabilities to the models according to their out-of-sample predictive performance using Bayesian model selection criteria (see Section 2.3). The goal is to develop a more flexible prediction model that explicitly takes into account the different sources of uncertainty—uncertainty in parameter estimation, uncertainty in the choice of predictive variables to use in the regression, uncertainty in the degree of variability of the regression coefficients, and uncertainty due to the error term in the observation equation (1). However, in this section we focus on outlining the characteristics of an individual dynamic linear prediction model.

Suppose we have made a certain choice for the set of predictive variables that we use in our dynamic linear regression model and denote this choice by M_i. We model the vector of time varying coefficients such that coefficients do not change in a systematic way but only through shocks, i.e., they follow a random walk:

$$\theta_t = \theta_{t-1} + \omega_t \quad \text{(system equation)}, \tag{2}$$

where ω_t is an error term that is normally distributed with zero mean and variance matrix W_t. While the regression coefficients θ_t are exposed to random shocks, Equation (2) implies that there is no systematic change in θ_t. If the system variance matrix W_t equals 0, the regression coefficients θ_t are constant over time and the system model is globally reliable. Thus, our model nests the specification of constant regression coefficients. If W_t is large, the intrinsic variability of the regression coefficients θ_t is vast and the system model becomes entirely unreliable and, consequently, useless.

We will now describe how, at some arbitrary time t, the observation of a new return realization leads to an update of the estimated system coefficients and the estimated variance V in the observation equation. Let D_{t-1} denote the information set available before the time t return is realized, i.e., this information set contains all returns and all corresponding realizations of the predictive variables up to time $t-1$.

We develop the updating recurrence in an inductive manner, more precisely, in a fully conjugate Bayesian analysis. For details on the Bayesian updating recurrence refer to West and Harrison (1997). Let us assume that after updating according to the information set D_{t-1}, the time $t-1$ posterior distribution of the system coefficients θ_{t-1} and of the observational variance V are jointly normal/inverse-gamma distributed. This assumption implies that if S_{t-1} is the current estimate of the observational variance V and n_{t-1} is the associated number of degrees-of-freedom, then $\phi = 1/V$ is gamma distributed according to

$$(\phi | D_{t-1}) \sim G\left[\frac{n_{t-1}}{2}, \frac{n_{t-1}S_{t-1}}{2}\right]. \tag{3}$$

The assumption about the $t-1$ posterior implies additionally that conditional on V, the set of system coefficients θ_{t-1} is multivariate normal. Let m_{t-1} denote the $t-1$ point estimate of θ_{t-1} and C^*_{t-1} the covariance of the coefficients normalized by the observational variance, then θ_{t-1} is distributed according to

$$(\theta_{t-1}|V,D_{t-1}) \sim N\left[m_{t-1}, VC^*_{t-1}\right]. \tag{4}$$

This implies further, that the coefficients θ_{t-1} are unconditionally multivariate t-distributed

$$(\theta_{t-1}|D_{t-1}) \sim T_{n_{t-1}}\left[m_{t-1}, S_{t-1}C^*_{t-1}\right]. \tag{5}$$

When taking the set of predictive variables F_t to make a forecast of the time t return r_t, we must consider that due to the random walk assumption in system equation (2) the variance in the coefficients will increase between time $t-1$ and time t by W_t. That means, conditional on V, the prior distribution of θ_t is

$$(\theta_t|V,D_{t-1}) \sim N\left[m_{t-1}, VC^*_{t-1}+W_t\right]. \tag{6}$$

The coefficients, however, do not change in a systematic way, thus, the distribution remains centered around the estimate m_{t-1}. With this joint prior distribution of the observational variance and the coefficients, we are able to calculate a forecast of the time t return r_t by integrating over the entire range of θ and ϕ. Let $\varphi(x;\mu,\sigma^2)$ denote the density of a (multivariate) normal distribution evaluated at x and $\gamma(\phi;a,b)$ the density of a $G[a,b]$ distributed variable evaluated at ϕ, then we have

$$\begin{aligned}
f(r_t|D_{t-1}) &= \int_0^\infty \int_\theta f(r_t|D_{t-1},\theta) f(\theta|D_{t-1}) d\theta\, d\phi \\
&= \int_0^\infty \left[\int_\theta \varphi\left(r_t; F'\theta, \frac{1}{\phi}\right) \varphi\left(\theta; m_{t-1}, \frac{1}{\phi}C^*_{t-1}+W_t\right) d\theta\right] \\
&\quad \times\gamma\left(\frac{n_{t-1}}{2}, \frac{n_{t-1}S_{t-1}}{2}\right) d\phi \\
&= \int_0^\infty \varphi\left(r_t; F_t' m_{t-1}, F_t'\left(\frac{1}{\phi}C^*_{t-1}+W_t\right)F_t+\frac{1}{\phi}\right) \\
&\quad \times\gamma\left(\frac{n_{t-1}}{2}, \frac{n_{t-1}S_{t-1}}{2}\right) d\phi \\
&= T_{n_{t-1}}[p_t, Q_t],
\end{aligned} \tag{7}$$

where

$$Q_t = F_t' R_t F_t + S_{t-1} \tag{8}$$

is the total, unconditional variance of the forecast r_t, and $R_t = S_{t-1}C^*_{t-1} + W_t$ is the unconditional variance of the time t-prior of the coefficient vector θ_t. The first term in (8) characterizes the variance coming from uncertainty in the estimation of θ_t, the second term S_{t-1} is the estimate of the variance of the error term in the observation equation. Since the prior belief about the coefficients θ_t is m_{t-1}, the point estimate of r_t is given by

$$p_t = F_t' m_{t-1}. \tag{9}$$

The Bayesian approach employs the following updating recurrence relationship

$$e_t = r_t - p_t \quad \text{(error in prediction).} \tag{10}$$

The prediction error is the essential signal conditioning learning. Whenever e_t equals zero, the observed return equals the forecast, and thus, there is no updating in the coefficients.

$$n_t = n_{t-1} + 1 \quad \text{(degrees of freedom).} \tag{11}$$

$$S_t = S_{t-1} + \frac{S_{t-1}}{n_t}\left(\frac{e_t^2}{Q_t} - 1\right) \text{(estimator of observational variance).} \tag{12}$$

Since the total variance of the forecast is given by Q_t, we have $E(e_t^2) = Q_t$. If the error in prediction coincides with its expectation, i.e., $e_t^2 = Q_t$, the estimate of the observational variance is not updated, i.e., $S_t = S_{t-1}$. A prediction error below the expected error leads to a reduction in the estimated observational variance, and vice versa. The adaptive vector

$$A_t = \frac{R_t F_t}{Q_t} \quad \text{(adaptive vector)} \tag{13}$$

measures the information content of the current observation in relation to the precision of the estimated regression coefficient and therefore characterizes the extent to which the posterior of θ_t reacts to the new observation. The point estimate m and the covariance matrix C^* are updated in the following way:

$$m_t = m_{t-1} + A_t e_t \quad \text{(estimator for expected coefficient vector)}, \tag{14}$$

$$C_t^* = \frac{1}{S_{t-1}}(R_t - A_t A_t' Q_t) \quad \text{(estimator for variance of coeff. vector)}. \tag{15}$$

The posteriors of the variance V and the coefficient vector θ_t are again jointly normal-gamma distributed as stated in the following equations:

$$(\phi|Dt) \sim G\left[\frac{n_t}{2}, \frac{n_t S_t}{2}\right], \tag{16}$$

$$(\theta_t|V, D_t) \sim T_{n_t}[m_t, S_t C_t^*]. \tag{17}$$

Thus, the normal-gamma distributions represent a conjugate class of distributions for this type of DLM, since the posterior is of the same class of distribution as the prior. And therefore, we choose our $t = 0$ prior distribution from this class.

As pointed out in West and Harrison (1997), using a system variance matrix W_t has several drawbacks. For example, it is ambiguous and not invariant to the scale of regression variables. To overcome these problems, we apply an approach that captures the variability of the regression coefficients by using a discount factor $\delta \in (0, 1]$. This approach relies on the assumption that the variance matrix W_t of the error term ω_t is proportional to the estimation variance $S_{t-1}C_{t-1}^*$ of the coefficient vector θ_t. More precisely, it is assumed that

$$W_t = \frac{1-\delta}{\delta} S_{t-1} C_{t-1}^* \tag{18}$$

and thus the expression for the variance of the forecasted coefficient vector simplifies to

$$R_t = S_{t-1}C_{t-1}^* + \frac{1-\delta}{\delta} S_{t-1}C_{t-1}^* = \frac{1}{\delta} S_{t-1}C_{t-1}^*, \tag{19}$$

which ensures analytical tractability of the model. This assumption implies that periods of high estimation error in the coefficients coincide with periods of high variability in coefficients.

A choice of δ equal to 1 corresponds to $W_t = 0$, i.e., to the assumption that the regression coefficients are constant over time, similar to the models analyzed in Avramov (2002) and Cremers (2002). Choosing a discount factor $\delta < 1$ explicitly assumes variability of the underlying regression parameters.

The prediction of one particular model depends not only on the choice of predictive variables but also on the choice of δ, thus we rewrite the point estimate of r_t as

$$p_{t,i}^{j} = E(r_t | M_i, \delta_j, D_{t-1}) = F_t' m_{t-1} | M_i, \delta_j, D_{t-1}. \tag{20}$$

Since the choice of δ is a critical model assumption, this parameter will be chosen within our Bayesian model averaging procedure.

2.2 Prior Information

To specify the prior information at time $t = 0$, we use a natural conjugate g-prior specification (see, e.g., Zellner (1986), this type of prior was also used in the studies by Cremers (2002) and Kandel and Stambaugh (1996)). This is a noninformative prior, which is consistent with the null-hypothesis of non-predictability. Thus, the $t = 0$ prior for the coefficient vector θ_t equals zero. The prior belief about the covariances among coefficients is a multiple of the OLS estimate of the variance in coefficients, i.e., $C_0^* = g(1/\phi_0)(F'F)^{-1}$ where g serves as the scaling factor that determines the confidence one assigns to the null-hypothesis of non-predictability.

2.3 Bayesian Model Selection

The empirical literature on asset price dynamics shows that there is considerable uncertainty about which factors contain significant information for predicting asset returns. This means that even if we restrict our attention to simple linear models as specified in (1) and (2) there is a high degree of model uncertainty due to the a priori choice of the set of predictive variables F_t used as regressors. Agreeing on k candidate regressors (including the constant) alone implies 2^k different possible linear regression models. The presumed variability in the regression coefficients θ_t (characterized by the choice of the discount factor δ) constitutes a further a priori specification. Considering a number of d different discrete values of δ leads to a total of $d \cdot 2^k$ possible dynamic linear models.

The arbitrary choice of one particular model from this substantial pool of possible models is always debatable. Bayesian model selection offers a systematic approach to this problem that tests the reliability of all $d \cdot 2^k$ models against the observed data. Starting from an uninformed prior, it allows for assigning a posterior probability to each model and, due to averaging across the models, it assigns a posterior probability of inclusion to the set of active predictors for each of the candidate variables. However, the determination of the universe of possible models together with the assumption of the prior probability leaves some room for discretion. We take a large number of candidate predictive variables and perform robustness checks with respect to different assumptions about the prior.

Let M_i denote a certain choice of predictive variables from the k candidates, then for each of the δ values considered, there are 2^k possible choices. Therefore, the exact specification of a DLM model comprises a certain selection of predictive variables and a choice of δ. We start out with the diffuse conditional prior $P(M_i|\delta) = 1/2^k$. For a given $\delta = \delta_j$, we use Bayes' rule to obtain the posterior probabilities

$$P_t(M_i|r_t, \delta_j) = \frac{f(r_t|M_i, \delta_j)P_{t-1}(M_i|\delta_j)}{f(r_t|\delta_j)} \tag{21}$$

where

$$f(r_t|\delta_j) = \sum_M f(r_t|M_i, \delta_j)P_t(M_i, \delta_j) \tag{22}$$

The crucial part is the conditional density

$$f(r|M_i, \delta_j) \sim \frac{1}{\sqrt{Q_i^j}}\ t_{n-1}\left(\frac{r - p_i^j}{\sqrt{Q_i^j}}\right) \tag{23}$$

where p_i^j and Q_i^j are the respective point estimates for the return and the variance of model M_i for a given value of $\delta = \delta_j$. The posterior model probability P_t is used as a prior for $t+1$. The time t return prediction of the average model for a given $\delta = \delta_j$, then equals

$$p_t^j = \sum_{i=1}^{2^k} P_{t-1}(M_i|\delta_j)p_{t,i}^j \tag{24}$$

where

$$P_{t-1}(M_i|\delta_j) = P_{t-1}(M_i|\delta_j, D_{t-1}) \tag{25}$$

is the time t prior probability (i.e., the time $t-1$ posterior probability) that is assigned to one particular selection of predictive variables for given δ.

Since a particular choice of δ cannot be done on an ad-hoc basis, we also perform Bayesian Model Averaging over different values of δ. If we consider d candidates for δ, we assign a prior probability of $1/d$ to each δ value. The time t posterior probability of a certain δ is then

$$P(\delta_j|r) = \frac{f(r|\delta_j)P(\delta_j)}{\sum_\delta f(r|\delta_j)P(\delta_j)}. \tag{26}$$

The total posterior of a certain model configuration (i.e., variable choice and choice of δ) is then given by

$$P(M_i, \delta_j|r) = P(M_i|r, \delta_j)P(\delta_j|r) \tag{27}$$

and the unconditional average prediction of the average model is

$$p_t = \sum_{j=1}^{d} P_{t-1}(\delta_j)p_t^j. \tag{28}$$

2.4 Variance Decomposition

Since the Bayesian Model Averaging approach keeps track of all possible sources of uncertainty regarding the prediction, we are able to decompose the variance of the return into four parts:

$$
\begin{aligned}
\mathrm{Var}(r_t) \;\; = \;\; & \sum_{\delta_j} \sum_{M_i} (S_{t-1}|M,\delta) P_{t-1}(M|\delta) P_{t-1}(\delta) + \\
& \sum_{\delta_j} \left[\sum_{M_i} (F_t R_{t-1} F_t|M,\delta) P_{t-1}(M|\delta) \right] P_{t-1}(\delta) + \\
& \sum_{\delta_j} \left[\sum_{M_i} (p^j_{t,i} - p^j_t)^2 P_{t-1}(M|\delta) \right] P_{t-1}(\delta) + \\
& \sum_{\delta_j} (p^j_t - p_t)^2 P_{t-1}(\delta). \qquad (29)
\end{aligned}
$$

Equation (29) can be deduced by decomposing the variance of the random variable r step by step into expected in-sample variances and inter-sample variances. Starting with the decomposition with respect to different values of δ, we can write

$$
\mathrm{Var}(r) = E_\delta(\mathrm{Var}(r|\delta)) + \mathrm{Var}_\delta(E(r|\delta)), \qquad (30)
$$

where E_δ and Var_δ denote the expected value and the variance with respect to δ. The term $E_\delta(\mathrm{Var}(r|\delta))$ represents the first three terms in Equation (29). The term $\mathrm{Var}_\delta(E(r|\delta))$ is the last term in (29). In a second step, the term $E_\delta(\mathrm{Var}(r|\delta))$ can be further decomposed into

$$
\mathrm{Var}(r|\delta) = E_M(\mathrm{Var}(r|M,\delta)) + \mathrm{Var}_M(E(r|M,\delta)), \qquad (31)
$$

which splits term three of Equation (29) from the remainder. The final variance decomposition as shown in (29) follows from simple rearrangements.

The individual terms of (29) can be interpreted in a very intuitive way. The first term is the expected observational variance, i.e., it is the time t prior probability weighted average over all estimations of the observational variance S_{t-1}. The second term states the expected variance from errors in the estimation of the coefficient vector θ, again as a prior probability weighted average over the estimation errors of all $d \cdot 2^k$ models. We will refer to this as estimation uncertainty. Both the third and the fourth term characterizes model uncertainty. The third term measures model uncertainty with respect to variable selection: the expected value of variance among predictors with identical δ. And the fourth term measures model

Table 1: **Summary Statistics.** Values in % (649 Observations).

	Mean	Standard deviation	Median	Minimum	Maximum
Momentum 1	0.585	4.177	0.730	-22.087	16.299
Momentum 2	0.592	4.184	0.730	-22.087	16.299
Dividend yield	3.469	1.261	3.320	1.080	7.440
Earnings yield	7.025	2.745	6.105	2.151	14.970
Turnover	37.682	25.471	22.760	12.160	99.660
Credit spread	0.979	0.479	0.941	0.244	3.306
T-bill rate	5.015	2.855	4.850	0.580	15.520
Change in T-bill	0.002	0.479	0.010	-3.850	2.400
Term spread	1.389	1.154	1.270	-1.910	4.390
Yield spread	0.590	1.072	0.365	-3.190	8.680
January dummy	8.30	27.6	0.000	0.000	100.0
Industrial Production Growth	3.487	5.501	3.964	-12.649	22.961
Inflation	3.930	2.993	3.157	-0.743	14.756
Change in Inflation	-0.005	0.381	-0.003	-1.996	2.153

uncertainty with respect to the time variability of the regression coefficients.

3 Data Description

We investigate monthly data from March 1951 to March 2005, resulting in 649 observations. The dependent variable is the monthly excess return of the S&P 500 Total Return Index, where we use the three month t-bill rate as risk-free interest rate. The choice of explanatory variables is guided by previous academic studies. In particular, we endeavor to make our analysis of Bayesian model averaging of dynamic prediction models comparable with previous studies that focus on Bayesian model averaging of prediction models with static coefficients. Therefore, we gather data on the following set of variables proposed by Cremers (2002):

1. **Momentum 1:** One month lagged excess return of the S&P 500 TR index,

2. **Momentum 2:** Two months lagged excess return of the S&P 500 TR index,

3. **Dividend yield;** One month lagged dividend yield of the S&P 500 index,

4. **Earnings yield:** 100 divided by the one month lagged price earnings ratio of the S&P 500 index,
5. **Turnover:** One month lagged NYSE share value turnover,
6. **Credit spread:** One month lagged difference in yields of 15 year bonds rated BBB and AAA,
7. **T-bill rate:** One month lagged three month t-bill rate,
8. **Change in T-bill:** One month lagged three month t-bill rate minus two month lagged t-bill rate,
9. **Term spread:** One month lagged difference of the yield of ten year US government bonds minus the three month t-bill rate,
10. **Yield spread:** One month lagged difference of the USA Federal Funds Market Rate and the three month t-bill rate,
11. **January dummy:** Dummy variable that equals one in January and zero in all other months,
12. **Growth in industrial production:** Two month lagged annual growth rate in US industrial production,
13. **Inflation:** Two month lagged annual rate of change in the US consumer price index,
14. **Change in inflation:** Two months lagged inflation minus three months lagged inflation,
15. **Constant**.

Data sources are Ecowin for US industrial production and Global Financial Data for all other time series. Table 1 provides some summary statistics on the used data.

4 Results

The approach outlined in Section 2 requires the choice of appropriate priors and the selection of adequate values of δ. We choose a conjugate prior, which ensures that the posterior distribution will be of the same family as the prior distribution. For the actual implementation, we perform the estimation procedure for a g-prior with $g = 50$. We repeat the analysis using a g-prior of ten. Finding our conclusions unchanged from this robustness check, we omit the results for the sake of brevity. The second choice is about delta, where we use the following values in our empirical implementation: 1.00, 0.98, and 0.96. We choose the values of deltas such that we cover the static case (δ=1.00), a rather noisy situation where coefficients are expected to change rapidly (δ=0.96), and an intermediate case (δ=0.98). As described in Section 2.1, the effect of δ strictly lower than 1.00 is to increase the variance of the unconditional prior of the coefficient vector by a factor of $1/\delta$. Ignoring other influencing factors on the estimated variance of the coefficient vector, the total effect of δ will be a 50 percent variance increase within 20 months for a δ of 0.98[2]. For a δ of 0.96, the 50 percent increase will be reached twice as fast, approximately in ten months. A value of 1.00 corresponds to the standard assumption used in the existing prediction literature, i.e.,that coefficients are static.

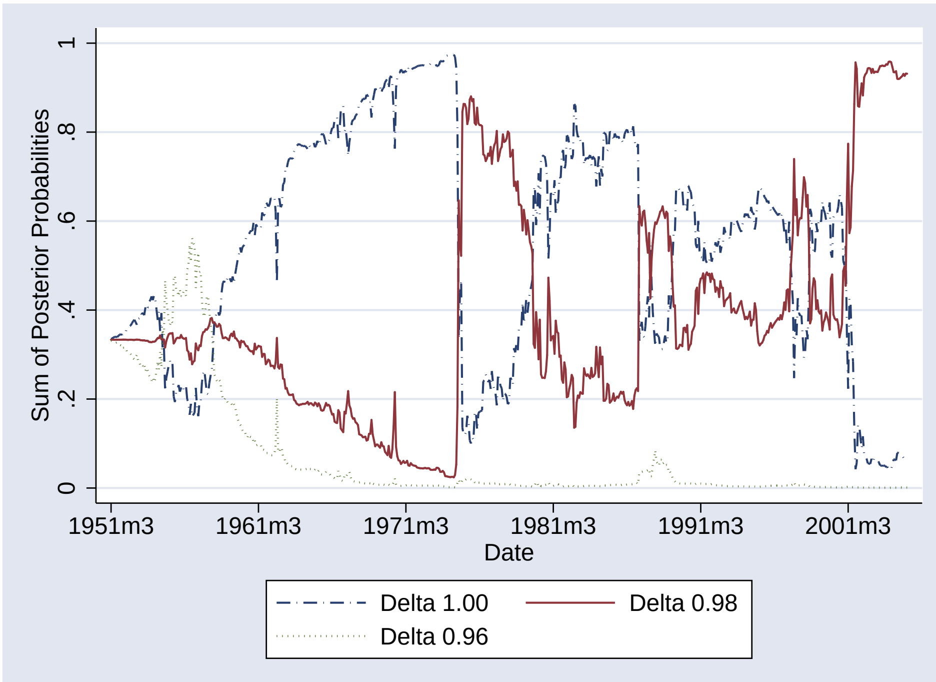
4.1 Are Coefficients Time Varying?

In a first step we analyze the importance of allowing for time variation of coefficients in the estimated models. Note that we start with an uninformed prior giving equal weight to each individual model and each individual δ value in the first step. Therefore, every model and every model class has the same chance to turn out to be important.

In Figure 1, we plot the total posterior probability of all models for each δ. The figure shows that the posterior probabilities start to depart significantly from the unconditional values (1/3 for each δ) after approximately five years. From the graph it becomes obvious that setting δ equal to 0.96 (i.e., high variability in regression coefficients) is clearly dominated by the other model specifications. The class of static regression models ($\delta = 1.00$) shows outstanding performance during the first decades of the horizon until the first oil price shock

[2]Since we estimate monthly returns, $\delta = 0.98$ corresponds to an annualized increase in estimation variance of 27 percent and $\delta = 0.96$ corresponds to 63 percent.

Figure 1: **Sum of Posterior Probabilities of Models with a Given δ.**



in 1974. During the oil crisis in 1974, static models are punished for their over-confidence by a sharp reduction in posterior probability. The dynamic models with $\delta = 0.98$ gain during this crisis. After the oil crisis in 1974, static models again outperform dynamic models and regain up to 80 percent weight, however, at the beginning of the 1980s this trend slows down and vanishes. From the stock price shock of 1987 until the burst of the Internet stocks bubble, static models and dynamic models with $\delta = 0.98$ compete. After the collapse of the Internet bubble, dynamic models clearly outperform static models and gain more than 90 percent of total posterior weight.

4.2 Sources of Prediction Uncertainty

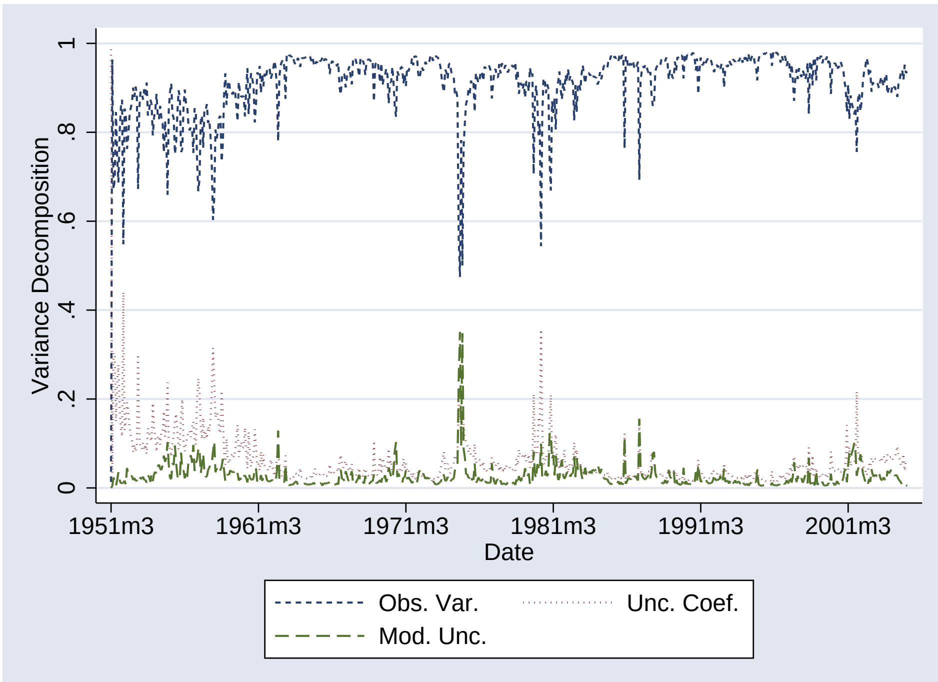
An especially attractive property of our approach is that it enables us to decompose the entire variance into several sources: (a) the standard observational variance, (b) the variance from estimating the coefficients, (c) uncertainty about the true model with respect to the combination of variables, and (d) uncertainty about the true model with respect to the amount of time variation in the coefficients. We think that it is important to analyze and understand the uncertainty of predictions and its sources. Under the Bayesian model averaging setup, the sources of prediction variance are orthogonal to each other. As described in Section 2.4, this allows attributing weights to the four components of prediction variance. In Figure 2, we plot the relative weights of these components of prediction variance over time.

In Panel A of Figure 2, we plot these components as a fraction of total variance. The dominant source of uncertainty is observational variance. This is not surprising: in efficient markets, stocks should fluctuate randomly around their expected values. This fluctuation around expected values will be especially pronounced for short prediction horizons. Therefore, in Panel B we zoom into the picture by ignoring observational variance and focusing only on the other three components. In most periods, the estimation uncertainty in coefficients captures more than half of the remaining variance. In periods of stress, model uncertainty peaks (e.g., in a couple of periods in the 1970s—oil price shocks—and around 1990—Iraq-Kuwait war). Uncertainty about the correct δ is relatively low in the first half of the sample but becomes more pronounced from 1974 onwards. In fact, the uncertainty about the true time variation of coefficients is—in the second half of our data—of a magnitude similar to the uncertainty about the true combination of variables.

While Figure 2 shows an aggregated picture of prediction uncertainty and its components, Figure 3 shows a similar analysis for $\delta = 0.98$ and $\delta = 1.00$. It is interesting to observe that the two graphs look very different. While in both situations the observational variance S dominates, the magnitude of uncertainty from the estimation uncertainty in the coefficients is, by construction of the model, much more pronounced in the case of $\delta = 0.98$. Therefore, these graphs illustrate very well the fundamental difference between these two models. In the case of $\delta = 1.00$, the uncertainty coming from the coefficients is continuously reduced. This is not the case for the dynamic models using $\delta = 0.98$, where the various components fluctuate over the whole sample period.

Figure 2: **Sources of Prediction Variance.**

Panel A: Including the observational variance.



Panel B: Excluding the observational variance.

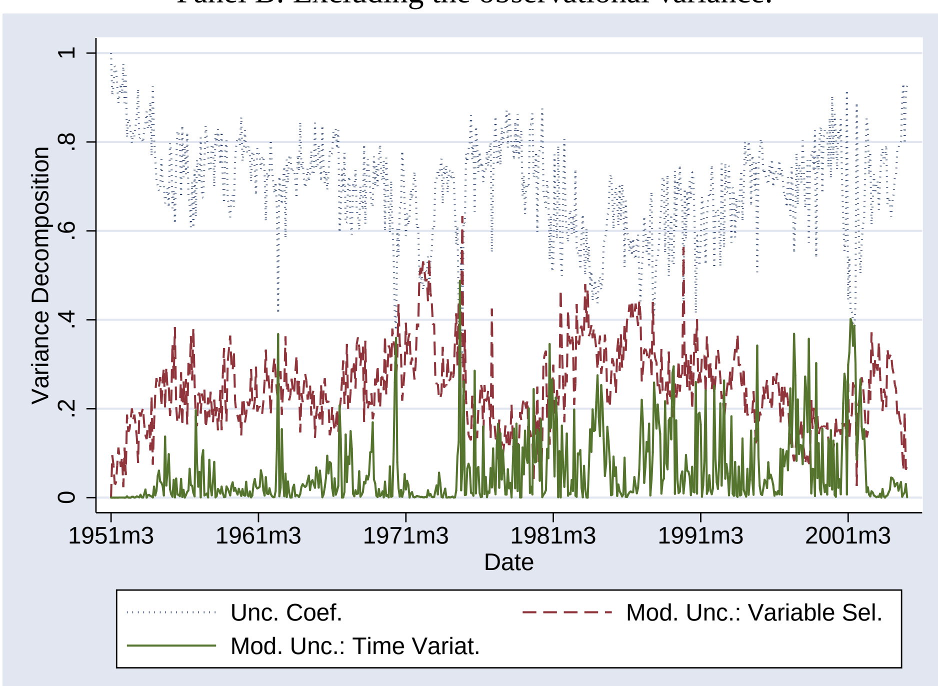
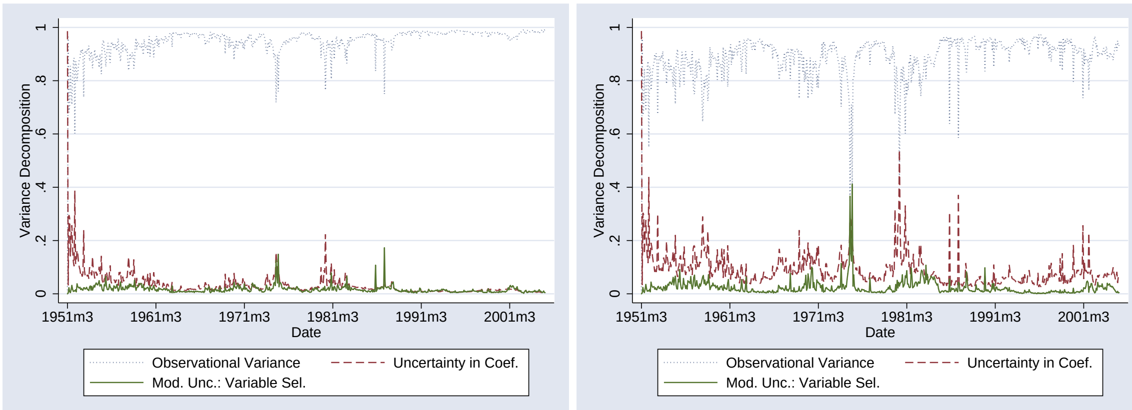


Figure 3: **Sources of Prediction Variance.** Left graph for $\delta = 1$, right graph for $\delta = 0.98$.

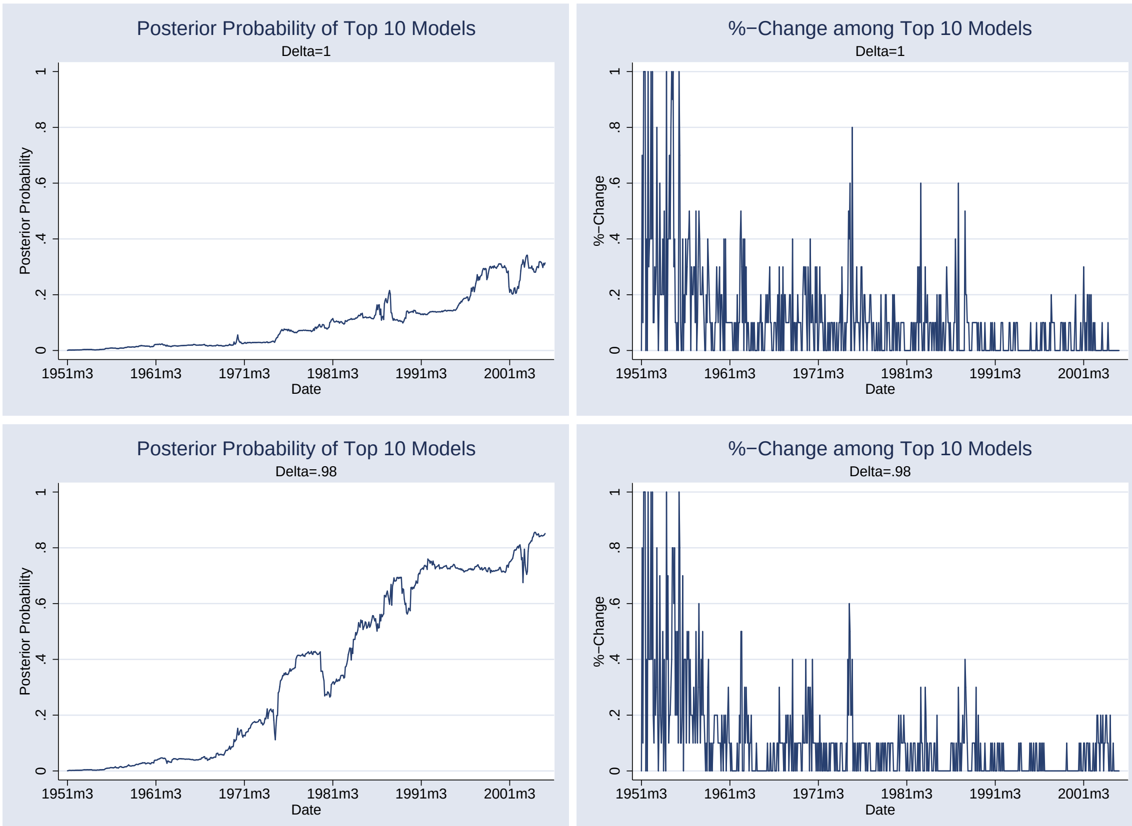


4.3 Characteristics of Prediction Models with Different Assumptions for the Time Variation of Coefficients

In a next step we aim at characterizing prediction models with different assumptions for the time variation of their coefficients. An important first analysis is to characterize the top models. Pesaran and Timmermann (1995) and Bossaerts and Hillion (1999), for example, select top performing models according to various statistical measures for their prediction analysis and report a large amount of variability among these top models. Note that in the remainder of this work we will focus—for the sake of simplicity and readability—on the comparison between $\delta = 0.98$ and $\delta = 1.00$ and drop models with $\delta = 0.96$ as these do not receive notable posterior probability (see Figure 1).

We begin this part of the empirical analysis by documenting the "importance" of individual models for a given δ. Thus, we focus on the Top 10 models for a given delta in a first step. Figure 4 shows how much posterior probability the Top 10 models of each δ receive among the entire set of models with a given δ plotted over time (left column). In the case of constant coefficients, the posterior probability assigned to the Top 10 models does not ultimately account for more than 30 percent. In contrast, the posterior probability assigned to the Top 10 models assuming time variation in the coefficients, ultimately increases to more than 80 percent. Consequently, in the case of constant coefficients Top 10 models are less distinct from other models than in the case of time varying coefficients. This is a potentially impor-

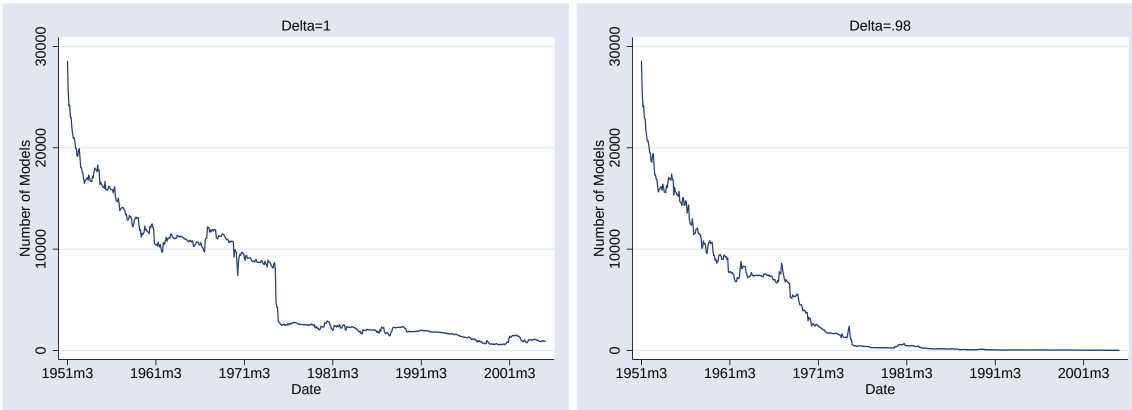
Figure 4: **Analysis of the Top 10 Models.** For δ equal to 1.00 and 0.98.



tant insight as it challenges the strategy followed in several papers on equity return prediction of selecting individual top models given the assumption of constant coefficients. Of course, one has to keep in mind that our analysis is based on a Bayesian framework.

Figure 4 also measures the stability of Top 10 models by counting—from month to month—the changes among Top 10 model specifications (right column). Figure 4 supports our observation that there is more variability among the Top 10 models in the situation of constant coefficients. Finally, Figure 5 summarizes an additional measure to capture the properties of individual models given different assumptions for the time variation of coefficients. It counts the number of models required to get a total amount of posterior probability of at least 90 percent. It turns out that if one assumes constant coefficients, then the number of models required to accumulate a posterior probability of 90 percent is considerably larger than in the case of time varying coefficients: 900 vs. 18 models in March 2005.

Figure 5: **Analysis of top models.** The Number of Individual Models whose Accumulated Posterior Probability Equals 90% (for δ equal to 1.00 and 0.98).



Pesaran and Timmermann (1995) and Bossaerts and Hillion (1999), among others, report that their individual top models changed considerably over time. They admit that their analysis suffers from variability in the top models specifications. Our analysis documents precisely this behavior—many different model specifications with similar posterior probabilities—for models assuming constant coefficients. On the other hand we show that this "stationarity-issue" can be resolved largely by explicitly allowing coefficients to vary over time.

Another model characteristic that we want to analyze in a second step is model size. In Table 2, we report for three points in time and for each delta the cumulative posterior probability of all models with a given number of predictive variables. In accordance with existing studies (e.g., Cremers (2002)), relatively small models obtain high probabilities. In the three months shown in the table, models including three to seven variables get highest posterior probabilities. This documents that parsimonious models are more useful for prediction than very general models, although small models might be biased from a theoretical viewpoint due to the exclusion of potentially important prediction variables. An even more interesting result is that the optimal model size depends on the underlying assumption regarding time variation of coefficients. Comparatively larger models with six and seven explanatory variables get highest posterior probabilities for the case of constant coefficients (i.e., $\delta = 1.00$). These model sizes receive considerably less weight when we account for time variation in coefficients. In our opinion, this indicates the very interesting fact that models assuming

Table 2: **Sum of Posterior Probabilities.** Tabulated per δ and model size.

	March 1985			March 1995			March 2005		
Model size	D=1	D=.98	D=.96	D=1	D=.98	D=.96	D=1	D=.98	D=.96
1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	0.000	0.007	0.063	0.000	0.003	0.101	0.001	0.007	0.127
3	0.004	0.109	0.400	0.004	0.072	0.357	0.021	0.017	0.090
4	0.045	0.352	0.477	0.069	0.396	0.446	0.187	0.298	0.722
5	0.174	0.351	0.056	0.212	0.402	0.093	0.308	0.576	0.061
6	0.289	0.142	0.003	0.297	0.112	0.002	0.265	0.102	0.000
7	0.263	0.033	0.000	0.240	0.014	0.000	0.145	0.006	0.000
8	0.149	0.005	0.000	0.124	0.001	0.000	0.055	0.000	0.000
9	0.056	0.001	0.000	0.043	0.000	0.000	0.015	0.000	0.000
10	0.015	0.000	0.000	0.010	0.000	0.000	0.003	0.000	0.000
11	0.003	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.000
12	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
13	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
14	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
15	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

static coefficients compensate for their lack of flexibility by adding additional explanatory variables. This is a first piece of evidence that predictive regressions will be systematically different depending on whether time variation in coefficients is allowed or ruled out.

To conclude this section, Table 3 focuses on another model characteristic, namely, the importance of individual predictive variables and their coefficients. We measure the importance by the sum of posterior probability of all models that include a specific variable and perform our analysis for all values of deltas at the same three points in time as in Table 2.

The models agree to a large extent on the most important variables, the T-bill rate and the change in T-bill rate. For models allowing for time variation in coefficients, these variables represent the most important variables at all three points in time. In addition to these two variables, the yield spread and the dividend yield seem to play an important role. Similar to the results observed in Table 2, Table 3 documents that the choice of δ has an important influence on the model structure. In the case of $\delta = 1$, the separation between important and less important variables is relatively blurred. The change in inflation is the least important variable but still receives a posterior probability of 9.4 percent in March 1985, followed by

Table 3: **Sum of Posterior Probabilities of Models Including Specific Variables (I) and Average Coefficients (C) within model classes of given time variability of regression coefficients**

Variable		March 1985			March 1995			March 2005		
		δ=1.00	δ=.98	δ=.96	δ=1.00	δ=.98	δ=.96	δ=1.00	δ=.98	δ=.96
Momentum 1	C	-0.009	-0.005	-0.000	-0.004	-0.002	0.000	-0.001	0.000	-0.000
	I	0.222	0.058	0.003	0.163	0.027	0.002	0.128	0.009	0.000
Momentum 2	C	-0.012	0.000	0.000	-0.016	-0.001	-0.000	-0.006	0.00	0.000
	I	0.239	0.045	0.005	0.271	0.022	0.001	0.171	0.007	0.000
Dividend yield	C	0.137	0.839	1.490	0.159	0.573	0.302	0.027	0.467	0.856
	I	0.425	0.695	0.845	0.390	0.728	0.868	0.235	0.593	0.747
Earnings yield	C	0.166	0.081	0.057	0.198	0.148	0.041	0.276	0.202	0.089
	I	0.667	0.258	0.102	0.707	0.271	0.114	0.878	0.416	0.225
Turnover	C	0.014	0.005	0.000	0.008	0.000	-0.000	0.001	-0.000	-0.000
	I	0.518	0.095	0.007	0.457	0.032	0.004	0.178	0.012	0.001
Credit spread	C	0.022	0.029	0.004	0.149	0.006	0.000	0.039	-0.007	0.000
	I	0.209	0.039	0.005	0.258	0.021	0.002	0.161	0.010	0.001
T-bill rate	C	-0.198	-0.416	-0.758	-0.209	-0.350	-0.040	-0.066	-597	-0.787
	I	0.736	0.838	0.928	0.765	0.937	0.961	0.377	0.955	0.958
Change in T-bill	C	-1.172	-0.659	-0.161	-0.913	-0.544	-1.500	-1.070	0.448	3.528
	I	0.975	0.995	0.946	0.909	0.991	0.888	0.966	0.994	0.863
Term spread	C	0.089	0.034	0.013	0.055	-0.027	-0.033	0.029	-0.059	-0.090
	I	0.313	0.109	0.041	0.262	0.074	0.098	0.196	0.067	0.049
Yield spread	C	-0.237	0.000	0.234	-0.008	0.444	0.163	-0.000	2.146	3.772
	I	0.536	0.712	0.516	0.138	0.821	0.486	0.126	0.969	0.825
January dummy	C	0.088	0.043	0.017	0.198	0.025	0.004	0.118	-0.002	-0.000
	I	0.185	0.036	0.010	0.254	0.020	0.004	0.197	0.006	0.000
Indust. Prod. Growth	C	-0.086	-0.068	-0.005	-0.100	-0.062	-0.000	-0.082	0.124	0.003
	I	0.915	0.475	0.050	0.946	0.479	0.068	0.862	0.632	0.035
Inflation	C	-0.052	-0.013	-0.001	-0.118	-0.026	-0.002	-0.244	-0.056	-0.000
	I	0.321	0.054	0.007	0.542	0.078	0.004	0.861	0.084	0.001
Change in Inflation	C	-0.020	-0.030	-0.002	-0.015	-0.008	-0.000	-0.029	-0.004	-0.000
	I	0.094	0.031	0.002	0.089	0.012	0.000	0.094	0.003	0.000
Constant	C	-0.054	-0.503	-0.059	-0.055	-0.201	-0.242	-0.010	-0.014	-0.044
	I	0.172	0.198	0.068	0.159	0.078	0.039	0.130	0.021	0.010

the constant, whose posterior probability amounts to 17.2 percent. These variables are also among the least important in March 1995 and March 2005. In the case of $\delta < 1$, the separation between important and unimportant variables is much more clear-cut. This picture matches the observations made in Table 2 exactly. Our results suggest that models that do not allow for time variation in coefficients have to add explanatory variables in order to get "enough flexibility". However, it is highly questionable if all these variables are, in fact, economically important in predicting equity returns.

A final notable observation includes the surprisingly dynamic behavior of the $\delta = 1$ models between 1995 and 2005. During this time period, the T-bill rate—still crucial in models with $\delta < 1$—loses much importance, which seems to be transferred to the earnings yield and inflation.

As far as the average coefficients (C) reported in Table 3 are concerned, there are two important observations. First, the signs of coefficients are, in most cases, stable across models using different values for δ. Secondly, note that coefficients of individual models for $\delta = 1$ are stable but unobservable by definition and are updated when new observations arrive. The Bayesian model averaging approach, however, introduces time variation of the regression coefficients of the average static model by assigning time varying weights to individual models.

In general, the behavior of models with constant coefficients closely matches patterns of predictive models as described in Pesaran and Timmermann (1995) and Bossaerts and Hillion (1999). Our analysis shows that the erratic changes in importance of predictive variables reported in these papers might be explained by the fact that the models estimated did not allow for time variation in coefficients. For us, the results presented in this section show that empirical approaches using models assuming constant coefficients and selecting individual models should only be relied on with reservations if one wants to identify important predictive variables.

4.4 Characteristics of the Average Prediction Model

While we have been focusing until now on the comparison of models with different assumptions for the time variation of coefficients, we analyze the resulting overall average model in this section. Table 4 reports the importance and the coefficients of explanatory variables in

Table 4: **Sum of Posterior Probabilities of Models Including Specific Variables and Average Coefficients.**

	March 1985		March 1995		March 2005	
	Coef.	Imp.	Coef.	Imp.	Coef.	Imp.
Momentum 1	-0.009	0.187	-0.003	0.118	-0.000	0.017
Momentum 2	-0.009	0.199	-0.011	0.191	-0.001	0.018
Dividend Yield	0.286	0.482	0.292	0.500	0.438	0.569
Earnings Yield	0.148	0.581	0.182	0.565	0.207	0.447
Turnover	0.012	0.431	0.006	0.319	-0.000	0.023
Credit spread	0.024	0.174	0.103	0.177	-0.004	0.020
T-bill Rate	-0.245	0.758	-0.254	0.820	-0.561	0.916
Change in T-bill	-1.063	0.979	-0.797	0.935	0.348	0.992
Term spread	0.078	0.271	0.029	0.201	-0.053	0.075
Yield spread	-0.186	0.571	0.137	0.358	2.002	0.912
January dummy	0.078	0.154	0.142	0.178	0.006	0.019
Indust. Prod. Growth	-0.082	0.821	-0.087	0.794	0.110	0.647
Inflation	-0.044	0.266	-0.088	0.392	-0.068	0.137
Change in Inflation	-0.022	0.081	-0.012	0.064	-0.006	0.010
Constant	-0.143	0.176	-0.102	0.133	-0.014	0.028

the overall average model at the same three points in time as analyzed in the previous tables.

Most important variables are stable across different points in time and include the T-bill rate change, the T-bill rate level, the yield spread, the industrial production, the dividend yield, and the earnings yield. As far as average coefficients are concerned, Table 4 shows that there is quite some variation over time. However, most variables' coefficients maintain their sign over time. One notable exception is the T-bill rate change, which is the most important variable at all three points in time but changes its coefficient's sign between March 1995 and March 2005.

Figure 6 illustrates for a selection of six variables—the most important variables in March 2005, the dividend yield, the January dummy, and the constant—the importance measured as the sum of posterior probabilities assigned to models including this specific variable over time. For the change in T-bill rate Figure 6 shows the dramatic influence of the oil crisis in 1974. The T-bill rate represents another very important variable that suffers to some extent from the oil crisis in 1974 but recovers relatively quickly through the 1980s and 1990s. Interestingly, the yield spread shows a similar dependence on the oil crisis in 1974 as compared

Figure 6: **Importance of Individual Variables.** Sum of posterior probabilities assigned to models including the T-bill rate change (tbcha), T-bill rate level (tbill), yield spread (yield-sprd), dividend yield (divyield), January dummy (jandum), and the constant (Kons) over time.

with the change in T-bill rate but does not succeed in maintaining its importance at a level equal to the change in T-bill rate. This shows that the dominance of the change in T-bill rate in our sample cannot be explained exclusively by the oil crisis.

Figure 6 also shows graphs for three more explanatory variables of special interest. The dividend yield represents the most traditional predictive variables and a large share of the existing literature is devoted to the question of whether the dividend yield predicts equity returns. Our graph shows that the importance of the dividend yield was unaffected by the oil crisis, and it also showed increasing predictive power through the 1980s and decreasing predictive ability during the 1990s. However, recently it seems that the dividend yield again became more important as a predictive variable. Our results fit well into the existing literature. Goyal and Welch (2005), Ang and Bekaert (2005), and Paye and Timmermann (2003) also report empirical evidence that the predictive power of dividend yields suffered during the 1990s. However, our empirical methodology enables us to illustrate the entire pattern of the importance of the dividend yield over time.

Another traditional predictive variable that we would like to analyze briefly is the January dummy. Haugen and Jorion (1996), for example, document that the January effect was still in place during the 1990s and that the magnitude of the effect had not been reduced significantly. Figure 6 shows a comparable pattern for the predictive importance of the January dummy in our empirical study. Although the January dummy does not seem to be a very important predictive variable, it keeps a stable and notable sum of posterior probabilities amounting to approximately 20 percent between 1960 and 2000. Just recently, however, the January dummy's importance dropped to close to zero indicating that nowadays the January effect should have been removed to a large extent for the S&P 500.

Finally, we analyze the constant term in Figure 6: the constant term to some extent represents a benchmark to test for predictability in general. If the constant term turned out to be the most important explanatory variable, it would be a strong signal against predictability. This is, however, not at all the case for our measure of importance. In contrast, the importance of the constant term decreases steadily. However, it is important to keep in mind that the current measure includes all models that include a constant term. In section 4.5 we actually focus in more detail on the model that consists only of the constant term to answer in more detail the question of whether we find predictability.

4.5 Influence of the 1974 Oil Crisis on our Results

The careful reader might have observed the considerable influence of the oil crisis in 1974 on some of the results presented thus far. In order to verify that our conclusions are not driven exclusively by this one event, we basically performed the entire analysis for a data set starting after the oil crisis. In this section, we report very briefly selected results in order to show that results change, of course, but do so quantitatively not qualitatively.

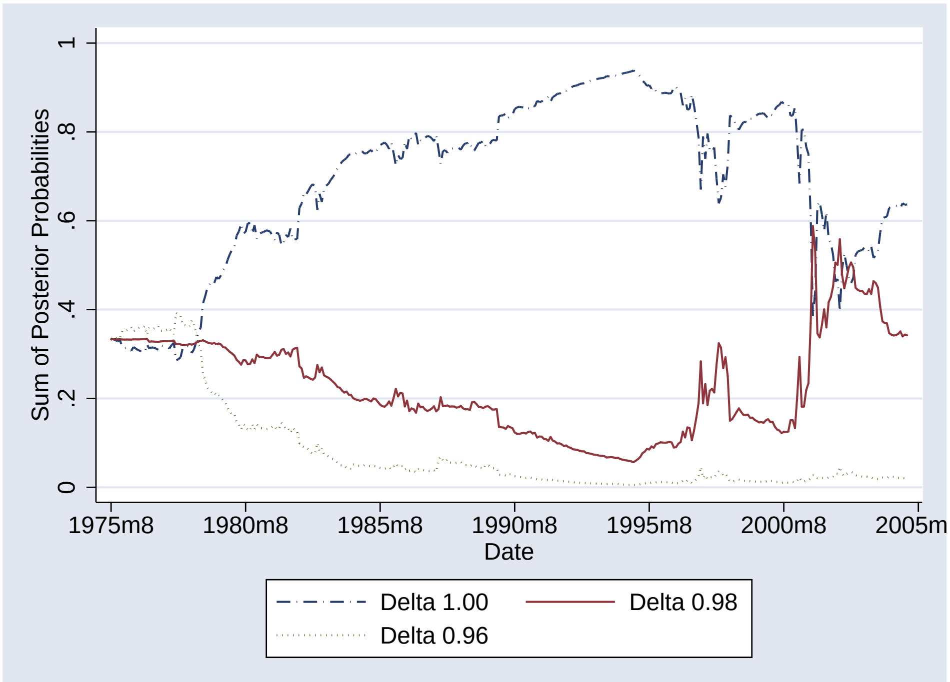
In Figure 7, we again plot the total posterior probability of all models for each δ. Similar to Figure 1, the graph documents that setting $\delta = 0.96$ (i.e., high variability in regression coefficients) is clearly dominated by the other model specifications. The class of static regression models ($\delta = 1.00$) shows outstanding performance until approximately 2000. The class of dynamic regression model with $\delta = 0.98$, however, does not vanish but keeps a certain and notable weight between 10 percent and 30 percent. After 2000—potentially after the burst of the Internet bubble—models with static and dynamic coefficients take an equal share of posterior probability. Therefore, we believe that our conclusion that coefficients in predictive regressions vary to some extent over time is confirmed.

In section 4.3 we documented that the T-bill rate change turns out to be the most important predictive variable after the oil crisis if we use our entire sample for estimation. Figure 8 shows the same variables as depicted in Figure 6. It is interesting to note that all our observations—except for the T-bill rate change—remain valid. As far as the T-bill rate change is concerned, Figure 8 shows that if one ignores the oil crisis then the T-bill rate change does not evolve as an especially important variable. It seems that its dominance if one uses the entire data set, in fact, comes from the oil crisis in 1974.

4.6 Analysis of Out-of-Sample Prediction Accuracy

First, we address the question of overall predictability in our Bayesian framework. Neglecting predictive power of any of the 14 predictive variables used in this study corresponds to taking the average equity premium as the best prediction for the following month's premium.

Figure 7: **Sum of Posterior Probabilities of Models with a Given δ.**



Correspondingly, the non-predictability benchmark model is the model that includes only the constant as a predictor and assumes that the coefficient of the constant does not vary over time (i.e., $\delta = 1.00$). We test the Bayesian performance of the average model as well as the performance of every individual model against this non-predictability benchmark. In Bayesian statistics, performance is measured in the form of likelihood, i.e., as the value of the predictive density function evaluated at the ex post realized return. Good performance of an individual model is rewarded in the form of an upward revision of the posterior probability. From Bayes principle, it follows that the relative change in posterior weights directly corresponds to the performance of the individual model. Thus, the performance of an individual model over one time step relative to the non-predictability benchmark can be determined as the Bayes factor

$$\frac{f(r|M_i,\delta_j)}{f(r|M_{\text{bench}},\delta=1)} = \frac{P(M_i,\delta_j|r)}{P(M_i,\delta_j)} \Big/ \frac{P(M_{\text{bench}},\delta=1|r)}{P(M_{\text{bench}},\delta=1)} \tag{32}$$

Whenever a certain model predicts with higher precision than the non-predictability benchmark model, i.e., its predictive density evaluated at the realized return exceeds the predictive

Figure 8: **Importance of Individual Variables.** Sum of posterior probabilities assigned to models including the T-bill rate change (tbcha), T-bill rate level (tbill), yield spread (yield-sprd), dividend yield (divyield), January dummy (jandum), and the constant (Kons) over time.

Table 5: **Performance of Individual Models Relative to the Non-predictability Benchmark Model.** The number of models that outperform the non-predictability benchmark over the respective time interval is reported.

Time period	Numb. of outperforming models	Time period	Numb. of outperforming models
1951-2005	13600	1951-1960	11065
1960-2005	19958	1960-1970	17211
1970-2005	23580	1970-1980	81382
1980-2005	175	1980-1990	3834
1990-2005	784	1990-2000	6048
2000-2005	8785		

density of the benchmark model, the relative increase in posterior probability exceeds the relative increase in probability of the benchmark model. Let $f_{i,j,t-1}(r_t)$ denote the predictive density given model M_i with time variability δ_j determined with information set D_{t-1} and evaluated at the ex post realized return r_t. The performance over some longer time interval $[\underline{T},\overline{T}]$ is then measured as the joint likelihood of the returns, i.e., the product of the single period performances

$$\frac{\Pi_{t=\underline{T}}^{\overline{T}} f_{i,j,t-1}(r_t)}{\Pi_{t=\underline{T}}^{\overline{T}} f_{\text{bench},t-1}(r_t)} = \frac{P(M_i,\delta_j|r_{\underline{T}},r_{\underline{T}+1},\ldots,r_{\overline{T}})}{P(M_i,\delta_j)} \Big/ \frac{P(M_{\text{bench}},\delta=1|r_{\underline{T}},r_{\underline{T}+1},\ldots,r_{\overline{T}})}{P(M_{\text{bench}},\delta=1)}, \quad (33)$$

which equals the individual models relative change in probability weight over the time interval in relation to the benchmark models relative change in probability. We evaluate all 98,302 individual models relative to the benchmark over different time intervals and report the number of models that outperform the benchmark in Table 5. From Table 5 it is apparent that the benchmark is not the single best performing model in any of the reported time. This clearly indicates in-sample predictability. To attest explicit out-of-sample predictability we must be able to identify ex-ante those models that outperform the benchmark ex-post. Applying Bayesian model averaging, the average model is the proper candidate. Taking into account that the prior as well as the posterior probability of the average model constantly equals one, the one period performance is evaluated as

$$\frac{f(r|\text{average model})}{f(r|M_{\text{bench}},\delta=1)} = \frac{P(M_{\text{bench}},\delta=1)}{P(M_{\text{bench}},\delta=1|r)}. \tag{34}$$

The multi-period performance is determined analogously to Equation (33). For the entire data sample from 1951 to 2005, the average model clearly outperforms the non-predictability benchmark. The average model gains relative weight by a factor of 368.4. Stated differently, the weight of the benchmark within the average model drops from its naive prior of $(1/3)(1/32768) = 1.02 \cdot 10^{-5}$ to $2.76 \cdot 10^{-8}$. Close examination, however, reveals that the performance of the non-predictability benchmark suffers disproportionately during the oil price shock between 1973 and 1975, when its probability weight drops by six orders of magnitude within two years. This is perfectly consistent with the observations of Goyal and Welch (2005) that the OLS out-of-sample significance of certain predictors crucially depends on the inclusion of the oil-price shock in the data sample. Looking at sub-samples, we discover that the average model outperforms the non-predictability benchmark in the 1970s and in the 1990s.

An alternative way to look at out-of-sample predictability is to investigate whether individual DLMs consistently outperform the benchmark. Out of the sample of 3,834 models that outperform the non-predictability benchmark in the 1980s, 551 also outperform the benchmark in the 1990s. Taking the 6,048 models that outperform the benchmark in the 1990s, 166 of them also outperform between 2000 and 2005. However, only eleven models outperform the benchmark in the 1980s, 1990s and between 2000 and 2005. All these models have a variability parameter of $\delta = 0.98$, i.e., only the dynamic models are able to consistently outperform the non-predictability benchmark. These top performing models are, further more, very similar with respect to the choice of predictive variables, e.g., they consist of combinations of the variables yield spread, T-bill rate, industrial production, dividend yield, and earnings yield. These results, of course, create the temptation to change the naive prior in which all models are equally weighted and give more weight to the dynamic coefficient model from the very beginning. We abstain from trying this because repeatedly adapting the prior to find better out-of-sample predictability is, in our belief, simply data mining. Therefore, we conclude

Table 6: **Out-of-sample Accuracy.** Decade 1950 has 106 observations, decades 1960, 1970, 1980, and 1990 have 120 observations, and decade 2000 has 63 observations.

	Hitrate					Squared Error by Total Variance			
Decade	Unc. Mean	Avg.	δ=1	δ=.98	δ=.96	Avg.	δ=1.00	δ=.98	δ=.96
1950	0.632	0.547	0.613	0.547	0.519	0.947	0.974	0.957	0.931
1960	0.583	0.558	0.567	0.542	0.533	0.999	1.004	0.998	0.972
1970	0.475	0.633	0.633	0.633	0.600	1.380	1.399	1.329	1.341
1980	0.583	0.567	0.592	0.550	0.550	1.447	1.450	1.445	1.447
1990	0.667	0.658	0.658	0.633	0.625	0.950	0.949	0.964	0.963
2000	0.508	0.508	0.476	0.587	0.508	1.328	1.409	1.303	1.318
All periods	0.579	0.585	0.599	0.582	0.561	1.167	1.184	1.159	1.153

that the issue of out-of-sample predictability is ambiguous and demands further investigation in future research.

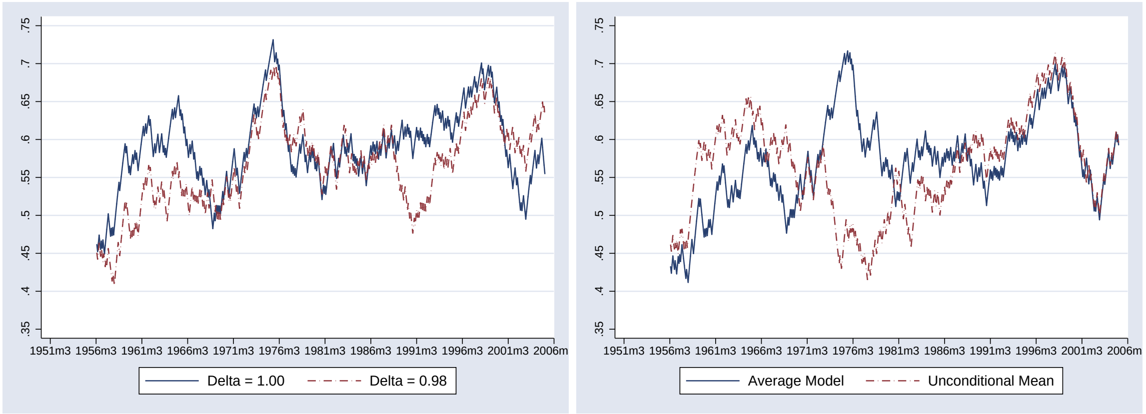
However, posterior probabilities need not necessarily conform to traditional measures of prediction performance. Therefore, we report—for the sake of comparability to other papers—several statistics regarding out-of-sample prediction performance of different model specifications in Table 6. Note, however, that none of these measures are used in estimating or optimizing our predictive models at any point.

We calculate two measures to assess out-of-sample prediction quality. First, we analyze the hit rate, defined as the proportion of periods where the sign of the prediction equals the sign of the realized return. This measure focuses on the direction of excess returns, completely ignoring magnitudes. Second, we investigate the ratio of squared errors and total expected variance of the prediction. This ratio should be close to one if the models predict the variance appropriately. A ratio larger than one can be interpreted as underestimation of true return variance, i.e., overconfidence of the prediction model. In Table 6, we report the averages of these measures per decade for the three values of δ. In addition, the column denoted "Avg." reports values for the overall average model, which represents an average of all models across all δs weighted by the appropriate posterior probabilities. For the hit rate we also include a column denoted "Unc. Mean" which states the hit rate based on predictions using the unconditional mean.

Table 6 shows that models with constant coefficients perform slightly better than models with time varying coefficients with respect to the hit rate. When interpreting this result, one must consider the fact that even for models with constant coefficients some amount of time variation is introduced due to the Bayesian model averaging approach. This observation changes when we focus on the most recent time period where models with time varying coefficients outperform models with static coefficients. Similar to Goyal and Welch (2005) we find that the unconditional mean is a decent predictor of future equity returns. Especially in the first two decades and during the 1990s (confirming observations of Goyal and Welch (2005) and Ang and Bekaert (2005)), it outperforms all other models that use conditioning information. However, the unconditional mean—not surprisingly—suffers strongly from the oil crisis in the 1970s and is also dominated by models using conditioning information in the 1980s and the most recent decade. Looking at the overall picture, the unconditional mean shows a slightly worse hit rate than average models with explanatory variables. As the hit rate seems to be a very popular measure among academics and practitioners, we investigate the development of the hit rate over time for several models using an exponentially weighted average (see Figure 9). The left graph in Figure 9 compares models with constant coefficients and models with time varying coefficients. One clearly recognizes that models with constant coefficients showed dominant performance in the 1950s and 1960s and also in the 1980s. However, models with time varying coefficients have outperformed since 2000. The right graph in Figure 9 compares the overall average model resulting from the Bayesian Model Averaging with the simple model that considers the unconditional mean as the best predictor for the future equity return. As already indicated in Table 6, the model based on the unconditional mean suffers much more from the oil crisis in the 1970s than the average model with explanatory variables.

The final measure summarized in Table 6—squared forecast error divided by total variance—should be close to 1 to indicate that the ex-post variance complies with the ex-ante prediction variance. The picture looks different with respect to this variable. The model with $\delta = 1.00$ is dominated by the models allowing for time variation. It seems that the model with static coefficients misestimates the uncertainty of its predictions. This is a promising result that indicates that asset allocation techniques that take uncertainty about the prediction into

Figure 9: **Comparison of Hit Rates.** The graph shows the exponentially weighted average of the hit rate of a given model up to this point in time. Left graph: $\delta = 1.00$ vs. $\delta = 0.98$; Right graph: hit rate of the average model using conditioning information vs. the unconditional mean.



account could take advantage of the improved estimation of prediction uncertainty of models with time varying coefficients.

5 Conclusion

This is the first paper to explicitly consider time variation and uncertainty about coefficients and test performance compared to models with static regression coefficients. We estimate dynamic linear models to predict S&P 500 returns and use Bayesian model averaging to address the issue of model selection. Our paper shows empirical evidence for the existence of time variation in regression coefficients. Using data for the period from March 1951 to March 2005, the total posterior probability assigned to models with static regression coefficients is less than 10 percent and this result is robust for changes in the prior. While static models dominate in the beginning of the analyzed horizon, dynamic models successfully compete with static models in the late twentieth century. In the period from 2000 to 2005, models with time varying coefficients are assigned approximately 90 percent of the posterior probability. It is further interesting to note that the uncertainty about the true level of time variation is of equal magnitude than the uncertainty about the true choice of predictive variables.

Comparing static models with highest posterior probability to best-performers among the class of dynamic models, we find that “good” static models tend to be larger than “good” dynamic models. Thus, static models seem to compensate for the lack of flexibility caused by assuming constant coefficients by adding additional variables into the models. Therefore, relatively large models play an important role and there is a steady turnover in the class of top-performing models. This offers an explanation for an important issue that has been raised, for example, in Pesaran and Timmermann (1995) and Bossaerts and Hillion (1999), namely, that variables included in top predictive models change erratically over time. We show that if we consider the fact that coefficients of predictive variables vary over time, we are able to identify top models with respect to their posterior probabilities that are comparatively stable over time. In contrast, if we force coefficients to be constant over time, we observe large fluctuations among top-performing models.

Furthermore, our results are able to shed new light on specific questions in equity return prediction. For example, we focus on the importance of the dividend yield as a predictive variable. We document that the dividend yield became a less important predictive variable during the 1990s but has regained large parts of its explanatory power since 2000. Another frequently discussed issue in equity return prediction is the January effect. We find that models including the January dummy receive a relatively constant amount of posterior probability of around 20 percent between the 1960s and the 1990s. However, according to our analysis, the January dummy has lost predictive importance for the S&P 500 since 2000.

The final result of our paper addresses the general issue of whether predictability exists. Using various measures and empirical analyses we identify in-sample predictability but fail to unambiguously identify out-of-sample predictability. However, we want to emphasize that only predictive models allowing explicitly for time varying coefficients succeed in consistently outperforming the non-predictability benchmark across all periods of time. We interpret this observation as another indication that allowing for time variation in coefficients is important in identifying models with predictive power.

References

Ang, A., and G. Bekaert, 2005, Stock return predictability:is it there?, working paper.

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

Barberis, N., 2000, Investing for the long run when returns are predictable, *Journal of Finance* 55, 225–264.

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

Campbell, J., and S. B. Thompson, 2005, Predicting the equity return premium out of sample: Can anything beat the historical average?, working paper.

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

Dimson, E., P. Marsh, and M. Staunton, 2002, *Triumph of the Optimists: 101 Years of Global Investment Returns* (Princeton University Press).

Goyal, A., and I. Welch, 2005, A comprehensive look at the empirical performance of equity return prediction, working paper.

Haugen, R. A., and P. Jorion, 1996, The january effect: Still there after all these years, *Financial Analysts Journal* 52, 27–31.

Jostova, G., and A. Philipov, 2005, Bayesian analysis of stochastic betas, *Journal of Financial and Quantitative Analysis* 40, 747–778.

Kandel, S., and R. F. Stambaugh, 1996, On the predictability of stock returns: An asset-allocation perspective, *Journal of Finance* 51, 385–424.

Kim, C.-J., J. C. Morley, and C. R. Nelson, 2000, Is there a structural break in the equity premium?, working paper.

MacKinlay, A. C., and L. Pastor, 2000, Asset pricing models: Implications for expected returns and portfolio selection, *The Review of Financial Studies* 13, 883–916.

Pastor, L., 2000, Portfolio selection and asset pricing models, *The Journal of Finance* 55, 179–223.

——— , and R. F. Stambaugh, 2000, Comparing asset pricing models: An investment perspective, *Journal of Financial Economics* 56, 335–381.

——— , 2001, The equity premium and structural breaks, *The Journal of Finance* 56, 1207–1239.

Paye, B. S., and A. Timmermann, 2003, Instability of return prediction models, working paper.

Pesaran, M. H., and A. Timmermann, 1995, Predictability of stock returns: Robustness and economic significance, *The Journal of Finance* 50, 1201–1228.

Pettenuzzo, D., and A. Timmermann, 2005, Predictability of stock returns and asset allocation under structural breaks, working paper.

Raftery, A. E., D. Madigan, and J. A. Hoeting, 1997, Bayesian model averaging for linear regression models, *Journal of the American Statistical Association* 92, 179–191.

Viceira, L. M., 1997, Testing for structural change in the predictability of asset returns, working paper.

West, M., and J. Harrison, 1997, *Bayesian Forecasting and Dynamic Models* (Springer Verlag).

Xia, Y., 2001, Learning about predictability: The effects of parameter uncertainty on dynamic asset allocation, *The Journal of Finance* 56, 205–246.

Zellner, A., 1986, *On assessing prior distribution and Bayesian regression analysis with g-prior distributions* (North Holland) In Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti, eds. P.K. Goel and A. Zellner.