

Predictive Regressions with Time-Varying Coefficients *

Thomas Dangl
Vienna University of Technology and
ISK

Michael Halling
University of Vienna

November 19th, 2006

Abstract

Papers in equity return prediction usually rely on the assumption of constant coefficients in linear predictive models. We question and relax this assumption and find strong empirical support for models with time-varying regression coefficients. Analyzing model uncertainty, we document that uncertainty about the level of time-variation in coefficients and uncertainty about the choice of predictive variables are similarly important sources of prediction variance. Furthermore, we document out-of-sample predictability of the average model and individual predictive models. Most importantly, only individual predictive models with time-varying coefficients show consistent out-of-sample predictability after the oil price shock in 1974.

JEL Classifications: G12, C11

Keywords: Empirical asset pricing, equity return prediction, Bayesian econometrics.

*Acknowledgment: This paper is based on Michael Halling's dissertation at the University of Vienna. Special thanks belong to Michael's committee members - Josef Zechner (chair), Lubos Pastor and Elroy Dimson. Furthermore, we thank M. Baltutis, P. Bossaerts, M. Cremers, H. Freitas Lopes, B. Hanke, K. Ritzberger, Z. Rachev, F. Schausberger, A. Stomper, A. Timmermann, C. Vega, M. Wagner, and Y. Wu for their helpful comments and participants of the 10th Symposium on Finance, Banking and Insurance (2005), the VGSF Research Seminar in Vienna (2006), the NHH Skinance Conference (2006) and the Meeting of the European Finance Association in Zurich (2006) for their discussions. The paper previously circulated under a different title, "Equity Return Prediction: Are Coefficients Time-Varying?", and with another co-author, Otto Randl, whom we particularly thank for his input to earliest versions of the paper.

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 existing results (e.g., Campbell and Thompson (2005), Ang and Bekaert (2006), Cooper and Gulen (2006) and Goyal and Welch (2006)).

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 estimation 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 $2^k - 1$ different models). A seminal paper is Pesaran and Timmermann (1995), which represents the first paper that acknowledges this problem, looks at all possible models and determines the best predictive model using standard diagnostic statistics. Similarly, Bossaerts and Hillion (1999) evaluate several approaches for the use of statistical criteria to select the best return forecasting model. In contrast to these papers which do not explicitly account for model uncertainty, Avramov (2002) and Cremers (2002) address model uncertainty in a Bayesian setup by determining an average model across all possible models using Bayesian Model Averaging rather than selecting an individual model (see Raftery, Madigan, and Hoeting (1997)).

In this paper we extend the literature in two important directions. First, we estimate predictive regressions that allow for time-varying coefficients.¹ For the returns of the S&P 500, we then compare the predictive performance of regressions with constant coefficients (i.e., the standard empirical models in the prediction literature) to the performance of models with time-varying coefficients using different out-of-sample measures. While static models dominate in the beginning of the analyzed horizon (i.e., March 1951 to March 2005), dynamic models successfully compete with static models in the late twentieth century. Models with time-varying coefficients also outperform models with static coefficients with respect to adjusted mean squared prediction errors. Finally and most importantly, only predictive models with time varying coefficients succeed in consistently outperforming the no-predictability benchmark, i.e., the unconditional mean return, after the oil price shocks of the 1970ies.

Second, we explicitly take into account that we are uncertain about the true level of time-variation. This uncertainty represents another dimension of model uncertainty that has so far been ignored in the prediction literature. As a consequence we distinguish the following sources of prediction uncertainty in this paper: (i) the variance in predictive variables, (ii) the estimation uncertainty in coefficients, (iii) the model uncertainty with respect to the choice of predictive variables, and (iv) the model uncertainty with respect to the time-variation in coef-

¹There are several reasons why coefficients might vary over time, e.g., due to changes in market sentiments, in monetary policies, in the institutional framework or in macroeconomic conditions.

ficients. To determine the importance of these different sources of uncertainty, we decompose the total prediction variance and measure the contribution of each of the four sources. Our empirical analysis illustrates that the first two sources, as expected, are most important. The two dimensions of model uncertainty are, however, non-negligible. Interestingly, we find that accounting for the uncertainty in time variation of coefficients is frequently as important as accounting for the uncertainty in variable selection.

Finally, we use our framework to shed new light on various results and puzzles found in the literature before. We document in this paper that incorrect inferences about the importance or non-importance of explanatory variables may occur if time variation in coefficients is ignored. 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 (2006) and Paye and Timmermann (2003)—that the dividend yield became a less important predictive variable during the 1990s. However, we show empirically that it has regained large parts of its explanatory power since 2000.

One specific puzzle in equity return prediction for which we offer an explanation is the behavior that the best predictive models identified in Pesaran and Timmermann (1995) and Bossaerts and Hillion (1999) have been so unstable. 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 the reason for this behavior might be the assumption of constant coefficients. Best models with constant coefficients fluctuate considerably from one time period to the next, while best models with time-varying coefficients are comparatively stable over time.

While our paper is the first to explicitly model time-varying coefficients in predictive regressions, there is ample descriptive evidence that economic relationships and, as a consequence, coefficients in predictive regressions vary over time.² 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. Ang and Bekaert (2006) test for time variation in coefficients in an ad-hoc way by splitting their entire sample into different sub-periods. They 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. In this paper, we go much further than these papers by providing empirical evidence on time-varying coefficients in a rigorous econometric framework.

Another stream of literature related to our paper consists of papers that estimate regime

²Barsky (1989) documents time-varying stock-bond correlations. Dimson, Marsh, and Staunton (2002) present empirical evidence on time-varying correlations between various economic variables. McQueen and Roley (1993) and Boyd, Hu, and Jagannathan (2005) find that the incorporation of news into stock prices varies with the business cycle.

switching models and search for structural breaks in the predictive relationship between equity returns and explanatory variables.³ 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 in the relation between the dividend yield and equity returns. Paye and Timmermann (2003), in contrast, identify several structural breaks in the coefficients of state variables. We differ from these papers as we do not approximate the time variation in coefficients by a step function, but allow the coefficients to change gradually over time.⁴ For this purpose we apply a Bayesian econometric method (following West and Harrison (1997)) that allows us to model time varying coefficients that are subject to random shocks. The uncertainty about the “level” of coefficients’ time-variation is treated as another dimension of model uncertainty, which we address in a consistent manner within the Bayesian Model Averaging approach (see Raftery, Madigan, and Hoeting (1997) for technical details and Avramov (2002) and Cremers (2002) for applications to return prediction).⁵

The paper is structured in the following way. Section 2 presents the empirical methodology. Section 3 describes the data sources and variables used in the empirical study. Section 4 reports empirical results and discusses their implications. Section 5 presents robustness checks and Section 6 concludes.

2 Prediction Models with Time-Varying Coefficients

Similar to the vast majority of papers on return prediction (see, for example, Pesaran and Timmermann (1995), Bossaerts and Hillion (1999), Avramov (2002), Cremers (2002), Goyal and Welch (2006), and Ang and Bekaert (2006)) we assume a linear relationship between predictive variables (chosen from a set of k candidate variables, including a constant) and the dependent variable, i.e., the excess return r of some asset. However, while these papers assume that the unobservable regression coefficients θ are constant over time, we doubt that this assumption is valid and model the coefficients in our Dynamic Linear Models to be

³Pastor and Stambaugh (2001) and Kim, Morley, and Nelson (2005) use Bayesian econometrics to identify structural breaks in equity premia. Both papers report that they identify empirical evidence for the existence of structural breaks. They differ, however, quite considerably in the timing of the breaks. Interestingly, Kim, Morley, and Nelson (2005) do not find evidence for structural breaks in the post-WWII period (i.e., the period that we analyze in this paper).

⁴Note 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 condition stock market betas on observables while we allow for time-varying coefficients when regressing an equity market index on a set of predictive variables.

⁵Another stream of literature that is to a lesser extent related to our paper is the one on portfolio selection under uncertainty. Kandel and Stambaugh (1996), Barberis (2000), and Xia (2001) 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 allocation than other sources of risk such as uncertainty in parameter estimation.

time-varying (see Section 2.1). The central contribution of our paper is to evaluate whether the data supports time-varying coefficients or whether it confirms the constant coefficient paradigm. After having estimated the $2^k - 1$ Dynamic Linear Models that result from all possible selections of predictive variables, we use a Bayesian model selection criterion to assign posterior probability weights to the individual models (similar to Avramov (2002) and Cremers (2002)) and determine an average prediction model (see Section 2.2).

The goal of this econometric approach is to provide a flexible prediction framework that explicitly accounts for the different sources of uncertainty that arise in the choice of predictive variables to be used in the regression, uncertainty in the estimation of coefficients, uncertainty in the degree of variability of the regression coefficients, and observational uncertainty. In the following Section 2.1 we focus on outlining the characteristics of an individual dynamic linear prediction model (i.e., for a given choice of predictive variables), in Section 2.2 we discuss the Bayesian model selection approach, and in Section 2.3 we present the decomposition of the prediction variance into the already mentioned building blocks.

2.1 Dynamic Linear Models

In this section we develop dynamic linear models (according to West and Harrison (1997)) that explicitly allow for a time varying nature of the linear relationship between the asset return r_{t+1} over the interval $(t, t + 1]$ and the vector X_t of realizations of the explanatory variables observed at time t .⁶ More specifically, we estimate models of the form

$$r_{t+1} = X_t' \theta_t + v_{t+1}, \quad v \sim N(0, V) \quad (\text{observation equation}), \quad (1)$$

$$\theta_t = \theta_{t-1} + \omega_t \quad \omega \sim N(0, W_t) \quad (\text{system equation}), \quad (2)$$

The vector θ_t consists of unobservable, time varying regression coefficients, and the observational error v is assumed to be normally distributed with mean 0 and constant but unknown variance V . While Equation (2) states that these coefficients are exposed to random shocks ω which are jointly normal (with mean 0 and variance matrix W_t), there are no systematic movements in θ assumed.

If the system variance matrix W_t equals 0, the regression coefficients θ_t are constant over time. Thus, our model nests the specification of constant regression coefficients. If W_t is large, the intrinsic variability of the regression coefficients θ_t increases the flexibility of the model. At the same time the out of sample prediction variance, however, increases and makes the predictions unreliable and, consequently, useless. The specific structure, which we impose on W_t and how we estimate the magnitude of time variation of the underlying coefficients will be explained in more detail below.

Let $D_t = [r_t, r_{t-1}, \dots, X_t, X_{t-1}, \dots, \text{Priors}_{t=0}]$ denote the information set available at time t . This information set contains all returns, all corresponding realizations of the predictive

⁶Observable variables have a subscript that indicates the time at which they are known. When speaking about beliefs regarding non-observable variables, like the regression coefficients and the variance V , we state the information set on which these beliefs are conditioned.

variables up to time t and our initial time zero choice of priors regarding θ and V . We will now describe how, at some arbitrary time $t + 1$, the observation of a new return realization leads to an update of the estimated system coefficients and the estimated observational variance V .

Following West and Harrison (1997)), we develop the updating recurrence in a fully conjugate Bayesian analysis ensuring that prior and posterior distributions come from the same family of distributions. Specifically, we use a normally distributed prior for the system coefficients θ_0 and an inverse-gamma distributed prior for the observational variance V . To specify the prior information at time $t = 0$, we use the following 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)).

$$V|D_0 \sim IG\left[\frac{1}{2}, \frac{1}{2}S_0\right], \quad (3)$$

$$\theta_0|D_0, V \sim N[0, gS_0(X'X)^{-1}]. \quad (4)$$

where

$$S_0 = \frac{1}{N-1}r'(I - X'(X'X)^{-1}X)r. \quad (5)$$

This is a noninformative prior, which is consistent with the null-hypothesis of no-predictability and where g serves as the scaling factor that determines the confidence assigned to the null-hypothesis of no-predictability. Thus, the prior for the coefficient vector $\theta_0|D_0$ is centered around zero and the covariances among coefficients are multiples of the OLS estimate of the variance in coefficients.

Suppose at some arbitrary time t we have already observed the current return r_t . Hence, we are able to form a posterior belief about the values of the unobservable coefficients $\theta_{t-1}|D_t$ and of the observational variance $V|D_t$. These posteriors are again jointly normally/inverse-gamma distributed of the form

$$V|D_t \sim IG\left[\frac{n_t}{2}, \frac{n_t S_t}{2}\right], \quad (6)$$

$$\theta_{t-1}|D_t, V \sim N[m_t, VC_t^*], \quad (7)$$

where S_t is the mean of the time t estimate of the observational variance V and n_t is the associated number of degrees-of-freedom. The vector m_t denotes the point estimate of the vector of coefficients θ_{t-1} conditional on D_t and V . C_t^* is the estimated, conditional covariance matrix of θ_{t-1} normalized by the observational variance. This assumption implies that unconditionally on V the posteriors of the coefficients are multivariate t -distributed given by

$$\theta_{t-1}|D_t \sim T_{n_t}[m_t, S_t C_t^*]. \quad (8)$$

When iteratively updating the estimates it must be regarded that due to varying regression coefficients the posterior distribution of $\theta_{t-1}|D_t$ does not automatically become the prior

distribution of $\theta_t|D_t$. According to Equation (2), the underlying regression coefficients are exposed to Gaussian shocks without a systematic component, which increase the variance but preserve the mean of the estimate,

$$\theta_t|D_t \sim T_{n_t} [m_t, S_t C_t^* + W_t]. \quad (9)$$

With this joint prior distribution of the observational variance and the coefficients, we are able to calculate a forecast of the time $t + 1$ return r_{t+1} by integrating over the entire range of θ and V . Let $\varphi(x; \mu, \sigma^2)$ denote the density of a (possibly multivariate) normal distribution evaluated at x and $ig(V; a, b)$ the density of a $IG[a, b]$ distributed variable evaluated at V , then the predictive density is

$$\begin{aligned} f(r_{t+1}|D_t) &= \int_0^\infty \left[\int_\theta \varphi(r_t; X_t' \theta, V) \varphi(\theta; m_t, V C_t^* + W_t) d\theta \right] \\ &\quad \times ig\left(\frac{n_t}{2}, \frac{n_t S_t}{2}\right) dV \\ &= \int_0^\infty \varphi(r_t; X_t' m_t, X_t' (V C_t^* + W_t) X_t + V) \\ &\quad \times ig\left(\frac{n_t}{2}, \frac{n_t S_t}{2}\right) dV \\ &= \mathbf{t}_{n_t}(r_{t+1}; \hat{r}_{t+1}, Q_{t+1}), \end{aligned} \quad (10)$$

where $\mathbf{t}(r_{t+1}; \hat{r}_{t+1}, Q_{t+1})$ is the density of a Student- t -distribution with n_t degrees of freedom, mean \hat{r}_{t+1} , variance Q_{t+1} , evaluated at r_{t+1} . The mean of the predictive distribution of r_{t+1} is given by

$$\hat{r}_{t+1} = X_t' m_t \quad (11)$$

since the prior of the regression coefficients is centered at m_t . The total unconditional variance of the predictive distribution is given by

$$Q_{t+1} = X_t' R_t X_t + S_t, \quad (12)$$

$$R_t = S_t C_t^* + W_t, \quad (13)$$

where R_t denotes the unconditional variance of the time t -prior of the coefficient vector θ_t . The first term in (12) characterizes the variance coming from uncertainty in the estimation of θ_t , the second term S_t is the estimate of the variance of the error term in the observation equation.

After the time $t + 1$ return r_{t+1} is observed, the priors about θ_t and V are updated using equations (14) to (19).

$$e_{t+1} = r_{t+1} - \hat{r}_{t+1} \quad (\text{error in prediction}). \quad (14)$$

The prediction error is the essential signal conditioning learning. Whenever e_{t+1} equals zero, the observed return equals the forecast, and thus, there is no updating in the coefficients.

$$n_{t+1} = n_t + 1 \quad (\text{degrees of freedom}). \quad (15)$$

$$S_{t+1} = S_t + \frac{S_t}{n_t} \left(\frac{e_{t+1}^2}{Q_{t+1}} - 1 \right) \quad (\text{estimator of observational variance}). \quad (16)$$

Since the total variance of the forecast is given by Q_{t+1} , we have $E(e_{t+1}^2) = Q_{t+1}$. If the error in prediction coincides with its expectation, i.e., $e_{t+1}^2 = Q_{t+1}$, the estimate of the observational variance is unchanged, i.e., $S_{t+1} = S_t$. A prediction error below the expected error leads to a reduction in the estimated observational variance, and vice versa. The adaptive vector

$$A_{t+1} = \frac{R_t X_t}{Q_{t+1}} \quad (\text{adaptive vector}) \quad (17)$$

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:

$$m_{t+1} = m_t + A_{t+1} e_{t+1} \quad (\text{estimator for expected coefficient vector}), \quad (18)$$

$$C_{t+1}^* = \frac{1}{S_t} (R_t - A_{t+1} A_{t+1}' Q_{t+1}) \quad (\text{estimator for variance of coeff. vector}). \quad (19)$$

It is still open to specify the system variance matrix W . To give structure to W_t we apply a discount factor approach. This approach relies on the assumption that the variance matrix W_t of the error term ω_t is proportional to the estimation variance $S_t C_t^*$ of the coefficient vector $\theta_t | D_t$. More precisely, it is assumed that

$$W_t = \frac{1 - \delta}{\delta} S_t C_t^*, \quad \delta \in \{\delta_1, \delta_2, \dots, \delta_d\}, \quad 0 < \delta_i \leq 1, \quad (20)$$

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

$$R_t = S_t C_t^* + \frac{1 - \delta}{\delta} S_t C_t^* = \frac{1}{\delta} S_t C_t^*, \quad (21)$$

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 evaluated in the vast majority on equity return prediction. Choosing a discount factor $\delta < 1$ explicitly assumes variability of the underlying regression parameters. As a consequence, the prediction of one particular dynamic linear model depends not only on the choice of the predictive variables but also

on the choice of δ . Both these choices represent model uncertainty, which we address in a Bayesian Model Averaging framework.

2.2 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 X_t used as regressors. Agreeing on k candidate regressors (including the constant) alone implies $2^k - 1$ 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 $d \cdot (2^k - 1)$ possible dynamic linear models.

The arbitrary choice of one particular model from this substantial pool of possible models is always debatable. Bayesian model selection (see Avramov (2002) and Cremers (2002)) offers a systematic approach to this problem that tests the reliability of all $d \cdot (2^k - 1)$ models against the observed data. Starting from an uninformed prior, it assigns posterior probabilities to each model. By averaging across models one can determine posterior probabilities of inclusion for each of the candidate variables and δ values. 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, different values of δ 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, and δ_j a certain selection from the set $\{\delta_1, \delta_2, \dots, \delta_d\}$. Certainly, these choices crucially influence the predictive density of the forecasts of the individual models, thus we rewrite the point estimate of r_{t+1} as

$$\hat{r}_{t+1,i}^j = E(r_{t+1}|M_i, \delta_j, D_t) = X_t' m_t | M_i, \delta_j, D_t. \quad (22)$$

When giving prior weights to the individual models, we start out with the diffuse conditional prior $P(M_i|\delta_j, D_0) = 1/(2^k - 1) \forall i$. We use Bayes' rule to obtain the posterior probabilities

$$P(M_i|\delta_j, D_t) = \frac{f(r_t|M_i, \delta_j, D_{t-1})P(M_i|\delta_j, D_{t-1})}{f(r_t|\delta_j, D_{t-1})} \quad (23)$$

where

$$f(r_t|\delta_j, D_{t-1}) = \sum_M f(r_t|M_i, \delta_j, D_{t-1})P(M_i, \delta_j, D_{t-1}) \quad (24)$$

The crucial part is the conditional density

$$f(r_t|M_i, \delta_j, D_{t-1}) \sim \frac{1}{\sqrt{Q_{t,i}^j}} \mathbf{t}_{n_{t-1}} \left(\frac{r_t - \hat{r}_{t,i}^j}{\sqrt{Q_{t,i}^j}} \right) \quad (25)$$

where $\mathbf{t}_{n_{t-1}}$ is the density of a Student- t -distribution and $\hat{r}_{t,i}^j$ and $Q_{t,i}^j$ are the respective point estimates and variance of the predictive distribution for model M_i and given $\delta = \delta_j$, see Equation (10). The time $t + 1$ return prediction of the average model for a given $\delta = \delta_j$, then equals

$$\hat{r}_{t+1}^j = \sum_{i=1}^{2^k-1} P(M_i|\delta_j, D_t) \hat{r}_{t+1,i}^j. \quad (26)$$

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

$$P(\delta_j|D_t) = \frac{f(r_t|\delta_j, D_{t-1})P(\delta_j|D_{t-1})}{\sum_{\delta} f(r_t|\delta, D_{t-1})P(\delta|D_{t-1})}. \quad (27)$$

Note that this posterior probability is going to be of key importance in our empirical analysis as it shows which assumptions on time-variation are supported by the data.

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

$$P(M_i, \delta_j|D_t) = P(M_i|\delta_j, D_t)P(\delta_j|D_t) \quad (28)$$

and the unconditional average prediction of the average model is

$$\hat{r}_{t+1} = \sum_{j=1}^d P(\delta_j|D_t) \hat{r}_{t+1}^j. \quad (29)$$

2.3 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 prediction variance of the return

into four parts:

$$\begin{aligned}
\text{Var}(r_{t+1}) = & \sum_j \left[\sum_i (S_t | M_i, \delta_j | D_t) P(M_i | \delta_j, D_t) \right] P(\delta_j | D_t) + \\
& \sum_j \left[\sum_i (X_t' R_t X_t | M_i, \delta_j, D_t) P(M_i | \delta_j, D_t) \right] P(\delta_j | D_t) + \\
& \sum_j \left[\sum_i (\hat{r}_{t+1,i}^j - \hat{r}_{t+1}^j)^2 P(M_i | \delta_j, D_t) \right] P(\delta_j | D_t) + \\
& \sum_j (\hat{r}_{t+1}^j - \hat{r}_{t+1})^2 P(\delta_j | D_t). \tag{30}
\end{aligned}$$

Equation (30) 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

$$\text{Var}(r) = E_\delta(\text{Var}(r|\delta)) + \text{Var}_\delta(E(r|\delta)), \tag{31}$$

where E_δ and Var_δ denote the expected value and the variance with respect to δ . The term $E_\delta(\text{Var}(r|\delta))$ represents the first three terms in Equation (30). The term $\text{Var}_\delta(E(r|\delta))$ is the last term in (30). In a second step, the term $E_\delta(\text{Var}(r|\delta))$ can be further decomposed into

$$\text{Var}(r|\delta) = E_M(\text{Var}(r|M, \delta)) + \text{Var}_M(E(r|M, \delta)), \tag{32}$$

which splits term three of Equation (30) from the remainder. The final variance decomposition as shown in (30) follows from simple rearrangements.

The individual terms of (30) can be interpreted in a very intuitive way. The first term is the expected observational variance, i.e., it is the time $t + 1$ prior probability-weighted average over all estimations of the observational variance S_t . 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 $d \cdot (2^k - 1)$ 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 uncertainty with respect to the time variability of the regression coefficients.

2.4 Performance Measures

Predictive regression models are usually evaluated against a non-predictability benchmark model. This non-predictability benchmark model, denoted M_{bench} , assumes that none of the predictors adds information on top of a long term market risk premium, i.e., it is the model that includes only a non time-varying constant term in the regression. In Bayesian statistics,

performance is measured in the form of ex post 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 relative performance of individual models. Thus, the performance of an individual model over the time step from time $t - 1$ to time t relative to the no-predictability benchmark can be determined as the Bayes factor

$$B_t^{i,j} = \frac{f(r_t|M_i, \delta_j, D_{t-1})}{f(r_t|M_{\text{bench}}, \delta = 1, D_{t-1})} = \frac{P(M_i, \delta_j, D_t)}{P(M_i, \delta_j, D_{t-1})} / \frac{P(M_{\text{bench}}, \delta = 1, D_t)}{P(M_{\text{bench}}, \delta = 1, D_{t-1})}. \quad (33)$$

Whenever a certain model predicts with higher precision than the no-predictability benchmark model, i.e., its predictive density evaluated at the realized return exceeds the predictive density of the benchmark model, the relative increase in posterior probability exceeds the relative increase in probability of the benchmark model. The performance over some longer time interval $[\underline{T}, \bar{T}]$ is consequently measured as the joint realized relative likelihood of the returns over this interval, i.e., the product of the single period performances

$$\Pi_{t=\underline{T}}^{\bar{T}} B_t^{i,j} = \frac{P(M_i, \delta_j, D_{\bar{T}})}{P(M_i, \delta_j, D_{\underline{T}})} / \frac{P(M_{\text{bench}}, \delta = 1, D_{\bar{T}})}{P(M_{\text{bench}}, \delta = 1, D_{\underline{T}})}, \quad (34)$$

which is the individual model's relative change in probability weight over the time interval in relation to the benchmark model's relative change in probability.

Taking into account that the prior as well as the posterior probability of the average model constantly equals one, the one period performance of the average model is evaluated as

$$B_t^{\text{average model}} = \frac{f(r_t|\text{average model}, D_{t-1})}{f(r_t|M_{\text{bench}}, \delta = 1, D_{t-1})} = \frac{P(M_{\text{bench}}, \delta = 1, D_t)}{P(M_{\text{bench}}, \delta = 1, D_{t-1})}. \quad (35)$$

The multi-period performance is determined analogously to Equation (34).

In addition to these Bayesian performance criteria we analyze the mean squared prediction error (MSPE) as a standard statistic of prediction accuracy. Clark and West (2006), however, point out that, for the purpose of comparing the performance of two nested prediction models, the MSPEs have to be adjusted for different model specifications.⁷ Each model's adjustment depends on the variation in the model's prediction⁸ and is calculated as (for any

⁷We are well aware of the fact that Clark and West (2006) do not directly address our class of models. Nevertheless, we are confident that their basic idea can be applied to the evaluation and performance comparison of our models.

⁸Consider the following example of comparing a model with only a constant to a model with several explanatory variables. The idea of this adjustment is to account for the fact that the explanatory variables considered in the second model introduce additional noise and consequently bias the MSPE.

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

individual or average model)

$$mspe_{adj} = \frac{1}{T} \sum_{t=0}^T [(r_{t+1} - \hat{r}_{t+1})^2 - \hat{r}_{t+1}^2]. \quad (36)$$

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 δ , where we use the following values in our empirical implementation: 1.00, 0.98, and 0.96. We choose the values of δ such that we cover the static case ($\delta=1.00$), a rather noisy situation where coefficients are expected to change rapidly ($\delta=0.96$), and an intermediate case ($\delta=0.98$). As described in Section 2.1, the effect of δ strictly lower than 1.00 is an increase in the variance 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

δ equal to 0.98.⁹ For δ equal to 0.96, the 50 percent increase will be reached twice as fast, in approximately 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 Performance of Models with Time-Varying Coefficients

In a first step we analyze the empirical support received by models with constant and time-varying coefficients, respectively. 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 value of δ considered. The figure shows that the posterior probabilities start to depart significantly from the unconditional weights (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.¹⁰ 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 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 posterior 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.

While Figure 1 documents the level of posterior probabilities for each class of models, Figure 2 analyzes the Bayes Factor (see Section 2.4 for a definition). I.e., it analyzes the change in posterior probability of the average model with time varying coefficients ($\delta = 0.98$) relative to the change in posterior probability of the average model with constant coefficients ($\delta = 1.00$) over the entire set of possible time intervals defined by Start Date and End Date. Whenever the model with time varying coefficients outperforms the model with constant coefficients over a certain interval (i.e., the Bayes Factor > 1) this is indicated by a black dot at the coordinates that characterize the respective time period. The particular strength of this representation is that it clearly illustrates the influence of the evaluation period on model performance.¹¹

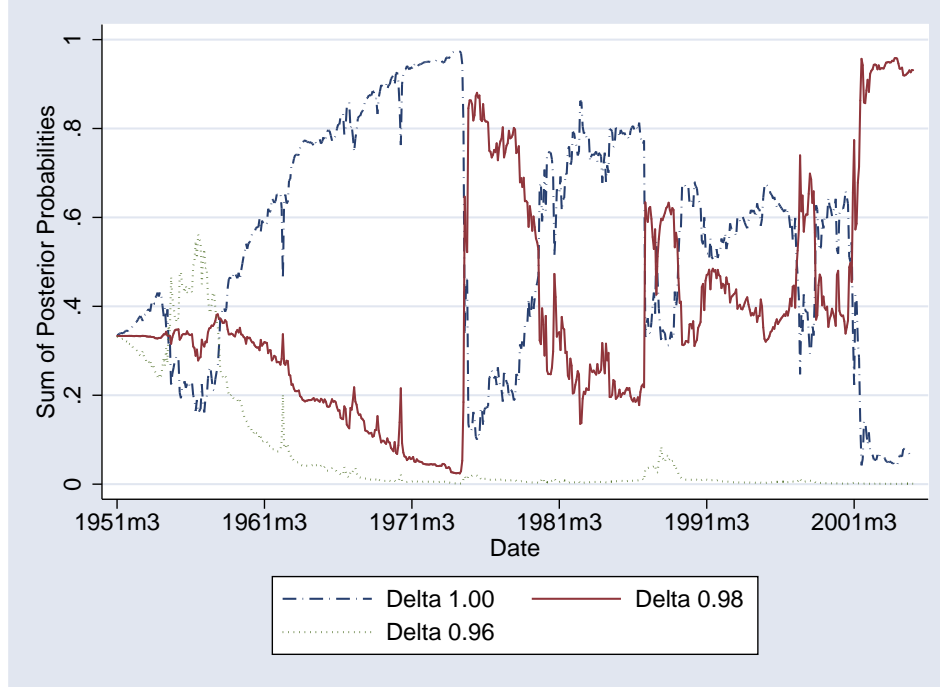
There are several interesting observations to be made. First, for nearly every start date the average model with time-varying coefficients outperforms the average model with constant

⁹Since 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.

¹⁰Consequently, we are going to concentrate on models with $\delta = 0.98$ and $\delta = 1.0$ in the remainder of the paper. Results for models with $\delta = 0.96$ can be obtained from the authors on request.

¹¹ Goyal and Welch (2006) and Cooper and Gulen (2006) have shown — only for models with constant coefficients — that the definition of the evaluation period can have an important influence on model performance.

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



coefficients if evaluated over the entire remaining sample period. Second, the average model with constant coefficients clearly dominates the average model with time-varying coefficients before the oil price shock in 1974. In contrast, the average model with time-varying coefficients outperforms the average model with constant coefficients frequently after the oil price shock. It seems that predictive relationships changed and became more dynamic after the oil price shock.

In addition to evaluating the Bayes Factor, we look at the Mean Squared Prediction Errors (MSPEs) to ensure comparability of our results to the literature. As explained in Section 2.4, we use adjusted MSPEs as our predictive models are nested (see Clark and West (2006)). Figure 3 shows the adjusted MSPEs over time, i.e., the mean of the adjusted squared prediction errors up to a given point in time. Although MSPEs are similar for the two average models, the average model with time-varying coefficients shows slightly lower MSPEs over large parts of the sample period. Again, this result confirms that the data supports models with time-varying coefficients and that the predictive performance increases if coefficients are allowed to vary over time.

4.2 Evidence on Out-of-Sample Predictability

From an investor's point of view, the most important question is whether we find empirical evidence for out-of-sample predictability. Neglecting predictive power of any of the 14 predictive variables used in this study corresponds to taking the average equity premium as the

Figure 2: **Model Comparison using the Bayes Factor.** A black dot indicates that the model with time-varying coefficients ($\delta = 0.98$) outperforms the model with constant coefficients ($\delta = 1.0$) with respect to the Bayes Factor over the period defined by Start Date and End Date.

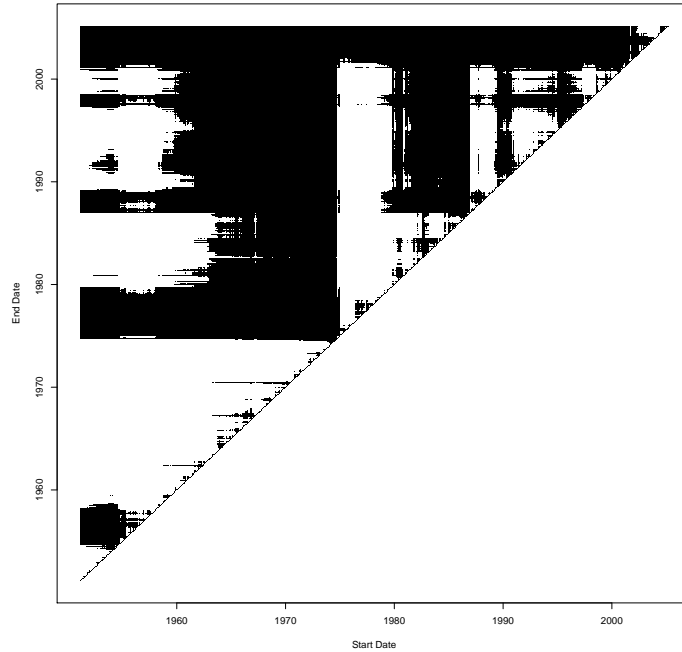
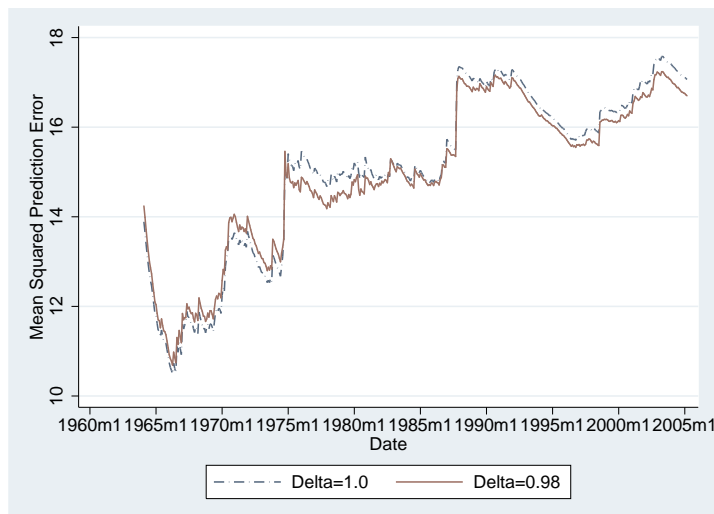


Figure 3: **Adjusted Mean Squared Prediction Errors over Time.** Comparison of average model with constant coefficients ($\delta = 1.0$) and average model with time-varying coefficients ($\delta = 0.98$).



best prediction for the following month's premium. Correspondingly, the no-predictability benchmark model is the single model that includes only the constant as a predictor and assumes that the constant does not vary over time (i.e., $\delta = 1.00$). If one believes, in contrast, in predictability, the posterior probability weighted average model (across different combinations of predictive variables and different choices of time-variation in coefficients) would be the appropriate choice. Therefore, we address the question of out-of-sample predictability by comparing the performance of these two models according to the Bayes Factor and the Mean Squared Prediction Errors.

Figure 4 summarizes our results for the Bayes Factor (a black dot indicates that the average model receives more support from the data than the no-predictability benchmark over the specific evaluation period defined by Start Date and End Date). The graph illustrates that the average model outperforms the no-predictability benchmark if the oil price shock in 1974 is included in the evaluation period. For the entire data sample from 1951 to 2005, for example, the average model gains relative weight by a factor of 368.4. Stated differently, the weight of the no-predictability benchmark within the average model drops from its naive prior of $(1/3)(1/32767) = 1.02 \cdot 10^{-5}$ to $2.76 \cdot 10^{-8}$. The no-predictability benchmark suffers disproportionately during the oil price shock between 1973 and 1975, as its probability weight drops by six orders of magnitude within two years. If this specific period of time is not included in the evaluation period, we find only limited evidence for out-of-sample predictability.¹²

An alternative way to look at out-of-sample predictability is to investigate whether individual predictive regressions consistently (i.e., over multiple subsequent evaluation periods) outperform the benchmark. We perform such an analysis for the 25 years from 1980 to 2005. Out of the set of 3,834 individual models that outperform the no-predictability benchmark in the 1980s, 551 also outperform the benchmark in the 1990s. However, only eleven models outperform the benchmark in the 1980s, 1990s and between 2000 and 2005. This result reinforces the strong performance of the no-predictability benchmark. The most important result, however, is the fact that all eleven models have a variability parameter of $\delta = 0.98$, i.e., only dynamic models are able to consistently outperform the no-predictability benchmark.¹³

Figure 5 reinforces our conclusions from analyzing the Bayes Factor. The no-predictability benchmark performs slightly better than the average model until the oil price shock in 1974. From there on, the average model shows a considerably lower adjusted MSPE.

¹²This is perfectly consistent with the observations of Goyal and Welch (2006) that the OLS out-of-sample significance of certain predictors crucially depends on the inclusion of the oil-price shock in the data sample. While they come to the conclusion that none of the models analyzed in their paper outperforms the no-predictability benchmark after the oil price shock, we discover that the average model including models with time-varying coefficients outperforms the no-predictability benchmark also during the 1990s.

¹³One could argue that, by looking at 98301 models, there is a large probability to find some models randomly that outperform the no-predictability benchmark consistently. This is, of course, true but the eleven models that we find do not look random at all. They agree to 100% on the time-variation parameter and, additionally, to a very large extent on their choice of predictive variables (i.e., yield spread, T-bill rate, industrial production, dividend yield, and earnings yield).

Figure 4: **Model Comparison using the Bayes Factor.** A black dot indicates that the average model outperforms the no-predictability model with respect to the Bayes Factor over the period defined by Start Date and End Date.

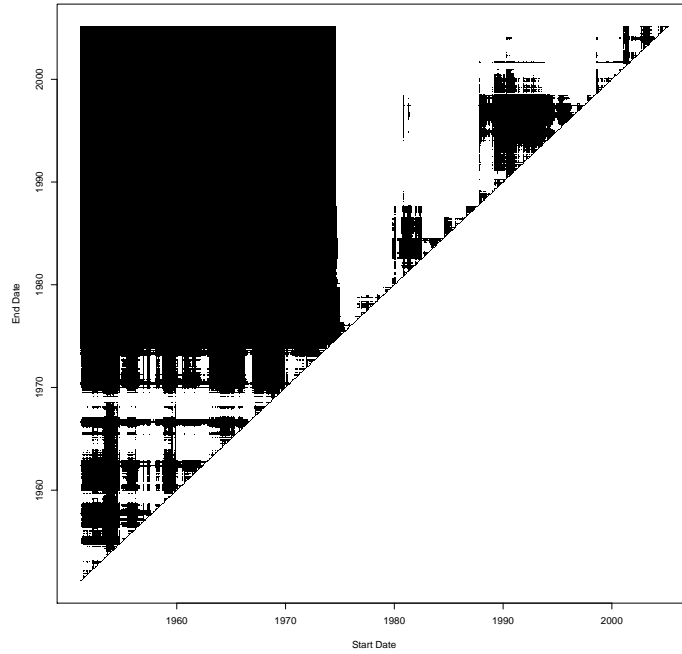
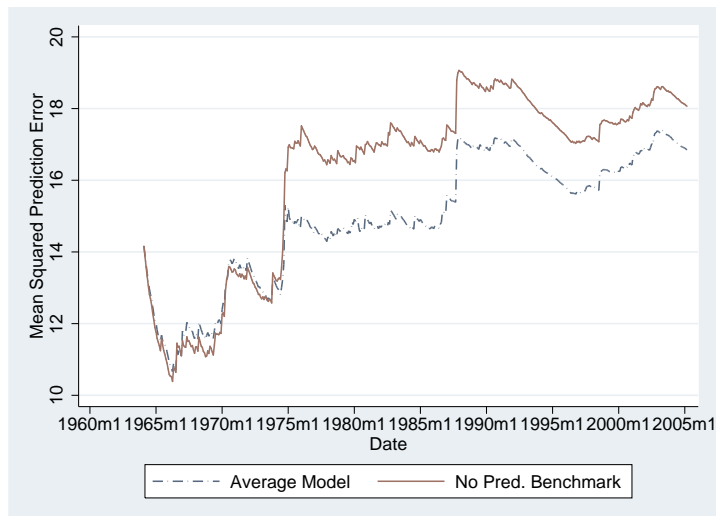


Figure 5: **Adjusted Mean Squared Prediction Errors over Time.** Comparison of the average model and the no-predictability benchmark.



4.3 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 2.3, this allows attributing weights to the four components of prediction variance. In Figure 6, we plot the relative weights of these components of prediction variance over time.

In Panel A of Figure 6, 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.¹⁴ 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.

4.4 Comparison of Model Characteristics

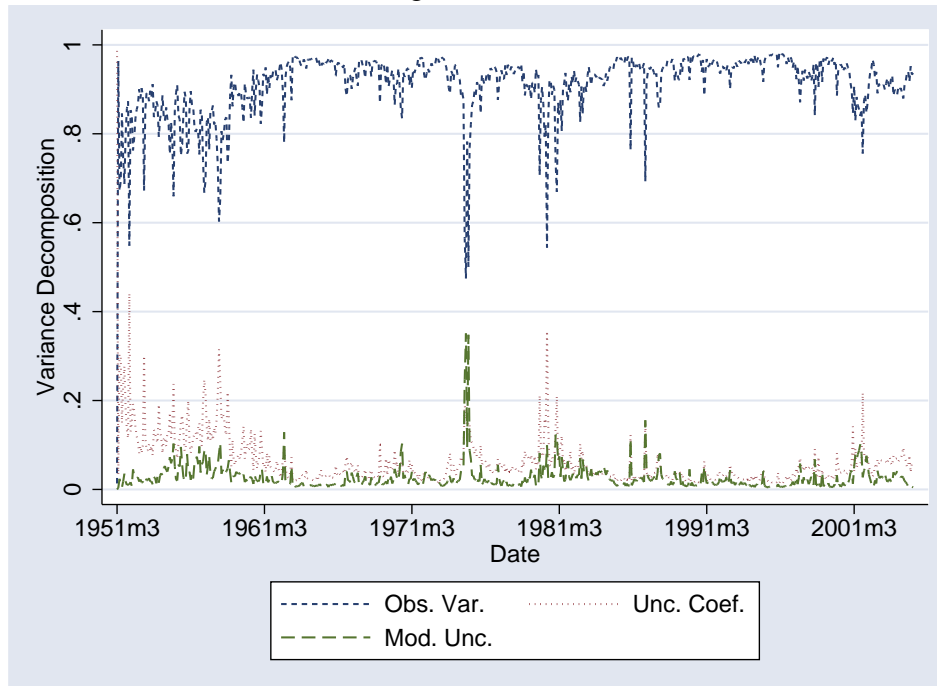
The final step of our empirical analysis aims at characterizing the predictive models and at evaluating the importance of individual explanatory variables. We analyze the following average models in detail: (i) the average model across different combinations of variables and different levels of time-variation of coefficients, (ii) the average model across different combinations of variables with constant coefficients ($\delta = 1.00$), and (iii) the average model across different combinations of variables with time-varying coefficients ($\delta = 1.00$).¹⁵ The goal of this analysis is two-fold. First, we want to understand what these predictive models look like and how they evolve over time. Second, we want to illustrate in how far misleading observations (e.g., on the importance of individual predictive variables) can result from exclusively analyzing models with constant coefficients.

¹⁴The fact that parameter uncertainty is most of the time more important than model uncertainty fits well to findings documented in Pastor and Stambaugh (1999). Interestingly, they find the same relationship for cost of capital estimations on the firm level while the results presented here are for cost of capital on the market level.

¹⁵Note that, also in this section, we 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 a notable posterior probability (see Figure 1). These results can be obtained from the authors upon request.

Figure 6: **Sources of Prediction Variance.**

Panel A: Including the observational variance.



Panel B: Excluding the observational variance.

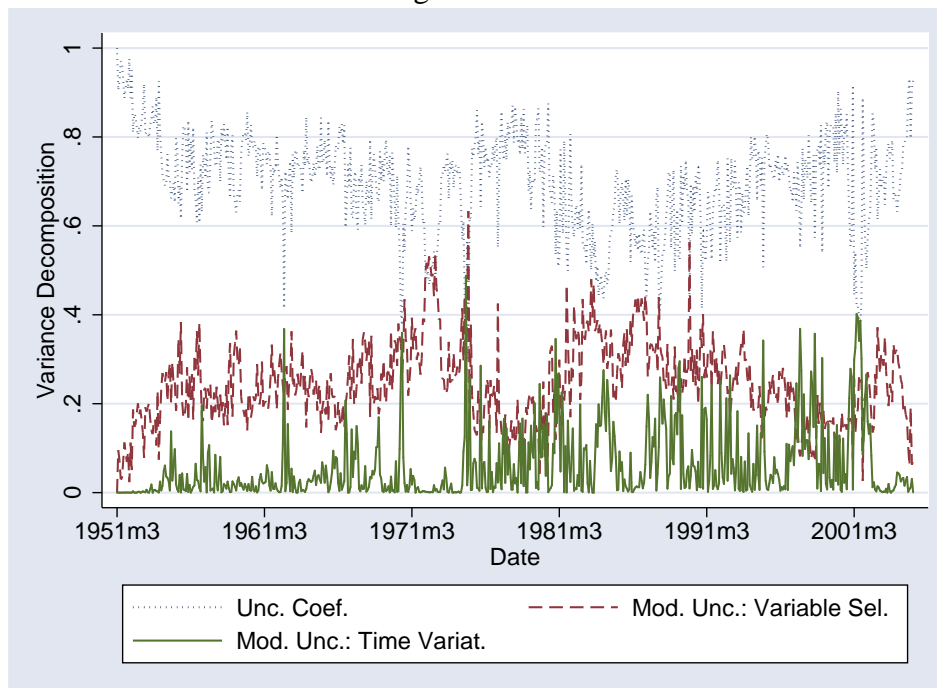
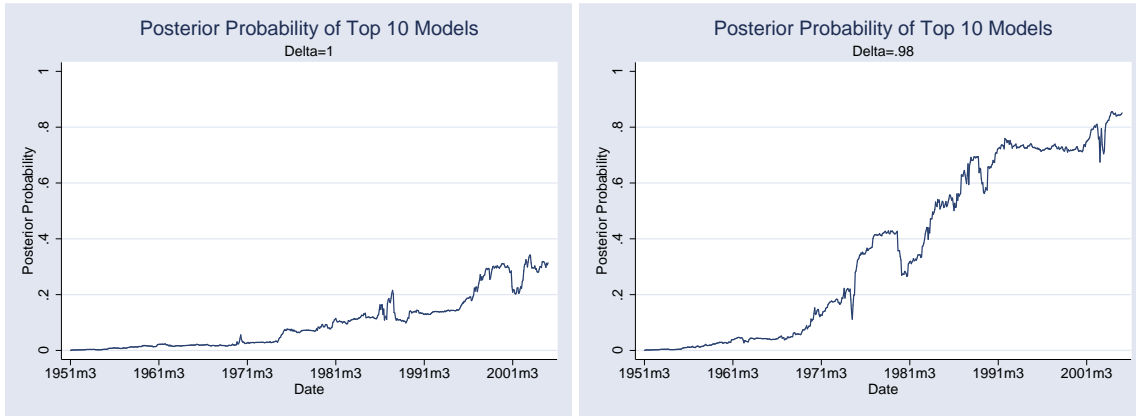


Figure 7: Analysis of Top Models.



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. For this purpose, we focus on the Top 10 models for a given δ in a first step. Figure 7 shows how much posterior probability the Top 10 models receive within each δ -class plotted over time. In the case of constant coefficients, the posterior probability assigned to the Top 10 models does not account for more than 30 percent at the end of the sample period. In contrast, the posterior probability assigned to the Top 10 models assuming time variation in the coefficients, increases to more than 80 percent over the sample period. Consequently, in the case of models with constant coefficients Top 10 models are less distinct from other models than in the case of models with time varying coefficients.

This is a potentially important insight, as it provides an explanation for the erratic behavior of best models reported in the literature before. 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. However, we show that this “stationarity-issue” can be resolved largely by allowing coefficients to vary over time.¹⁶

Another model characteristic that we want to analyze in this section is model size. In Table 2, we report for three points in time 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

¹⁶An additional measure to capture the properties of individual models counts the number of models required to get a total amount of posterior probability of at least 90 percent. If one assumes constant coefficients, 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.

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

	March 1985			March 1995			March 2005		
Model size	$\delta = 1$	$\delta = .98$	Avg.	$\delta = 1$	$\delta = .98$	Avg.	$\delta = 1$	$\delta = .98$	Avg.
1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	0.000	0.007	0.002	0.000	0.003	0.001	0.001	0.007	0.007
3	0.004	0.109	0.028	0.004	0.072	0.027	0.021	0.017	0.017
4	0.045	0.352	0.109	0.069	0.396	0.175	0.187	0.298	0.291
5	0.174	0.351	0.208	0.212	0.402	0.273	0.308	0.576	0.557
6	0.289	0.142	0.258	0.297	0.112	0.237	0.265	0.102	0.113
7	0.263	0.033	0.215	0.240	0.014	0.167	0.145	0.006	0.015
8	0.149	0.005	0.119	0.124	0.001	0.084	0.055	0.000	0.004
9	0.056	0.001	0.045	0.043	0.000	0.029	0.015	0.000	0.001
10	0.015	0.000	0.012	0.010	0.000	0.007	0.003	0.000	0.000
11	0.003	0.000	0.002	0.002	0.000	0.001	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

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 predictive variables.

An even more interesting result is that the optimal model size depends on the underlying assumption regarding the 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., $\delta = 1.00$). When accounting for time variability of coefficients, smaller models are favored. In our opinion, this indicates the very interesting fact that models assuming static coefficients compensate 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.

Furthermore, this result raises the question whether all predictive variables identified to be important in models with constant coefficients are truly important. We address this issue in more detail in Table 3 and Figure 8. We measure the importance of a specific explanatory variable by the sum of posterior probability of all models that include this variable. While Table 3 reports the importance and average coefficients for the same three points in time as analyzed in Table 2, Figure 8 illustrates the development of the importance over time for each of our explanatory variables. Most important predictive variables over the entire sample period are the T-bill rate change, the T-bill rate level, the yield spread, the industrial production,

Table 3: **Importance of Variables and Average Coefficients.** Sum of Posterior Probabilities of Models Including Specific Variables (I) and Average Coefficients (C).

Variable		March 1985			March 1995			March 2005		
		$\delta=1.00$	$\delta=.98$	Avg.	$\delta=1.00$	$\delta=.98$	Avg.	$\delta=1.00$	$\delta=.98$	Avg.
Momentum 1	C	-0.009	-0.005	-0.009	-0.004	-0.002	-0.003	-0.001	0.000	-0.000
	I	0.222	0.058	0.187	0.163	0.027	0.118	0.128	0.009	0.017
Momentum 2	C	-0.012	0.000	-0.009	-0.016	-0.001	-0.011	-0.006	0.00	-0.001
	I	0.239	0.045	0.199	0.271	0.022	0.191	0.171	0.007	0.018
Dividend yield	C	0.137	0.839	0.286	0.159	0.573	0.292	0.027	0.467	0.438
	I	0.425	0.695	0.482	0.390	0.728	0.500	0.235	0.593	0.569
Earnings yield	C	0.166	0.081	0.148	0.198	0.148	0.182	0.276	0.202	0.207
	I	0.667	0.258	0.581	0.707	0.271	0.565	0.878	0.416	0.447
Turnover	C	0.014	0.005	0.012	0.008	0.000	0.006	0.001	-0.000	-0.000
	I	0.518	0.095	0.431	0.457	0.032	0.319	0.178	0.012	0.023
Credit spread	C	0.022	0.029	0.024	0.149	0.006	0.103	0.039	-0.007	-0.004
	I	0.209	0.039	0.174	0.258	0.021	0.177	0.161	0.010	0.020
T-bill rate	C	-0.198	-0.416	-0.245	-0.209	-0.350	-0.254	-0.066	-0.597	-0.561
	I	0.736	0.838	0.758	0.765	0.937	0.820	0.377	0.955	0.916
Change in T-bill	C	-1.172	-0.659	-1.063	-0.913	-0.544	-0.797	-1.070	0.448	0.348
	I	0.975	0.995	0.979	0.909	0.991	0.935	0.966	0.994	0.992
Term spread	C	0.089	0.034	0.078	0.055	-0.027	0.029	0.029	-0.059	-0.053
	I	0.313	0.109	0.271	0.262	0.074	0.201	0.196	0.067	0.075
Yield spread	C	-0.237	0.000	-0.186	-0.008	0.444	0.137	-0.000	2.146	2.002
	I	0.536	0.712	0.571	0.138	0.821	0.358	0.126	0.969	0.912
January dummy	C	0.088	0.043	0.078	0.198	0.025	0.142	0.118	-0.002	0.006
	I	0.185	0.036	0.154	0.254	0.020	0.178	0.197	0.006	0.019
Indust. Prod. Growth	C	-0.086	-0.068	-0.082	-0.100	-0.062	-0.087	-0.082	0.124	0.110
	I	0.915	0.475	0.821	0.946	0.479	0.794	0.862	0.632	0.647
Inflation	C	-0.052	-0.013	-0.044	-0.118	-0.026	-0.088	-0.244	-0.056	-0.068
	I	0.321	0.054	0.266	0.542	0.078	0.392	0.861	0.084	0.137
Change in Inflation	C	-0.020	-0.030	-0.022	-0.015	-0.008	-0.012	-0.029	-0.004	-0.006
	I	0.094	0.031	0.081	0.089	0.012	0.064	0.094	0.003	0.010
Constant	C	-0.054	-0.503	-0.143	-0.055	-0.201	-0.102	-0.010	-0.014	-0.014
	I	0.172	0.198	0.176	0.159	0.078	0.133	0.130	0.021	0.028

the dividend yield, and the earnings yield. The graphs of Figure 8 show that the importance of individual variables varies quite considerably across time and, partly, across average models with different assumptions on the time-variation of coefficients. In the following, we are going to comment on some, especially interesting variables and observations in more detail.

The dividend yield represents the most traditional predictive variable and a large share of the existing literature is devoted to the question of whether the dividend yield predicts equity returns. Our analysis confirms that the dividend yield showed increasing predictive power through the 1980s and decreasing predictive ability during the 1990s. However, recently the dividend yield again became more important as a predictive variable. Our results fit very well into the existing literature. Goyal and Welch (2006), Ang and Bekaert (2006), 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 important observation is that models with constant coefficients and models with time-varying coefficients differ quite considerably regarding the importance of the dividend yield. If one analyzed only models with constant coefficients, the dividend yield's importance during the 80ties and, most importantly, during the last few year would be overlooked or underestimated.

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 8 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 an especially important predictive variable, it keeps a stable sum of posterior probabilities amounting to approximately 20 percent between 1960 and 2000. Just recently, however, the January dummy's importance dropped close to zero indicating that nowadays the January effect is of minor importance for predicting the S&P 500 return. Again, we emphasize that this recent decrease in importance would have been overlooked if only models with constant coefficients were analyzed.¹⁷

Finally, there are two more interesting observations with respect to the importance of predictive variables. First, the oil price shock influences the importance of some predictive variables considerably. While the change in the T-Bill rate and the yield spread became more important, the T-Bill rate's importance suffered during the oil price shock. Second, the separation between important and less important variables is relatively blurred in the case of $\delta = 1$. The change in inflation, for example, is the least important variable but still receives a posterior probability of 9.4 percent in March 1985, followed by 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 $\delta = 0.98$, the separation between important and

¹⁷In fact, models with constant and time-varying coefficients differ considerably in their quantification of several variables' importance (e.g., T-Bill rate, inflation and turnover). The turnover variable is an especially interesting example, as Cremers (2002) documents the importance of this variable. Figure 8, however, illustrates that this assessment is driven by the assumption of constant coefficients.

unimportant variables is much more clear-cut.

As far as the average coefficients are concerned, Table 3 shows that there is quite some variation over time.¹⁸ The most important observation in this context is that coefficients do not “trend away” to extreme values although we model them as random walks. In fact, 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.

5 Robustness Checks

5.1 Influence of the 1974 Oil Crisis on our Results

From the previous section the considerable influence of the oil crisis in 1974 on some of the results presented thus far is evident. In order to verify that our main conclusions are not driven exclusively by this single event, we basically performed the entire analysis for a data set starting after the oil crisis. For reasons of brevity, we concentrate, however, on the question whether the data supports predictive models with time-varying coefficients.

In Figure 9, 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 models with $\delta = 0.98$, however, does not vanish but keeps a certain and notable weight between 10 percent and 30 percent. After 2000, 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 general, our results change quantitatively but not qualitatively if one uses the post 1974 sample period. There is, however, one notable exception concerning the importance of the T-bill rate change to predict equity returns. In Section 4.4 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. The (unreported) equivalent analysis for the shorter sample period shows that the T-bill rate change loses its importance. In this case it does not evolve as an especially important variable. Thus, its dominant role as a predictor can be attributed to the oil crisis in 1974.

¹⁸Note that coefficients of individual models for $\delta = 1$ are constant by definition and are updated when new observations arrive. The Bayesian model averaging approach, however, introduces time variation to the regression coefficients of the average static model by assigning time varying weights to individual models.

Figure 8: Importance of Individual Variables.

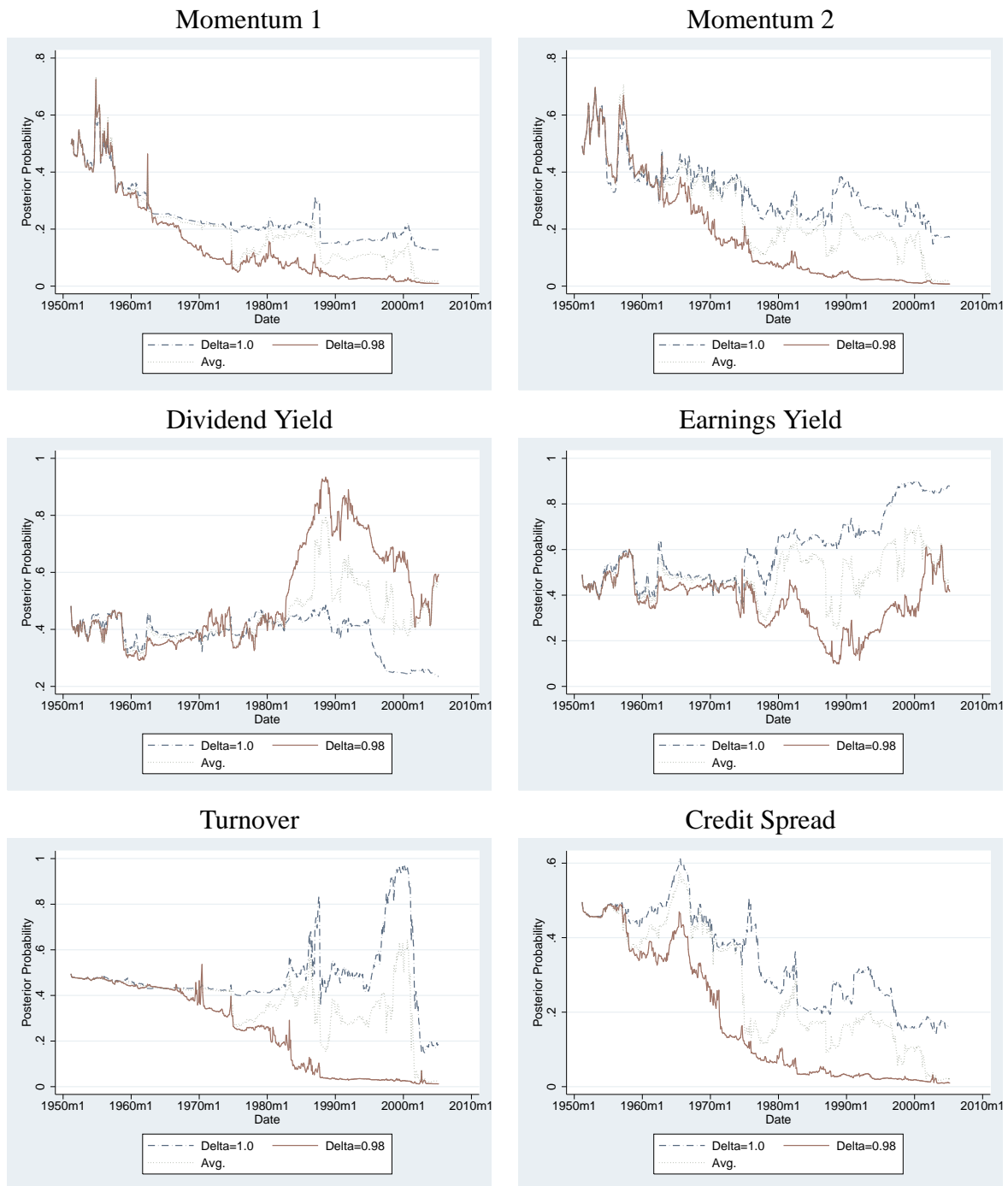


Figure 8: (Cont.) Importance of Individual Variables.

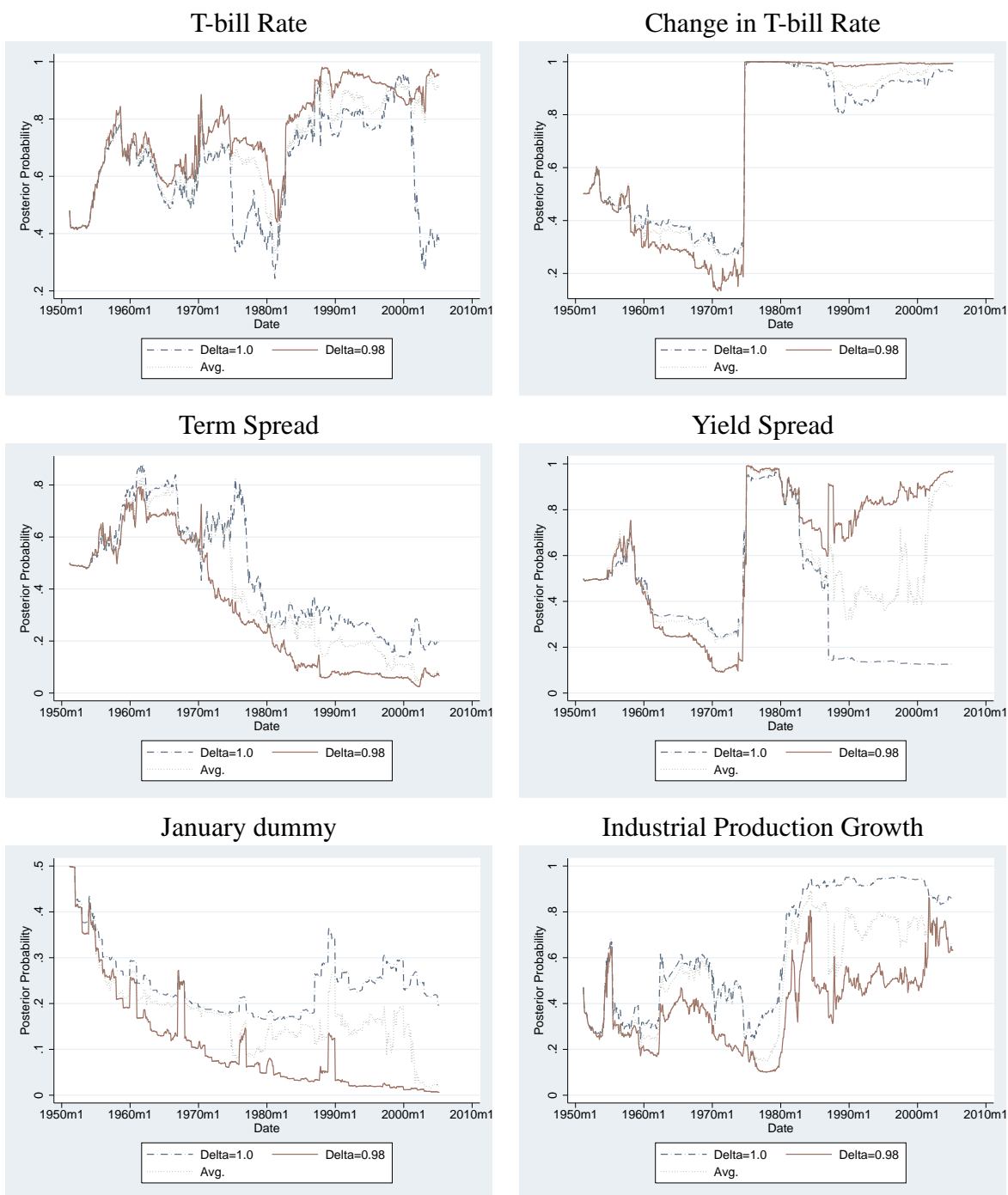


Figure 8: (Cont.) Importance of Individual Variables.

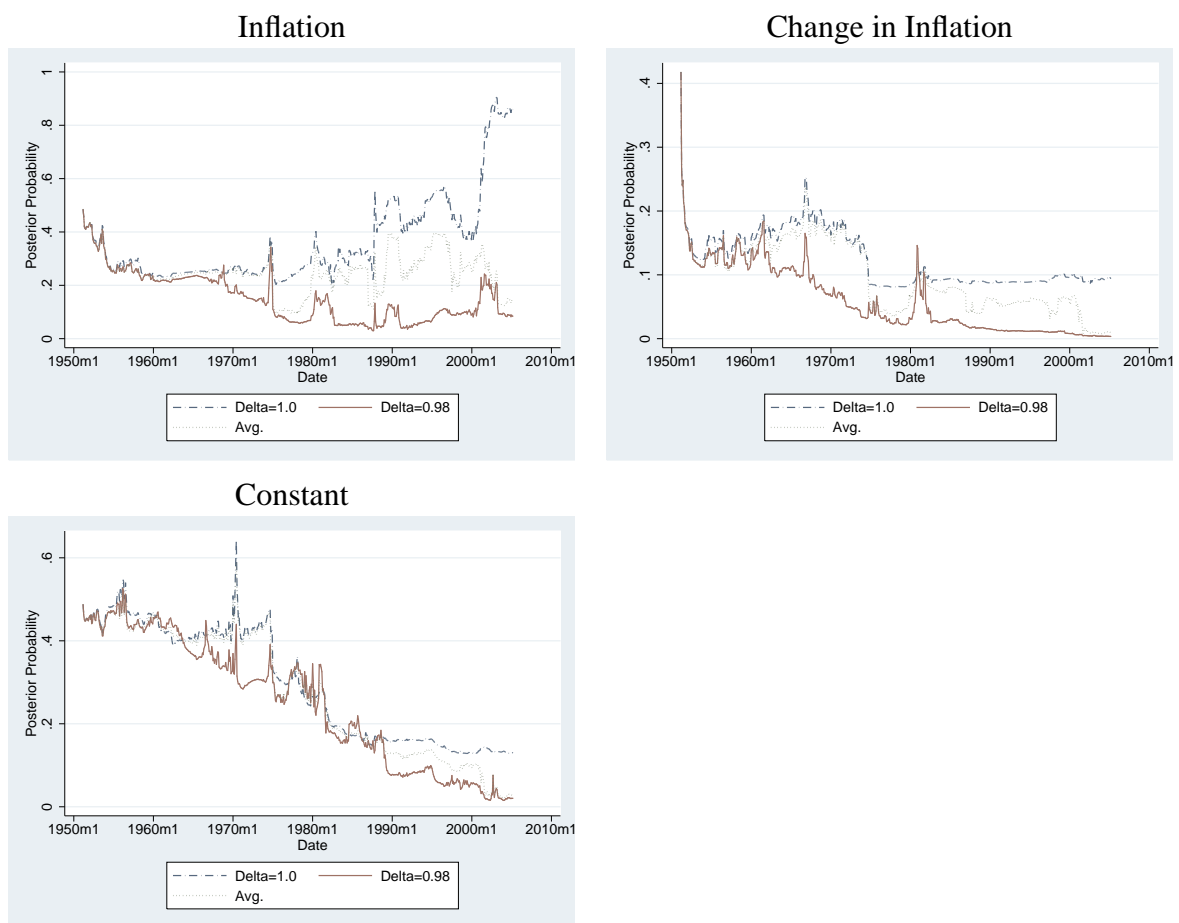
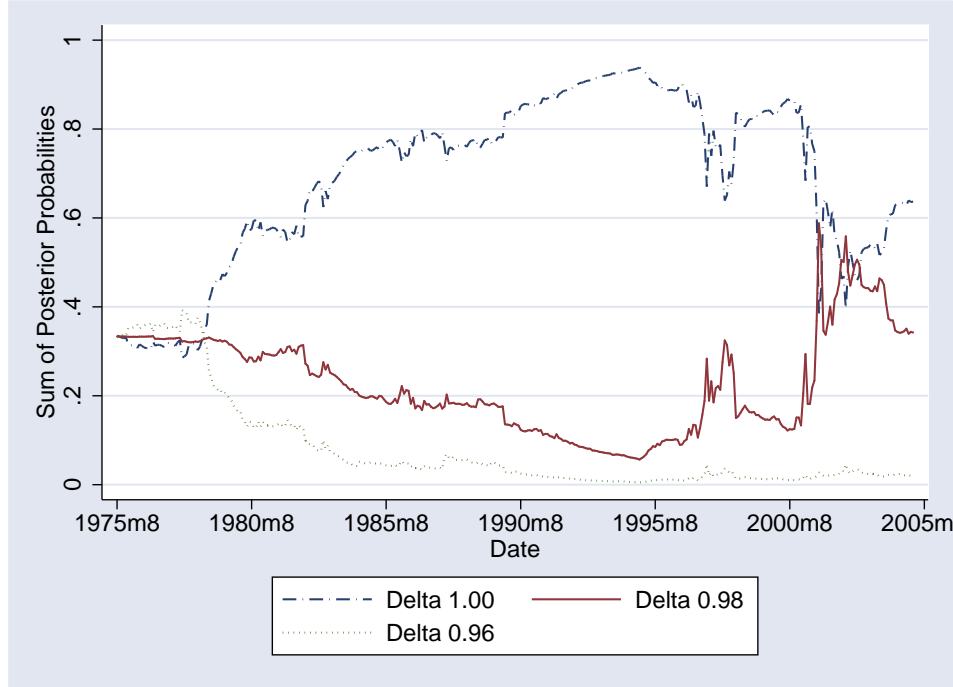


Figure 9: **Sum of Posterior Probabilities of Models with a Given δ .** We only use data from August 1975 to March 2005 to estimate and evaluate these models.

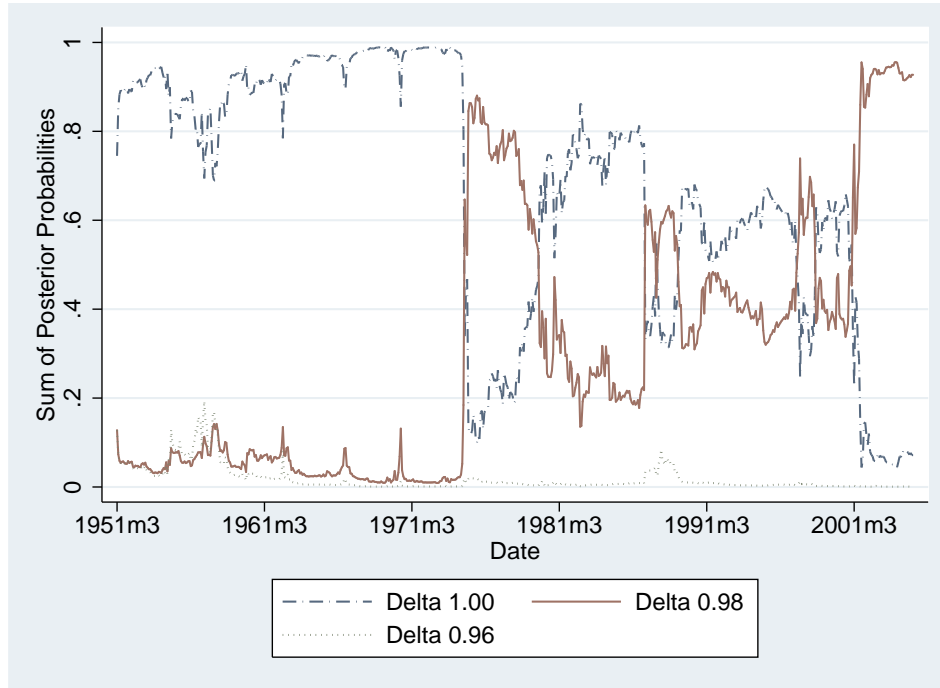


5.2 Changed Prior Probability for the No-Predictability Benchmark Model

Another obvious dimension of robustness checks concerns our choice of priors. As mentioned in the beginning of the results section we replicated our analysis using a different g-prior. Our results are robust to this change of prior information. In this section, we want to focus on a different piece of prior information, namely the prior probability of each individual model. So far, we assumed an equal prior probability for each model amounting to $1/(d \cdot (2^k - 1))$. One could, however, take a conservative and sceptical point of view and doubt the existence of predictability. Consequently, one would attribute a larger prior probability to the no-predictability benchmark. In this section, we assume a prior probability of 50% for the no-predictability benchmark model. The remaining models receive an equal prior probability amounting to $0.5 \cdot 1/(d \cdot (2^k - 2))$.

Figure 10 shows again the posterior probability of each model class over time. One clearly observes the influence of the changed prior in the beginning of the period. The posterior probability assigned to models with $\delta = 1.0$ starts out at a higher level and keeps this level. Over time, however, the influence of the changed prior fades out and Figure 10 converges to Figure 1 (especially after the oil crisis in 1974). Consequently, our assumptions on prior model probabilities do not seem to influence our results much.

Figure 10: **Sum of Posterior Probabilities of Models with a Given δ .** The no-predictability benchmark model has a prior probability of 50%. The remaining models receive equal prior probabilities of $0.5 \cdot 1/(d \cdot (2^k - 2))$.



6 Conclusion

This paper contributes in several dimensions to the literature on equity return prediction. First, it is the first paper to use and evaluate predictive regressions with time varying coefficients. Using a flexible and powerful Bayesian framework we find strong empirical support for models with time-varying coefficients, both with respect to the Bayes Factor and with respect to Mean Squared Prediction Errors. We find out-of-sample predictability if the oil price shock in 1974 is included in the evaluation period. If this is not the case, the no-predictability benchmark performs spectacularly well and is only consistently outperformed by eleven predictive models with time-varying coefficients. The oil price shock in 1974 plays, in general, an important role according to our analysis. Models with constant coefficients clearly dominate before the oil price shock in 1974. Thereafter, predictive relationships become more dynamic resulting in considerably more empirical support for models with time-varying coefficients.

Second, we decompose total prediction uncertainty and quantify how much of it can be attributed to one of the following four sources of uncertainty: (i) the variance in predictive variables, (ii) the estimation uncertainty in coefficients, (iii) the model uncertainty with respect to the choice of predictive variables, and (iv) the model uncertainty with respect to the time-variation in coefficients. Most interestingly, we find that the uncertainty about the level of time-variation in coefficients is, frequently, of similar importance than the uncertainty about the choice of predictive variables.

Third, we are able to refine several results presented in the equity return literature before. 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 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, we evaluate the importance of individual variables to predict equity returns. According to our analysis, the most important variables are the T-bill rate change, the T-bill rate level, the yield spread, the industrial production, the dividend yield, and the earnings yield. The importance of predictive variables varies, however, over time and depends on the choice of time-variation of coefficients. We clearly document that conclusions about the importance of individual variables drawn only from models with constant coefficients can be misleading. Our analysis reveals, for example, that the dividend yield became recently more

important as a predictive variable. Similarly, we document that the January dummy lost its predictive power, recently. Both developments would have been overlooked or underestimated if only models with constant coefficients were analyzed.

From an investor's point of view our paper offers several interesting insights. First, it documents that predictive relationships are currently quite dynamic and vary over time. Second, if one relies on the set of variables analyzed in this paper and uses only predictive models with constant coefficients, it will most probably be impossible to outperform the no-predictability benchmark, i.e., the unconditional mean. Third, the set of important predictive variables is dynamic as well. Variables that are key predictors over a certain period might become less relevant in the subsequent months. Fourth, we believe that our results, together with the evidence reported in papers on regime switching models, suggest that predictive relationships change rather continuously and gradually than in rare, discrete jumps.

While our paper provides several contributions to the literature on equity return prediction, it also raises new questions. Most importantly, what are the economic forces that cause time-varying predictive relationships? Possible mechanisms include changes in the monetary policy, the macroeconomic conditions or the institutional environment. Another possible explanation could be related to market volatility (similar to Kim, Morley, and Nelson (2005) who relate market volatility to regime shifts). Answering these interesting questions is, however, left for further research.

References

- Ang, A., and G. Bekaert, 2006, Stock return predictability: is it there?, forthcoming, *The Review of Financial Studies*.
- 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.
- Barsky, R. B., 1989, Why don't the prices of stocks and bonds move together, *The American Economic Review* 79, 1132–1145.
- 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.
- Boyd, J. H., J. Hu, and R. Jagannathan, 2005, The stock market's reaction to unemployment news: Why bad news is usually good for stocks, *The Journal of Finance* LX, 649–672.
- Campbell, J., and S. B. Thompson, 2005, Predicting the equity return premium out of sample: Can anything beat the historical average?, working paper.
- Clark, T.E., and K. D. West, 2006, Using out-of-sample mean squared prediction errors to test the martingale difference hypothesis, *Journal of Econometrics* 135, 155–186.
- Cooper, M., and G. Gulen, 2006, Is time-series-based predictability evident in real time?, *Journal of Business* 79(3), 1263–1292.
- 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, 2006, A comprehensive look at the empirical performance of equity premium prediction, forthcoming, *The Review of Financial Studies*.

- 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, 2005, The structural break in the equity premium, *Journal of Business and Economic Statistics* 23(2), 181–191.
- 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.
- McQueen, G., and V. V. Roley, 1993, Stock prices, news, and business conditions, *The Review of Financial Studies* 6, 683–707.
- Pastor, L., 2000, Portfolio selection and asset pricing models, *The Journal of Finance* 55, 179–223.
- , and R. F. Stambaugh, 1999, Costs of equity capital and model mispricing, *Journal of Finance* 54(1), 67–121.
- , 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.