Predictive Regressions with Time-Varying Coefficients *

Thomas Dangl
Vienna University of Technology

Michael Halling
University of Utah

September, 2009

Abstract

We evaluate predictive regressions that explicitly consider the time-variation of coefficients in a comprehensive Bayesian framework. This allows for fast and consistent adjustment of regression coefficients to changes in the underlying economic relationships. For monthly returns of the S&P 500 index, we demonstrate statistical as well as economic evidence of out-of-sample predictability: relative to an investor using the historic mean, an investor using our methodology could have earned consistently positive utility gains (between 1.8 and 5.8% p.a. over different time periods). We also show that predictive models with constant coefficients are dominated by models with time-varying coefficients. Finally, we document a strong link between out-of-sample predictability and the business cycle.

JEL Classifications: G12, C11
Keywords: Empirical asset pricing, equity return prediction, Bayesian econometrics.

*Acknowledgments: We thank K. Abadir, G. Bekaert, C. Bienz, P. Bossaerts, M. Cooper, M. Cremers, E. Dimson, H. Freitas Lopes, A. Goyal, B. Hanke, F. de Jong, H. Liu, K. Nyborg, M. Panayides, L. Pastor, K. Ritzberger, Z. Rachev, A. Stomper, A. Timmermann, S. Van Nieuwerburgh, C. Vega, M. Wagner, Y. Wu, J. Zechner and G. Zhou for their helpful comments. We also thank participants of the 10th Symposium on Finance, Banking and Insurance (2005), the VGSF Research Seminar in Vienna (2006), the NHH Skinance Conference (2006), the Meeting of the European Finance Association in Zurich (2006), the Utah Junior Faculty Research Group, the CEPR ESSFM (2009, Asset Pricing Week, Informal Session), the UBC Summer Finance Conference (2009, 3-slide session) and research seminars at NHH Bergen, National University of Singapore, Imperial College, Washington University St. Louis, University of Utah, Stockholm School of Economics and Tilburg University 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 the earliest versions of the paper.

Predictive Regressions with Time-Varying Coefficients

Abstract

We evaluate predictive regressions that explicitly consider the time-variation of coefficients in a comprehensive Bayesian framework. This allows for fast and consistent adjustment of regression coefficients to changes in the underlying economic relationships. For monthly returns of the S&P 500 index, we demonstrate statistical as well as economic evidence of out-of-sample predictability: relative to an investor using the historic mean, an investor using our methodology could have earned consistently positive utility gains (between 1.8 and 5.8% p.a. over different time periods). We also show that predictive models with constant coefficients are dominated by models with time-varying coefficients. Finally, we document a strong link between out-of-sample predictability and the business cycle.

JEL Classifications: G12, C11

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

1 Introduction

The issue of predicting equity returns is one of the most widely discussed topics in financial economics (see Campbell (2007) for a recent survey article). In-sample, numerous studies find evidence of predictability (see, for example, Stambaugh (1999), Ang and Bekaert (2007), Lettau and Van Nieuwerburgh (2008) and Pastor and Stambaugh (2009b)). Out-of-sample, however, little consensus exists on the fundamental questions of whether predictability exists and which variables have the best predictive performance (see, for example, Goyal and Welch (2008), Campbell and Thompson (2008), Cooper and Gulen (2006) and Rapach, Strauss, and Zhou (2009)). Given the conflicting points of view in the literature, Spiegel (2008) asks whether academics can *"produce an empirical model that allows for economic changes over time that is also capable of determining the 'right'parameter values in time to help investors?"* This is precisely the question that we address in this paper.

The literature agrees that parameter instability (i.e., time-variation in coefficients) represents a major challenge and that it might influence many of the results in the literature.[1] 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." Similarly, Cremers

[1] There are several reasons coefficients might vary over time, e.g., due to changes in regulatory conditions, in market sentiments, in monetary policies, in the institutional framework or in macroeconomic interrelations. Barsky (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. Veldkamp (2005) and Van Nieuwerburgh and Veldkamp (2006), among others, relate learning asymmetries and the flow of information to the business cycle.

(2002) claims in his conclusion that his model is limited by the assumption of parameter stability. Ang and Bekaert (2007) test for time variation in coefficients 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 as it is if estimated from the entire sample.[2]

The literature, however, is inconclusive about the true degree of time-variation in coefficients, and, despite the agreement on the issue, there is lack of systematic evidence. We identify the following important questions that have not been addressed in the literature and that we address in this paper: What degree of time-variation is supported by the data? How important is the issue of parameter instability (e.g., relative to the issue of choosing the right predictive variables)? By how much do current results (e.g., on out-of-sample predictability and the importance of individual predictive variables) change once parameter instability is taken into account?

We analyze these questions by estimating predictive regressions for S&P 500 returns that explicitly allow for time-variation of regression coefficients. For this purpose we apply a Bayesian econometric method[3] that enables us to model time-varying coefficients that are subject to random shocks. The two dimensions of model uncertainty— the choice

[2]Most existing papers in equity return prediction use rolling window regressions and/or perform sub-period investigations to take care of time-varying coefficients. Both approaches are ad-hoc and depend on exogenous parameters (like the window length or the dates of sub-periods) that, in many cases, lack both economic and statistical motivation.

[3]See, for example, Avramov (2002), Cremers (2002), Pastor (2000), Wachter and Warusawitharana (2007) and Johannes, Korteweg, and Polson (2009) for recent applications of Bayesian econometrics and inference in asset pricing.

of predictors and the degree of coefficients' time-variation— are addressed in a consistent manner within a Bayesian model averaging approach (see Raftery, Madigan, and Hoeting (1997) for technical details and Avramov (2002) and Cremers (2002) for applications to return prediction).

There is a stream of literature that addresses the issue of parameter instability by estimating regime-switching models and by searching for structural breaks in the predictive relationship between equity returns and explanatory variables. Pastor 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 of the existence of structural breaks. They differ quite considerably, however, in the timing of the breaks. Viceira (1997) is to our knowledge the first to search for structural changes in predictive relationships, but does not find evidence of structural breaks in the relationship between the dividend yield and equity returns. Paye and Timmermann (2006), in contrast, identify several structural breaks in the coefficients of state variables such as the lagged dividend yield or the term spread. All of these studies focus on in-sample predictability and ignore the question of whether an investor would have been able to detect these regime shifts in real-time (i.e., out-of-sample). Lettau and Van Nieuwerburgh (2008) represent a notable exception as they also perform out-of-sample tests. They conclude that regime-shifting models perform very poorly out-of-sample because of unreliable estimates of the timing of breaks and of the size of the shift.

We differ from these papers because we do not assume, ex ante, that the time variation in coefficients follows a step function. In contrast, the methodology proposed in this paper allows for gradual changes of coefficients. The methodology is also simple and parsimonious enough to enable us to evaluate out-of-sample predictability[4] for a comprehensive set of predictive variables. As shown in our empirical analysis, models with gradually varying coefficients are strongly supported by the data.[5]

Using monthly returns of the S&P 500 from May 1937 to December 2002, we compare the predictive out-of-sample performance (using statistical and economic measures) of models with time-varying coefficients to two benchmark models: (i) regressions with constant coefficients, and (ii) the unconditional mean of past returns, which constitutes the no-predictability benchmark.[6] Our most important result is that we find strong and consistent empirical support for models with time-varying coefficients. These models significantly outperform the two benchmark models across different time periods as far as prediction accuracy is concerned. This gain in prediction accuracy is also important in economic terms resulting in consistent utility gains between 1.8 and 5.8% p.a. for different time periods, relative to an investor using the historic mean. In comparison, investors

[4] We define out-of-sample in a strict sense; i.e., all results reported and discussed in this paper are based on predictions that an investor could calculate and use in real-time (without knowing the full sample).

[5] Note that there is an extensive literature (see 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. Another stream of literature that is to a lesser extent related to our paper is the one focussing 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.

[6] The benchmark models with constant coefficients used in the paper are equal to OLS regressions with an extending window. We are aware that, in the literature, regressions with constant coefficients use rolling windows and thus mimic time-varying coefficients in an ad-hoc way. The methodology proposed in this paper, in contrast, accounts for time-varying coefficients in a systematic and statistically consistent way.

using the predictions of models with constant coefficients realize a utility gain of .2% only in one sub-period (i.e., 1965 to 2002) and utility losses between -1.9% and -5.8% in all other periods. The findings of other researchers put these results in further perspective: following the same approach to calculating utility gains and comparable data sets, Rapach, Strauss, and Zhou (2009) find utility gains in the order of .5% to 1.5%, and Campbell and Thompson (2008) report maximum utility gains of .3%.

Most interestingly, we find a strong relationship between out-of-sample predictability and the business cycle. Although we find evidence of predictability during recessions as well as during expansions (in contrast to Henkel, Martin, and Nardari (2008) who do not find any evidence of in-sample predictability during expansions), the evidence is much stronger during recessions. In general, models with time-varying coefficients generate return predictions that match business cycle related patterns implied by asset pricing theory (e.g., Campbell and Cochrane (1999)) very well. On average, predicted equity risk premia are negative at the beginning of and increase during a recession (and peak around the trough). During expansions, predicted risk premia decrease and reach their lowest levels around the peak of the business cycle. Finally, an investor who relies on these predictions times the market very well, reducing her exposure around the peak of the business cycle and moving back into the market before the trough.

In the next step we analyze the models with time-varying coefficients in more detail to get a better understanding of the sources of their outperformance. Specifically, we decompose prediction uncertainty into four components, (i) the observational variance (i.e., the variance assigned to the random disturbance term in the predictive relationship),

(ii) the estimation uncertainty in coefficients, (iii) the model uncertainty with respect to the choice of predictive variables (see Avramov (2002) and Cremers (2002)), and (iv) the model uncertainty with respect to the time-variation in coefficients. Empirically, we find that the first two sources are most important, as expected. The two dimensions of model uncertainty are, however, non-negligible, especially when the stock market is under stress (e.g., during the oil price shock in the 70s).

Finally, we investigate the importance of individual predictive variables within the models with time-varying coefficients. We find that the relative valuation of high- and low-beta stocks (i.e., the cross-sectional premium) plays a dominant role among our set of predictive variables. We also find that the dividend yield receives considerable empirical support. Even more importantly, we document that, in the case of the dividend yield, our model with time-varying coefficients is able to learn the structural break due to the initiation of SEC rule 10b-18 in November 1982[7] — in contrast to constant coefficients or even regime-switching models (see Goyal and Welch (2008) and Lettau and Van Nieuwerburgh (2008)). Thus, while previous studies report a steady decline of the importance of the dividend yield as a predictive variable during the 80s and 90s, we document an increase.

The paper is structured as follows. Section 2 presents the empirical methodology. Section 3 describes the variables used in the empirical study. Section 4 reports empirical results and discusses their implications. Section 5 concludes.

[7]This rule enabled firms to legally buy back shares under certain circumstances (see Grullon and Michaely (2002) for details).

2 Prediction Models with Time-Varying Coefficients

Like 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 (2008), and Ang and Bekaert (2007)), 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 model the coefficients in our dynamic linear models to be time-varying (see Section 2.1). An important contribution of our paper is to evaluate whether the data supports time-varying coefficients or whether it confirms the constant coefficient paradigm. For each degree of time-variation of coefficients, we estimate the $2^k - 1$ dynamic linear models that result from all possible combinations of predictive variables. Then, we use a Bayesian model selection criterion to assign posterior probability weights across individual models that differ in the selected variables and degree of time-variation (similar to Avramov (2002) and Cremers (2002)). Finally, we use these posterior probabilities to 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: uncertainty in the choice of predictive variables, uncertainty in the estimation of coefficients, uncertainty in the degree of time-variation of the regression coefficients, and the general disturbance term. In 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), and in Section 2.2 we discuss the Bayesian model selection approach.

2.1 Dynamic Linear Models

In this section we introduce 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.[8] More specifically, we estimate models of the form

$$
\begin{aligned}
r_{t+1} &= X_t' \theta_t + v_{t+1}, \quad v \sim N(0, V) \quad \text{(observation equation)}, && (1)\\
\theta_t &= \theta_{t-1} + \omega_t \qquad \omega \sim N(0, W_t) \quad \text{(system equation)}. && (2)
\end{aligned}
$$

The vector θ_t consists of unobservable, time-varying regression coefficients, and the observational disturbance v is assumed to be normally distributed with mean 0 and constant but unknown variance V. In what follows we call V the observational variance. While Equation (2) states that these coefficients are exposed to random shocks ω that are jointly normal (with mean 0 and variance matrix W_t), we do not assume systematic movements

[8] We are performing an out-of-sample analysis, where out-of-sample is to be interpreted in a strict sense; i.e., for predicting the return at time $t+1$, we use only information that is available at or before time t. This will be made more precise in the following paragraphs when we define information sets on which estimates are conditioned.

Observable 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.

in θ.[9] We are aware that Equation (2) implies that theoretically coefficients might drift to arbitrarily high values, hence causing returns to be non-stationary. This simplifying assumption is made to keep the model tractable as a multivariate model. The structure in the dynamics of the coefficients is fed into the estimation by new observations which arise at a monthly frequency in our application. As long as Equation (2) is thought to apply for a finite period of time and not forever, this specification should be safe (see also Primiceri (2005)). Any assumption on the coefficients' dynamics beyond Equation (2) would either be an ad-hoc constraint or require a drastic reduction of dimensionality, both of which are undesirable.

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 increases, the intrinsic variability of the regression coefficients θ_t increases the flexibility of the model. At the same time, however, the out-of-sample prediction variance increases and, consequently, reduces the precision of the prediction. The specific structure we impose on W_t and how we estimate the magnitude of time variation of the underlying coefficients will be explained below.

Let $D_t = [r_t, r_{t-1}, ..., X_t, X_{t-1}, ..., \text{Priors}_{t=0}]$ denote the information set available at time t. This information set contains all returns, all corresponding realizations of the predictive variables up to time t and our initial time zero choice of priors regarding θ and V. In Appendix A.1 we will describe in detail 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

[9] See Primiceri (2005) and Cogley and Sargent (2003) for a similar model specification with an application to monetary policy and Brown, Song, and McGillivray (1997) for an application to house prices.

estimated observational variance. The essential result is that using a normally distributed prior for the system coefficients θ_0 and an inverse-gamma distributed prior for the observational variance V leads to a fully conjugate Bayesian analysis, which ensures that prior and posterior distributions come from the same family of distributions. For the time $t = 0$ specification of the prior information we use a natural conjugate g-prior specification (see, e.g., Zellner (1986); this type of prior was, for example, also used in the study by Cremers (2002)):

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

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

where

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

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.

The forecast of the time $t+1$ return r_{t+1} (i.e., the predictive density) can be found by integrating the conditional density of r_{t+1} over the range of θ and V. It is a Student-t-distribution, as illustrated by Equation (11) in the Appendix.

To specify W_t, the system variance matrix, we apply a discount factor approach (West and Harrison (1997)). This approach relies on the assumption that the variance matrix W_t of the error term ω_t is proportional to the estimation variance/covariance matrix of the coefficients $\theta_t|D_t$. More precisely, the scaling factor is assumed to be $\frac{1-\delta}{\delta}$ for $\delta \in \{\delta_1, \delta_2, \ldots, \delta_d\}$ with $0 < \delta_i \leq 1$.

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 of studies 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 ex ante 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 choice. Considering a number of d different discrete values of δ leads to a total of $d \cdot (2^k - 1)$ possible dynamic linear models.[10]

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. 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 different values of δ into account. Further, we perform robustness checks with respect to different assumptions about the prior.

The posterior probability of each of the $d \cdot (2^k - 1)$ models is updated month by month according to Bayes rule; i.e., based on the realized likelihood of the model's return prediction. Appendix A.2 provides more details on the Bayesian model averaging approach. The overall average model's predictive density is then the posterior-probability weighted average predictive density of all $d \cdot (2^k - 1)$ models in our universe. The beauty of this approach is its flexibility. If we want to analyze, for example, the empirical support for models including a specific predictive variable or having a certain degree of time-variation, we simply average across all models with this specific characteristic.

[10] We assume the same degree of time-variation for all coefficients included in a specific model. The proposed framework would be flexible enough to allow for variable-specific degrees of time-variation. Given the lack of theoretical predictions for the level of time-variation of individual variables and the enormous number of degrees of freedom, we have to make this simplifying assumption.

3 Empirical Study Design

3.1 Data Description

We calibrate and test the proposed methodology using total excess returns of the S&P 500 Index from May 1937 to December 2002. The choice of equity returns and explanatory variables is guided by previous academic studies and by the goal of ensuring the comparability of our results with these studies. In particular, we want to relate our results to those reported in Goyal and Welch (2008) and, thus, reuse their dataset in our study.[11] For the sake of brevity, we include only a short description of the predictive variables here (see Goyal and Welch (2008) for a more extensive discussion of the data set and the data sources):

- Dividends: Dividend Yield **(dy)** is the difference between the log of dividends on the S&P 500 Index and the log of one-month-lagged prices.
- Earnings: Earnings to Price Ratio **(ep)** is the difference between the log of earnings and the log of prices. Dividend Payout Ratio **(dpayr)** is the difference between the log of dividends and the log of earnings.
- Variance: As a measure of Stock Variance **(svar)** the sum of squared daily returns on the S&P 500 is used.
- Cross-sectional premium: Cross-Sectional Beta Premium **(csp)** quantifies the relative valuation of high- and low-beta stocks according to Polk, Thompson, and

[11] We particularly thank Amit Goyal for providing their dataset.

Vuolteenaho (2006).[12]

- Book value: Book to Market Ratio **(bmr)** is the ratio of book value at the end of the previous year[13] divided by the end-of-month market value, both taken from the Dow Jones Industrial Average.

- Net issuing activity: Net Equity Expansion **(ntis)** is the ratio of twelve-month moving sums of net issues by NYSE listed stocks to the total market capitalization of NYSE stocks.

- T-bills: T-bill Rate **(tbl)** is the secondary market rate of 3-month US treasury bills.

- Long-Term Yield: Long-term Government Bond Yields **(lty)** and Long-term Government Bond Returns **(ltr)** are the yields and returns of long-term US treasury bonds, respectively.

- Corporate Credit: Default Return Spread **(dfr)** is the difference between returns on long-term corporate bonds and returns on long-term government bonds. Default Yield Spread **(dfy)** is the difference between BAA-rated and AAA-rated corporate bond yields.

- Inflation **(inf)** is the Consumer Price Index (all urban consumers) from the Bureau of Labor Statistics, lagged by one additional month.

From the dataset of Goyal and Welch (2008) we exclude the predictive variables "Investment to Capital Ratio", "Percent Equity Issuing"and "Consumption, Wealth, Income

[12]The availability of this variable limits our dataset both in early and in later years.

[13]For the months January and February, the book value is additionally lagged by one year.

Table 1: **Summary Statistics (788 Observations).**

	Mean	Standard deviation	Minimum	Maximum	Median
dy	-3.295	.422	-4.525	-2.406	-3.291
ep	-2.650	.402	-3.839	-1.775	-2.674
dpayr	-.650	.197	-1.183	.063	-.624
svar	.002	.004	.000	.071	.001
csp	.000	.002	-.004	.008	-.000
bmr	.614	.237	.121	1.207	.617
ntis	.018	.015	-.031	.054	.020
tbl	.042	.032	.000	.163	.039
lty	.057	.030	.018	.148	.056
ltr	.005	.024	-.084	.152	.003
dfy	.010	.005	.003	.032	.008
dfr	.000	.011	-.051	.070	.000
inf	.003	.005	-.014	.057	.003

Ratio", since they are not available at a monthly frequency. "Dividend to Price Ratio"is excluded from our multivariate study since it is almost perfectly correlated to **dy**. The "Term Spread"is also excluded for collinearity reasons since it is the difference between the variables **lty** and **tbl**. Furthermore, we consider a constant term in our predictive models. Table 1 provides summary statistics of the used data.[14]

3.2 Parameter Choices

The approach outlined in Section 2 requires the choice of appropriate priors and the selection of adequate values of δ. For the actual implementation, we perform the estimation

[14]We also performed the empirical analysis on a set of predictive variables proposed by Cremers (2002). For reasons of brevity, we decided to report only the results based on the Goyal and Welch dataset. Our results are qualitatively the same for the two datasets. There are, however, potentially interesting quantitative differences that we might investigate in future work. Detailed results on the Cremers dataset are available from the authors upon request.

procedure for a g-prior with $g = 50$.[15] The second choice is about δ, where we use the following values in our empirical implementation: 1.00, .98, and .96.[16] We choose the values of δ such that we cover the constant case ($\delta = 1.00$), a rather noisy situation where coefficients are expected to change rapidly ($\delta = .96$), and an intermediate case ($\delta = .98$). As described in Section 2.1, the effect of δ strictly lower than 1.00 corresponds to an increase in the variance of the coefficient vector by a factor of $1/\delta$. If we ignore 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 .98. For δ equal to .96, a 50 percent increase in variance will be reached twice as fast, in approximately ten months.

As far as prior probabilities of individual models and model families are concerned, we start out with an uninformed prior giving equal weight to each individual model (i.e., $1/(d \cdot (2^k - 1)) = 1/(3 \cdot (2^{14} - 1)))$ and each individual δ-value (i.e., $1/3$) at the beginning of the estimation horizon. Therefore, every model and every model family has the same chance to turn out to be important.[17]

[15] We repeat the analysis using a g-prior of ten. Finding our conclusions unchanged after this robustness check, we omit the results for the sake of brevity.

[16] Note that we also estimated models with δ equal to .99 and .97. Including these models in the Bayesian Model Averaging does not change the results notably. In order to keep the discussion simple and the number of parameters tractable, we focus on the three delta values mentioned in the text.

[17] To perform a robustness check, we take an even more conservative and skeptical point of view with respect to the existence of predictability. For this reason we attribute a larger prior probability amounting to 50% to the no-predictability benchmark; i.e., the model consisting only of a non-time-varying constant. The remaining models receive equal prior probability amounting to $.5 \cdot 1/(d \cdot (2^k - 2))$. Our results are robust to this change of prior information. The authors will provide detailed results for this specific case upon request.

4 Results

In the result section, we first concentrate on determining whether there is evidence of out-of-sample predictability and whether including models with time-varying coefficients improves predictability. In addition to statistical tests, we investigate if simple trading strategies would have been able to exploit the observed degree of out-of-sample predictability. Further, we evaluate the relationship between predictability and the business cycle in order to get a better understanding of the sources of predictability. After documenting that time-varying coefficients significantly improve prediction quality, we investigate the characteristics of these models in more detail. Finally, we illustrate how our models with time-varying coefficients adjust using a case study that examines the dividend yield as a predictive variable before and after release of Rule 10b-18 in November 1982.

4.1 Out-of-Sample Predictability

To test for out-of-sample predictability, we analyze the differences in mean squared prediction errors (MSPE) between the no-predictability benchmark and a predictive model. The no-predictability benchmark is the unconditional model that neglects the predictive power of any of the 13 predictive variables and takes the historical long-term average equity premium as the best prediction for the following month's premium.[18] We find broad support — over different subsamples, using statistical and economic measures — for the conclusions that predictive regressions with time-varying coefficients predict market re-

[18]This no-predictability benchmark model is thus nested in our universe of predictive regressions and corresponds to the model that includes only the constant as a predictor and assumes that the coefficient of the constant does not vary over time.

turns significantly better than the unconditional mean and that they perform significantly better than regressions with constant coefficients. More specifically, we consider the following different predictive models in this analysis:

- *BMA-Model incl. (excl.) TVar-Coeff.:* this model represents the Bayesian model average across all individual models including (excluding) models with time-varying coefficients.

- *Univariate Models incl. (excl.) TVar-Coeff.:* These models consider only one predictive variable at a time. In the cases where we include time-varying coefficients, we still use Bayesian model averaging to average across models with different assumptions of the degree of time-variation of the coefficient.

- *MOST-Model incl. (excl.) TVar-Coeff.:* The MOST-Models represent the individual models that receive most posterior probability — among all individual models including (excluding) models with time-varying coefficients — at the end of the month before the evaluation period starts. Then we keep this model specification (the variable selection and degree of time-variation of the coefficients) constant during the evaluation period, but we update the coefficient estimates.

- *MEDIAN-Model incl. (excl.) TVar-Coeff.:* The MEDIAN-Model represents the individual model that includes all variables that receive more than 50% posterior probability — among all individual models including (excluding) models with time-varying coefficients — at the end of the month before the evaluation period starts. In this case, we keep only the selection of variables constant during the evaluation period. For the MEDIAN-Model including time-varying coefficients, we use Bayesian model averaging again to average across different degrees of time-variation.

The motivation to look at univariate models, the MOST-Model and the MEDIAN-Model is to obtain a cleaner test of the importance of time-varying coefficients. The advantage of these models in contrast to the BMA-Models is that we fix the selection of variables. Any performance differences we find for these models between the versions including and excluding time-varying coefficients can, thus, be unambiguously related to the influence of time-varying coefficients.

4.1.1 Statistical Evaluation

For each predictive model, Table 2 reports differences in mean squared prediction errors relative to the no-predictability benchmark. Furthermore, we report p-values of tests that the reported differences in MSPEs are significantly larger than zero (i.e., implying that the predictive model predicts more accurately than the benchmark) and that unreported differences in MSPEs between models including and excluding time-varying coefficients are significantly larger than zero (last column). We properly account for the fact that these tests compare models that are nested and, therefore, correct the statistics (the differences in MSPEs) according to Clark and West (2006). Table 2 consists of four panels that summarize our results for four different sample periods: 1947+, 1965+, 1976+, and 1988+. This choice of sample periods is mainly driven by issues of comparability to other studies (especially, Goyal and Welch (2008) and Rapach, Strauss, and Zhou (2009)).

Furthermore, a common result of recent studies is that the evidence of out-of-sample predictability is largely driven by a few exceptional return observations. For this reason two of the sub-periods start immediately after periods of distress, the oil price shock of 1975 and the stock market crash in 1987.

We start with the analysis of the BMA-Model. Only if time-varying coefficients are considered, the resulting BMA-Model outperforms the no-predictability benchmark significantly in all sub-samples (the BMA-Model with constant coefficients succeeds only in the "1947+" period).[19] Furthermore, the BMA-Model including time-varying coefficients consistently and significantly improves the performance relative to the BMA-Model excluding time-varying coefficients (with p-values of 1% or lower across all sample periods).

Next, we focus on the 13 univariate models nested in the universe of models we consider. We find a significant improvement in prediction accuracy after including time-varying coefficients in many cases. Univariate models with time-varying coefficients significantly outperform the ones with only constant coefficients in 28 out of 52 cases across all sub-periods.[20] Furthermore, in only 2 out of 52 cases the model with constant coefficients tends to predict more accurately (indicated by a p-value that exceeds 50% in the last column of each table). Relative to the no-predictability benchmark, however, few univariate models perform consistently well. In the case of models excluding time-varying

[19]The results we find for differences in MSPEs are confirmed when we look at the Bayes Factors, which represent alternative Bayesian statistics. For the entire data sample from 1937 to 2003, for example, the weight of the no-predictability benchmark within the BMA-Model including time-varying coefficients drops from its naive prior of $(1/3)(1/16383) = 2.03 \cdot 10^{-5}$ to $2.66 \cdot 10^{-11}$. This result compares well to Cremers (2002). Detailed results are available from the authors upon request.

[20]At a 10% significance level.

coefficients, not a single univariate model outperforms the no-predictability benchmark significantly across all sample periods (hence, our results perfectly support the findings of Goyal and Welch (2008)). If coefficients are also modeled dynamically, we find one variable that consistently beats the historic average in a univariate framework, namely **csp**.

Finally, we further confirm the evidence that time-varying coefficients improve prediction accuracy for individual models by looking at the MOST and MEDIAN-Model. In both cases, the consideration of time-varying coefficients results in a significant performance enhancement across all sample periods.[21] In the case of these two models, the performance relative to the no-predictability benchmark also increases significantly once time-varying coefficients are considered. Except for the MEDIAN-Model in the 1988+ sub-sample (p-value of .16), they beat the historic mean consistently. Therefore, these models seem to represent quite reasonable alternatives to the BMA-Model. Note, however, that this is not at all the case if coefficients are restricted to be constant.

From these results we conclude that the inclusion of time-varying coefficients dramatically improves the out-of-sample predictability — across all model specifications and across all sub-periods. If time-varying coefficients are considered, the overall best performing model is the BMA-Model, as it shows the clearest performance advantage relative to the no-predictability mean. It is followed by the MOST-Model and the univariate model based on **csp**, which also show consistent, strong, out-of-sample predictive performance.

[21] In the case of the MOST-Model, the model with highest posterior probability in December 1946 is a

Table 2: **Statistical Evaluation:** This table summarizes the differences in MSPEs (multiplied by 100) between the no-predictability benchmark and a predictive model. It also provides the p-values of one-sided tests that the difference is larger than zero. The last column reports the p-values of one-sided tests that use the corresponding model with constant coefficients as benchmark. Given that we compare prediction quality with respect to a nested model, we apply the definitions of Clark and West (2006) for the statistics of the differences of MSPEs.

Sample Period: 1947+

	Models incl. TVar-Coeff.		**Models excl. TVar-Coeff.**		**Comp**
	Diff. in MSPE	p-value	Diff. in MSPE	p-value	p-value
BMA-Model	.0192	.00	.0046	.08	.00
ep	.0055	.02	.0026	.04	.12
svar	.0004	.41	-.0005	.95	.30
bmr	.0031	.21	.0031	.14	.05
tbl	.0055	.02	.0033	.08	.01
ltr	.0050	.01	.0018	.14	.01
dfy	.0035	.02	-.0001	.64	.02
inf	.0027	.09	.0009	.16	.18
dy	.0044	.01	.0036	.02	.08
dpayr	.0027	.17	-.0001	.54	.12
csp	.0094	.01	.0052	.02	.01
ntis	.0043	.01	.0020	.10	.04
lty	.0035	.08	.0025	.13	.06
dfr	.0008	.16	-.0005	.81	.02
MOST-Model	.0054	.10	.0054	.10	.
MEDIAN-Model	.0220	.00	.0028	.22	.00

Sample Period: 1965+

	Models incl. TVar-Coeff.		Models excl. TVar-Coeff.		Comp
	Diff. in MSPE	p-value	Diff. in MSPE	p-value	p-value
BMA-Model	.0261	.00	.0039	.11	.00
ep	.0067	.04	.0023	.11	.07
svar	.0007	.39	-.0007	.94	.29
bmr	.0018	.37	.0013	.35	.09
tbl	.0063	.05	.0047	.08	.11
ltr	.0056	.02	.0036	.06	.14
dfy	.0048	.02	-.0003	.79	.02
inf	.0029	.15	.0008	.24	.19
dy	.0046	.03	.0038	.05	.18
dpayr	.0041	.16	.0004	.38	.17
csp	.0106	.03	.0048	.03	.08
ntis	.0049	.03	.0031	.07	.18
lty	.0044	.12	.0035	.15	.13
dfr	.0008	.23	-.0003	.65	.08
MOST-Model	.0141	.06	.0015	.35	.00
MEDIAN-Model	.0018	.02	.0003	.47	.00

Sample Period: 1976+

	Models incl. TVar-Coeff.		Models excl. TVar-Coeff.		Comp
	Diff. in MSPE	p-value	Diff. in MSPE	p-value	p-value
BMA-Model	.0304	.00	.0038	.18	.00
ep	.0018	.21	.0017	.22	.08
svar	-.0007	.87	-.0008	.91	.28
bmr	-.0006	.61	-.0005	.56	.02
tbl	.0015	.31	.0013	.35	.04
ltr	.0030	.16	.0030	.16	.49
dfy	.0027	.10	-.0000	.54	.10
inf	-.0003	.61	-.0002	.57	.66
dy	.0018	.26	.0016	.30	.02
dpayr	-.0001	.52	-.0002	.58	.31
csp	.0097	.04	.0031	.16	.05
ntis	.0039	.07	.0037	.08	.22
lty	.0005	.43	.0005	.43	.11
dfr	.0002	.41	-.0003	.71	.08
MOST-Model	.0252	.00	-.0012	.60	.00
MEDIAN-Model	.0307	.00	-.0004	.54	.00

Sample Period: 1988+

	Models incl. TVar-Coeff.		Models excl. TVar-Coeff.		Comp
	Diff. in MSPE	p-value	Diff. in MSPE	p-value	p-value
BMA-Model	.0210	.08	-.0012	.60	.01
ep	.0036	.12	.0036	.13	.05
svar	-.0006	.76	-.0007	.76	.30
bmr	.0009	.39	-.0006	.57	.06
tbl	-.0012	.85	-.0016	.89	.09
ltr	.0008	.37	.0008	.37	.34
dfy	.0006	.27	-.0003	.80	.19
inf	-.0005	.70	-.0006	.71	.42
dy	.0013	.38	.0007	.44	.01
dpayr	-.0018	.92	-.0019	.92	.06
csp	.0148	.05	.0029	.18	.05
ntis	.0003	.46	.0002	.47	.21
lty	-.0007	.81	-.0008	.82	.08
dfr	-.0014	.94	-.0010	.91	.97
MOST-Model	.0280	.06	.0004	.48	.00
MEDIAN-Model	.0134	.16	.0062	.18	.01

4.1.2 Economic Evaluation

So far, we have documented that, statistically speaking, models with time-varying coefficients represent a significant improvement. In a further step, we test whether the identified levels of out-of-sample predictability of monthly S&P 500 returns are sufficient such that an investor might rationally use the predicted return (and its estimated variance) for portfolio optimization (see Kandel and Stambaugh (1996) and Campbell and Thompson (2008)). To test for economic evidence that a trading strategy could have exploited this degree of out-of-sample predictability in a profitable way, we follow Campbell and Thompson (2008) and Rapach, Strauss, and Zhou (2009) and consider an investor with a single-period horizon and mean-variance preferences. We analyze the gain in realized

model with constant coefficients; thus, there is no p-value for the comparison.

utility of an investor who uses any of the predictive models in comparison to the no-predictability benchmark.

Table 3 summarizes these results for the same set of predictive models and sample periods.[22] These utility gains very convincingly confirm and even strengthen our previous results. Overall, the BMA-Model and the MEDIAN-Model show best performance, i.e., consistently positive and large utility gains, if time-varying coefficients are included. These utility gains are statistically significant during all evaluation periods except the 1988+ period. The differences, however, between the models including time-varying coefficients and the ones excluding time-varying coefficients are statistically significant during all periods (also for the MOST-Model). Regarding univariate models, the inclusion of time-varying coefficients improves the performance of each individual model across all sub-periods.[23] However, only **csp** generates positive utility gains consistently across all sub-periods (only the one during the 1965+ period is significant).

4.2 Return Predictability and the Business Cycle

In the previous section we have documented statistically significant and economically important levels of predictability for models with time-varying coefficients. In this section we aim to analyze the sources of predictability in more detail. In particular, we relate predictability to the business cycle.

[22] We determine monthly realized mean-variance utility where we use daily S&P 500 returns within a month to estimate the monthly variance. Average realized utility gains and significance levels are inferred from these time series. Tests are omitted for the 1947+ period since our set of daily index returns starts only in 1964 (we use data from Datastream for this purpose).

[23] The only exceptions are the univariate model based on **dfr** in sub-period 1988+ and the one based on **inf** in sub-period 1976+. Several of these improvements are also statistically significant.

Table 3: **Economic Evaluation:** We assume an investor with a single-period horizon, mean-variance preferences and a relative risk aversion equal to 3. Further we limit the share invested into the S&P 500 to be between 0% and 150%. The table shows utility gains p.a. of an investor using any of the predictive models relative to an investor following the no-predictability benchmark. Significance tests are based on the monthly time series of realized utility gains where daily index returns within a month are used to estimate the monthly return variance. Tests are omitted for the 1947+ period since our set of daily index returns starts only in 1964 (we use data from Datastream for this purpose). ***, ** and * indicate standard significance levels of the utility gain relative to the no-predictability benchmark. Bold utility gains in columns 3 to 5 indicate that the models including time-varying coefficients perform significantly better than the models excluding time-varying coefficients at least at the 10% level.

	Models incl. TVar-Coeff.				**Models excl. TVar-Coeff.**			
	1947+	1965+	1976+	1988+	1947+	1965+	1976+	1988+
BMA-Model	2.57	**5.75*** **	**4.82*** **	**1.76**	-1.97	.16	-1.87	-5.79
ep	-.61	1.07	-1.36	-1.77	-1.60	-.22	-1.40	-1.79
svar	-.77	**.35**	-1.67	-2.95	-1.69	-.46	-1.80	-2.97
bmr	-1.74	-.71	-3.67	-3.27	-3.13	-2.07	-3.83	-6.41
tbl	-.55	**.63**	**-2.72**	**-4.04**	-1.47	.14	-3.30	-4.66
ltr	.38	1.60	-.73	-2.80	-.96	1.14	-.82	-2.89
dfy	.01	**1.36**	-1.01	-2.27	-1.46	-.46	-1.67	-3.12
inf	-.23	.74	-2.05	-3.13	-.75	.51	-1.98	-3.14
dy	-.77	.74	-2.27	-4.84	-1.47	.38	-2.43	-5.08
dpayr	-.82	**.36**	-2.47	-4.56	-1.81	-.26	-2.58	-4.72
csp	.88	**2.90***	**1.22**	**2.94**	-1.11	1.20	-1.71	-1.98
ntis	-.13	1.16	**-1.13**	-3.15	-1.66	-.05	-1.34	-3.22
lty	-1.03	.37	-2.79	-3.82	-1.33	.31	-2.84	-3.88
dfr	-.82	**.51**	-1.38	-3.69	-1.87	-.46	-1.69	-3.41
MOST-Model	-2.30	**3.55***	**4.24****	**2.93**	-2.30	-2.44	-4.06	-4.36
MEDIAN-Model	2.68	**4.04****	**4.92****	**2.26**	-2.87	-3.06	-3.97	-3.31

4.2.1 Financial Returns and the Real Economy

From a theoretical point of view, Campbell and Cochrane (1999) provide a foundation for the link between time-varying expected rates of returns and the business cycle. Simply speaking, the argument is as follows (see Cochrane (2007)): investors have a slow-moving external habit; if the economy slides into a recession, the risk of falling short of the minimum level of consumption increases and investors become more risk averse; thus, the risk premium of equity has to go up during a recession; in order to make that happen, stock prices have to decrease at the beginning of a recession. The time-variation in risk premium is, therefore, linked to the time-variation in investors' risk aversion.

In this section, we are going to link these theoretical predictions to the empirical results of our models. Specifically, we expect the estimated risk premium to behave according to the dynamics implied by the Campbell and Cochrane (1999) model: it has to be negative (on a monthly frequency) during the beginning of a recession, increase during the recession, and be larger at the end of the recession than at the end of the expansion. In such a framework, predictability would arise if our predictive models are able to anticipate the business cycle (see Henkel, Martin, and Nardari (2008) and Rapach, Strauss, and Zhou (2009) for initial empirical support).

Models with dynamic coefficients should outperform models with constant coefficients (as we documented for the entire sample in the previous section) if the relationships between individual predictive variables and the risk premium depend on the business cycle, as well. This link between the business cycle and the time-variation in coefficients can be motivated by different economic theories. Veldkamp (2005) and Van Nieuwerburgh and Veldkamp (2006), among others, relate learning asymmetries caused by a varying rate of information flow to the business cycle. In these models, the information content of economic signals varies across the business cycle. Chakley and Lee (1998) offer a different mechanism to cause the asymmetries in learning by claiming that during recessions the fraction of noise traders increases. McQueen and Roley (1993) and Boyd, Hu, and Jagannathan (2005) find empirical evidence for these asymmetric learning patterns, as the incorporation of news into stock prices varies with the business cycle. It is exactly this variation in learning[24] and in the information flow that we try to capture with our time-varying coefficients.

4.2.2 Predictive Performance Across Business Cycles

We use the NBER dates of peaks and troughs to identify recessions and expansions ex-post; i.e., this information is not used at any time during the estimation of the predictive models. It is currently not our goal to predict business cycles. The idea of this analysis is to see how closely the level of predictability and the dominance of models with time-varying coefficients is related to the business cycle.

Table 4 summarizes our main two statistics — differences in mean squared prediction errors (Diff. MSPE) and utility gains — across models for different periods related to the business cycle. Consistent with other recent papers (Henkel, Martin, and Nardari (2008)

[24]Theories about learning fit, in general, very nicely into our empirical Bayesian framework, as our investors basically learn about the coefficients and other parameters over time and update their beliefs every period.

and Rapach, Strauss, and Zhou (2009)[25]), we find significantly stronger evidence for predictability during recessions than during expansions (third row of Table 4) using both measures.[26] It is interesting to highlight that utility gains relative to the no-predictability benchmark are huge during recessions. This is primarily because the no-predictability benchmark is overly optimistic about the monthly equity premium and thus suffers from severe losses during recessions. Another important result is that the dominance of models with time-varying coefficients prevails during both recessions and expansions (see last two columns of Table 4). Finally, we also find statistically significant levels of out-of-sample predictability during expansions, but only for models including time-varying coefficients, albeit at a much smaller scale. This result is in contrast to the findings of Henkel, Martin, and Nardari (2008), who conclude that there is even no in-sample predictability during expansions using their predictive variables and econometric technique.

In the next step we look more closely at economic turning points; i.e., peaks and troughs of the business cycle. For this purpose, we split the business cycle into 4 periods of 3 months each:[27] (i) Late Expansion: 3 months before a peak, (ii) Early Recession: 3 months after a peak, (iii) Late Recession: 3 months before a trough and (iv) Early Expansion: 3 months after a trough. The last four rows of Table 4 report the results for these sub-periods. The BMA-Model incl. TVar-Coeff. outperforms the no-

[25]Henkel, Martin, and Nardari (2008) analyze market return predictability in the G7 countries. In contrast to our study, they perform an in-sample analysis, look at only 4 predictive variables and use regime-switching models. Rapach, Strauss, and Zhou (2009) focus on the US market and analyze basically the same set of predictive variables as we do. However, they assume constant coefficients and apply a rather ad-hoc way of averaging across individual predictive models.

[26]The only exception, as it is not significantly different from zero, is the difference in MSPEs for models excluding time-varying coefficients.

[27]Our results are robust to different choices of windows around peaks and troughs.

Table 4: **Business Cycle Analysis:** This table summarizes our main statistics across recessions (123 monthly observations) and expansions (665 monthly observations) and across 4 business cycle (BC) periods (33 monthly observations per period): Late Expansion: 3 months prior to peak, Early Recession: 3 months after peak, Late Recession: 3 months before trough, Early Expansion: 3 months after trough. The statistics include differences in mean squared prediction errors relative to the no-predictability benchmark (Diff. MSPE) and utility gains relative to an investor using the unconditional mean return (Util. Gain). Significance tests are relative to the no-predictability benchmark except for the columns labeled "Model Comparison"(in this case, the significance tests is across models) and the row labeled "Diff."(in this case, the test is between recessions and expansions). Significance tests for differences between values of specific statistics across individual stages of the business cycle are discussed and reported in the text.

	Models incl. TVar-Coeff.		**Models excl. TVar-Coeff.**		**Model Comparison**	
	Diff. MSPE	Util. Gain	Diff. MSPE	Util. Gain	Diff. MSPE	Util. Gain
Rec.	.0600***	24.967***	.0168**	9.378*	.0478***	15.589***
Exp.	.0121**	1.622	.0052	-1.990	.0152***	3.613**
Diff.	-.0479***	-23.345***	-.0116	-11.369**		
Late Exp.	.0327***	16.597*	.0148	3.841	.0184*	12.756
Early Rec.	.1156***	47.849***	.0707***	30.543**	.0401**	17.306**
Late Rec.	.0490**	6.4206	-.0092	-6.810	.0752***	13.230
Early Exp.	.0075	-2.273	-.0039	-8.321*	.0266***	6.048

predictability benchmark significantly for all stages except early Expansion; i.e., shortly after the trough. Even the BMA-Model excl. TVar-Coeff. shows predictability around the peak of the business cycle. A closer look at the utility gains relative to an investor using the no-predictability benchmark reveals that the naive investor performs relatively well towards the end of a recession and early in an expansion, because of the nearly constant and high weight on the risky asset. These utility gains, however, do not offset the huge losses such an investor suffers from during the beginning of a recession.

Figure 1 shows the predicted equity premium (first row) and the equity market weight of a mean-variance optimizing investor (second row) across peaks and troughs. It shows that the predictions from the BMA-Model incl. TVar-Coeff. fit the theoretical pattern implied by Campbell and Cochrane (1999) very well: in Early Recession, predicted returns are on average negative[28] (the difference between predicted returns in Late Expansion and in Early Recession is statistically significant). Towards the end of the recession, however, the predicted risk premium increases and peaks in Late Recession, reflecting the fact that investors become more risk-averse during a recession. During expansion, predicted risk premia decrease again; hence, the difference between the expected risk premia in Late Expansion and in Late Recession is statistically significant. In contrast, the predictions from BMA-Model excl. TVar-Coeff. do not match this pattern at all. In this case, the expected risk premium stays at a relatively constant, positive but low level during the entire recession. We conclude that these predictions are, thus, less economically meaningful.

As far as portfolio weights are concerned (second row of illustrations in Figure 1), we find that the asset allocation strategy of an investor relying on the BMA-Model incl. TVar-Coeff. seems to time the market very well. On average, the investor withdraws from the market quickly at the beginning of a recession (the drop in portfolio weight is statistically significant), and then moves back in (even more than before) towards the end of it. In contrast, an investor using predictions from the BMA-Model excl. TVar-Coeff. pulls out of the market after a peak but completely fails to move into the market again towards the end of the recession.

Our model is very nicely consistent with the implications of asset pricing models that use time-varying risk aversion to generate time-varying risk premia (e.g., see Campbell

[28]Although one would expect the long-term equity premium always to be positive, the short-term premium could be negative in a world with time-varying equity premia. In a business cycle model, prices have to drop in the beginning of the recession resulting in short-term negative expected returns (see also Pastor and Stambaugh (2009b)).

and Cochrane (1999)). This agreement between our empirical predictions and asset pricing theory suggests the notion that time-varying risk aversion along the business cycle is related to the existence of out-of-sample predictability. Thus, we conclude that predictability reflects business cycle risk rather than market inefficiency. Therefore, it is also not surprising that predictability is not driven away over time. This view is somewhat supported by the literature on fund manager skills that finds that fund managers perform statistically and economically better during recessions than during expansions (see, for example, Kacperczyk, Van Nieuwerburgh, and Veldkamp (2009)). Thus, we conjecture that fund managers actively exploit the higher levels of market return predictability during recessions, but they are not able to eliminate it because of the risks involved.

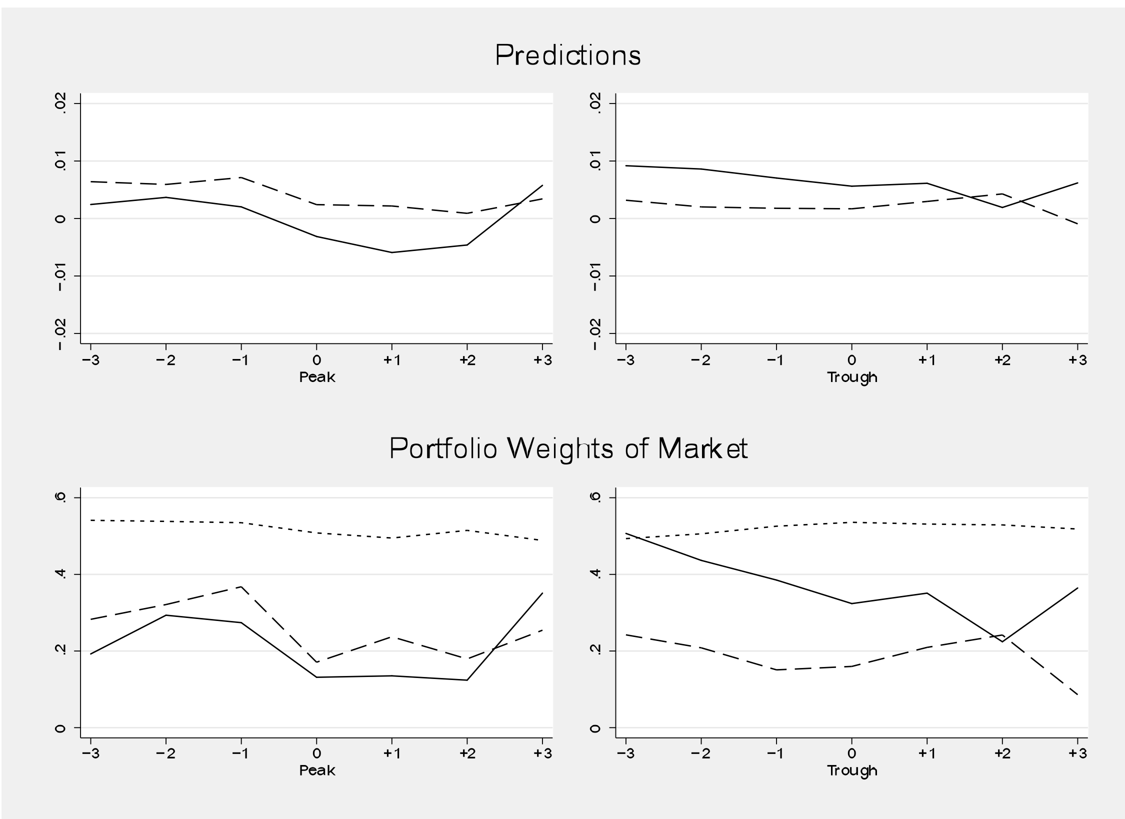
4.3 Characterization of the BMA-Model

The previous section described empirical results that confirm that the BMA-Model including time-varying coefficients performs consistently well at predicting market returns. Given that this model is a fairly sophisticated combination of many individual models, we want to shed some more light on it and evaluate its characteristics in more detail.

4.3.1 Variance Decomposition and the Degree of Time-Variation

As a first step, we perform a variance decomposition. Since the Bayesian model averaging approach keeps track of all possible sources of uncertainty regarding the prediction, we

Figure 1: **Equity Premium Predictions and Portfolio Weights Around Peaks and Troughs:** The two graphs in the first row show the predicted monthly equity premium using BMA-Model incl. TVar-Coeff. (solid line) and BMA-Model excl. TVar-Coeff. (long dashed line). The two graphs in the second row show the portfolio weights of a mean-variance optimizing investor who uses forecasts from BMA-Model incl. TVar-Coeff. (solid line), uses forecasts from BMA-Model excl. TVar-Coeff. (long dashed line), or does not believe in predictability and uses the historic mean and standard deviation (short dashed line). Each graph shows averages across the 11 recessions of our sample period.



can decompose the prediction variance of the return into four parts:

$$
\begin{aligned}
\mathrm{Var}(r_{t+1}) \;=\; & \sum_j \left[\sum_i (S_t|M_i,\delta_j|D_t) P(M_i|\delta_j,D_t) \right] P(\delta_j|D_t) + \\
& \sum_j \left[\sum_i (X_t' R_t X_t|M_i,\delta_j,D_t) P(M_i|\delta_j,D_t) \right] P(\delta_j|D_t) + \\
& \sum_j \left[\sum_i (\hat{r}^j_{t+1,i} - \hat{r}^j_{t+1})^2 P(M_i|\delta_j,D_t) \right] P(\delta_j|D_t) + \\
& \sum_j (\hat{r}^j_{t+1} - \hat{r}_{t+1})^2 P(\delta_j|D_t). \qquad (6)
\end{aligned}
$$

Equation (6) can be deduced by decomposing the variance of the random variable r step by step into expected in-sample variances and inter-sample variances.[29]

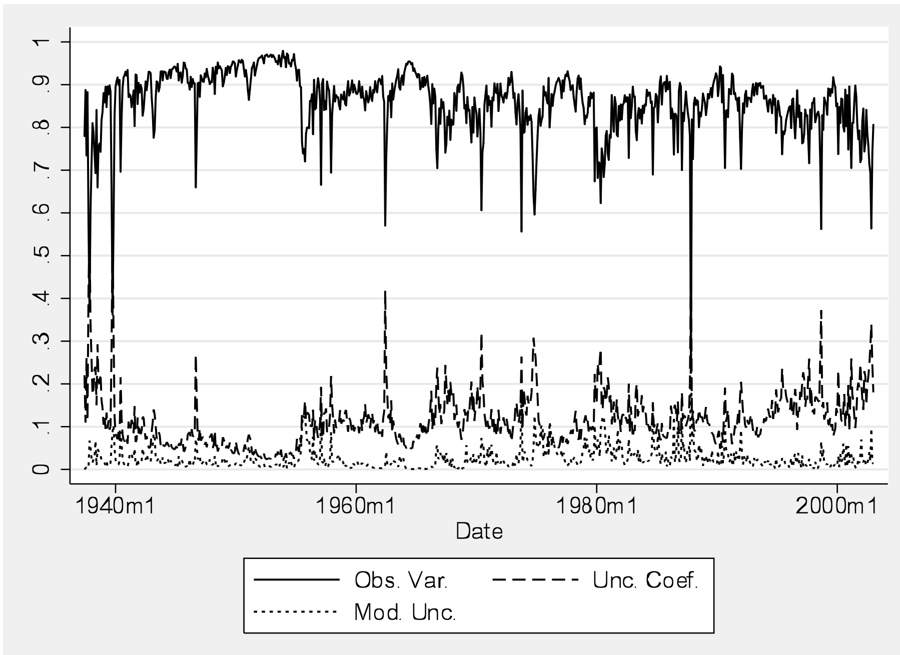
The individual terms of (6) can be interpreted in a very intuitive way. The first term is the expected observational variance; i.e., the variance assigned to the random disturbance term. The second term states the expected variance from errors in the estimation of the coefficient vector. We will refer to it as estimation uncertainty. Both the third and the fourth term characterize model uncertainty. The third term measures model uncertainty with respect to variable selection, and the fourth term measures model uncertainty with respect to the time variability of the regression coefficients.

In Figure 2, we plot the relative weights of these components of prediction variance over time. Panel A shows these components as a fraction of total variance. The dominant source of uncertainty is observational variance. This is not surprising, since over short

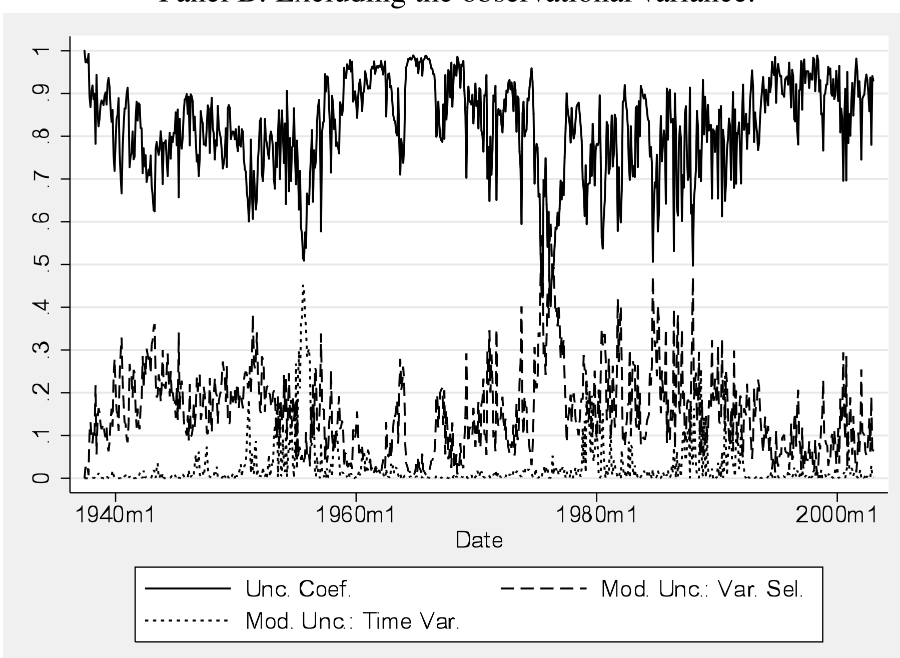
[29] Starting with the decomposition with respect to different values of δ, we can write $\mathrm{Var}(r) = E_\delta(\mathrm{Var}(r|\delta)) + \mathrm{Var}_\delta(E(r|\delta))$, where E_δ and Var_δ denote the expected value and the variance with respect to δ. The term $E_\delta(\mathrm{Var}(r|\delta))$ represents the first three terms in Equation (6). The term $\mathrm{Var}_\delta(E(r|\delta))$ is the last term in (6). In a second step, the term $E_\delta(\mathrm{Var}(r|\delta))$ can be further decomposed into $\mathrm{Var}(r|\delta) = E_M(\mathrm{Var}(r|M,\delta)) + \mathrm{Var}_M(E(r|M,\delta))$, which splits term three of Equation (6) from the remainder. The final variance decomposition as shown in (6) follows from simple rearrangements.

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

Panel A: Including the observational variance.



Panel B: Excluding the observational variance.



prediction horizons, random fluctuations are expected to dominate the uncertainty in the predicted trend component.

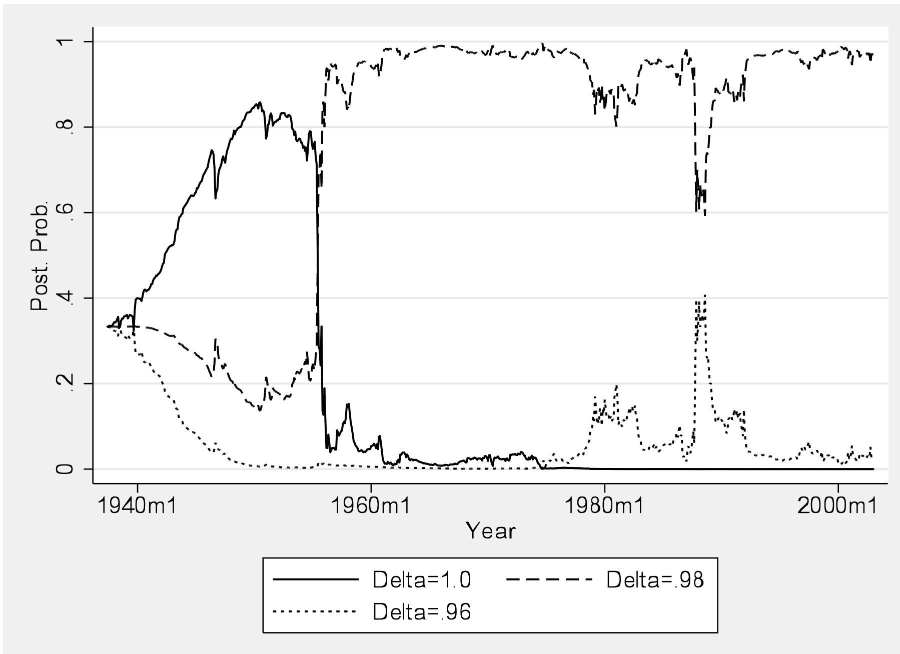
Therefore, Panel B masks out observational variance and focuses only on the other three components. In most periods, the estimation uncertainty in coefficients captures more than half of the remaining variance.[30] In periods of stress, model uncertainty peaks (e.g., in a couple of periods in the 1970s due to oil price shocks, and around 1990 due to the Iraq-Kuwait war). Uncertainty about the correct degree of time-variation (δ) is, in general, relatively low except for individual periods (e.g., in the mid-50s, in the end of the 80s, and in the beginning of the 90s).

Figure 2 shows that there is little uncertainty about the degree of time-variation, but it does not reveal the empirically estimated degree of time-variation. Given the results discussed before, we expect to find that models with time-varying coefficients play an important role within the BMA-Model. To address this question, we plot the total posterior probability of all models for each value of δ considered (see Figure 3).

Figure 3 draws an unambiguous picture. Models with moderately time-varying coefficients (i.e., $\delta = .98$) consistently accumulate more than 80% of posterior probability. Constant coefficient models (i.e., $\delta = 1.0$) perform well over the first 15 years but lose support from the data in and after 1955. Note that the cumulative posterior probability of constant coefficient models basically drops to 0 and stays there from 1974 onwards. In contrast, the very dynamic models with $\delta = .96$ play no role during the 50s and 60s

[30]The fact that parameter uncertainty is more important than model uncertainty most of the time fits well with 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.

Figure 3: **Sum of Posterior Probabilities of Models with a Given δ.** For the BMA-Model including time-varying coefficients (BMA-Model incl. TVar-Coeff.), this figure reports cumulative posterior probabilities of models with a specific degree of time-variation of coefficients.

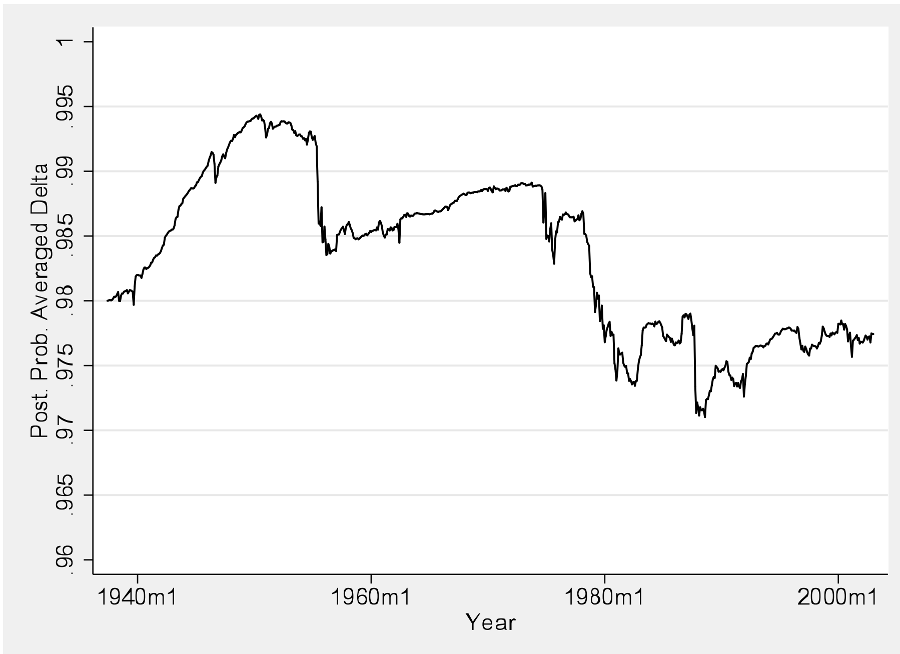


but receive considerable support over some later time periods: especially notable is the short blip following the stock market crash in October 1987. Given the dominance of the models with $\delta = .98$ in Figure 3, it is not surprising that we find little uncertainty about the degree of time-variation in Figure 2.

Similarly, Figure 4 shows the posterior probability weighted average value of δ; i.e. the estimated degree of time-variation in coefficients across time.[31] We see that the degree of time-variation itself changes over time: periods with relatively stable estimates of

[31] Note that in order to obtain a more precise picture of the average delta, we re-estimated the BMA model with a set of five different values of δ, i.e., $\delta \in \{.96, .97, .98, .99, 1.00\}$. To perform a robustness check, we recalculated mean squared prediction errors and utility gains from this more precise model and found our previously reported and discussed results basically unchanged.

Figure 4: **Posterior Probability Weighted Average δ (i.e., degree of time-variation).** For the BMA-Model including time-varying coefficients, this figure reports the posterior probability weighted average degree of time-variation. In order to get a more precise estimate of this average δ we consider five specific delta values in the estimation, namely $\delta \in \{.96, .97, .98, .99, 1.00\}$.



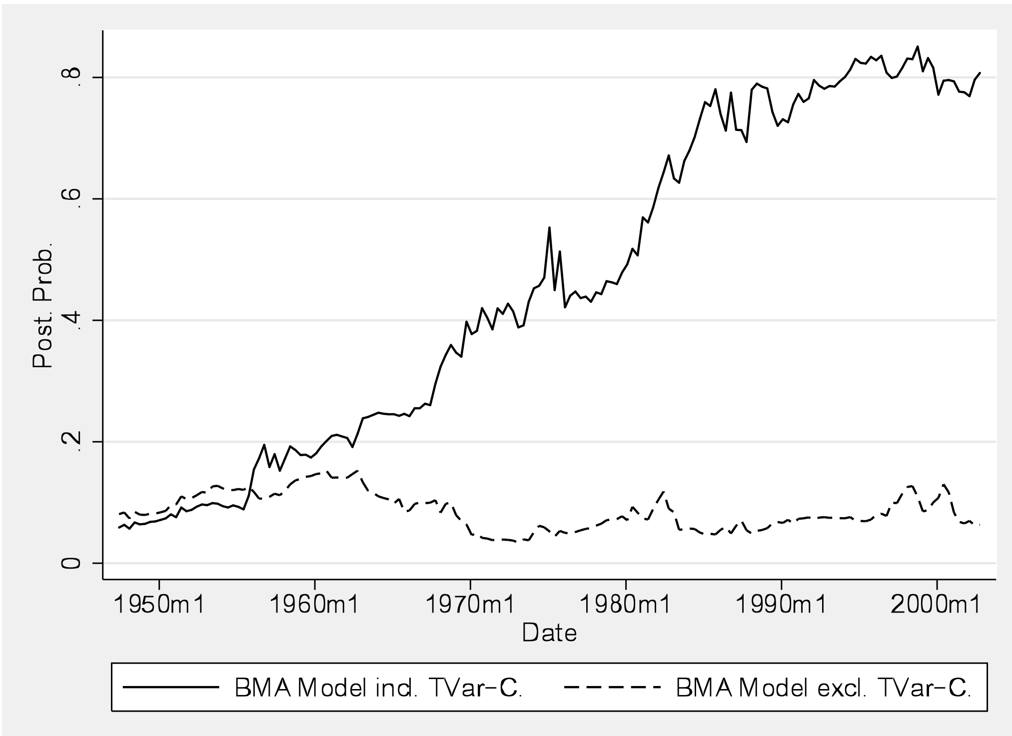
δ (e.g., from the mid-50s to the mid-70s) alternate with periods showing sharp changes, mostly steep drops. These sharp drops in average δ (i.e., increases in the estimated variability of the regression coefficients) can in many cases be associated with crises like the oil price shock of the mid-70s or the stock market crash of 1987. A potential future research question is to more precisely relate the dynamics of the estimated degree of time-variation to the economic cycle or other economic events (see Henkel, Martin, and Nardari (2008) for evidence that parameter instability is related to cyclical economic conditions).

4.3.2 Analysis of Individual Coefficients and Models

Another interesting analysis is to characterize the top performing 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 individual models within the BMA-Model excluding time-varying coefficients as well as within the BMA-Model including time-varying coefficients. Figure 5 shows how much posterior probability the Top 10 models receive over time. In the case of the BMA-Model excluding time-varying coefficients, the posterior probability assigned to the Top 10 models does not account for more than 7 percent at the end of the sample period and never exceeds 16 percent. In contrast, the posterior probability assigned to the Top 10 models of the BMA-Model including time-varying coefficients increases to more than 80 percent over the sample period. Consequently, in the case of the BMA-Model excluding time-varying coefficients, the Top 10 individual models are less distinct from other individual models.

This is a potentially important insight, as it provides an explanation for the erratic behavior of the best models reported in the literature to date. 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 as-

Figure 5: **Sum of Posterior Probabilities of Top 10 Models.** This figure reports the sum of the posterior probabilities of the Top 10 Models (i.e., the 10 models whose posterior probabilities are largest at a given point in time) within two groups of models: (1) the BMA-Model including time-varying coefficients and (2) the BMA-Model excluding time-varying coefficients.



suming constant coefficients. However, we show that this "stationarity issue" can be largely resolved by allowing coefficients to vary over time.

In the next step, we evaluate the importance of individual predictive variables in the BMA-Models. For each variable, we use the sum of posterior probabilities of all models that include this variable as our measure of importance. This measure is the natural choice in a Bayesian framework and allows us to evaluate ex-post how much support individual variables receive from the data. The limitation of this measure is, however, that it does not directly analyze the predictive power of individual variables.

Table 5 evaluates this measure of importance at four points in time (Dec. 1964, Dec. 1975, Dec. 1987, Jan. 2003) and shows a few interesting results. First, the dividend yield, the cross-sectional premium and the book to market ratio consistently receive the highest posterior probabilities. These variables receive weights that are larger than 50% (i.e., the unconditional prior value) across all four points in time (in most cases, their posterior probability exceeds 90%). Second, in contrast to the previous result, we find that no single variable consistently exceeds the prior of 50% if we limit our analysis to the BMA-Model excluding time-varying coefficients. The cross-sectional premium performs best and falls short only of the unconditional prior in December 1975 with a value of 49%. Together, these results further emphasize the previous observation that the assumption of constant coefficients results in instability of models, i.e., in instability of the assessment of importance of predictive variables.

Putting this section's results together, we conclude that the BMA-Model including time-varying coefficients is more successful in identifying important variables and models (i.e., combinations of variables) than the BMA-Model excluding time-varying coefficients. We think that a possible explanation for this observation is that models with constant coefficients flip between individual variables or models to compensate for the lack of variation in the coefficients.

4.4 Case Study: The Dividend Yield as a Predictive Variable

In this section, we perform a case study. We focus on the dividend yield as a predictive variable and analyze how its predictive performance changed due to release of Rule 10b-

Table 5: **Importance of Individual Variables:** This table measures the sum of posterior probabilities across all models that include a specific explanatory variable at 4 points in time. Columns 2 to 5 cover all models, and columns 6 to 9 focus on models with constant coefficients. See section 3.1 for the definition of the variables and their abbreviations.

	Models incl. TVar-Coeff.				**Models excl. TVar-Coeff.**			
	1964.12	1975.12	1987.12	2003.1	1964.12	1975.12	1987.12	2003.1
dy	.94	.80	.93	.91	.72	.40	.31	.25
ep	.25	.16	.46	.77	.29	.30	.27	.25
dpayr	.38	.21	.58	.78	.32	.45	.46	.22
svar	.17	.07	.01	.01	.22	.22	.44	.25
csp	.99	1.00	1.00	1.00	.71	.49	.55	.89
bmr	1.00	1.00	1.00	1.00	.98	.96	.96	.39
ntis	.25	.08	.13	.40	.35	.31	.83	.69
tbl	.37	.26	.95	.71	.38	.34	.44	.37
lty	.41	.70	.08	.37	.39	.39	.27	.41
ltr	.27	.21	.19	.12	.30	.50	.59	.63
dfy	.12	.06	.01	.18	.39	.50	.32	.21
dfr	.06	.02	.13	.03	.21	.16	.43	.28
inf	.31	.19	.07	.03	.49	.71	.47	.33

18 by the SEC in November 1982. We do this case study for two important reasons: (i) to discuss the adaptation of dynamic linear models to changes in the economic relationships (in this case to changes in the regulatory framework), and (ii) to compare the performance of dynamic linear models to regime-switching models. Rule 10b-18 facilitated share repurchases under certain circumstances (see Grullon and Michaely (2002) for details on Rule 10b-18). As a consequence of this change in regulation, individual firms' dividend and payout policies adjusted, resulting in a significant reduction in aggregate dividend yield combined with an apparent change in the information content of dividend payments (see Boudoukh, Michaely, Richardson, and Roberts (2007) for empirical evidence).

Lettau and Van Nieuwerburgh (2008) provide strong in-sample evidence for regime shifts in the long-term mean of the dividend-price ratio. Allowing for one regime shift in the mean dividend yield, their in-sample analysis dates the shift to the year 1991. If two shifts are allowed, these shifts are dated to the years 1954 and 1994. The authors fail, however, to link these dates to specific economic events causing these regime shifts. While most regime-shifting models concentrate only on ex-post predictability and in-sample detection of shifts, Lettau and Van Nieuwerburgh (2008) also explicitly analyze the out-of sample properties of their regime-shifting model. They find poor predictive quality which is dominated by their no-predictability benchmark. This is so because of non-reliable real-time results in (i) dating regime shifts and more severely (ii) the estimation of the size of the shift in the steady state. That is, they find that regime-shifting models have considerable difficulty in learning out-of-sample whether a shift has occurred recently.

How, in contrast, does our methodology perform in detecting and learning this regulatory change in real-time? The BMA-Model including time-varying coefficients does very well in handling the regime shift. Figure 6 shows the dividend yield's importance over time, measured as the sum of the posterior probabilities assigned to individual predictive models including the dividend yield.[32] Two different models are compared: the BMA-Model including time-varying coefficients, and the BMA-Model excluding time-varying coefficients. The vertical line in the graph indicates the date of the release of rule 10b-18.

The BMA-Model including time-varying coefficients views the dividend yield as a consistently important variable. In a reaction to the structural change caused by the release of rule 10b-18, the BMA-Model including time-varying coefficients increases the

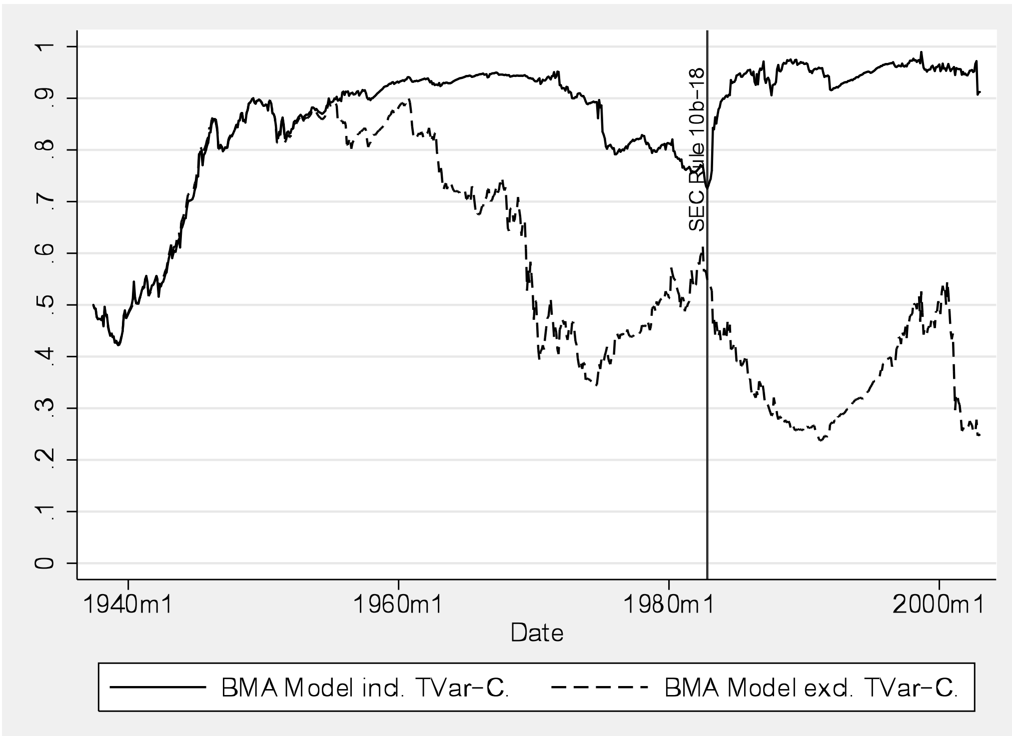
[32] We decided to analyze the importance of models including the dividend yield because this captures the importance of the economic link between the dividend yield and the risk premium. Alternatively, one could also analyze the average coefficient of the dividend yield. Qualitatively, this would yield a similar result.

overall weight of the dividend yield. This reaction is immediate and suggests that the information content of dividend payments increased, although overall dividend payments declined. This is so because (i) the dynamic linear models adapt their coefficients to the new situation and (ii) due to Bayesian learning, models with good out-of-sample performance receive (step by step) higher weights.

In contrast, models with constant coefficients (BMA-Model excluding time-varying coefficients) cannot properly handle the update in the regulatory framework that obviously changed the predictive impact of the dividend yield on equity returns. Since constant-coefficients models are, by definition, only slowly adapting estimated sensitivities, the only possible reaction to bad calibration is that the BMA procedure weights down models that include the dividend yield as a predictor (the importance of the dividend yield as a predictor drops by 19.9% in March 1983). This pattern can also explain the results reported in Goyal and Welch (2008) and Ang and Bekaert (2007), who detect instability of prediction models using the dividend yield.

To conclude, this small case study shows that our framework with time-varying coefficients can quickly learn — in real-time — changes in economic relationships, even if these changes are discrete jumps such as the release of Rule 10b-18. In contrast, regime-switching models that focus exclusively on the dividend yield as a predictor seem to perform much worse out-of-sample (see, for example, Lettau and Van Nieuwerburgh (2008)). We interpret this case-study evidence as supportive of our choice of econometric technique, especially considering that the goal of our study is to evaluate out-of-sample predictability of a comprehensive set of 13 predictive variables rather than only the divi-

Figure 6: **The Dividend Yield as a Predictive Variable:** This figure reports the sum of posterior probabilities of all models including the dividend yield as a predictive variable for two groups of models: (1) the BMA-Model incl. TVar-Coefficients and (2) the BMA-Model excl. TVar-Coefficients.



dend yield.

5 Conclusion

Although the literature on equity return prediction is growing quickly, it is still quite inconclusive about two fundamental questions: Does out-of-sample predictability exist, and what are the important predictive variables? The literature agrees, however, that parameter instability represents a major challenge in this area. Most papers address it by using rolling window regressions and/or by performing sub-period investigations. Both approaches are ad-hoc, non systematic and unhelpful in understanding the true degree of parameter instability. In contrast, we propose a systematic way to take time-variation of coefficients into account.

Coming back to the fundamental questions in return prediction, we find large, significant and consistent improvements in the accuracy of out-of-sample predictions if models with time-varying coefficients are considered. These gains in prediction accuracy also result in considerable economic profits for an investor who uses the predictions of our framework with time-varying coefficients. Such an investor outperforms both an investor who uses constant coefficient models and an investor who uses the unconditional mean and variance.

Furthermore, we find that predictability is closely related to the business cycle. Our empirical methodology predicts on average a decreasing (increasing) equity risk premium during expansions (recessions) — exactly as implied by asset pricing theory (e.g., Campbell and Cochrane (1999)). In this theory, the driving force behind this pattern is time-varying risk aversion. Thus, we view our study's results as consistent with a story in which time-varying risk aversion is responsible (at least partly) for out-of-sample predictability of equity returns.

In contrast to the existing literature, we do not find that predictability exists exclusively during recessions. We also document evidence for out-of-sample predictability during expansions — on a smaller scale and only if time-varying coefficients are taken into consideration. We also analyze the potential sources of this outperformance and find that it is directly related to the inclusion of time-varying coefficients: models with constant coefficients receive basically no support from the data. Further, even if we abstract from the issue of variable selection, we find significant gains in prediction performance for individual models (e.g., univariate models) including time-varying coefficients. Finally, we show that our simple way of modeling time-variation can quickly learn changes in the underlying relationships, such as changes in the regulatory environment in the case of the dividend yield.

While we are confident that our paper provides several contributions to the literature on equity return prediction, it also raises new questions. Most importantly, it raises a question about the economic forces that cause time-varying predictive relationships. In this respect, we would need both more theoretical and more empirical research. In a broader context, our results have important implications for the portfolio optimization and asset allocation literature. Our findings imply that predictive relationships vary considerably over time. Thus, predictions of the equity premium beyond a monthly horizon become more uncertain relative to monthly predictions (see Pastor and Stambaugh (2009a)). How investors should optimally account for this information in their long-term asset allocation decisions is an interesting question for future research.

A Appendix

A.1 The Mathematics of Dynamic Linear Models

From the specification of the dynamic linear model in Equations (1) and (2) in Section 2.1, we develop the recurrence for updating the belief about the system coefficients and the observational variance in response to observing a new return realization (see West and Harrison (1997)). Given a normally distributed prior for the system coefficients θ_0 and an inverse-gamma distributed prior for the observational variance V, this can be done in a fully conjugate Bayesian analysis ensuring that prior and posterior distributions come from the same family of distributions. As a time $t = 0$-prior we use the natural conjugate g-prior specification stated in Equations (3) to (5).

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

$$V|D_t \sim IG\left[\frac{n_t}{2}, \frac{n_t S_t}{2}\right], \tag{7}$$

$$\theta_{t-1}|D_t, V \sim N\left[m_t, VC_t^*\right], \tag{8}$$

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

$$\theta_{t-1}|D_t \sim T_{n_t}\left[m_t, S_t C_t^*\right]. \tag{9}$$

When iteratively updating the estimates, we must remember 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, which increase the variance but preserve the mean of the estimate,

$$\theta_t|D_t \sim T_{n_t}\left[m_t, S_t C_t^* + W_t\right]. \tag{10}$$

As mentioned in Section 2.1, we can find the predictive density of the time $t+1$ return r_{t+1} by integrating the conditional density of r_{t+1} over the 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. The predictive density is then

$$\begin{aligned} f(r_{t+1}|D_t) &= \int_0^\infty \left[\int_\theta \varphi\left(r_t; X_t'\theta, V\right)\varphi\left(\theta; m_t, VC_t^* + W_t\right) d\theta\right] \\ &\quad \times ig\left(\frac{n_t}{2}, \frac{n_t S_t}{2}\right) dV \\ &= \int_0^\infty \varphi\left(r_t; X_t' m_t, X_t'\left(VC_t^* + W_t\right) X_t + V\right) \\ &\quad \times ig\left(\frac{n_t}{2}, \frac{n_t S_t}{2}\right) dV \\ &= \mathbf{t}_{n_t}(r_{t+1}; \hat{r}_{t+1}, Q_{t+1}), \end{aligned} \tag{11}$$

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

$$\hat{r}_{t+1} = X_t' m_t \tag{12}$$

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

$$\begin{aligned} Q_{t+1} &= X_t' R_t X_t + S_t, & (13) \\ R_t &= S_t C_t^* + W_t, & (14) \end{aligned}$$

where R_t denotes the unconditional variance of the time t-prior of the coefficient vector θ_t. The first term in (13) 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 (15) to (20).

$$e_{t+1} = r_{t+1} - \hat{r}_{t+1} \quad \text{(error in prediction).} \tag{15}$$

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.

$$
\begin{aligned}
n_{t+1} &= n_t + 1 \quad \text{(degrees of freedom).} && (16) \\
S_{t+1} &= S_t + \frac{S_t}{n_t}\left(\frac{e_{t+1}^2}{Q_{t+1}} - 1\right) \quad \text{(estimator of observational variance).} && (17)
\end{aligned}
$$

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

$$
A_{t+1} = \frac{R_t X_t}{Q_{t+1}} \quad \text{(adaptive vector)} \tag{18}
$$

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 as follows:

$$
\begin{aligned}
m_{t+1} &= m_t + A_{t+1} e_{t+1} \quad \text{(estimator for expected coefficient vector),} && (19) \\
C_{t+1}^* &= \frac{1}{S_t}(R_t - A_{t+1} A_{t+1}' Q_{t+1}) \quad \text{(estimator for variance of coeff. vector).} && (20)
\end{aligned}
$$

The discount factor approach that we use to give structure to W_t assumes 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

$$W_t = \frac{1-\delta}{\delta} S_t C_t^*, \quad \delta \in \{\delta_1, \delta_2, \ldots, \delta_d\}, \quad 0 < \delta_i \leq 1, \tag{21}$$

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

$$R_t = S_t C_t^* + \frac{1-\delta}{\delta} S_t C_t^* = \frac{1}{\delta} S_t C_t^*, \tag{22}$$

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. The nested family of models with constant regression coefficients corresponds to a specification of $\delta = 1$. Reducing δ below the value of 1 introduces time variation to the set of regression coefficients. The choice of δ is, in addition to the selection of the set of predictive variables, a further dimension of model uncertainty that is treated in the Bayesian model averaging framework presented in Section 2.2.

A.2 Bayesian Model Selection

Let M_i denote a certain choice of predictive variables from the k candidates, and δ_j a certain selection from the set $\{\delta_1, \delta_2, \ldots, \delta_d\}$. 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

$$\hat{r}^{j}_{t+1,i} = E(r_{t+1}|M_i, \delta_j, D_t) = X_t' m_t|M_i, \delta_j, D_t. \tag{23}$$

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

$$P(M_i|\delta_j, D_t) = \frac{f(r_t|M_i, \delta_j, D_{t-1})P(M_i|\delta_j, D_{t-1})}{f(r_t|\delta_j, D_{t-1})}, \tag{24}$$

where

$$f(r_t|\delta_j, D_{t-1}) = \sum_{M} f(r_t|M_i, \delta_j, D_{t-1})P(M_i, \delta_j, D_{t-1}). \tag{25}$$

The crucial part is the conditional density

$$f(r_t|M_i, \delta_j, D_{t-1}) \sim \frac{1}{\sqrt{Q^{j}_{t,i}}}\ \mathbf{t}_{n_{t-1}}\left(\frac{r_t - \hat{r}^{j}_{t,i}}{\sqrt{Q^{j}_{t,i}}}\right), \tag{26}$$

where $\mathbf{t}_{n_{t-1}}$ is the density of a Student-t-distribution and $\hat{r}^{j}_{t,i}$ and $Q^{j}_{t,i}$ are the respective point estimates and variance of the predictive distribution of model M_i and given $\delta = \delta_j$; see Equation (11). The time $t+1$ return prediction of the average model for a given $\delta = \delta_j$

then equals

$$\hat{r}^{j}_{t+1} = \sum_{i=1}^{2^k-1} P(M_i|\delta_j, D_t)\hat{r}^{j}_{t+1,i}. \tag{27}$$

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

$$P(\delta_j|D_t,) = \frac{f(r_t|\delta_j, D_{t-1})P(\delta_j|D_{t-1})}{\sum_\delta f(r_t|\delta, D_{t-1})P(\delta|D_{t-1})}. \tag{28}$$

Note that this posterior probability is going to be of key importance in our empirical analysis, as it indicates 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

$$P(M_i, \delta_j|D_t) = P(M_i|\delta_j, D_t)P(\delta_j|D_t) \tag{29}$$

and the unconditional average prediction of the average model is

$$\hat{r}_{t+1} = \sum_{j=1}^{d} P(\delta_j|D_t)\hat{r}^{j}_{t+1}. \tag{30}$$

References

Ang, A., and G. Bekaert, 2007, Stock return predictability: Is it there?, *The Review of Financial Studies* 20, 651–707.

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, *The 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.

Boudoukh, J., R. Michaely, M. Richardson, and M. R. Roberts, 2007, On the importance of measuring payout yield: Implications for empirical asset pricing, *The Journal of Finance* 62, 877–915.

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.

Brown, J. P., H. Song, and A. McGillivray, 1997, Forecasting uk house prices: A time-varying coefficient approach, *Economic Modelling* 14, 529–548.

Campbell, J., and S. B. Thompson, 2008, Predicting the equity return premium out of sample: Can anything beat the historical average?, *The Review of Financial Studies* 21, 1509–1531.

Campbell, J. Y., 2007, Estimating the equity premium, *Canadian Journal of Economics* 41, 1–21.

——— , and J. H. Cochrane, 1999, By force of habit: A consumption-based explanation of aggregate stock market behavior, *The Journal of Political Economy* 107, 205–251.

Chakley, M., and I.-H. Lee, 1998, Learning and asymmetric business cycles, *Review of Economic Dynamics* 1, 623–645.

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.

Cochrane, J. H., 2007, *Financial Markets and the Real Economy* (in R. Mehra (ed.) Handbook of the Equity Premium, Amsterdam: Elsevier).

Cogley, Timothy, and Thomas J. Sargent, 2003, Drifts and volatilities: Monetary policies and outcomes in the post wwii u.s., working paper.

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, 2008, A comprehensive look at the empirical performance of equity premium prediction, *The Review of Financial Studies* 21, 1455–1508.

Grullon, G., and R. Michaely, 2002, Dividends, share repurchases, and the substitution hypothesis, *The Journal of Finance* 57, 1649–1684.

Henkel, S. J., J. S. Martin, and F. Nardari, 2008, Time-varying short-horizon predictability, working paper.

Johannes, M., A. Korteweg, and N. Polson, 2009, Sequential learning, predictive regressions, and optimal portfolio returns, working paper.

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

Kacperczyk, M., S. Van Nieuwerburgh, and L. Veldkamp, 2009, Attention allocation over the business cycle: Evidence from the mutual fund industry, working paper.

Kandel, S., and R. F. Stambaugh, 1996, On the predictability of stock returns: An asset-allocation perspective, *The 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.

Lettau, M., and S. Van Nieuwerburgh, 2008, Reconciling the return predictability evidence, *The Review of Financial Studies* 21, 1607–1652.

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, *The 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.

——— , 2009a, Are stocks really less volatile in the long-run?, working paper.

———, 2009b, Predictive systems: Living with imperfect predictors, The Journal of Finance, Forthcoming.

Paye, B. S., and A. Timmermann, 2006, Instability of return prediction models, *Journal of Empirical Finance* 13, 274–315.

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.

Polk, C., S. Thompson, and T. Vuolteenaho, 2006, Cross-sectional forecasts of the equity premium, *Journal of Financial Economics* 81, 101–141.

Primiceri, G. E., 2005, Time varying structural vector autoregressions and monetary policy, *Review of Economic Studies* 72, 821–852.

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.

Rapach, D. E., J. K. Strauss, and G. Zhou, 2009, Out-of-sample equity premium prediction: Combination forecasts and links to the real economy, The Review of Financial Studies, forthcoming.

Spiegel, M., 2008, Forecasting the equity premium: Where we stand today, *The Review of Financial Studies* 21, 1453–1454.

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

Van Nieuwerburgh, S., and L. Veldkamp, 2006, Learning asymmetries in real business cycles, *Journal of Monetary Economics* 53, 753–772.

Veldkamp, L., 2005, Slow boom, sudden crash, *Journal of Economic Theory* 124, 230–257.

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

Wachter, J. A., and M. Warusawitharana, 2007, What is the chance that the equity premium varies over time? evidence from predictive regressions, 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.