Predictive Regressions with Time-Varying Coefficients *

Thomas Dangl
Vienna University of Technology

Michael Halling
University of Utah

October 28, 2011

Abstract

We evaluate predictive regressions that explicitly consider the time-variation of coefficients in a comprehensive Bayesian framework. 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 find that predictive models with constant coefficients are dominated by models with time-varying coefficients. Finally, we show 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. 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 find that predictive models with constant coefficients are dominated by models with time-varying coefficients. Finally, we show 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, 2008, 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; Pastor and Stambaugh, 2009). 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; Rapach, Strauss, and Zhou, 2010). 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. 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. For example, Barsky (1989) documents time-varying stock-bond correlations, Dimson et al. (2002) present empirical evidence on time-varying correlations between various economic variables, McQueen and Roley (1993) and Boyd et al. (2005) find that the incorporation of news into stock prices varies with the business cycle, and Veldkamp (2005) and Van Nieuwerburgh and Veldkamp (2006), among others, relate learning asymmetries and the flow of information to the business cycle. Bossaerts and Hillion (1999) state 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 (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 show 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.

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 (see, for example, Avramov, 2002; Cremers, 2002; Pastor, 2000; Wachter and Warusawitharana, 2011; and Johannes, Korteweg, and Polson,

2011) that enables us to model time-varying coefficients that follow a random walk. The random walk assumption is, obviously, unattractive from a theoretical point of view. Empirically, however, it is common and frequently outperforms more general models with autocorrelated coefficient dynamics (see, for example, Meese and Rogoff, 1983a,b).[1] The two dimensions of model uncertainty— the choice 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; 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 relation between equity returns and explanatory variables. Pastor and Stambaugh (2001) and Kim et al. (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 relations, but does not find evidence of structural breaks in the relation 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

[1] In Appendix A.3 we verify for our specific application that models with random walk coefficients perform significantly better than models with autocorrelated coefficients.

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 for a comprehensive set of predictive variables. 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). As shown in our empirical analysis, models with gradually varying coefficients are strongly supported by the data.[2]

Methodologically, the paper closest to our study is Johannes et al. (2011) who in some specifications allow for drifting regression coefficients. There are, however, important conceptual differences. For example, our focus is on evaluating whether time-varying coefficients improve the predictive power of a standard set of 13 predictive variables; i.e., in a multivariate setup. Johannes et al. (2011), in contrast, focus on the issue of stochastic volatility and only consider models with a single predictive variable (i.e., pay-

[2]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 (2011) address the issue of model instability (i.e., structural breaks in predictive relations) and show that it can have a larger impact on optimal asset allocation than other sources of risk such as uncertainty in parameter estimation.

out measures). Despite these methodological differences, we arrive at the same main result, namely that out-of-sample predictability exists.

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 that constitutes the no-predictability benchmark.[3] 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 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 et al. (2010) find utility gains in the order of 0.5% to 1.5%, and Campbell and Thompson (2008) report maximum utility gains of 0.3%.

Most interestingly, we find a strong relation between out-of-sample predictability and

[3]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. In Appendix A.3 we report results from rolling window regressions. These results show that our proposed methodology also clearly dominates rolling window OLS regressions.

the business cycle. Although we find evidence of predictability during recessions as well as during expansions (in contrast to Henkel, Martin, and Nardari, 2011, 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 are consistent with business cycle related patterns implied by asset pricing theory (e.g., Campbell and Cochrane, 1999; Menzly, Santos, and Veronesi, 2004). On average, predicted equity risk premia 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 relation), (ii) the estimation uncertainty in coefficients, (iii) the model uncertainty with respect to the choice of predictive variables (see Avramov, 2002; 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 show 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 (this rule enabled firms to legally buy back shares under certain circumstances, see Grullon and Michaely, 2002, for details) — in contrast to constant coefficients or even regime-switching models (see Goyal and Welch, 2008; 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 show 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.

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 relation 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 support 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; 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 relation 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. We are performing an out-of-sample analysis, in which 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.

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

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.[4] In what follows we call V the observational variance. The concept of the predictive regression expressed in Equation (1), i.e., that time t observable variables predict time $t+1$ returns is not necessarily in contradiction to the efficient markets hypothesis. Time variation of expected returns can arise in efficient markets, e.g., as a consequence of time variation in risk aversion, see Campbell and Cochrane (1999), or of the presence of long-run consumption risk as in Bansal and Yaron (2004). 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

[4]See Johannes et al. (2011) for a Bayesian framework with stochastic volatility.

in θ, i.e., we consider changes in θ as unpredictable.[5]

We are aware that Equation (2) implies that coefficients follow a random walk and that theoretically they might drift to arbitrarily high or low values, hence causing returns – without regularly updating the system equation to new observations – to be non-stationary. To avoid this undesirable property one must impose some structure on the system equation. Our investigations, however, show that any deviation from the assumption of no predictability in the shocks on coefficients reduces the predictive power of our regression system. In Appendix A.3, we extend Equation (2) to allow for autocorrelation in the coefficients and then perform a horse race within our predictability framework that shows that random walk coefficients significantly outperform autocorrelated coefficients with respect to all analyzed statistics. We argue that the superior predictive performance of the parsimonious model stems from avoiding estimation errors while allowing to calibrate coefficients to observed data.

We share this finding with a series of empirical studies on predictability. For example, the literature on exchange rate prediction relies extensively on random walk models although the assumption of non-stationary exchange rates is theoretically unfounded. Meese and Rogoff (1983a) and Meese and Rogoff (1983b) find that random walk models outperform more sophisticated models such as structural models in predicting exchange rates. They started a large literature that, until today, has unsuccessfully tried to find predictive models that outperform random walk models (see, for example, Kilian and Taylor, 2003, for a more recent paper). In the literature on equity return predictability, random

[5] See Primiceri (2005) and Cogley and Sargent (2003) for a similar model specification with an application to monetary policy and Brown et al. (1997) for an application to house prices.

walk assumptions are also common. When modeling predictors, Ferreira and Santa-Clara (2011), for example, assume that the dividend-price ratio follows a random walk in their predictive setup (see also Campbell, 2008). Similar to our argument above, they find that the main advantage of this assumption is a substantial reduction in estimation error as fewer parameters have to be estimated.

Given that we perform the estimation on a monthly basis, this high frequency of observations drives the dynamics of the estimated coefficients and, thus, mitigates any concerns about the random walk assumption. Such concerns, in contrast, might become more important when updating at a lower frequency. But as the variance in coefficients gradually increases over time the estimation procedure becomes more responsive to incoming observations (see Equations 18 and 19 in Appendix A.1), avoiding unbounded drifts of estimated coefficients even in this case.

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 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):

$$
\begin{aligned}
V|D_0 &\sim IG\left[\frac{1}{2}, \frac{1}{2}S_0\right], & (3)\\
\theta_0|D_0, V &\sim N\left[0, gS_0(X'X)^{-1}\right], & (4)
\end{aligned}
$$

where

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

This is a diffuse prior centered around 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.[6]

[6]For the paper's main results the g-prior is derived from the entire sample. As is shown in Equation (4) the g-prior multiplies the variance/covariance structure of θ by a large scalar. This makes the prior

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 Appendix A.1.

To keep the model tractable, we need to give some structure to the system variance matrix W_t. Equations (1) and (2) show that an increase in the system variance negatively affects the precision of the coefficients' estimates $\theta_t|D_t$ (see Appendix A.1 for the analysis, especially Equation (10)). In periods of low system variance, the estimation error of the coefficient vector θ tends toward zero as the sample size increases. In times of high system variance, the estimate of θ loses precision. Hence, the estimation error in determining θ is positively related to the underlying system variance. A simple way to capture this relation is to assume that W_t is proportional to the variance/covariance matrix of $\theta|D_t$. More precisely, the scaling factor is assumed to be $\frac{1-\delta}{\delta}$ with $0 < \delta_i \leq 1$. West and Harrison (1997) call δ a discount factor and, consequently, this model setup a discount factor approach.

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

essentially uninformative which means that initially the estimation procedure is highly responsive to the observations and adapts quickly to the empirical patterns. This is a very common procedure in Bayesian econometrics (see, for example, Zellner, 1986; Koop, Poirier, and Tobias, 2007) and does not invalidate our claim that results are out-of-sample. In Appendix A.3 we show results in which we derive the OLS estimate from a specific burn-in period of 60 months in the beginning of our sample period. The results are basically unaffected by this choice of prior.

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) represents another a priori specification choice. We consider values on a grid $\delta \in \{\delta_1, \delta_2, \ldots, \delta_d\}$ where $0 < \delta_i \leq 1$ and d captures the number of discrete values of δ considered. Since the Bayesian updating approach integrates (sums) over the range of the discount factor δ, the specific choice of the set of δs is not critically influencing the result as long as the relevant range of time variability is adequately covered. Considering a number of d different discrete values of δ leads to a total of $d \cdot (2^k - 1)$ possible dynamic linear models.[7]

The arbitrary choice of one particular model from this substantial pool of possible models is always debatable. Bayesian model selection (see Avramov, 2002; Cremers,

[7]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.

2002) offers a systematic approach to this problem that tests the reliability of all $d \cdot (2^k - 1)$ models against the observed data (see Appendix A.2 for details). Starting from an uninformed prior that gives equal weight to each individual model, 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.

3 Empirical Study Design

3.1 Data Description

We calibrate and test the proposed methodology using monthly total excess returns of the S&P 500 Index from May 1937 to December 2002.[8] 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 data set in our study.[9] 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.

[8] The sample period is driven by concerns of data availability; specifically, the cross-sectional beta premium variable is only available for this specific time period. In Appendix A.3 we extend our predictive framework such that variables can enter or drop from the data set at any time and reevaluate the predictive performance of our models for the longer time period of January 1927 to December 2008 (see the Appendix for some empirical results). None of our main results is affected by this extension of the sample period.

[9] We particularly thank Amit Goyal for providing their data set.

- Cross-sectional premium: Cross-Sectional Beta Premium **(csp)** quantifies the relative valuation of high- and low-beta stocks according to Polk et al. (2006).
- Book value: Book to Market Ratio **(bmr)** is the ratio of book value at the end of the previous year (for the months January and February, the book value is additionally lagged by one year) 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.

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

	Mean	Standard deviation	Minimum	Maximum	Median
dy	-3.295	0.422	-4.525	-2.406	-3.291
ep	-2.650	0.402	-3.839	-1.775	-2.674
dpayr	-0.650	0.197	-1.183	0.063	-0.624
svar	0.002	0.004	0.000	0.071	0.001
csp	0.000	0.002	-0.004	0.008	-0.000
bmr	0.614	0.237	0.121	1.207	0.617
ntis	0.018	0.015	-0.031	0.054	0.020
tbl	0.042	0.032	0.000	0.163	0.039
lty	0.057	0.030	0.018	0.148	0.056
ltr	0.005	0.024	-0.084	0.152	0.003
dfy	0.010	0.005	0.003	0.032	0.008
dfr	0.000	0.011	-0.051	0.070	0.000
inf	0.003	0.005	-0.014	0.057	0.003

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

3.2 Parameter and Prior 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 procedure for a g-prior with $g = 50$.[10] The second choice is the range of δ to be covered in the BMA when integrating over the degree of variability in θ, thereby determining an empirical estimate of δ. We choose $\delta \in [0.96, 1.00]$ and evaluate different levels of granularity.[11] The lower boundary of the relevant range for δ is derived from the following considerations. 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 $1/\delta$ per month. If we ignore other determinants of this variance, the total effect of δ equal to .98 will be a 50 percent increase within 20 months. For δ equal to .96, a 50 percent increase in variance will be reached twice as fast, in approximately ten months. The latter case describes a world, in which coefficients are extremely unstable and, thus, we select it as the lower boundary.

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., $1/(d \cdot (2^k - 1))$ with d as the number of supporting points in the range of δ) and each individual δ-value (i.e., $1/d$) at the beginning of the estimation horizon. Therefore, every model and every model family has the same chance to turn out to be important.[12]

[10] 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.

[11] Specifically, we conduct our estimation with two choices of granularity in δ, $\delta \in \{.96, .98, 1.00\}$ and $\delta \in \{.96, .97, .98, .99, 1.00\}$. The increase in supporting points for the BMA integration over δ does not change our results notably.

[12] 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 $.5 \cdot 1/(d \cdot (2^k - 1) - 1)$. 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 relation between predictability and the business cycle 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. 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. We find broad support — over different subsamples, using statistical and economic measures — for the conclusions that predictive regressions with time-varying coefficients predict market returns 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, in which 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 is determined in the following way. At the end of the month before the evaluation period starts we identify all predictive variables that receive more than 50% posterior probability in the BMA-Model incl. TVar-Coeff. (the 50%-threshold is basically an ad-hoc way to determine variables that show decent predictive performance). Then we focus on the model that includes these predictive variables. The *MEDIAN-Model incl. (excl.) TVar-Coeff.* is the model that includes these predictive variables and has *time-varying (constant)* coefficients.

The motivation to look at univariate models, the MOST-Model, and the MEDIAN-Model is to differentiate the effect of repeated updating of dynamic coefficients within a model from the effect of the BMA that shifts weights between models depending on historical performance. In contrast to the BMA-Models, MOST and MEDIAN models fix a certain 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.

In this section we distinguish four different sample periods: 1947+, 1965+, 1976+, and 1988+. These sample periods only affect the selection of predictions that are analyzed in a specific statistical test or economic evaluation. They do not affect the estimation of the models; i.e., predictive models are still updated monthly. This choice of sample periods is mainly driven by issues of comparability to other studies (especially, Goyal and Welch, 2008; Rapach, Strauss, and Zhou, 2010). 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.

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).

We start with the analysis of the BMA-Model (*BMA-Model incl. (excl.) TVar-Coeff.*). 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).[13] 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 sig-

[13]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 by a factor of 10^{-6}. This result compares well to Cremers (2002). Appendix A.3 presents and discusses the time-series dynamics of the Bayes Factor over our sample period. Even more detailed results are available from the authors upon request.

nificantly (at a 10% significance level) outperform the ones with only constant coefficients in 28 out of 52 (i.e., 13 univariate models times four time periods) cases across all sub-periods. 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 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.[14] 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

[14] In the case of the MOST-Model, the model with highest posterior probability in December 1946 is a model with constant coefficients; thus, there is no p-value for the comparison.

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. We distinguish four different evaluation periods: 1947+, 1965+, 1976+, and 1988+.

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	0.0192	0.00	0.0046	0.08	0.00
ep	0.0055	0.02	0.0026	0.04	0.12
svar	0.0004	0.41	-0.0005	0.95	0.30
bmr	0.0031	0.21	0.0031	0.14	0.05
tbl	0.0055	0.02	0.0033	0.08	0.01
ltr	0.0050	0.01	0.0018	0.14	0.01
dfy	0.0035	0.02	-0.0001	0.64	0.02
inf	0.0027	0.09	0.0009	0.16	0.18
dy	0.0044	0.01	0.0036	0.02	0.08
dpayr	0.0027	0.17	-0.0001	0.54	0.12
csp	0.0094	0.01	0.0052	0.02	0.01
ntis	0.0043	0.01	0.0020	0.10	0.04
lty	0.0035	0.08	0.0025	0.13	0.06
dfr	0.0008	0.16	-0.0005	0.81	0.02
MOST-Model	0.0054	0.10	0.0054	0.10	.
MEDIAN-Model	0.0220	0.00	0.0028	0.22	0.00

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.

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	0.0261	0.00	0.0039	0.11	0.00
ep	0.0067	0.04	0.0023	0.11	0.07
svar	0.0007	0.39	-0.0007	0.94	0.29
bmr	0.0018	0.37	0.0013	0.35	0.09
tbl	0.0063	0.05	0.0047	0.08	0.11
ltr	0.0056	0.02	0.0036	0.06	0.14
dfy	0.0048	0.02	-0.0003	0.79	0.02
inf	0.0029	0.15	0.0008	0.24	0.19
dy	0.0046	0.03	0.0038	0.05	0.18
dpayr	0.0041	0.16	0.0004	0.38	0.17
csp	0.0106	0.03	0.0048	0.03	0.08
ntis	0.0049	0.03	0.0031	0.07	0.18
lty	0.0044	0.12	0.0035	0.15	0.13
dfr	0.0008	0.23	-0.0003	0.65	0.08
MOST-Model	0.0141	0.06	0.0015	0.35	0.00
MEDIAN-Model	0.0018	0.02	0.0003	0.47	0.00

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	0.0304	0.00	0.0038	0.18	0.00
ep	0.0018	0.21	0.0017	0.22	0.08
svar	-0.0007	0.87	-0.0008	0.91	0.28
bmr	-0.0006	0.61	-0.0005	0.56	0.02
tbl	0.0015	0.31	0.0013	0.35	0.04
ltr	0.0030	0.16	0.0030	0.16	0.49
dfy	0.0027	0.10	-0.0000	0.54	0.10
inf	-0.0003	0.61	-0.0002	0.57	0.66
dy	0.0018	0.26	0.0016	0.30	0.02
dpayr	-0.0001	0.52	-0.0002	0.58	0.31
csp	0.0097	0.04	0.0031	0.16	0.05
ntis	0.0039	0.07	0.0037	0.08	0.22
lty	0.0005	0.43	0.0005	0.43	0.11
dfr	0.0002	0.41	-0.0003	0.71	0.08
MOST-Model	0.0252	0.00	-0.0012	0.60	0.00
MEDIAN-Model	0.0307	0.00	-0.0004	0.54	0.00

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	0.0210	0.08	-0.0012	0.60	0.01
ep	0.0036	0.12	0.0036	0.13	0.05
svar	-0.0006	0.76	-0.0007	0.76	0.30
bmr	0.0009	0.39	-0.0006	0.57	0.06
tbl	-0.0012	0.85	-0.0016	0.89	0.09
ltr	0.0008	0.37	0.0008	0.37	0.34
dfy	0.0006	0.27	-0.0003	0.80	0.19
inf	-0.0005	0.70	-0.0006	0.71	0.42
dy	0.0013	0.38	0.0007	0.44	0.01
dpayr	-0.0018	0.92	-0.0019	0.92	0.06
csp	0.0148	0.05	0.0029	0.18	0.05
ntis	0.0003	0.46	0.0002	0.47	0.21
lty	-0.0007	0.81	-0.0008	0.82	0.08
dfr	-0.0014	0.94	-0.0010	0.91	0.97
MOST-Model	0.0280	0.06	0.0004	0.48	0.00
MEDIAN-Model	0.0134	0.16	0.0062	0.18	0.01

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; 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 et al. (2010) and consider an investor with a single-period horizon and mean-variance preferences.

We analyze the gain in realized utility of an investor who uses any of the predictive models in comparison to the no-predictability benchmark. We determine monthly realized mean-variance utility where we use daily S&P 500 returns within a month to estimate the monthly variance. The utility function is $E(Rp) - (\gamma/2)Var(Rp)$, where Rp is the portfolio return and $\gamma = 3$. Average realized utility gains and significance levels are inferred from these time series.

Table 3 summarizes these results for the same set of predictive models and sample periods. 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 although not all of these improvements are statistically significant. The only exceptions are the univariate model based on **dfr** in sub-period 1988+ and the one based on **inf** in sub-period 1976+. However, only **csp** generates positive utility gains consistently across all sub-periods (only the one during the 1965+ period is 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. (monthly utility changes are annualized) 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. We distinguish four different evaluation periods: 1947+, 1965+, 1976+, and 1988+. ***, **, 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	0.16	-1.87	-5.79
ep	-0.61	1.07	-1.36	-1.77	-1.60	-0.22	-1.40	-1.79
svar	**-0.77**	**0.35**	-1.67	-2.95	-1.69	-0.46	-1.80	-2.97
bmr	-1.74	-0.71	-3.67	-3.27	-3.13	-2.07	-3.83	-6.41
tbl	-0.55	**0.63**	**-2.72**	**-4.04**	-1.47	0.14	-3.30	-4.66
ltr	**0.38**	1.60	-0.73	-2.80	-0.96	1.14	-0.82	-2.89
dfy	**0.01**	**1.36**	-1.01	-2.27	-1.46	-0.46	-1.67	-3.12
inf	-0.23	0.74	-2.05	-3.13	-0.75	0.51	-1.98	-3.14
dy	-0.77	0.74	-2.27	-4.84	-1.47	0.38	-2.43	-5.08
dpayr	**-0.82**	**0.36**	-2.47	-4.56	-1.81	-0.26	-2.58	-4.72
csp	**0.88**	**2.90***	**1.22**	**2.94**	-1.11	1.20	-1.71	-1.98
ntis	**-0.13**	1.16	**-1.13**	-3.15	-1.66	-0.05	-1.34	-3.22
lty	-1.03	0.37	-2.79	-3.82	-1.33	0.31	-2.84	-3.88
dfr	**-0.82**	**0.51**	-1.38	-3.69	-1.87	-0.46	-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 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.

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 also 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. 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 should 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, 2011; Rapach, Strauss, and Zhou, 2010, 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 relations 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 et al. (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 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 are 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, 2011; Rapach, Strauss, and Zhou, 2010), we find significantly stronger evidence for predictability during recessions than during expansions (third row of Table 4) using both measures. The only exception, as it is not significantly different from zero, is the difference in MSPEs for models excluding time-varying coefficients. 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 et al. (2011), 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: (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-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

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 are across Models incl. TVar-Coeff. and Models excl. TVar-Coeff.) 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.	0.0600***	24.967***	0.0168**	9.378*	0.0478***	15.589***
Exp.	0.0121**	1.622	0.0052	-1.990	0.0152***	3.613**
Diff.	-0.0479***	-23.345***	-.0116	-11.369**		
Late Exp.	0.0327***	16.597*	0.0148	3.841	0.0184*	12.756
Early Rec.	0.1156***	47.849***	0.0707***	30.543**	0.0401**	17.306**
Late Rec.	0.0490**	6.421	-0.0092	-6.810	0.0752***	13.230
Early Exp.	0.0075	-2.273	-0.0039	-8.321*	0.0266***	6.048

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 in 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): towards the end of the recession the predicted risk premium increases and peaks in Late Recession, potentially 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 (see equation (10) in Campbell and Thompson, 2008) 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 consistent with the implications of asset pricing models that use time-varying risk aversion to generate time-varying risk premia (e.g., see Campbell 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, 2011). 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.

4.2.3 Return Predictions and Macrovariables

To shed more light on this relation between return predictability and the business cycle, we compare the time-series of our predictions to three macrovariables: the consumption-wealth ratio (*cay*) of Lettau and Ludvigson (2001), the wealth-consumption ratio of Lustig et al. (2010), and quarterly GDP growth rates.[15]

We observe interesting time-series patterns in Figure 2. In the case of *cay* and predictions from the BMA-Model incl. TVar-Coefficients, there are periods, in which the two series move in the same direction (e.g., in the late 1960s, early 1970s, and late 1980s), and periods, in which they move in opposite directions (e.g., in the late 1950s and early 1990s). In the case of predictions from the BMA-Model excl. TVar-Coefficients, the graph reveals several periods with pronounced inverse dynamics (especially during the 1970s and in the early 1990s).

If we calculate time-series correlations between these data series, we find a slightly negative but insignificant correlation between predictions from the BMA-Model incl.

[15] All data are quarterly. Data of *cay* are from Martin Lettau's webpage. Data of the wealth-consumption ratio of Lustig et al. (2010) are from Stijn van Nieuwerburgh's webpage. We use the quarterly log wealth-consumption ratio, *wc*, in our analysis. GDP growth rates are obtained from Datastream.

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.

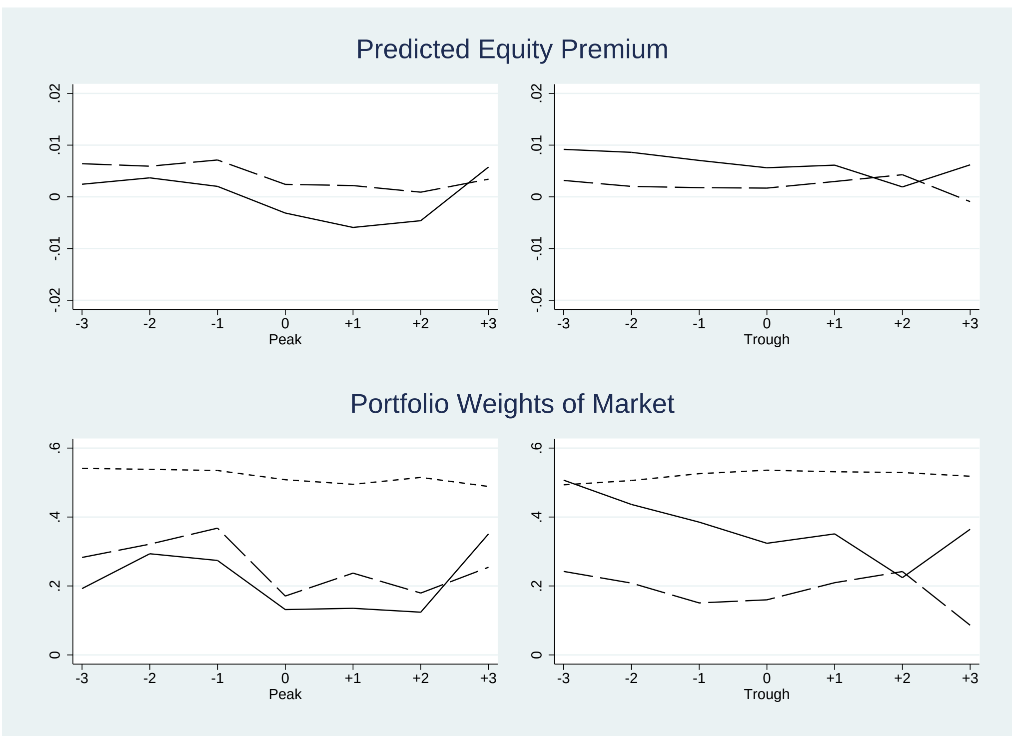
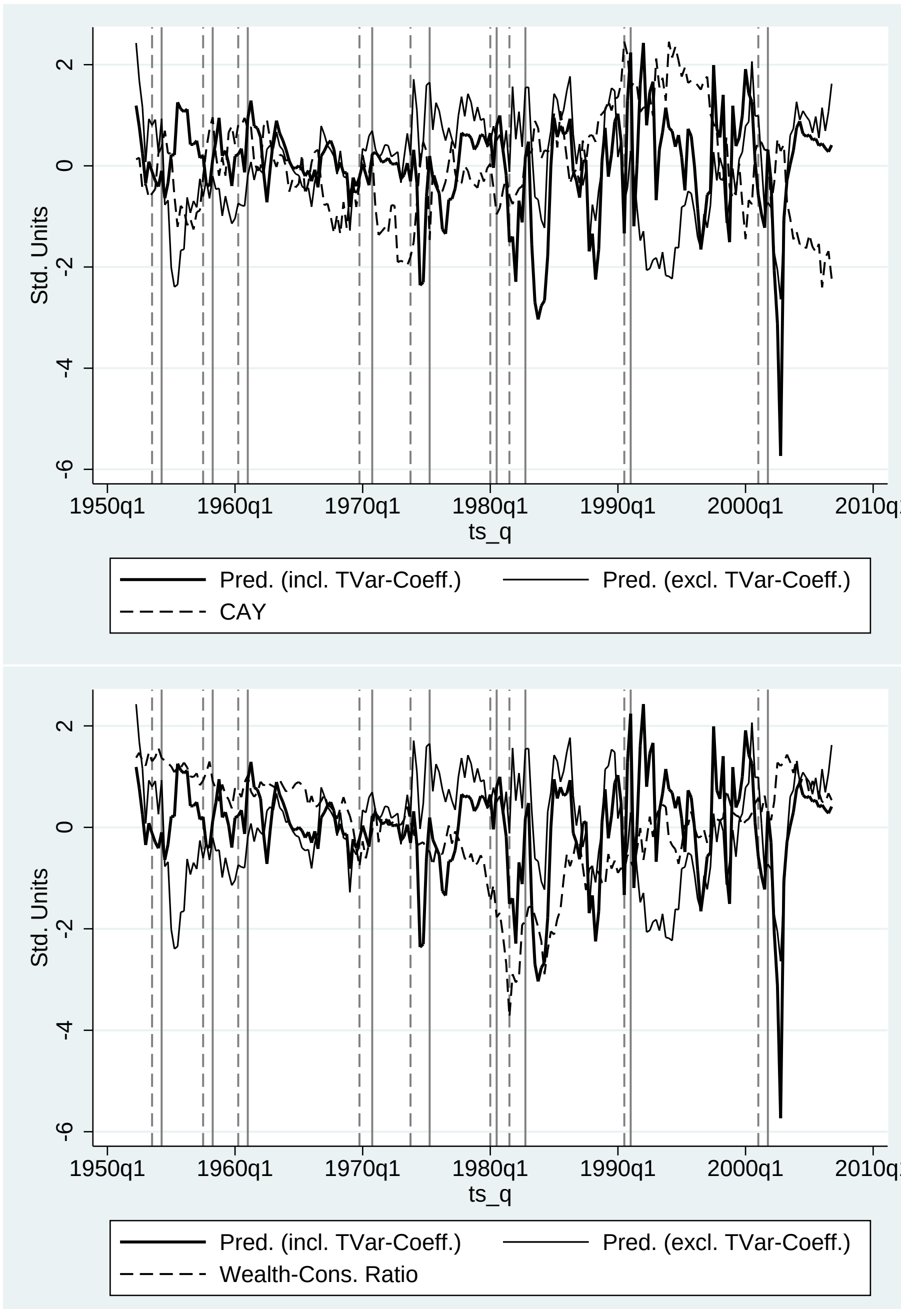
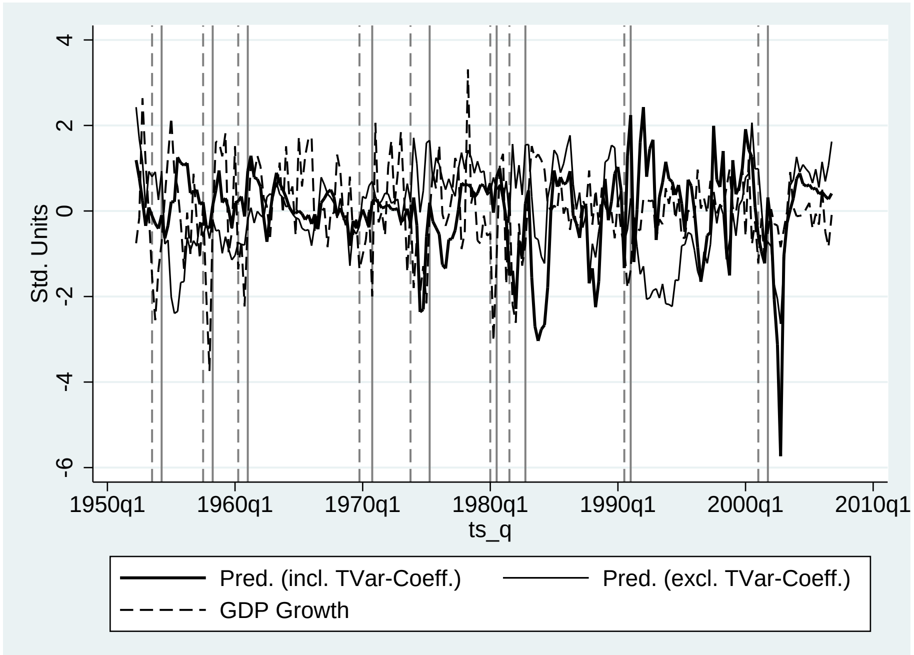


Figure 2: **Predictions vs. Macrovariables:** the graph consists of three pictures that show time-series graphs of predictions from the BMA-Model incl. TVar-Coeff., the BMA-Model excl. TVar-Coeff., and a selected macrovariable. The macrovariable in the first graph is *cay*, in the second graph is *wc*, and in the third graph it's GDP-growth. Vertical dashed (solid) lines indicate peaks (troughs). All data are quarterly and normalized to standard deviation of unity. Quarterly return predictions are calculated as compounded monthly predictions.





TVar-Coefficients and *cay*. In contrast, we find a significantly negative correlation of -0.42 between predictions from the BMA-Model excl. TVar-Coefficients and *cay*. Given that Lettau and Ludvigson (2001) argue that *cay* is positively related to expected future returns this substantial, negative correlation is surprising but consistent with the poor predictive performance of models with constant coefficients.

The time-series of *wc* shows quite different dynamics than *cay*, which is not surprising given the low correlation of 0.16 between these time-series as reported in Lustig et al. (2010). The wealth-consumption ratio *wc* slowly decreases over time, reaches its minimum in the early 1980s, and then steadily increases during the remainder of the sample period. Predictions from the BMA-Model incl. TVar-Coefficients show a somewhat related general pattern albeit much noisier; predictions from the BMA-Model excl. TVar-Coefficients, however, tend to follow the opposite pattern with high predicted returns during the late 1970s and early 1980s.

These observations are also confirmed if we look at the time-series correlations: we find a significantly positive correlation in the case of predictions from the BMA-Model incl. TVar-Coefficients (0.22) and a significantly negative correlation in the case of predictions from the BMA-Model excl. TVar-Coefficients (-0.19). Similar to the case of *cay*, this negative correlation of predictions of models with constant coefficients is counter-intuitive (Lustig, Van Nieuwerburgh, and Verdelhan, 2010, find a *positive* correlation of 12% between total wealth returns and the value-weighted CRSP stock return).

Finally, in the case of GDP-growth the visual inspection of the graph does not yield any obvious insights. Correlations are positive but insignificant for predictions from the BMA-Model incl. TVar-Coefficients (0.07) and significantly negative for predictions from the BMA-Model excl. TVar-Coefficients (-0.13). If we zoom into business cycle recessions, we find that correlations between GDP-growth and predictions from the BMA-Model incl. TVar-Coefficients increase substantially to 0.18. This is consistent with our previous result that, on average, predicted expected returns increase during recessions as does GDP-growth.

To conclude, the time-series comparison of return predictions to a selection of macrovariables shows reasonable patterns for predictions from the BMA-Model incl. TVar-Coefficients. Predictions from the BMA-Model excl. TVar-Coefficients, in contrast, show correlations with macrovariables that are at odds with existing results regarding the link between these macrovariables and future expected returns.

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

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

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.[16]

[16]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.

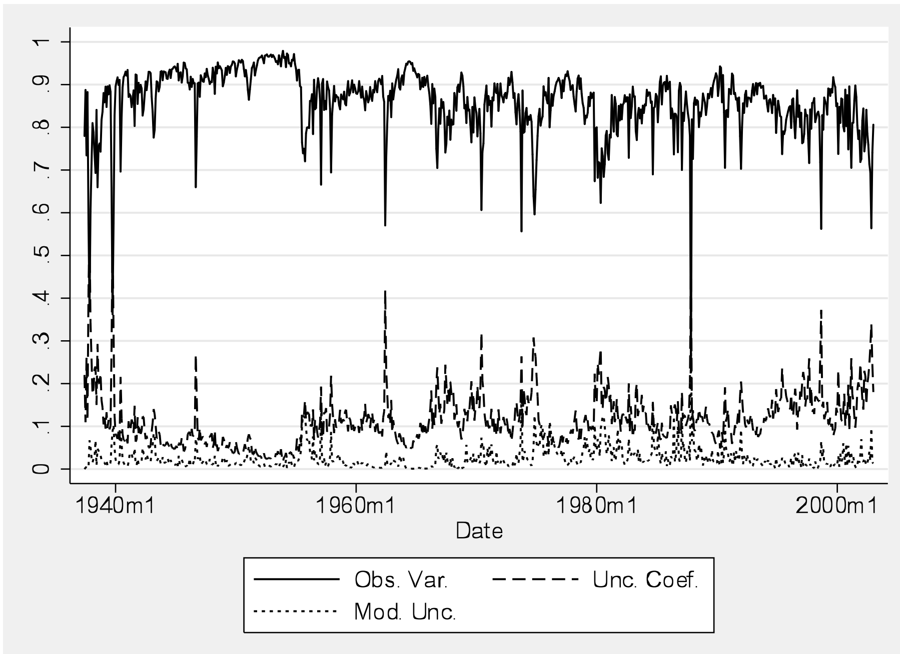
The individual terms of (6) can be interpreted in a very intuitive way. The first term is the expected observational variance ($S_t|M_j, \delta_i$ is the time-t estimate of the observational variance V in (1) conditional on model M_j and time variation coefficient δ_i). The second term states the expected 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 3, 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 prediction horizons, random fluctuations are expected to dominate the uncertainty in the predicted trend component.

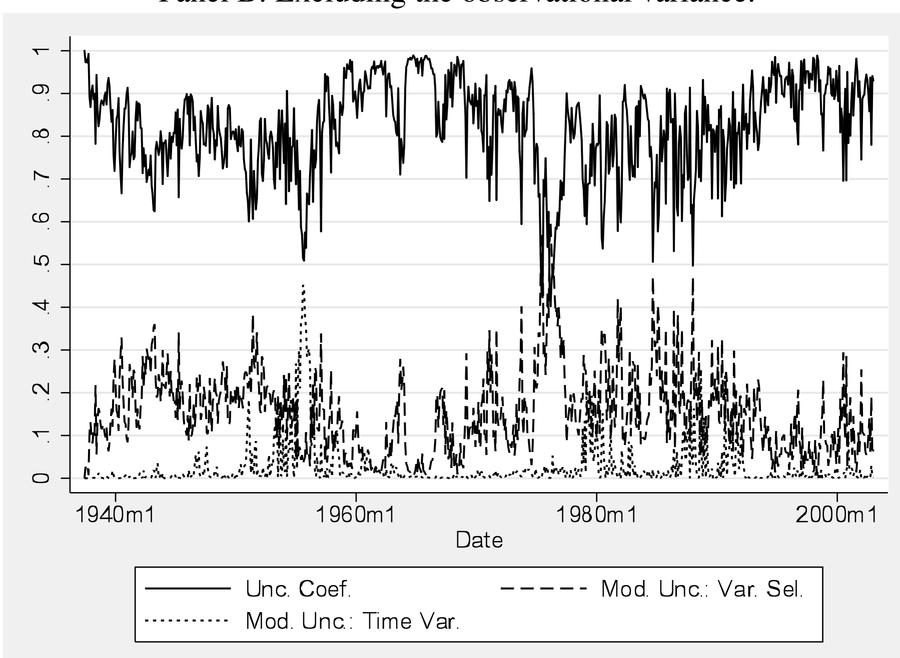
Therefore, Panel B masks out observational variance and focuses only on the other three components. In most periods, the estimation uncertainty in coefficients captures more than half of the remaining variance. This fits well with findings documented in Pastor and Stambaugh (1999). Interestingly, they find the same relation for cost of capital estimations on the firm level, while the results presented here are for cost of capital on the market level. 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

Figure 3: **Sources of Prediction Variance.**

Panel A: Including the observational variance.



Panel B: Excluding the observational variance.



the 90s).

Figure 3 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 4).

Figure 4 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, very dynamic models with $\delta = .96$ play no role during the 50s and 60s 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 4, it is not surprising that we find little uncertainty about the degree of time-variation in Figure 3.

Similarly, Figure 5 shows the posterior probability weighted average value of δ; i.e. the estimated degree of time-variation in coefficients across time. We see that the degree of time-variation itself changes over time: periods with relatively stable estimates of 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

Figure 4: **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.

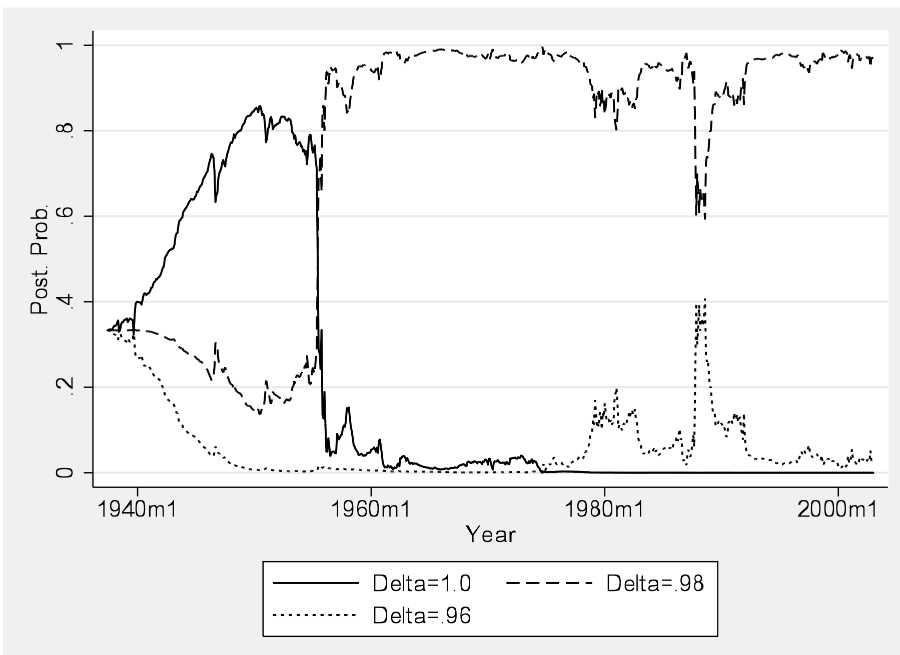
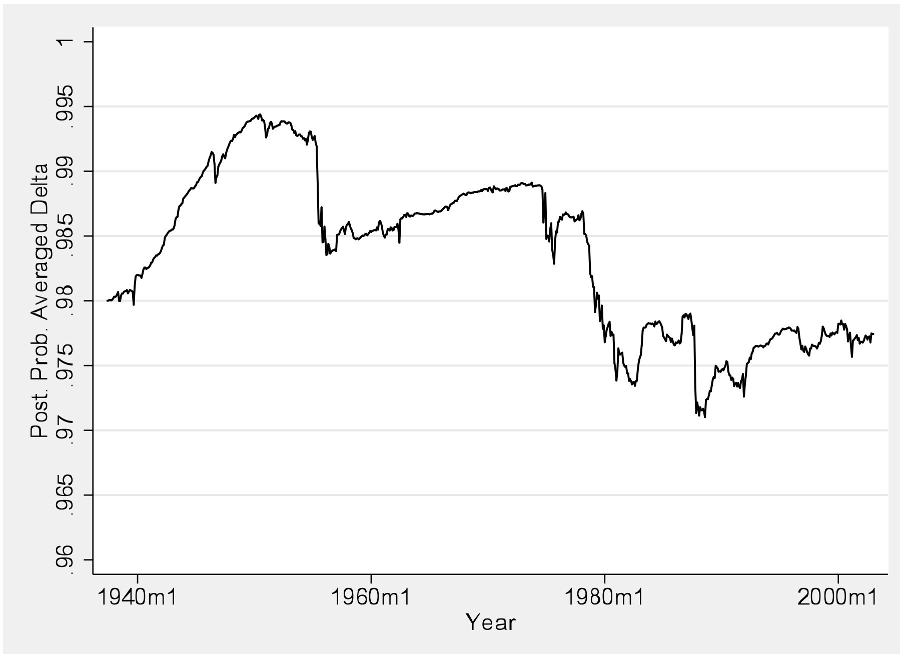


Figure 5: **Posterior Probability Weighted Average δ (i.e., degree of time-variation).** For the BMA-Model including time-varying coefficients, this figure reports the posterior probability weighted average degree of time-variation. To get a more precise estimate of this average δ we consider five specific delta values in the estimation, namely $\delta \in \{.96, .97, .98, .99, 1.00\}$.



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, 2011, for evidence that parameter instability is related to cyclical economic conditions).

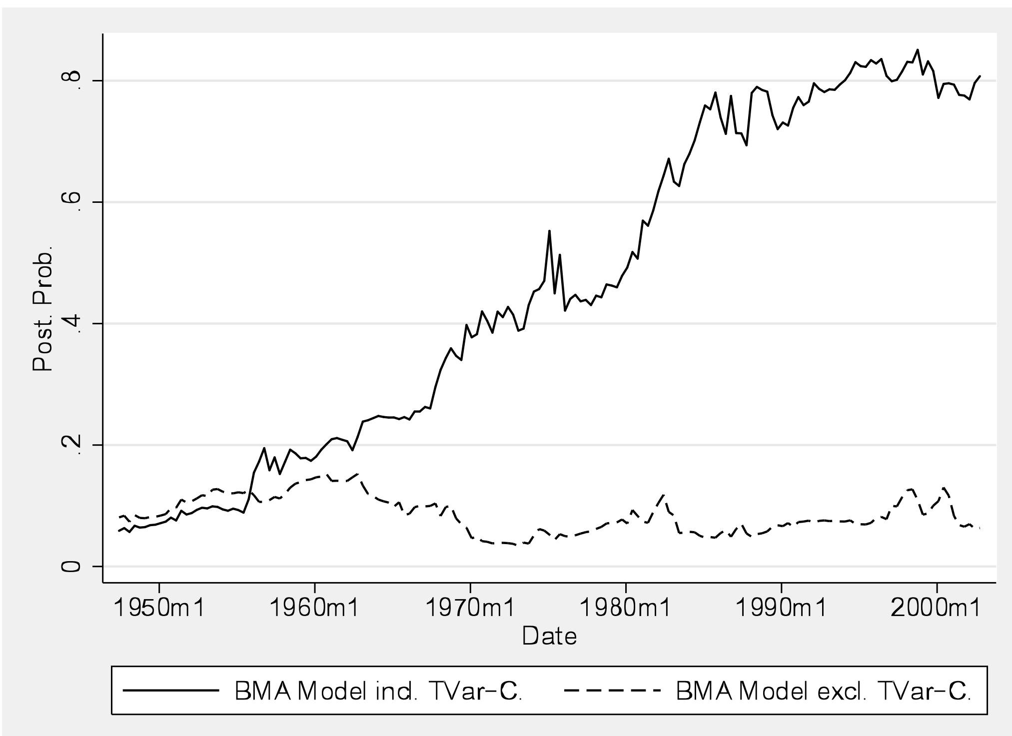
4.3.2 Analysis of Individual Coefficients and Models

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 6 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 assuming 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

Figure 6: **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.



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.

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	0.94	0.80	0.93	0.91	0.72	0.40	0.31	0.25
ep	0.25	0.16	0.46	0.77	0.29	0.30	0.27	0.25
dpayr	0.38	0.21	0.58	0.78	0.32	0.45	0.46	0.22
svar	0.17	0.07	0.01	0.01	0.22	0.22	0.44	0.25
csp	0.99	1.00	1.00	1.00	0.71	0.49	0.55	0.89
bmr	1.00	1.00	1.00	1.00	0.98	0.96	0.96	0.39
ntis	0.25	0.08	0.13	0.40	0.35	0.31	0.83	0.69
tbl	0.37	0.26	0.95	0.71	0.38	0.34	0.44	0.37
lty	0.41	0.70	0.08	0.37	0.39	0.39	0.27	0.41
ltr	0.27	0.21	0.19	0.12	0.30	0.50	0.59	0.63
dfy	0.12	0.06	0.01	0.18	0.39	0.50	0.32	0.21
dfr	0.06	0.02	0.13	0.03	0.21	0.16	0.43	0.28
inf	0.31	0.19	0.07	0.03	0.49	0.71	0.47	0.33

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-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 relations (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 that 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.

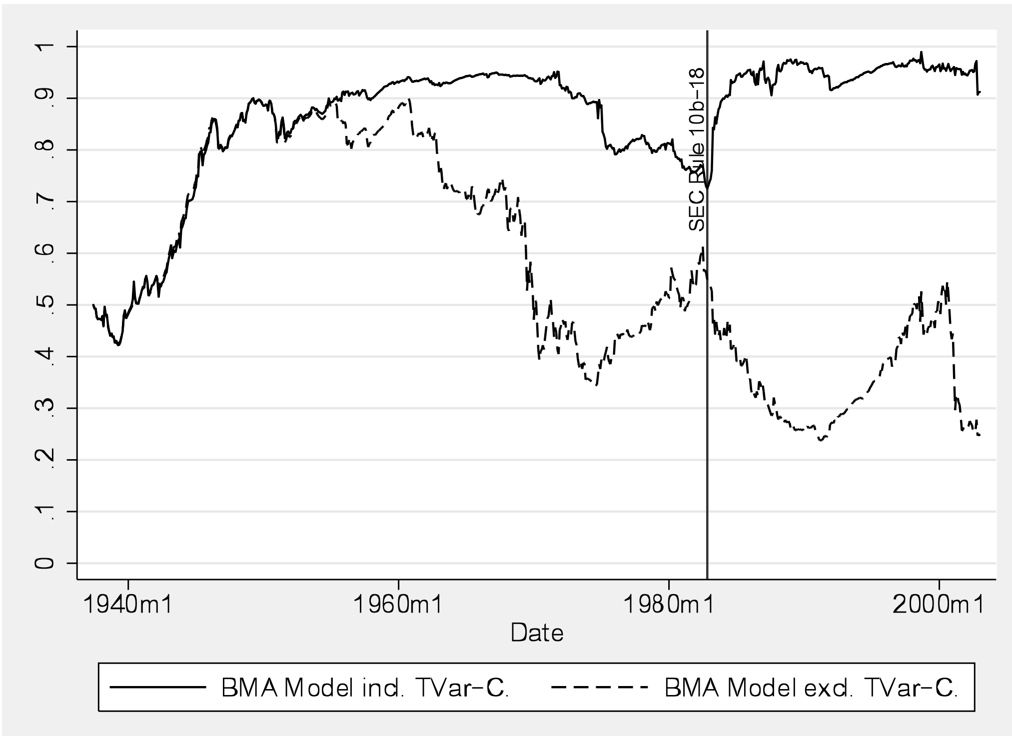
How, in contrast, does our methodology perform in detecting and learning this regulatory change in real-time? The BMA-Model including time-varying coefficients does very well in handling the regime shift. Figure 7 shows the dividend yield's importance over time, measured as the sum of the posterior probabilities assigned to individual predictive models including the dividend yield. 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 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 relations, 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 tech-

Figure 7: **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.



nique, 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 dividend 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) — broadly consistent with 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 show 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 in coefficients — namely, as a random walk — can quickly learn changes in the underlying relations, such as changes in the regulatory environment in the case of the dividend yield. While the simplifying assumption of random walk coefficients is theoretically unappealing, it seems to be empirically important, as we find that models with autocorrelated coefficients are outperformed by models using random walk coefficients. One possible explanation of this result is that the random walk assumption reduces estimation errors by not imposing a special autocorrelation structure on the coefficients' dynamics. Another possible explanation is that we implicitly favor the random walk assumption in our setup by forcing the same autocorrelation structure on all coefficients within a model.

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 relations. 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 relations 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, 2011). 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 after integrating out 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

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

where $\mathbf{t}(r_{t+1};\hat{r}_{t+1},Q_{t+1})$ is the density of a 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

$$Q_{t+1} = X_t'R_tX_t + S_t, \tag{13}$$

$$R_t = S_tC_t^* + W_t, \tag{14}$$

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 $S_tC^*_t$ of the coefficient vector $\theta_t|D_t$. More precisely, it is assumed that

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

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

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

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

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

A.3 Empirical Robustness Tests

In this subsection of the Appendix we evaluate the robustness of our main results on predictability along several dimensions. First, we change the way in which we determine the uninformative prior for the coefficient vector $\theta_0|D_0$. While our main results are based on full-sample OLS estimates of the variance in coefficients, the results discussed in this section use an explicit burn-in period of 60 months. Second, we want to exploit the available time-series dimension to a larger extent. The sample period used to derive our main results is based on the period of time, for which we observe all explanatory variables, namely May 1937 to December 2002. But except for one variable, *csp* (i.e., the cross-sectional premium), all data are available for a much longer time period, namely January 1927 to December 2008. To extend our sample period, we change the predictive technology such that it allows variables to enter and exit the sample at arbitrary points in time. Third, we also provide information about the Bayes Factor (i.e., the performance measure underlying our BMA approach) of the average model with constant coefficients relative to the full BMA model. Fourth, we report results from rolling-window OLS regressions, in which we consider a kitchen-sink specification that includes all predictive variables (following Goyal and Welch, 2008 we use a rolling window size of 60 months).

Fifth, the last and potentially the most important robustness test is to relax the random walk assumption for time-varying coefficients (see Equation 2, i.e., the system equation of our model, that specifies that coefficients follow a random walk, i.e., $\theta_t = \theta_{t-1} + \omega_t$). This assumption is not completely consistent with asset pricing theory, since without regularly linking the model to empirical data, expected asset returns are not stationary. One way to adapt the model to address the stationarity issue is to formulate the system equation as an autoregressive process. To keep the model tractable we introduce autoregression to the system equation in the following simple form

$$\theta_t = G\,\mathbf{I}\,\theta_{t-1} + \omega_t, \tag{31}$$

where $\mathbf{I}$ is the identity matrix and $0 < G \leq 1$ is a scalar.[17] If G equals 1, Equation (31) equals Equation (2), hence, our main model is a boundary case of the more general version presented here. To get an assessment of the importance of the stationarity issue we measure the predictive performance of models with autoregressive dynamics in θrelative to the random walk specification ($G = 1.0$) according to their MSPE-statistics (we use $G = 0.99$, 0.98, 0.97, and 0.96). The goal is to check if any individual parameter choice of G performs better than our current choice of random walk coefficients.

Table 6 (similar to Table 2) reports the prediction accuracy of BMA-Models incorporating these extensions/robustness checks. For simplicity, we ignore univariate models, as we did not find any evidence of predictability in our main results. The most important implication of these results is that in the horse race between random walk coefficients and mean-reverting coefficients the random walk hypothesis works by far best from an em-

[17]Given the restricted number of observations we have available, the full estimation of a general autoregressive model for θ in our Bayesian framework is empirically not feasible. This problem of limited empirical data availability is also discussed in Kilian and Taylor (2003) for the case of exchange rate predictions. They conclude that "Our analysis suggests that this difficulty of beating the random walk model in real time need not reflect an inherent shortcoming of forecasting models based on economic fundamentals. Instead, we showed that this stylized empirical fact appears to be a natural consequence of the small time span of data available for empirical work."

pirical point of view. P-values of out-of-sample predictability tests are always lowest for the BMA-Model with random walk coefficients; any evidence of predictability actually disappears across nearly all sub-samples if coefficients are modeled to be mean-reverting.

Further, comparing the results of Table 2 and 6 shows that the extensions to model an explicit burn-in phase in the beginning and to increase our sample period do not alter any of our main results: we still find that models with time-varying coefficients work well and that there is evidence of out-of-sample predictability. As far as simple rolling-window OLS predictions are concerned, we confirm the existing evidence of no-predictability.

Table 7 (similar to Table 3) reports utility gains of investors using any of these prediction technologies. Again, we find that only dynamic linear models with random walk coefficients generate significantly positive utility gains among the models with time-varying coefficients. Interestingly, the rolling-window regressions work well in terms of utility gains[18] — though, not as well as our suggested predictive model with time-varying, random walk coefficients. Further more, our results are basically unaffected by the burn-in phase and the extended sample period.

Another important result of this study is that equity premia predicted via models with time-varying coefficients seem to be economically reasonable. Figure 8 (similar to Figure 1) shows the dynamics of the prediction over the business cycle. In the case of $G = 1$, i.e., random walk coefficients (the two graphs in the top row), we find a similar pattern to Figure 1 (around the peak of the business cycle predicted returns decrease and increase again

[18]This tension between statistical evidence of predictability and economic gains in the case of rolling-window regressions is consistent with Goyal and Welch (2008). Goyal and Welch (2008) attribute it to characteristics of the evaluated sample period:“Put differently, some strategy CEV gains are due to the fact that the risky equity investment was a better choice than the risk-free rate in our data.”

Table 6: **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. 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.

	1947+		**1965+**		**1976+**		**1988+**		**Expansions**		**Recessions**	
	ΔMSPE	p-value	ΔMSPE	p-value	ΔMSPE	p-value	ΔMSPE	p-value	ΔMSPE	p-value	ΔMSPE	p-value
$G = 1.0$	0.018	0.00	0.023	0.00	0.030	0.00	0.030	0.01	0.013	0.24	0.045	0.04
$G = 0.99$	0.003	0.11	0.003	0.14	0.002	0.20	0.002	0.20	-0.007	0.70	0.010	0.14
$G = 0.98$	0.002	0.18	0.003	0.13	0.002	0.19	0.003	0.17	-0.000	0.51	0.011	0.13
$G = 0.97$	0.000	0.30	0.003	0.19	0.002	0.27	0.002	0.27	0.008	0.33	0.010	0.17
$G = 0.96$	0.000	0.32	0.003	0.19	0.002	0.22	0.003	0.21	0.004	0.39	0.011	0.17
Rolling OLS	-0.051	0.99	-0.054	0.99	-0.049	0.98	-0.067	0.99	-0.140	1.00	-0.062	0.87
Months	744		528		396		252		799		136	

Table 7: **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. ***, **, and * indicate standard significance levels of the utility gain relative to the no-predictability benchmark.

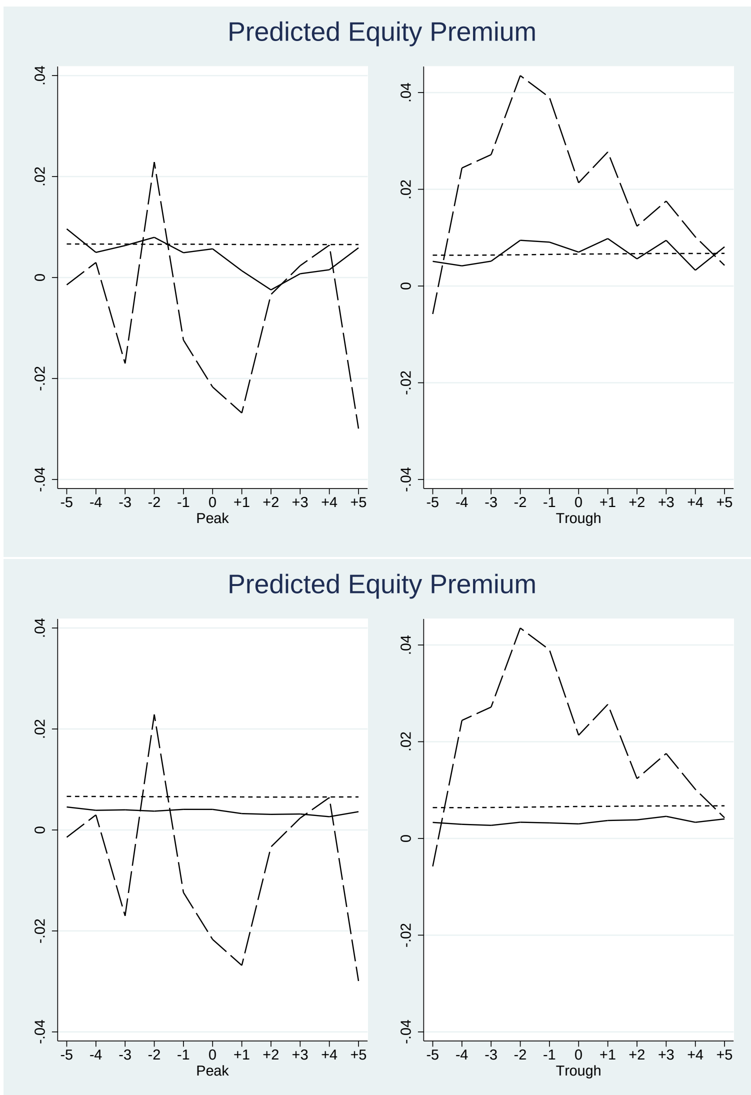
	1947+	**1965+**	**1976+**	**1988+**	**Expansions**	**Recessions**
$G = 1.0$	4.57***	7.31***	8.23***	6.99**	1.87	17.11***
$G = 0.99$	-0.84	1.50	0.15	0.94	-3.13	14.12***
$G = 0.98$	-1.53	1.15	-0.29	0.84	-3.99	14.18***
$G = 0.97$	-1.85	1.18	-0.22	1.02	-3.62	13.75**
$G = 0.96$	-2.01	1.25	-0.11	1.20	-3.78	13.74**
Rolling OLS	4.25***	6.89***	5.87***	3.97*	-2.44	22.18***

around the trough of the business cycle.); i.e., these dynamics are robust to the extensions implemented in this section. Comparing these figures to the two figures in the bottom row clearly shows that the dynamics of $G = 0.99$ (i.e., a small degree of autoregression in the coefficients) are quite different and do not show the same, economically intuitive, pattern. In the latter case predictive returns become very smooth and hardly show any reaction to the business cycle. Overall, this comparison illustrates nicely that random walk coefficients are required to match the patterns of realized returns over the business cycle.

Finally, while we concentrate on evaluation criteria that are standard in the equity premium literature in the main text of the paper, we also look at the Bayesian performance statistic, on which our model selection approach is built, namely the Bayes Factor, in this section. The idea is to illustrate that our previous conclusions still hold and are not driven by our choice of evaluation criteria.

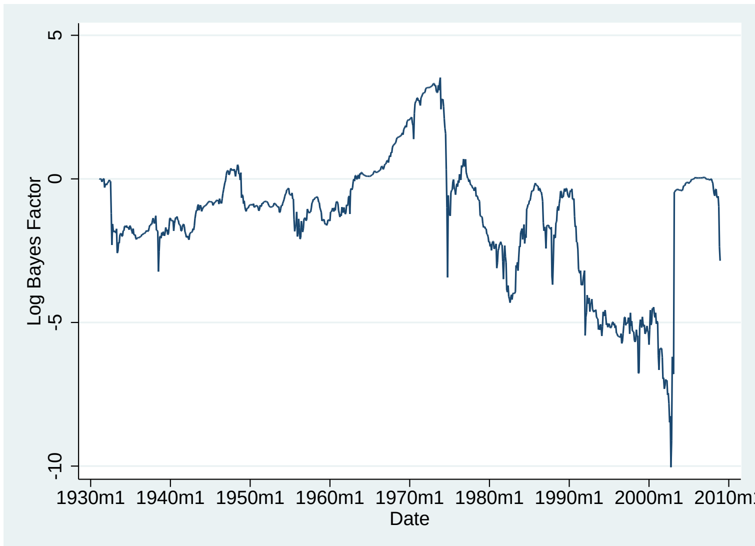
Figure 9 shows the appropriate Bayes Factor of the *BMA-Model excl. TVar-Coeff.*

Figure 8: **Equity Premium Predictions Around Peaks and Troughs:** The two graphs in the first row show the predicted monthly equity premium using the BMA-Model incl. TVar-Coeff. with $G = 1$ (solid line), the realized returns (long dashed line), and the no predictability benchmark (short dashed line). The two graphs in the second row show the predicted monthly equity premium using the BMA-Model incl. TVar-Coeff. with $G = 0.99$ (solid line), the realized returns (long dashed line), and the no predictability benchmark (short dashed line). Each graph shows averages across the 11 recessions of our sample period.



relative to the *BMA-Model incl. TVar-Coeff.* on a log scale. The Bayes Factor at a certain point in time is the ratio of the relative posterior weights to the relative prior weights of the compared models. Thus, on the log scale, a value of 0 corresponds to equal marginal support of both types of models during the current time interval; positive values are in support of models with constant coefficients and negative values are in support of models with time-varying coefficients. The picture confirms our overall conclusion: models with time-varying coefficients receive much more support from the data. Models with constant coefficients only receive somewhat more support for a small time period around 1970. The dominance of models with time-varying coefficients grows especially towards the end of our sample. Also, note the sudden jump back up to 0 in 2004. This jump occurs because Variable *csp* drops from the sample. Obviously, this variable is a key predictive variable and also a variable that drives the outperformance of models with time-varying coefficients. We leave it for further research to understand this regularity in more detail.

Figure 9: **Bayes Factor:** The graph shows the log Bayes Factor of the *BMA-Model excl. TVar-Coeff.* relative to the *BMA-Model incl. TVar-Coeff.* over time. On the log scale, a value of 0 corresponds to equal marginal support of both types of models during the current time interval; positive values are in support of models with constant coefficients and negative values are in support of models with time-varying coefficients.



References

Ang, A., Bekaert, G., 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.

Bansal, R., Yaron, A., 2004. Risks for the long-run: A potential resolution of asset pricing puzzles. Journal of Finance 59, 14811509.

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

Boudoukh, J., Michaely, R., Richardson, M., Roberts, M. R., 2007. On the importance of measuring payout yield: Implications for empirical asset pricing. The Journal of Finance 62 (2), 877–915.

Boyd, J. H., Hu, J., Jagannathan, R., 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., Song, H., McGillivray, A., 1997. Forecasting uk house prices: A time-varying coefficient approach. Economic Modelling 14, 529–548.

Campbell, J., Thompson, S. B., 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., 2008. Estimating the equity premium. Canadian Journal of Economics 41, 1–21.

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

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

Clark, T., West, K. D., 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, T., Sargent, T. J., 2003. Drifts and volatilities: Monetary policies and outcomes in the post wwii u.s., working paper.

Cooper, M., Gulen, G., 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., Marsh, P., Staunton, M., 2002. Triumph of the Optimists: 101 Years of Global Investment Returns. Princeton University Press.

Ferreira, M. A., Santa-Clara, P., 2011. Forecasting stock market returns: The sum of the parts is more than the whole. Journal of Financial Economics 100, 514–537.

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

Grullon, G., Michaely, R., 2002. Dividends, share repurchases, and the substitution hypothesis. The Journal of Finance 57 (4), 1649–1684.

Henkel, S. J., Martin, J. S., Nardari, F., 2011. Time-varying short-horizon predictability. Journal of Financial Economics 99, 560–580.

Johannes, M., Korteweg, A., Polson, N., 2011. Sequential learning, predictive regressions, and optimal portfolio returns, working paper.

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

Kacperczyk, M., Van Nieuwerburgh, S., Veldkamp, L., 2011. Attention allocation over the business cycle: Evidence from the mutual fund industry, working paper.

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

Kilian, L., Taylor, M. P., 2003. Why is it so difficult to beat the random walk forecast of exchange rates? Journal of International Economics 60, 85–107.

Kim, C.-J., Morley, J. C., Nelson, C. R., 2005. The structural break in the equity premium. Journal of Business and Economic Statistics 23(2), 181–191.

Koop, G., Poirier, D., Tobias, J., 2007. Bayesian Econometric Methods. Cambridge University Press.

Lettau, M., Ludvigson, S., 2001. Consumption, aggregate wealth, and expected stock returns. The Journal of Finance LVI, 815–849.

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

Lustig, H., Van Nieuwerburgh, S., Verdelhan, A., 2010. The wealth-consumption ratio, working paper.

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

McQueen, G., Roley, V. V., 1993. Stock prices, news, and business conditions. The Review of Financial Studies 6, 683–707.

Meese, R., Rogoff, K., 1983a. Empirical exchange rate models of the seventies: do they fit out of sample? Journal of International Economics 14, 3–24.

Meese, R., Rogoff, K., 1983b. The out-of-sample failure of empirical exchange rate models: sampling error or misspecification. In: Frenkel, J.A. (Ed.), Exchange Rates and International Macroeconomics. Chicago University Press and National Bureau of Economic Research, Chicago and Boston.

Menzly, L., Santos, T., Veronesi, P., 2004. Understanding predictability. The Journal of Political Economy 112, 1–47.

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

Pastor, L., Stambaugh, R. F., 1999. Costs of equity capital and model mispricing. The Journal of Finance 54(1), 67–121.

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

Pastor, L., Stambaugh, R. F., 2001. The equity premium and structural breaks. The Journal of Finance 56, 1207–1239.

Pastor, L., Stambaugh, R. F., 2009. Predictive systems: Living with imperfect predictors. The Journal of Finance 64, 1583–1628.

Pastor, L., Stambaugh, R. F., 2011. Are stocks really less volatile in the long-run? Journal of Finance forthcoming.

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

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

Pettenuzzo, D., Timmermann, A., 2011. Predictability of stock returns and asset allocation under structural breaks. Journal of Economitrics 164, 60–78.

Polk, C., Thompson, S., Vuolteenaho, T., 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., Madigan, D., Hoeting, J. A., 1997. Bayesian model averaging for linear regression models. Journal of the American Statistical Association 92, 179–191.

Rapach, D. E., Strauss, J. K., Zhou, G., 2010. Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. The Review of Financial Studies 23, 821–862.

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., Veldkamp, L., 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., Warusawitharana, M., 2011. What is the chance that the equity premium varies over time? evidence from predictive regressions, working paper.

West, M., Harrison, J., 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.