Predictability of Stock Returns and Asset Allocation under Structural Breaks*

Davide Pettenuzzo
Bocconi University

Allan Timmermann
University of California, San Diego

September 27, 2005

Abstract

An extensive literature in finance has found that return predictability can have important effects on optimal asset allocations. While some papers have also considered the portfolio effects of parameter and model uncertainty, model instability ('breaks') has received far less attention. This poses an important concern when the parameters of return prediction models are estimated on data samples spanning several decades during which the parameters are unlikely to remain stable. In this paper we adopt a new approach that accounts for breaks to return prediction models both in the historical estimation period and at future (out-of-sample) points. The analysis covers optimal asset allocation under parameter uncertainty, model uncertainty and uncertainty about the stability of the return forecasting model, as captured by the number of potential breaks. Our empirical findings suggest that model instability has a large effect on the asset allocation when compared to the effect of parameter estimation and model uncertainty. The possibility of future breaks has its largest effect at long investment horizons, but historical (in-sample) breaks can significantly change investment decisions even at short horizons through its effect on current parameter estimates.

*We thank Jun Liu, Ross Valkanov, Jessica Wachter and Mark Watson as well as seminar participants at Princeton, UCSD and Studiencenter Gerzensee for helpful comments on the paper.

I. Introduction

That stock returns are predictable is now widely accepted by the finance profession.[1] Such predictability has been used extensively for purposes of testing asset pricing models (e.g. Ferson and Harvey (1991), Harvey (1989)), assessing the performance of mutual funds (e.g. Ferson and Schadt (1996), Farnsworth, Ferson, Jackson, and Todd (2002)) and, most notably, in a large literature on optimal asset allocation under time-varying investment opportunities.[2]

Investors attempting to exploit such predictability in returns encounter several sources of uncertainty. Most obviously, the parameters of return prediction models are typically estimated with considerable uncertainty - a point emphasized by Kandel and Stambaugh (1996) and Barberis (2000) who propose Bayesian methods for integrating out this type of uncertainty. Moreover, since finance theory often does not identify which particular predictor variables to use, how to measure such variables and which functional form to use, investors also face model uncertainty. This point has been explored by Avramov (2002) and Cremers (2002).

One aspect of return predictability that has received far less attention is model instability. Despite the positive historical correlation between stock returns and the lagged dividend yield, the 1990s saw an unprecedented bull market with large stock returns and historically low values of the dividend yield. This brought many to question the stability of the relation between the dividend yield and stock returns. Lettau and Ludvigsson (2001) report evidence of a breakdown in a prediction model based on their *cay* variable in the mid-nineties. Lettau, Ludvigsson, and Wachter (2004) explain the run-up in stock prices during the nineties by means of a downward shift in the volatility of consumption growth. These findings lend credibility to long-lasting views among finance practitioners−indeed, as pointed out by Pastor and Stambaugh (2001), p. 1207 "Finance practitioners and academics often elect to rely on more recent data ... motivated in part by concerns that the probability distribution of excess returns changes over time, experiencing shifts known as "structural breaks"".

There are many reasons for questioning the model stability assumption. Structural instability is known to affect the majority of macroeconomic and financial variables, c.f. Stock and Watson (1996). Natural candidates for explanations of structural shifts such as institutional, legislative and technological change, large macroeconomic (oil price) shocks or changes in monetary targets or tax policy are known to occur in samples spanning long periods of time. This is important since predictability in stock returns is generally rather weak, necessitating the use of long spans of data in order to obtain reasonably precise estimates of the underlying regression coefficients. For example, Barberis (2000) uses monthly data from 1927 to 1995 to estimate the coefficient of the dividend yield in a return forecasting model. However, it is unlikely that this coefficient remained constant through a sample spanning the Great Depression, World War II, the stagflation period of

[1]See, e.g. Campbell (1987), Fama and French (1988), Kandel and Stambaugh (1996), Aït-Sahalia and Brandt (2001), Pesaran and Timmermann (1995).

[2]See, e.g., Aït-Sahalia and Brandt (2001), Avramov (2002), Barberis (2000), Brandt (1999), Brandt, Goyal, Santa-Clara, and Stroud (2002), Brennan, Schwartz, and Lagnado (1997), Brennan and Xia (2002), Campbell and Viceira (1999), Campbell and Viceira (2001), Kandel and Stambaugh (1996) and Xia (2001).

the seventies and the run-up in stock prices during the 1990s.

The model stability assumption is particularly important to asset allocation decisions since these rely on forecasts of future returns, often at long horizons. Suppose that it is found that there are structural breaks in the parameters of the return prediction model over some historical sample. If such breaks occurred in the past it would seem plausible to assume that they could also appear in the future. This introduces an extra source of risk related to when the next break(s) will occur, how long new 'regimes' will last and how large any shifts in the parameters of the return equation will be.

Despite these arguments, asset allocation exercises invariably assume that although the parameters of the return prediction model or the identity of the "true" model may not be known to investors, the parameters of the data generating process remain constant through time. To extend this analysis, we propose in this paper an approach that accounts for structural breaks in return forecasting models. Our approach builds on Chib (1998), Pastor and Stambaugh (2001) and Pesaran, Pettenuzzo, and Timmermann (2004) in proposing a changepoint model driven by an unobserved discrete state variable. This allows us to characterize structural breaks in the historical data sample. To forecast future returns under breaks, we introduce a meta distribution that sits on top of the parameters drawn for the individual regimes and characterizes how the parameters of the return model vary across different break segments. The model nests as special cases both a pooled scenario where the similarity between the parameters in the different 'regimes' is very strong (corresponding to a narrow dispersion of the distribution of these parameters across regimes) as well as a more idiosyncratic scenario where these parameters have little in common and can be very different (corresponding to a wide dispersion). Which of these cases is most in line with the data is reflected in the posterior distribution of the parameters across regimes.

Our approach is very general and allows for uncertainty about the timing (dates) of historical breaks as well as uncertainty about the number of breaks. We also extend our setup to allow for uncertainty about the identity of the predictor variables (model uncertainty) using Bayesian model averaging techniques as proposed by Avramov (2002) and Cremers (2002). Hence, investors are not assumed to know the true model or its parameter values, nor are they assumed to know the number and timing of past or future breaks. Instead, they come with prior beliefs about the "meta" distribution from which current and future values of the parameters of the return model are drawn and update these beliefs efficiently according to Bayes' rule as new data is observed.

Our empirical analysis investigates predictability of US stock returns using two popular predictor variables, namely the dividend yield and the short interest rate. We find evidence of seven breaks in return models based on either predictor variable in a data sample covering the period 1926-2003. Moreover, many of the break dates coincide with major events such as changes in the Fed's operating procedures (1979, 1982), the Great Depression, World War II and the growth slowdown following the oil price shocks in the early 1970s.

Structural breaks are found to have a large effect on optimal asset allocations. We find empirically that model instability can have an even larger effect on the asset allocation than sources of risk such as parameter estimation uncertainty and can lead to a steep negative slope in the relationship between the investment horizon and the proportion of wealth that a buy-and-hold investor allocates to stocks. This reflects the extent to which the coefficients of the predictor variables in the return equation were subject to change. Once rebalancing opportunities are introduced, as expected the asset allocation becomes less sensitive to the investment horizon, but we continue to find that breaks have an important effect on the asset allocation through two separate channels. First, past historical breaks increase the importance of parameter estimation uncertainty since the parameters from the current regime are typically surrounded by larger uncertainty than the full-sample parameters. For example, if in a sample with 30 years of data the most recent break took place ten years ago, then 10 years rather than 30 years of data can be used to estimate the model parameters. Second, the possibility of future breaks affects investors' optimal allocations even under rebalancing due to incomplete learning since they can only be detected with a lag. As a result, our estimates of the certainty equivalence of returns under rebalancing suggest that the cost of ignoring breaks can be large in economic terms.

Our study is related to that of Pastor and Stambaugh (2001) who model structural breaks in the equity premium using a univariate approach that is based on priors about the risk-return trade-off. This provides a way to deal with the problem that stock returns are typically so noisy that it is difficult to identify breaks in univariate return models based on first moments alone. Conversely, there is considerably more structure−and persistence−in the volatility of returns. Assuming that there is a trade-off between risk and returns, volatility can be used as an instrument that has power to detect breaks in the equity premium. This idea also reveals intuition for the approach we propose here, which is to increase the power to detect breaks in the model for expected returns by using as conditioning information variables that have been found empirically to be strongly correlated with returns.

There are also important differences between our approach and that proposed by Pastor and Stambaugh (2001). Most obviously, the model used here is multivariate whereas Pastor and Stambaugh use a univariate framework built on the relation between the mean and variance of returns. Furthermore, whereas the focus of Pastor and Stambaugh is on characterizing the presence of structural breaks in the equity premium in a long historical sample, we model the predictive distribution of future (out-of-sample) returns since our interest lies in the asset allocation implications of breaks. This means that we need to characterize the full breakpoint process, including the duration between future breaks and the size of possible breaks in the parameters that affect future stock returns.

Another literature that is related to our paper assumes that the parameters of the return equation are driven by a Markov switching process with a small number of states, as in Ang and Bekaert (2002), Ang and Chen (2002), Guidolin and Timmermann (2004) and Perez-Quiros and Timmermann (2000). The assumption of a fixed number of states amounts to imposing a restriction that 'history repeats'. For example, most papers on Markov switching in stock returns assume only two states so the mean and variance of returns can either be high or low depending on which state the model is in. This approach is well suited to identify patterns in returns that are linked to repeated events such as recessions and expansions. It is less clear that it is able to capture the effects of institutional and technological changes over long spans of time. These are more likely to lead to genuinely new and historically unique regimes.

The paper is organized as follows. Section II introduces the basic breakpoint methodology and Section III presents empirical estimates for return prediction models based on the dividend yield or the short interest rate. Section IV shows how investors' optimal asset allocation can be computed while accounting for past and future breaks. Section V considers the asset allocation effect empirically for a buy-and-hold investor while Section VI introduces rebalancing. Section VII proposes various extensions to the our approach and finally Section IIX concludes. Technical details are provided in an Appendix at the end of the paper.

II. **Methodology**

Studies of asset allocation in the presence of return predictability (e.g., Barberis (2000), Campbell and Viceira (2001), Campbell, Chan, and Viceira (2003) and Kandel and Stambaugh (1996)) have mostly used vector autoregressions (VARs) to capture the relation between asset returns and predictor variables, many of which are known to be highly persistent. We follow this literature and focus on a simple model with a single risky asset and a single predictor variable. This gives rise to a bivariate model relating returns (or excess returns) on the risky asset to a predictor variable, x_t. Zero-constraints on the coefficients of the lagged returns are often imposed and we shall do so here. This reflects the common finding that stock returns are not strongly serially correlated and reduces the number of parameters to be estimated. The resulting model takes the form

$$z_t = B'\tilde{x}_{t-1} + u_t, \tag{1}$$

where $z_t = (r_t, x_t)'$, $\tilde{x}_{t-1} = (1, x_{t-1})'$, r_t is the stock return at time t in excess of a short risk-free rate, while x_{t-1} is the lagged predictor variable and $E[u_t u_t'] = \Sigma$ is the covariance matrix. We refer to μ_r and μ_x as the intercepts in the equation for the return and predictor variable, respectively, while β_r and β_x are the coefficients on the predictor variable in the two equations:

$$\begin{aligned} r_t &= \mu_r + \beta_r x_{t-1} + \varepsilon_{rt} \\ x_t &= \mu_x + \beta_x x_{t-1} + \varepsilon_{xt}. \end{aligned} \tag{2}$$

A. *Predictive Distributions of Returns under Breaks*

Asset allocation decisions require the ability to evaluate the expected utility associated with the realization of future payoffs of risky assets. This, in turn, requires computing expectations over the predictive distribution of returns during an $h-$period investment horizon $[T, T+h]$ conditional on information available at the time of the investment decision, T, which we denote by $\mathcal{Z}_T$. To compute the predictive distribution of returns under breaks, we need to make assumptions about the probability that future breaks occur, their likely timing as well as the size of such breaks. Most obviously, we need an estimate of the probability of staying in the current regime. If more than one break can occur over the course of the investment horizon, we also need to model the distribution from which future regime durations are drawn. We next explain how this is done.

To capture instability in the parameters in (2), we build on the multiple change point model proposed by Chib (1998) and extended by Pastor and Stambaugh (2001). Shifts to the parameters of the return model are captured through an integer-valued state variable, S_t, that tracks the regime from which a particular observation of returns and the predictor variable, x_t, are drawn. For example, $s_t = l$ indicates that z_t has been drawn from $f\left(z_t \middle| \mathcal{Z}_{t-1}, \Theta_l\right)$, where $\mathcal{Z}_{t-1} = \{z_1, ..., z_t\}$ is the period-$t-1$ information set, while a change from $s_t = l$ to $s_{t+1} = l+1$ shows that a break has occurred at time $t+1$. The location and scale parameters in regime l are collected in $\Theta_l = (B_l, \Sigma_l)$. Allowing for K breaks or, equivalently, $K+1$ break segments, our model takes the form

$$\begin{array}{lll} z_t = B_1'\tilde{x}_{t-1} + u_t, & E[u_t u_t'] = \Sigma_1 & \text{for } \tau_0 \leq t \leq \tau_1 \quad (s_t = 1) \\ z_t = B_2'\tilde{x}_{t-1} + u_t, & E[u_t u_t'] = \Sigma_2 & \text{for } \tau_1 + 1 \leq t \leq \tau_2 \quad (s_t = 2) \\ \vdots & \vdots & \vdots \\ z_t = B_l'\tilde{x}_{t-1} + u_t, & E[u_t u_t'] = \Sigma_l & \text{for } \tau_{l-1} + 1 \leq t \leq \tau_l \quad (s_t = l) \\ \vdots & \vdots & \vdots \\ z_t = B_{K+1}'\tilde{x}_{t-1} + u_t, & E[u_t u_t'] = \Sigma_{K+1} & \text{for } \tau_K + 1 \leq t \leq T \quad (s_t = K+1) \end{array} \tag{3}$$

Here $\Upsilon_K = \{\tau_0, ..., \tau_K\}$ is the collection of break points with $\tau_0 = 1$. Within each regime we decompose the covariance matrix, Σ_j, into the product of a diagonal matrix representing the standard deviations of the variables, $diag(\psi_j)$, and a correlation matrix, Λ_j, each of which is modeled separately:

$$\Sigma_j = diag(\psi_j) \times \Lambda_j \times diag(\psi_j). \tag{4}$$

This specification allows both volatilities and correlations to vary across regimes.

The state variable S_t is assumed to be driven by a first order hidden Markov chain whose transition probability matrix is designed so that, at each point in time, S_t can either remain in the current state or jump to the subsequent state. The one-step-ahead transition probability matrix therefore takes the form

$$P = \left(\begin{array}{ccccc|c} p_{11} & p_{12} & 0 & \cdots & 0 & \\ 0 & p_{22} & p_{23} & \cdots & 0 & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \\ 0 & \cdots & 0 & p_{KK} & p_{K,K+1} & \\ 0 & 0 & \cdots & 0 & p_{K+1,K+1} & p_{K+1,K+2} \\ \hline 0 & 0 & \cdots & & 0 & p_{K+2,K+2} \\ & & & & & \ddots \end{array} \right). \tag{5}$$

Here $p_{j-1,j} = Pr\left(s_t = j \middle| s_{t-1} = j-1\right)$ is the probability of moving to regime j at time t given that we are in state $j-1$ at time $t-1$ so $p_{j,j} + p_{j,j+1} = 1$. K is the number of breaks in the historical sample up to time T so the $(K+1) \times (K+1)$ sub-matrix in the upper left corner of P describes possible breaks in the historical data sample, $\{z_1, ..., z_T\}$. The remaining part of P describes the breakpoint dynamics over the future out-of-sample investment period from T to $T+h$.[3] The special case without breaks corresponds to $K = 0$ and $p_{11} = 1$.

[3]Following Chib (1998), estimation proceeds under the assumption of K breaks in the historical sample ($1 \leq t \leq T$).

Notice that the persistence parameters in (5) are regime-specific. This assumption means that regimes can differ in their expected duration−the closer is p_{jj} to one, the longer the regime is expected to last. Furthermore, $p_{j,j}$, are assumed to be independent of $p_{i,i}$, for $j \neq i$, and are drawn from a beta distribution:

$$p_{j,j} \sim Beta\,(a,b)\,. \tag{6}$$

This means that breaks are difficult to predict ahead of time, but still can have very important implications for portfolio allocations.

B. *Meta Distributions*

Since we are interested in forecasting future returns out-of-sample, we follow Pastor and Stambaugh (2001) and Pesaran, Pettenuzzo, and Timmermann (2004) and adopt a hierarchical prior formulation, but extend those studies to allow for structural breaks in a multivariate setting.[4] To this end we assume that the location and scale parameters within each regime, $(B_j,\ \Sigma_j)$, are drawn from common distributions. We refer to these as meta distributions since they sit on top of the parameter distributions within each regime and characterize the degree of similarity in the parameters across different regimes. Suppose for example that the mean parameters do not vary much across regimes but that the variance parameters do. Then this will show up in the form of a wide dispersion in the meta distribution of the scale parameters and a narrow dispersion in the meta distribution for the location parameters.

The assumption that the parameters are drawn from a common meta distribution implies that data from previous regimes carry information relevant for current data and for the new parameters after a future break. An alternative approach of entirely discarding pre-break data tends to lead to imprecise estimates and also goes against the intuition that pre-break data contains some information about the parameters in the new regime. By using meta distributions that pool information from different regimes our approach makes sure that historic information is used efficiently in estimating the parameters of the current regime.

We next describe the meta distributions in more detail. We use a random coefficient model to introduce a hierarchical prior for the regime coefficients in (3) and (4), $\{B_j, diag(\psi_j), \Lambda_j\}$. Let m be the number of equations in the prediction model (1) and assume that the $m^2 \times 1$ vector of location coefficients are independent draws from a normal distribution, $vec(B)_j \sim N\,(b_0, V_0)$, $j = 1, ..., K+1$, while the m error term precisions $\psi_{j,i}^{-2}$ are independent and identical draws (IID) from a Gamma distribution, $\psi_{j,i}^{-2} \sim Gamma\,(v_{0,i}, d_{0,i})$, $i = 1, ..., m$. Finally, the $m\,(m-1)\,/2$ correlations, $\lambda_{j,ic}$, are

This assumption greatly simplifies estimation. We show later that uncertainty about the number of in-sample breaks can be integrated out using Bayesian model averaging techniques.

[4] Bai, Lumsdaine, and Stock (1998) apply a deterministic procedure to detect breaks in multivariate time series models and find that when break dates are common across equations, the resulting breaks are estimated more precisely. The power to detect breaks can also increase when the breaks are estimated from a multivariate model. Their framework is not well suited for our purpose, however, since asset allocation exercises build on the predictive distribution of future returns and thus require modeling the stochastic process underlying the breaks.

IID draws from a normal distribution, $\lambda_{j,ic} \sim N(\mu_{\rho,ic}, \sigma^2_{\rho,ic})$, $i, c = 1, ..., m$, $i < c$.[5] In this context, $b_0, v_{0,i}$ and $\mu_{\rho,ic}$ represent the location parameters, while V_0, $d_{0,i}$ and $\sigma^2_{\rho,ic}$ are the scale parameters of the three meta distributions.

The pooled scenario (all parameters are identical across regimes) and the regime-specific scenario (the parameters of each regime are unrelated) can be seen as extreme special cases each of which is nested in our framework. Which scenario most closely represents the data can be inferred from the estimates of the location parameters of the meta distribution V_0, $d_{0,i}$ and $\sigma^2_{\rho,ic}$.

To characterize the parameters of the meta distribution, we assume that[6]

$$\begin{aligned} b_0 &\sim N\left(\underline{\mu_\beta}, \underline{\Sigma_\beta}\right) \\ V_0^{-1} &\sim W\left(\underline{v_\beta}, \underline{V_\beta^{-1}}\right), \end{aligned} \tag{7}$$

where $W(.)$ is a Wishart distribution and $\underline{\mu_\beta}$, $\underline{\Sigma_\beta}$, $\underline{v_\beta}$ and $\underline{V_\beta^{-1}}$ are prior hyperparameters that need to be specified. The hyperparameters $v_{0,i}$ and $d_{0,i}$ of the error term precision are assumed to follow an exponential and Gamma distribution, respectively, c.f. George, Makov, and Smith (1993) with prior hyperparameters $\underline{\rho_{0,i}}$, $\underline{c_{0,i}}$ and $\underline{d_{0,i}}$:

$$v_{0,i} \sim Exp\left(\underline{\rho_{0,i}}\right) \tag{8}$$

$$d_{0,i} \sim Gamma\left(\underline{c_{0,i}}, \underline{d_{0,i}}\right). \tag{9}$$

Following Liechty, Liechty, and Müller (2004), we specify the following distributions for the hyperparameters of the correlation matrix (truncated to lie in the $(-1, 1)$ interval):

$$\mu_{\rho,ic} \sim N\left(\underline{\mu_{\mu,ic}}, \underline{\tau^2_{ic}}\right) \tag{10}$$

$$\sigma^{-2}_{\rho,ic} \sim Gamma\left(\underline{a_{\rho,ic}}, \underline{b_{\rho,ic}}\right), \tag{11}$$

where again $\underline{\mu_{\mu,ic}}$, $\underline{\tau^2_{ic}}$, $\underline{a_{\rho,ic}}$ and $\underline{b_{\rho,ic}}$ are prior hyperparameters for each element of the correlation matrix. We finally specify a prior distribution for the hyperparameters a and b of the transition probabilities,

$$\begin{aligned} a &\sim Gamma\left(\underline{a_0}, \underline{b_0}\right), \\ b &\sim Gamma\left(\underline{a_0}, \underline{b_0}\right). \end{aligned} \tag{12}$$

These are all standard choices of distributions.

C. *Prior elicitation*

To the extent possible, choice of priors in the breakpoint model must be guided by economic theory and intuition. Here we explain the choices made for the baseline results. In section VII we conduct a sensitivity analysis to shed light on the importance of these choices.

[5] Given the symmetry of the correlation matrix Λ_j, we only model elements above the main diagonal.

[6] Throughout the paper we use underscore bars (e.g. $\underline{a_0}$) to denote prior hyperparameters.

We impose two constraints on the parameters in the return prediction model, (2). First, to rule out explosive behavior in the driving variable (and consequently in stock returns), we impose that $\beta_x < 1$. Second, since neither the dividend yield nor the short interest rate can go negative, we impose that the unconditional mean in each state is non-negative, i.e. $0 \leq \mu_x/(1-\beta_x) \leq \bar{\mu}_x$, where $\bar{\mu}_x$ is an upper limit chosen so the unconditional mean of the predictor variable lies in the centre of the interval. This ensures that the predictive densities of all variables are well-behaved even at very long investment horizons.[7]

Starting with the prior hyperparameters for the mean of the regression coefficient, b_0, we set $\underline{\mu_\beta} = 0_{m^2}$ and $\underline{\Sigma_\beta} = sc \times I_{m^2}$ where sc is a scale factor set to ∞ to reflect uninformative priors. The hyperparameters for the prior variance of the regression coefficient, V_0, are set at $\underline{\nu_\beta} = \left(m^2\right) + 2$ and $\underline{V_\beta} = diag(0.1, 10, 0.01, 0.1)$. This is sufficient to preserve the variation in the regression coefficients across regimes and ensures that the mean of the Wishart distribution exists. As we shall see, this choice reflects the large variation in the slope coefficient of the predictor variable in the return equation (β_r), the very small variation in μ_x and the somewhat larger variation in μ_r and β_x.

Moving to the variance hyperparameters, we maintain uninformative priors and set $\underline{c_{0,i}} = 1$, $\underline{d_{0,i}} = 1/\infty$ and $\underline{\rho_{0,i}} = \infty$ in all equations, hence specifying a very large variance. We also use uninformative priors for the correlation coefficient, $r_{j,12}$, i.e. $\underline{\mu_{\mu,12}} = 0$, $\underline{\tau_{12}^2} = \infty$, $a_{\rho,12} = 1$ and $b_{\rho,12} = 0.01$. Finally, we assume a more informative prior for the diagonal elements, p_{ii} in (5), centered at 0.98, namely $\underline{a_0} = 1$ and $\underline{b_0} = 0.02$. This ensures that breaks do not occur too frequently.

III. **Breaks in Return Forecasting Models: Empirical Results**

Using the approach from Section 2, we next report empirical results for two commonly used return prediction models based on the dividend yield or the short interest rate.

A. *Data*

Following common practice in the literature on predictability of stock returns, we use as our dependent variable the continuously compounded return on a portfolio of US stocks comprising firms listed on the NYSE, AMEX and NASDAQ in excess of a 1-month T-bill rate. Data is monthly and covers the period 1926:12-2003:12. All data is obtained from the Center for Research in Security Prices (CRSP).

As forecasting variables we include a constant and either the dividend-price ratio−defined as the ratio between dividends over the previous twelve months and the current stock price−or the short interest rate measured by the 1-month T-bill rate obtained from the Fama-Bliss files. The dividend yield has been found to predict stock returns by many authors including Campbell (1987), Campbell and Shiller (1988), Keim and Stambaugh (1986) and Fama and French (1988). It has played a key role in the literature on asset allocation implications of return predictability, c.f. Kandel and Stambaugh (1996) and Barberis (2000). Furthermore, due to its persistence and the large negative

[7] If the upper constraint on the mean of the predictor variable is ignored while negative values are ruled out, the mean of the predictive density tends to increase too much at very long investment horizons.

correlation between shocks to the dividend yield and shocks to stock returns, the dividend yield is known to generate a large hedging demand for stocks, particularly at long investment horizons. The short interest rate has also been found to reliably predict stock returns, c.f. Campbell (1987) and Ang and Bekaert (2002).

Consistent with results in the literature, full-sample estimates of the parameters in the return equation (2) reveal (mean) coefficients on the dividend yield or the T-bill rate slightly less than two standard errors away from zero, although the sign of the regression coefficients differ−positive for the yield, negative for the T-bill rate. Both predictor variables are highly persistent with autoregressive parameters close to 0.98.

B. *Predictability from the Dividend Yield*

Determining whether the return prediction models are subject to breaks and, if so, how many breaks the data support, is the first step in our analysis. For a given choice of number of breaks, K, we get a new model, M_K, with its own set of parameters. Table 1 provides a comparison of models with different numbers of breaks by reporting various measures of model fit such as the conditional log-likelihood (which does not penalize for additional parameters as more breaks are considered), the marginal log-likelihood (which penalizes for overfitting in case too many breaks are identified) and the posterior model probability computed for models with up to eight breaks. The latter (computed under equal priors on each of the possible number of breaks) is a conventional way to summarize the evidence in support of a particular model.

First consider the return model based on the dividend yield. For this model there is strong support for structural breaks−the posterior odds ratios for the models with multiple breaks relative to a model with no breaks are all very high. Among all models with up to eight breaks, the fourth column in Table 1 shows that the model with seven breaks obtains a posterior probability weight of two-thirds. Although this may appear to be a large number of breaks, it is consistent with the evidence reported by Pastor and Stambaugh (2001) of 15 break points in the equity premium over a sample (1834-1999) a bit longer than twice the period covered here. Of course we cannot be sure that there are seven breaks in the sample, nevertheless the data strongly supports this specification.[8]

Table 1 also shows for each model the time of the associated breaks. More precisely, these are the posterior modes for the break dates since our model only provides probability estimates of the break dates. Accordingly, Figure 1 shows posterior probabilities of the break locations for the model with seven breaks. The break dates are quite precisely identified in the form of single spikes with probabilities ranging from 0.27 to more than 0.50. This suggests that there is not much uncertainty about the locations of the break dates for this model.

Five of the break locations are associated with major events and occurred around the Great Depression (1932), the beginning of World War II (1940), the major oil price shocks of the early seventies and resulting growth slowdown (1974), the end of the change in the Fed's operating

[8] When specifications with more than eight breaks were considered, the same break dates were essentially chosen twice, thus suggesting that further breaks are not warranted.

procedures (1982) and, more recently, at the beginning of the bull market of the nineties (1992). The two remaining break dates (1952 and 1958) are harder to interpret, but the first coincided with the Korean war while the break in the late fifties matches a similar increase in the posterior probability of a break in the equity premium model identified by Pastor and Stambaugh (2001) (their Figure 1, top panel).

These break dates suggest that changes to the conditional equity premium are associated with events such as major wars, changes to monetary policy and important slowdowns in economic activity as caused, e.g., by major supply shocks.

Parameter estimates for the model with seven breaks (eight regimes) are reported in Table 2. The mean of the dividend yield coefficient in the return equation ranges from a low of 0.39 prior to the Great Depression to values around two during the regimes from 1958-1974 and again from 1974-1982 before declining to 1.40 in the last regime. The standard deviation parameters of the return equation also vary considerably over the sample, from a high of 10% per month around the Great Depression to a low of only 3.4% per month during the 1950s. Since the mid-seventies, return volatility appears to have been quite low at 4-5% per month.

Turning to the parameter estimates for the dividend yield equation, it is clear that this process is highly persistent in all regimes with a mean autoregressive parameter that varies from 0.90 to 0.98. The variance of the dividend yield is again highest in the first two regimes and much lower after the final break around 1992. Correlation estimates for the innovations to stocks and the lagged dividend yield are large and negative in all regimes with mean values ranging from -0.96 to -0.85. Similarly, transition probabilities are high with means that always exceed 0.98 and go as high as 0.992, corresponding to mean durations ranging from 70 to 140 months.

One of the questions we set out to address in our paper was how dissimilar the parameters of the return equation are across the various regimes. To address this question, information on the posterior estimates of the hyperparameters of the meta distribution is provided in Table 3. To preserve space we only report the values of the mean parameters which are easiest to interpret. The parameter tracking the mean of the slope of the dividend yield in the return equation across different regimes is centered on 1.2 with a standard deviation centered at 0.50, giving rise to a 95% confidence interval $[0.23, 2.21]$. The autoregressive slope β_x in the dividend yield equation is centered on a value of 0.92 with a much smaller standard deviation of only 0.033 and a 95% confidence interval $[0.84, 0.97]$. Similarly, the hyperparameter tracking the correlation between shocks to returns and to the dividend yield is centered on -0.92 with a modest standard deviation of 0.04. The posterior distributions of the hyperparameters of the transition probability, a_0 and b_0, are surrounded by greater uncertainty as indicated by their relatively large standard deviations. This is consistent with the considerable difference in the duration of the various regimes identified by our model.

These findings suggest that the greatest variability in parameters across regimes is associated with the effect of the dividend yield on stock returns and the duration of the regimes. There is considerably less variability in the persistence of the dividend yield or in the correlation between shocks to returns and shocks to the dividend yield.

C. *Predictability from the Short Interest Rate*

Turning to the return model based on the short interest rate, Table 1 also suggests the presence of seven breaks in this model. Some of these breaks again appear around the time of major historical events such as the Great Depression (1934), the end of World War II (1947), the Vietnam War (1968), the beginning and end of the change to the Fed's operating procedures (1979 and 1982) and the beginning of the protracted bull market of the 1990s (1990). The two remaining breaks are estimated to have occurred around 1952 and 1968.[9]

Figure 2 shows the posterior probabilities for the breakpoint locations. These are more dispersed than those found for the prediction model based on the dividend yield, but define narrow ranges for most break dates, nevertheless. For example, the breaks in 1934 and 1947 are confined to one or two months while the break dates around 1979 and 1982 are also quite well determined. The dating of the breaks occurring around 1952 and 1990 is surrounded by the greatest uncertainty.

Compared to the findings for the dividend yield there is considerable overlap in the break dates of the two models. Both identify breaks in the early part of the Great Depression (1932/34), the Korean War (1952), the end of the Fed's monetarist experiment (1982) and at the beginning of the 1990s bull market (1990/92). This suggests that at least some of the breaks may be common across predictor variables.

Parameter estimates for the return model with seven breaks are displayed in Table 4. The mean of the coefficient on the lagged T-bill rate in the return equation is always negative but varies significantly over time, ranging from only -0.44 in the first regime to -9.98 in the very volatile regime from 1979 to 1982 when the Fed changed its monetary policy. Furthermore, the estimates of the slope on the T-bill rate within each regime are surrounded by large standard errors, particularly in the regimes from 1934-1947 and 1947-1952[10].

The process for the short interest rate is highly persistent with the mean of the persistence coefficient ranging from a low of 0.83 to a high of 0.97 after the most recent break. Although the correlation between shocks to returns and shocks to the short rate is closer to zero than was found for the dividend yield model, it also varies much more across regimes, ranging from a low of -0.31 during 1979-1982 to a high of 0.30 during 1982-1990. These changes appear not simply to reflect random sample variations since the standard deviations of the correlations are mostly quite low. All states continue to be highly persistent with mean transition probability estimates varying from 0.976 to 0.992, resulting in state durations between 40 and more than 160 months.

Turning finally to the meta distribution parameters for the short rate model shown in Table 5, once again the chief source of uncertainty is the slope coefficient of the interest rate in the return

[9] The conditional log-likelihood function of the return equation declines as we move from five to six breaks. This can happen even as the number of breaks increases as long as the likelihood of the joint distribution of stock returns and the short interest rate increases, which indeed it does uniformly as the number of breaks (and hence the number of parameters) goes up.

[10] Notice that periods such as during the Accord where interest rates were tightly controlled are picked up by our procedure which identifies 1934-1947 as a regime with very low interest rate volatility and an insignificant effect of T-bill rates on stock returns.

equation. For example, $b_0(\beta_r)$ has a mean of -4.6 and a standard deviation of 4.8, giving a very long 95% confidence interval that ranges from -14.8 to 4.4. Compared with the model based on the dividend yield, there is now also greater uncertainty about the correlation between shocks to returns and shocks to the T-bill rate as indicated by the higher standard deviation of μ_ρ.

IV. **Asset Allocation under Structural Breaks**

Investors care about instability in the return model because breaks affect future asset payoffs and therefore may alter their optimal asset allocation. To study the economic importance of structural breaks in the return model, we next consider the optimal asset allocation under a range of alternative modeling assumptions for a buy-and-hold investor (later generalized to allow for rebalancing) with power utility over terminal wealth and coefficient of relative risk aversion, γ:

$$u(W_{T+h}) = \frac{W_{T+h}^{1-\gamma}}{1-\gamma}, \quad \gamma > 0. \tag{13}$$

Following Kandel and Stambaugh (1996) and Barberis (2000), we assume that the investor has access to a risk-free asset with constant return, r_f, and a risky stock market portfolio with returns in excess of the risk-free rate, $r_{T+1}, ..., r_{T+h}$, where h is the investment horizon. All returns are continuously compounded.

A. *The Asset Allocation Problem*

Without loss of generality we set initial wealth at one, $W_T = 1$, and let ω be the allocation to stocks. Terminal wealth is then given by

$$W_{T+h} = (1-\omega)\exp(r_f h) + \omega \exp(r_f h + r_{T+1} + ... + r_{T+h}). \tag{14}$$

Defining the cumulative excess returns over h periods as

$$R_{T+h} = r_{T+1} + r_{T+2} + ... + r_{T+h},$$

subject to the no short-sales constraints $0 \leq \omega < 1$, the buy-and-hold investor solves the following program

$$\max_{\omega} E_T \left(\frac{((1-\omega)\exp(r_f h) + \omega \exp(r_f h + R_{T+h}))^{1-\gamma}}{1-\gamma} \right), \tag{15}$$

where E_T is the conditional expectation given information at time $T, \mathcal{Z}_T$. How this expectation is computed reflects the modeling assumptions made by the investor.

The predictive density for the h-period cumulative returns, R_{T+h}, can be constructed using the iterative scheme of Barberis (2000) that takes advantage of the assumed VAR structure. To see how this works, rewrite (2) as $z_t = \mu_t + \beta_{0t} z_{t-1} + u_t$ where $\mu_t = (\mu_{rt}\ \mu_{xt})'$ and

$$\beta_{0t} = \left[\begin{pmatrix} 0 & \beta_{rt} \\ 0 & \beta_{xt} \end{pmatrix} \right].$$

The time subscripts on the parameters reflects the possibility of breaks. Iterating forward on this model, we have

$$
\begin{aligned}
z_{T+1} &= \mu_{T+1} + \beta_{0T+1} z_T + u_{T+1} \\
z_{T+2} &= \mu_{T+2} + \beta_{0T+2}\mu_{T+1} + \beta_{0T+2}\beta_{0T+1} z_T + u_{T+2} + \beta_{0T+2} u_{T+1} \\
z_{T+3} &= \mu_{T+3} + \beta_{0T+3}\mu_{T+2} + \beta_{0T+3}\beta_{0T+2}\mu_{T+1} + \beta_{0T+3}\beta_{0T+2}\beta_{0T+1} z_T \\
&\quad + u_{T+3} + \beta_{0T+3} u_{T+2} + \beta_{0T+3}\beta_{0T+2} u_{T+1} \\
&\ \vdots \\
z_{T+h} &= \mu_{T+h} + u_{T+h} + \sum_{j=1}^{h-1}\left(\prod_{i=j+1}^{h}\beta_{0T+i}\right)(\mu_{T+j} + u_{T+j}) + \left(\prod_{i=j+1}^{h}\beta_{0T+i}\right) z_T.
\end{aligned} \tag{16}
$$

Assuming that $u_T \sim (0, \Sigma)$ is normally distributed, in the special case where μ and β_0 do not vary over time, we get the results reported by Barberis (2000), namely $z_{T+h} \sim N\left(\mu_{z_{T+h}}, \Sigma_{z_{T+h}}\right)$, where

$$
\begin{aligned}
\mu_{z_{T+h}} &= \sum_{s=0}^{h-1}\beta_0^s \mu + \beta_0^h z_T, \\
\Sigma_{z_{T+h}} &= \left(\sum_{s=0}^{h-1}\beta_0^s\right)\Sigma\left(\sum_{s=0}^{h-1}\beta_0^s\right)'.
\end{aligned}
$$

Furthermore, the sum $z_{T:T+h} = z_{T+1} + z_{T+2} + ... + z_{T+h}$ is distributed as a multivariate normal variable with mean vector μ_{sum} and variance-covariance matrix Σ_{sum}

$$
\begin{aligned}
\mu_{sum} &= h\mu + (h-1)\beta_0\mu + (h-2)\beta_0^2\mu \\
&\quad + ... + \beta_0^{h-1}\mu + (\beta_0 + \beta_0^2 + ... + \beta_0^h) z_T
\end{aligned} \tag{17}
$$

$$
\begin{aligned}
\Sigma_{sum} &= \Sigma + (I+\beta_0)\Sigma(I+\beta_0)' \\
&\quad + (I+\beta_0+\beta_0^2)\Sigma(I+\beta_0+\beta_0^2)' \\
&\quad + (I+\beta_0+...+\beta_0^{h-1})\Sigma(I+\beta_0+...+\beta_0^{h-1})'.
\end{aligned} \tag{18}
$$

Comparing (16) to (17) and (18), clearly the possibility of shifts in either the persistence parameters, β_0, the intercepts, μ, or the covariance matrix, Σ, can lead to important changes in the return distribution over an $h-$period investment horizon and must be carefully taken into account. Furthermore, such changes make it far more complicated to evaluate the predictive return distribution and require the use of numerical methods. We next consider a range of assumptions about how investors deal with their limited knowledge of the parameters of the return model.

B. *No Breaks*

First consider the asset allocation problem for an investor who ignores parameter estimation uncertainty and breaks. Once the predictor variables have been specified, the VAR parameters

$\Theta = (\mu, \beta_0, \Sigma)$ can be estimated and, using (17) and (18), the model can be iterated forward conditional on these parameter estimates. This generates a distribution for future stock returns, $p(R_{T+h}|\widehat{\Theta}, S_{T+h} = 1, \mathcal{Z}_T)$ where $S_{T+h} = 1$ shows that past and future breaks are ignored. The investor therefore solves the problem

$$\max_{\omega} \int u(W_{T+h}) p(R_{T+h}|\widehat{\Theta}, S_{T+h} = 1, \mathcal{Z}_T) dR_{T+h}. \tag{19}$$

Here we used that, from (14), the only part of W_{T+h} that is uncertain is R_{T+h}. This of course ignores that Θ is not known precisely but typically is estimated with considerable uncertainty.[11]

Next consider the decision of an investor who accounts for parameter estimation uncertainty but ignores both past and future breaks, i.e., assumes that $S_{T+h} = 1$. In the absence of breaks the posterior distribution $\pi(\Theta|S_{T+h} = 1, \mathcal{Z}_T)$ summarizes the uncertainty about the parameters given the historical data sample.[12] Integrating over this distribution leads to the predictive distribution of returns conditioned only on the observed sample (and not on any fixed Θ) and the assumption of no breaks prior to time $T + h$:

$$p(R_{T+h}|S_{T+h} = 1, \mathcal{Z}_T) = \int p(R_{T+h}|\Theta, S_{T+h} = 1, \mathcal{Z}_T) \pi(\Theta|S_{T+h} = 1, \mathcal{Z}_T) d\Theta. \tag{20}$$

This investor therefore solves the asset allocation problem

$$\max_{\omega} \int u(W_{T+h}) p(R_{T+h}|S_{T+h} = 1, \mathcal{Z}_T) dR_{T+h}. \tag{21}$$

Comparing stock holdings in (19) and (21) gives a measure of the economic importance of parameter estimation uncertainty. Both solutions ignore model instability, however. To illustrate the effect of breaks in the parameters of the return prediction model, we separately consider scenarios with past breaks but no future breaks as well as scenarios allowing for both past and future breaks.

C. *Past Breaks Only*

Suppose that there are K historic breaks in the sample up to time T but that no new breaks occur prior to the end of the investment horizon, $T + h$. Returns can then be modeled using only the posterior distribution of the parameters from the last regime, $\{B_{K+1}, \Sigma_{K+1}\}$. Hence the asset allocation is computed under the predictive density $p(R_{T+h}|S_{T+h} = S_T = K + 1, \mathcal{Z}_T)$. Under this assumption the predictive density of returns becomes

$$\begin{aligned} & p(R_{T+h}|S_{T+h} = S_T = K + 1, \mathcal{Z}_T) \\ = & \int p\left(R_{T+h}|\, \Theta_{K+1}, S_{T+h} = K + 1, S_T = K + 1, \mathcal{Z}_T\right) \pi\left(\Theta_{K+1}|\, H, p, \mathcal{S}_T, \mathcal{Z}_T\right) d\Theta_{K+1}, \end{aligned}$$

[11] To be more precise, we could condition also on M_{K_x} i.e. the return prediction model based on the predictor variable x and conditional on K historical breaks. The importance of M_{k_x} will become clear when we integrate out uncertainty about the number of in-sample breaks and uncertainty about the predictor variables.

[12] Throughout the paper, $\pi(\cdot|\cdot, \mathcal{Z}_T)$ refers to posterior distributions conditioned on information contained in $\mathcal{Z}_T$.

where $\mathcal{S}_T = (s_1, ..., s_T)$ is the collection of values of the state variable underlying the breakpoint process up to period T,

$$\Theta_{K+1} = \left(vec(B)_{K+1}, \psi_{K+1}, \Lambda_{K+1}\right)$$

are the parameters from the $K+1-$th regime (regression coefficients, error term variances and correlations), and

$$H = \left(b_0, V_0, v_{0,1}, d_{0,1}, ..., v_{0,m}, d_{0,m}, \mu_{\rho,12}, \sigma^2_{\rho,12}, ..., \mu_{\rho,m-1m}, \sigma^2_{\rho,m-1m}, a, b\right)$$

are the hyperparameters of the meta distribution. This investor therefore solves the following portfolio problem

$$\max_{\omega} \int u(W_{T+h}) p(R_{T+h} | S_{T+h} = S_T = K+1, \mathcal{Z}_T) dR_{T+h}. \tag{22}$$

Even though the possibility of future breaks is ignored in this scenario, historic breaks still affect the asset allocation since the parameters assumed to compute the predictive distribution in (22) are from the last regime, as opposed to being based on the full data sample.[13]

D. *Past and Future Breaks*

To allow for multiple breaks over the investment horizon $[T, T+h]$, we need to know not only the probability of shifting to a new regime but also the probability of staying in future regimes $p_{K+j,K+j}$, $j \geq 2$. Our hierarchical prior setup is ideally suited to address this question. Under our model, $p_{K+j,K+j}-$values are drawn from the conditional beta posterior

$$p_{K+j,K+j} | \mathcal{S}_T \sim Beta(a + l_j, b + 1),$$

where $l_j = \tau_j - \tau_{j-1}$ is the duration of the lth break segment.

The number of possible out-of-sample break point scenarios becomes very large as the number of possible breaks increases. To see this, consider the case with only two breakpoints in the out-of-sample period. This gives $\sum_{i=1}^{h-1} (h - i) = h(h-1)/2$ possible break point locations. For a five-year investment horizon ($h = 60$) this gives 1,770 possible break locations. Uncertainty about future breakpoint scenarios is captured through the variable S_{T+h} which must be integrated out. We do this by first conditioning on the maximum number of future breaks between T and $T+h$. Conditional on this we then compute the probability of the break locations. We finally sum over both the number of breaks and break locations.

For example, conditioning on at most n_b breaks over the investment horizon $[T, T+h]$, and letting j track the date where a break occurs, we can compute the probability of zero, one, two, up

[13] This is an over-simplification since these parameters depend on other regimes through the meta distribution. Furthermore, it should also be recalled that we allow for uncertainty about the time of the most recent break, τ_K. Later we show how to integrate out uncertainty about the number of in-sample breaks, K.

to n_b breaks as follows:

$$
\begin{aligned}
p(S_{T+h} &= K+1|S_T = K+1, \mathcal{Z}_T) = p^h_{K+1,K+1} \\
p(S_{T+h} &= K+2|S_T = K+1, \mathcal{Z}_T) = \sum_{j_1=1}^{h} (1-p_{K+1,K+1})\, p^{j_1-1}_{K+1,K+1} \\
p(S_{T+h} &= K+3|S_T = K+1, \mathcal{Z}_T) = \sum_{j_1=1}^{h-1} \sum_{j_2=j_1+1}^{h} p^{j_1-1}_{K+1,K+1} (1-p_{K+1,K+1})\, p^{j_2-j_1-1}_{K+2,K+2} (1-p_{K+2,K+2}) \\
&\vdots \\
p(S_{T+h} &= K+n_b+1|S_T = K+1, \mathcal{Z}_T) = \sum_{j_1=1}^{h-n_b+1} \ldots \sum_{j_{n_b}=j_{n_b-1}+1}^{h} \left(\prod_{j=1}^{n_b} p^{d_j}_{K+j,K+j} (1-p_{K+j,K+j}) \right).
\end{aligned}
$$

With these probabilities in place we can finally integrate out uncertainty about the future number of breaks:

$$
\begin{aligned}
p\left(R_{T+h}\right| S_T = K+1, \mathcal{Z}_T) &= \sum_{j=1}^{n_b+1} p\left(R_{T+h}\right| S_{T+h} = K+j, S_T = K+1, \mathcal{Z}_T) \\
\times p(S_{T+h} &= K+j|S_T = K+1, \mathcal{Z}_T).
\end{aligned} \tag{23}
$$

We refer to this as the composite-meta return distribution as it allows for past and future breaks, weighting the various scenarios by their respective probabilities according to our changepoint model. An investor who considers the uncertainty about the number of out of sample breaks but conditions on K historical (in-sample) breaks therefore solves

$$
\max_{\omega} \int u(W_{T+h}) p\left(R_{T+h}\right| S_T = K+1, \mathcal{Z}_T)\, dR_{T+h}. \tag{24}
$$

Notice that this expression does not restrict the number of future breaks (as long as n_b is set reasonably large), nor does it take the parameters as known. It does, however, take the number of historic breaks as fixed and also ignores uncertainty about the forecasting model itself. We next relax our assumptions about these points.

E. *Uncertainty about the number of historical breaks*

The predictive densities computed so far have conditioned on the number of in-sample breaks (K) by setting $S_T = K+1$. This is of course a simplification since the true number of in-sample breaks is unknown. To deal with this, we adopt a simple Bayesian model averaging method that computes the predictive density of returns as a weighted average of the predictive densities conditional on different numbers of historical (in-sample) breaks. For each choice of the number of breaks, k, and predictor variable, x, we get a model M_{k_x} with predictive density $p_x(R_{T+h}|S_T = k+1, \mathcal{Z}_T)$. Integrating over the number of breaks (but keeping the choice of predictor variable, x, fixed), the predictive density under the Bayesian model average is

$$
p_x(R_{T+h}|\mathcal{Z}_T) = \sum_{k_x=0}^{\bar{K}_x} p_x(R_{T+h}|S_T = k_x+1, \mathcal{Z}_T) p(M_{k_x}|\mathcal{Z}_T), \tag{25}
$$

where $\bar{K}_x$ is an upper limit on the largest number of in-sample breaks that is entertained. The weights used in the average are proportional to the posterior probability of model M_{k_x} and are hence given by the product of the prior for model $M_{k_x}, p(M_{k_x})$, and the marginal likelihood, $f(\mathcal{Z}_T | M_{k_x})$,

$$p(M_{k_x} | \mathcal{Z}_T) \propto f(\mathcal{Z}_T | M_{k_x}) \, p(M_{k_x}). \tag{26}$$

F. *Model uncertainty*

In addition to not knowing the parameters of a given return forecasting model and not knowing the potential number of historical breaks, it can reasonably be argued that investors do not know the true identity of the return model. This point has been emphasized by Pesaran and Timmermann (1995) and, more recently in a Bayesian setting, investigated by Avramov (2002) and Cremers (2002). The analysis of Avramov and Cremers treats model uncertainty by considering all possible combinations of a large range of predictor variables.

We follow this analysis by integrating across the two return prediction models considered in this study based on the dividend yield and the short interest rate. This is simply an illustration of how to handle model uncertainty and our analysis could of course be extended to a much larger set of variables. However, to keep computations feasible, we simply combine the return models based on these two predictor variables, in each case accounting for uncertainty about the number of past and future breaks:

$$p(R_{T+h}|\mathcal{Z}_T) = \sum_{x=1}^{\overline{X}} \sum_{k=0}^{\bar{K}_x} p_x(R_{T+h}|S_T = k_x + 1, \mathcal{Z}_T) p(M_{k_x}|\mathcal{Z}_T). \tag{27}$$

Here $p(M_{k_x}|\mathcal{Z}_T)$ is the posterior probability of the model with k breaks using x as the predictor variable and $\overline{X}$ is the number of different combinations of predictor variables used to forecast stock returns.

V. **Empirical Asset Allocation Results**

This section uses the methods from Section 4 to assess empirically the effect of structural breaks on a buy-and-hold investor's optimal asset allocation. We use the Gibbs sampler to evaluate the predictive distribution of returns under breaks. Details of the numerical procedure used to compute the distributions are provided in the Appendix.

Before moving to the results, it is worth recalling two important effects on asset allocation under predictability from variables such as the dividend yield. First, the dividend yield identifies a mean-reverting component in stock returns which means that the risk of stock returns grows more slowly than in the absence of predictability, creating a large hedging demand for stocks, c.f. Campbell, Chan, and Viceira (2003). Negative shocks to returns are bad news in the period when they occur but tend to increase subsequent values of the dividend yield and thus become associated with higher future expected stock returns. This effect is particularly important at long investment horizons. Second, parameter estimation uncertainty generally reduces a risk averse investor's demand for stocks. For example, if new information leads the investor to revise downward his belief about mean stock returns shortly after the investment decision is made, this will affect returns along the entire investment horizon in a similar way to a permanent negative dividend shock.

In our breakpoint model there is an interesting additional interaction between parameter estimation uncertainty and structural breaks. In the absence of breaks, parameter estimation uncertainty has a greater impact on returns in the sense that parameter values are fixed and not subject to change. The presence of breaks means that bad draws of the parameters of the return model will eventually cease to affect returns as they get replaced by new parameter values following future breaks. On the other hand, the presence of breaks to the parameters tends to lower the precision of current parameter estimates and thus increases the importance of parameter estimation uncertainty. Which effect dominates depends on the extent of the variability in the parameter values across regimes as well as on the average duration of the regimes.

A. *Results Based on the Dividend Yield*

Figures 3 and 4 plot the allocation to stocks under the four scenarios discussed in Section 4, namely (i) no breaks, ignoring parameter estimation uncertainty; (ii) no breaks, accounting for parameter estimation uncertainty; (iii) breaks in the historical sample, but no future breaks (so the last regime remains in effect over the course of the investment horizon); (iv) both past and future breaks allowed. The first two scenarios ignore breaks and so use full-sample parameter estimates. They correspond to the cases covered by Barberis (2000). We compute the optimal weight on stocks under two values for the coefficient of relative risk aversion, namely $\gamma = 5$ and $\gamma = 10$.

Figure 3 starts the dividend yield off from its value at the end of sample (2003:12) which is 1.5%. Under the models that assume no (past or future) breaks, the weight on stocks rises from a level near 10% at short investment horizons to 30% at the five-year horizon. The assumed absence of a break means that a very long data sample (1926-2003) is available for parameter estimation. This reduces parameter estimation uncertainty and leads to an increasing weight on stocks, the longer the investment horizon. This interpretation is confirmed by the finding that stock holdings are very similar irrespective of whether parameter estimation uncertainty is accounted for. In contrast, the model estimated under the parameters of the last regime−which acknowledges past breaks but ignores future breaks−implies a short-run allocation to stocks around 30% that is essentially independent of the investment horizon. In this case parameter estimation uncertainty is much larger due to the shortness of the data sample after the most recent break which occurred in 1992.[14] However, the greater risk of future stock returns due to parameter estimation uncertainty largely cancels out against the mean reversion in returns identified by the dividend yield. The higher allocation to stocks at short horizons compared to under no breaks can be explained by the higher mean stock return during the last regime than during the full sample.

[14] Differences observed between the stock holdings under this scenario and that under no breaks and no parameter estimation uncertainty are similar to the differences in asset allocation shown by Barberis (2000) for a short data sample (1985-1996).

Allowing for both past and future breaks−with the latter weighted by their probabilities computed under the assumed changepoint model−the weight on stocks starts out at 30% at the short horizon and declines to a level below 10% at the five-year horizon. Parameter instability and estimation uncertainty now dominates the hedging demand for stocks induced by return predictability from the dividend yield.

When the coefficient of risk aversion is increased from five to ten, the weight on stocks declines uniformly and stays well below 20% in all scenarios. This may seem low but is mainly driven by the assumed initial value of the dividend yield which, at 1.5%, is close to its historical minimum.

To demonstrate this point, Figure 4 shows the allocation to stocks when the parameters and the initial value of the dividend yield are set at their mean values during the regime prevailing from 1958-1974. In this regime, the values of the parameters and the dividend yield are closer to their overall average although, at 3.1%, the mean dividend yield is slightly below the historical average of 4%. Comparing Figures 3 and 4 it is clear that the level of the optimal stock holding can be quite sensitive to the initial value of the dividend yield. The allocation to stocks under the no-break model now starts at a level close to 40% at short investment horizons and increases to nearly 45% at the five-year horizon - or 50% if parameter estimation uncertainty is ignored. Stock holdings under the model that accounts for past breaks but ignores future breaks are now increasing in the horizon. Finally, the allocation to stocks under past and future breaks is non-monotonic in the investment horizon, first increasing from around 55% to 60% before declining to around 10% at the five-year horizon.

These findings suggest that the allocation to stocks is increasing in the horizon if the initial value of the dividend yield is very low and (past and future) breaks are ignored. If past breaks are accounted for but future breaks are ignored, the asset allocation can be flat or increasing as a function of the horizon. Finally, if both past and future breaks are modeled, we see a non-monotonic or sometimes strongly declining allocation to stocks, the longer the investment horizon.

Parameter instability can therefore have a larger effect on a buy-and-hold investor's optimal asset allocation than parameter estimation uncertainty. This can be seen by comparing the full sample (no break) plots in Figures 3 and 4 with and without estimation error. In both cases these are very similar. This is to be expected since investors have access to 75 years of data. In fact, the large effect of parameter estimation error documented by Barberis (2000) was found in a sample where the parameters were estimated using a much shorter data set from 1986 to 1995. Presumably investors would only want to use such a short sample if they thought that the parameters of the return model had changed over time.

B. *Results based on the Short Interest Rate*

Optimal stock holdings under the return prediction model based on the short rate are shown in Figures 5 and 6. First consider the results in Figure 5 when the short rate is set at its terminal value in 2003:12 (0.83%). The allocation to stocks is downward sloping as a function of the investment horizon irrespective of the assumed breakpoint scenario. There are two reasons for this. First, while shocks to the dividend yield and stock returns are strongly negatively correlated and thus give rise to a strong hedging demand for stocks, shocks to the short rate and stock returns are−on average−largely uncorrelated, c.f. Table 4. Second, as was the case for the dividend yield, the T-bill rate ended up far below its historical average in 2003:12. However, since the coefficient on the T-bill rate in the return prediction model is negative, this raises the expected stock return, whereas the low terminal value of the dividend yield reduces the expected stock return.

We continue to find that the level and slope of the stock holdings as a function of the investment horizon are sensitive to assumptions about breaks. Under the assumption of no breaks (but accounting for parameter estimation uncertainty), the weight on stocks starts from a level near 70% and declines to around 50% at the five year horizon. The allocation is only marginally higher if parameter estimation uncertainty is ignored.

Suppose next that we consider historical breaks but ignore future breaks. The far greater uncertainty about current parameter values under the return model based on the last regime (which uses 13 years of data after 1990) means that the allocation to stocks drops more precipitously from a level near 90% at the shortest investment horizon to 40% at the longest horizon. The drop is even sharper under the model that allows for both past and future breaks. This model sees the weight on stocks decline from nearly 100% to a level around 10% at the five year horizon, suggesting that the possibility of breaks has an even greater impact on the optimal allocation in the return prediction model based on the T-bill rate than in the model based on the dividend yield. This is a reflection of the greater dispersion of parameter estimates found for this model, as can be seen by comparing Tables 4 and 5 to Tables 2 and 3.

Once again we computed allocations under a higher coefficient of risk aversion ($\gamma = 10$). Stock holdings continue to slope downwards as a function of the investment horizon and decrease by roughly half compared to their level when $\gamma = 5$.

To account for the fact that the T-bill rate at the end of our sample (0.83%) is unusually low, we consider the asset allocations in a second regime with state and parameter values closer to the overall sample mean. Figure 6 shows the allocation to stocks under the parameters and mean interest rate value prevailing during the regime from 1952-1968. The allocation to stocks is now flat under the no-break models (with and without parameter estimation uncertainty) but continues to decline under the models that allow for breaks, particularly when both past and future breaks are considered. This suggests again that model instability is more important to the asset allocation decision than parameter estimation uncertainty.

C. *Uncertainty about the number of historical breaks*

Return models that allow for breaks include a larger number of parameters than the conventional full-sample model so one might be concerned that they overfits the data. We do not believe that this is a particular cause for concern since the number of breaks is selected using a criterion (posterior odds) that is known to penalize large models very heavily. In addition, by using a meta distribution that characterizes commonalities in the model parameters across different regimes, effectively the parameters are being shrunk towards a common prior which tends to reduce the effect of parameter estimation uncertainty.

One way to address this issue is to integrate out uncertainty about the number of historical (in-sample) breaks. The effect of using this approach is illustrated in Figure 7. This figure compares plots of optimal stock holdings under a forecasting model that assumes seven historical breaks and under Bayesian Model Averaging which considers between zero and eight breaks. The plots assume identical prior weights on each of the models in Table 1. The two sets of allocations are quite similar, particularly for the dividend yield model (left panel). The reason is easy to explain from Table 1: Nearly 70% of the total probability is allocated to a model with seven breaks while the remainder is allocated to the model with eight breaks whose break dates and parameter estimates are very similar to those for the model with seven breaks. Turning to the return equation based on the T-bill rate (right panel), the model with seven breaks gets a somewhat smaller weight, with the remainder going to models with five or eight breaks. Overall, however, the effect on asset allocations of conditioning on seven historical breaks is quite small.

D. *Model Uncertainty*

Figure 8 shows the results of accounting for model uncertainty in a simple experiment that combines the return forecasting models with up to seven breaks and includes either the dividend yield or the T-bill rate as a predictor variable, weighting these according to equation (27). Predictive densities across 18 different models are considered here (namely two predictor variables, each with between zero and eight breaks). Optimal stock holdings most resemble the allocation under the forecasting model based on the T-bill rate. It is easy to understand why: The return forecasting model based on the T-bill rate gives a better fit than the model based on the dividend yield irrespective of the number of breaks and hence gets a greater weight in the forecast combination.

VI. **Asset Allocation with Rebalancing**

Asset allocations reported so far considered a buy-and-hold investor who could not rebalance during the holding period. The assumption of no rebalancing serves the purpose of highlighting the importance of breaks to the return process for a long-term investor whose transaction costs are high. In practice, investors can rebalance broadly diversified stock portfolios at low costs (e.g. through futures contracts). We therefore next assume that the investor can adjust the portfolio weights every $\varphi = \frac{T}{B}$ months at B equally spaced points T, $T + \frac{T}{B}$, $T + 2\frac{T}{B}$, ..., $T + (B-1)\frac{T}{B}$. Setting $B = 1$ gives the buy-and-hold problem as a special case. For notational convenience, in the following we drop the subscript T, and simply refer to $T + b\frac{T}{B}$ as b.

Let ω_b, $b = 1, ..., B-1$, be the weights on the stock portfolio at the rebalancing times. Then $1 - \omega_b$ is the weight on the risk free asset at time $T + b\frac{T}{B}$ and the investor's optimization problem becomes

$$\max_{\{\omega_j\}_{j=0}^{B-1}} E_b \left[\frac{W_B^{1-\gamma}}{1-\gamma} \right] \tag{28}$$

$$s.t.\ W_{b+1} = W_b \left\{ (1-\omega_b) \exp\left(\varphi r^f\right) + \omega_b \exp\left(\varphi r^f + R_{b+1}\right) \right\}.$$

The derived utility of wealth is defined by

$$J\left(W_b, r_b, x_b, \Upsilon_b, \pi_b, T_b\right) \equiv \max_{\{\omega_j\}_{j=b}^{B-1}} E_b \left[\frac{W_B^{1-\gamma}}{1-\gamma} \right],$$

where r_b and x_b are the excess return and predictor variable at time T_b, Υ_b is the collection of model parameters, including the regime specific parameters, the meta hyperparameters and the elements of the transition probability matrix. Finally, π_b is the probability of remaining in the last historical regime at time T_b (i.e. the probability of not having observed a break in the out-of-sample period between T and T_b.)

Under power utility this expression simplifies to

$$J\left(W_b, r_b, x_b, \Upsilon_b, \pi_b, T_b\right) = \frac{W_b^{1-\gamma}}{1-\gamma} Q\left(r_b, x_b, \Upsilon_b, \pi_b, T_b\right).$$

We solve the investor's dynamic asset allocation problem under structural breaks and rebalancing by means of Monte Carlo simulation methods. Suppose the investor's allocation problem has been solved backward at the rebalancing points $B-1, B-2, ..., b+1$ so that $Q\left(r_b^i, x_b^i, \Upsilon_b, \pi_b^i, T_b\right)$ is known for all values $i = 1, 2, ..., G_x \times G_\pi$, where G_x is the number of grid points for the predictor variable x and G_π is the number of grid points used to evaluate the posterior probability $\Pr\left(S_{T_{b+1}} = K+1 \middle| S_{T_b} = K+1, \mathcal{Z}_{T_b}\right)$, i.e. the posterior probability of not observing a structural break between T_b and T_{b+1}. At each point $\left(x_b = x_b^i, \pi_b = \pi_b^i\right)$ we then find $Q\left(r_b^i, x_b^i, \Upsilon_b, \pi_b^i, T_b\right)$ by Monte Carlo integration of the expression

$$\max_{\omega_b} E_b \left[\left\{ (1-\omega_b) \exp\left(\varphi r^f\right) + \omega_b \exp\left(\varphi r^f + R_{b+1}\right) \right\}^{1-\gamma} Q\left(r_b^i, x_b^i, \Upsilon_b, \pi_b^i, T_b\right) \right].$$

Results from the analysis that accounts for rebalancing are displayed in Figure 9 for the dividend yield and in Figure 10 for the T-bill rate. Asset allocations under the no-break scenarios become marginally more flat as a function of the investment horizon independently of whether parameter estimation uncertainty is considered. Rebalancing has a much greater effect on the models that allow for breaks. Under the model specification that allows for both past and future breaks and predictability from the dividend yield, the allocation to stocks goes from being decreasing under no rebalancing (see Figure 4) to being marginally increasing under rebalancing. Rebalancing has such a large effect on the asset allocation under breaks since it provides an efficient way for investors to adjust their asset allocation in case a future adverse shock hits the parameters of the stock return equation. Still, the optimal asset allocations with and without breaks continue to be very different due to differences in the parameter estimates, thus showing that breaks have a very significant effect on asset allocations even under rebalancing.

Rebalancing also continues to have strong effects on a long-term investor's optimal asset allocation in the model with predictability of returns from the T-bill rate. This can be seen by comparing Figure 6 to Figure 10. Rather than being strongly downward sloping, the allocation to stocks now becomes mildly downward sloping in the model that considers past and future breaks.

A. *Learning Effects*

So far we ignored learning effects that arise when investors update their beliefs as captured by the state probabilities in the light of the arrival of future information. For a buy-and-hold investor this is not important since no decisions will be affected by future learning. However, under rebalancing, new information available at the rebalancing points can lead to important changes in the sequence of optimal portfolio choices and may affect the current investment decision. To account for future learning effects, let $\pi_{b+1} = \Pr\left(S_{T_{b+1}} = K+1 \middle| S_{T_b} = K+1, \mathcal{Z}_{T_{b+1}}\right)$ be the state probability at the first rebalancing point conditional on the initial information (at time T_b) as well as the new information arriving during the φ periods between time T_b and time T_{b+1}. Investors' learning can then be incorporated by letting them update their beliefs about the probability of observing structural breaks in the out-of-sample period using Bayes' rule at each point in time:

$$\pi_{b+1} = \frac{(\pi_b)^{\varphi} \times f\left(z_{T_b+1} + ... + z_{T_b+\varphi} \middle| S_{T_{b+1}} = S_{T_b} = K+1, \mathcal{Z}_{T_b}\right)}{f\left(z_{T_b+1} + ... + z_{T_b+\varphi} \middle| S_{T_b} = K+1, \mathcal{Z}_{T_b}\right)}. \tag{29}$$

Here $f\left(z_{T_b+1} + ... + z_{T_b+\varphi} \middle| S_{T_{b+1}} = S_{T_b} = K+1, \mathcal{Z}_{T_b}\right)$ is the predictive distribution under no breaks between T_b and T_{b+1} while $f\left(z_{T_b+1} + ... + z_{T_b+\varphi} \middle| S_{T_b} = K+1, \mathcal{Z}_{T_b}\right)$ is the composite predictive distribution integrating out uncertainty about the timing and the size of any breaks between T_b and T_{b+1}, both evaluated at $z_{T_{b+1}}$. If learning is ignored, the investor does not update the estimate of π_{b+1} from $\pi_b = \Pr\left(S_{T_{b+1}} = K+1 \middle| S_{T_b} = K+1, \mathcal{Z}_{T_b}\right)$. Differences between π_{b+1} and π_b therefore reflect learning effects due to the arrival of new information between T_b and T_{b+1}. Learning effects are important since optimal portfolio choices obviously depend not only on future values of asset returns and predictor variables, but also on future perceptions at the rebalancing points.

As a way to quantify the effect of learning on the optimal asset allocation, we compare in Figure 11 the allocation under the composite model that accounts for investors' updating of their state beliefs or ignores it. While the effect of accounting for future learning in the context of this model is smaller than, say, the effect of accounting for breaks in the first instance, it nevertheless has a systematic effect on the optimal asset allocation which is generally quite a bit lower under future learning. The reason for this finding is that learning makes asset returns riskier in the sense that negative shocks to the investment opportunity set make investors more pessimistic about the future and hence further decrease future returns. This effect is relatively minor at short investment horizons but becomes quite significant at longer horizons as can be seen from the figure.

B. *Welfare Costs of ignoring Breaks*

Our analysis so far suggests that structural instability in return forecasting models has a large impact on the investor's asset allocation. This does not, on its own, imply high expected utility costs from ignoring breaks, however. To investigate to which extent differences in asset allocations under different forecasting models translate into changes in expected utility, we follow Kandel and Stambaugh (1996) and compute the cost of ignoring breaks through the certainty equivalent return. This equates the expected utility of an investor that accounts for past and future breaks with that of an investor who either ignores breaks or allows for in-sample (historical) breaks but rules out future out-of-sample breaks. Our calculations allow for rebalancing every 12 months.

To see how this approach works, let the optimal weights of an $h-$period investor who believes in a particular model, say M_i corresponding to no breaks, be given by $\omega_{i,h}$. These weights give rise to the following certainty equivalent return:

$$\frac{\left[\exp(h \times cer_{i,h})\right]^{1-\gamma}}{1-\gamma} = E\left(\left. \frac{\left((1-\omega_{i,h})\exp(r_f h) + \omega_{i,h}\exp(r_f h + R_{T+h})\right)^{1-\gamma}}{1-\gamma} \right| p\left(\left. R_{T+h} \right| Y_T\right) \right),$$

where $p\left(\left. R_{T+h} \right| Y_T\right)$ is the predictive distribution over which future returns, R_{T+h}, is being evaluated. If model M_i is not correctly specified, this distribution will differ from $p\left(\left. R_{T+h} \right| M_i, Y_T\right)$. We next repeat the analysis under a different model, M_j, that in our case allows for past and future breaks. This gives rise to another set of portfolio weights, ω_h, and a certainty equivalent return $cer_{j,h}$:

$$\frac{\left[\exp(h \times cer_{j,h})\right]^{1-\gamma}}{1-\gamma} = E\left(\left. \frac{\left((1-\omega_{j,h})\exp(r_f h) + \omega_{j,h}\exp(r_f h + R_{T+h})\right)^{1-\gamma}}{1-\gamma} \right| p\left(\left. R_{T+h} \right| Y_T\right) \right).$$

Evaluating these expression under the assumption that the predictive return distribution $p\left(\left. R_{T+h} \right| Y_T\right)$ allows for breaks, we can compute the annualized differential certainty equivalence return simply as $1,200 \times (cer_{j,h} - cer_{i,h})$. For example, when model j allows for breaks while model i ignores breaks, this measures the cost in certainty equivalence terms of ignoring past and future breaks.

Results from this analysis are presented in Figure 12 in the form of the certainty equivalent measured in returns per annum. For the model based on the dividend yield, the certainty equivalent return lies between 0.1 and zero percent when computing the expected utility under past and future breaks versus under no breaks. However, the certainty equivalence return grows to 1.5 percent at short horizons under predictability from the T-bill rate, assuming a coefficient of risk aversion of five. When comparing the expected utility under past and future breaks against that under past breaks only, the picture is reversed as the certainty equivalence returns are highest (0.2-0.4%) under predictability from the yield and generally smaller under predictability from the T-bill rate.

VII. **Sensitivity Analysis and Extensions**

A. *Robustness to Priors*

To investigate the robustness of our empirical results with regard to the assumed priors, we conducted a sensitivity analysis. The greatest sensitivity of our results is related to the specification of $\underline{V_\beta}$. This matrix controls variations in the regression coefficients across regimes. In the basic results, we set $\underline{V_\beta} = sd \times I_{m\times k}$ with $sd = 10$. We experimented with different values of $\underline{V_\beta}$ and found that the variation in the parameters across regimes under the non-hierarchical model is preserved when the diagonal elements of $\underline{V_\beta}$ exceed the values chosen here. Our choice of $\underline{V_\beta}$ thus respects the variation in the original coefficient estimates and allows us to have reasonable meta distributions for the regression coefficients that can capture the values taken by the coefficients in the various regimes identified by our model.

Imposing constraints that the driving variable, x_t, is stationary ($0 < \beta_x < 1$) and that the unconditional mean of this variable is non-negative with symmetric constraints around its historical average ($0 \leq \mu_x/(1-\beta_x) \leq 0.08$ in the case of the dividend yield) did not have much effect on the results. Nor did imposing an additional parameter constraint which requires that the unconditional mean excess stock return within each regime lies between zero and 1% per month ($0 \leq \mu_r + \frac{\beta_r \mu_x}{1-\beta_x} \leq 0.01$) have much effect on the results.

B. *Dependence in Parameters Across Regimes*

So far we have assumed that location coefficient vector, $vec(B_j)$, the error term precision, $\psi_{j,i}^{-2}$, and correlation elements, $\lambda_{j,ic}$, in each regime j are independent draws from common distributions. However, it is possible that these parameters may be correlated across regimes, so we consider a specification that allows for autoregressive dynamics in the scale parameters across neighboring regimes, $vec(B_j)_i \sim (\mu_i + \rho_i vec(B_{j-1})_i, \sigma^2_{\eta,i})$, $i = 1, .., m^2$. Once again we adopt a hierarchical prior for the regime coefficients $\{B_j, \psi_{j,i}^{-2}, \lambda_{j,ic}\}$ so the regime-specific coefficients, $vec(B_j)$, $j = 1, ..., K+1$ are correlated across regimes, i.e. $vec(B_j) \sim (\boldsymbol{\mu} + \boldsymbol{\rho} vec(B_{j-1}), \boldsymbol{\Sigma}_\eta)$, where $\boldsymbol{\rho}$ is a diagonal matrix, while the error term precisions $\psi_{j,i}^{-2} \sim Gamma\,(v_{0,i}, d_{0,i})$ are independently and identical (IID) draws from a Gamma distribution and the correlation elements, $\lambda_{j,ic}$, are IID draws from a normal distribution, $\lambda_{j,ic} \sim N(\mu_{\rho,ic}, \sigma^2_{\rho,ic})$. At the next level of the hierarchy we assume that

$$(\boldsymbol{\mu}, \boldsymbol{\rho}) \sim N\left(\underline{\boldsymbol{\mu}_\beta}, \underline{\boldsymbol{\Sigma}_\beta}\right) \tag{30}$$

$$\boldsymbol{\Sigma}_\eta^{-1} \sim W\left(\underline{v_\beta}, \underline{\mathbf{V}_\beta^{-1}}\right), \tag{31}$$

where $W(.)$ is the Wishart distribution and $\underline{\boldsymbol{\mu}_\beta}$, $\underline{\boldsymbol{\Sigma}_\beta}$, $\underline{v_\beta}$ and $\underline{\mathbf{V}_\beta^{-1}}$ are hyperparameters that need to be specified a priori. We continue to assume that the hyperparameters of the error term precision $v_{0,i}$ and $d_{0,i}$ follow exponential and Gamma distributions, c.f. (8)-(9), while the hyperparameters of the correlation matrix follow a Normal-Inverted gamma distribution, c.f. (10)-(11).

Results for the Dividend Yield model are shown in Table 6, which should be compared to Table 3. There is only mild evidence of dependence in the location parameters across different regimes, with persistence parameters of the return and predictor variable equations that have means below one-half and are less than two standard deviations away from zero.

C. *Time-varying Volatility of Returns*

It is a well-known empirical fact that the volatility of stock returns varies over time. Ideally this should be captured by a return forecasting model used for asset allocation. In fact, since our model allows for breaks to the covariance matrix of returns, it is capable of accounting for heteroskedasticity in returns insofar as this coincides with the identified regimes. This is an important consideration since stock market returns were clearly far more volatile during periods such as the Great Depression.

To see how the volatility of stock returns changes over time in our model, Figure 13 provides a time-series plot of the standard deviation of the predictive density of returns. Since the standard deviation of returns (and of the yield) is allowed to vary across regimes in the break model, volatility follows a step function that tracks the various regimes. In fact, the mean value of the standard deviation of returns varies significantly from a level around 10% around the Great Depression to a level near 3-4% in the middle of the sample. This means that the asset allocations we computed earlier account not simply for shifts to the conditional equity premium but, equally importantly, also for shifts to the volatility of stock market returns.

D. *Sharpe Ratios Within Each of the Regimes*

Under the Bayesian approach adopted in this paper, the parameters are random variables. We can therefore consider the distribution of functions of these parameters such as the Sharpe ratio of returns within each regime. To this end we plot in Figure 14 the posterior predictive distributions of the 12-month Sharpe ratios in each of the eight regimes identified by our model. These plots set the predictor variable (in this case the dividend yield) at its mean value within a given regime. This ensures that the plots provide the typical distribution of Sharpe ratios emerging within a given regime. With one exception, the Sharpe ratios generally have a positive mean centered between 0.1 and 1 with very little probability mass on negative values. These values appear quite sensible, and reveal considerable variation in the Sharpe ratio over the almost eighty years covered by our sample.

The one regime (1927-1932) that generates a negative mean for the Sharpe ratio is dominated by the effects of the Great Depression. It is not surprising that returns behaved quite differently during this regime and that the posterior mean was negative for this regime. This observation does not violate the basic premise that ex-ante expected risk premia be non-negative since the ex-ante expected return from the meta distribution is positive. Another way of saying this is that although our model can be used *ex post* to document regimes with negative Sharpe ratios, the emergence of such regimes cannot be predicted in advance.

VIII. **Conclusion**

Optimal asset allocations are always derived contingent upon a set of assumptions about investors' knowledge of the underlying return forecasting model (model uncertainty), its parameters (parameter uncertainty) and its stability (structural breaks). Kandel and Stambaugh (1996) and Barberis (2000) pioneered the analysis of the effect of parameter estimation uncertainty on optimal asset allocations under predictability in returns. Subsequently, Avramov (2002) and Cremers (2002) extended their results to account for model uncertainty. In this paper we proposed a further step that allows for model instability. This is particularly relevant given the long data samples typically used to estimate the parameters of return prediction models and the sequence of institutional and technological changes witnessed in the twentieth century. Hence our analysis provides a method that accounts for

1. model uncertainty
2. parameter uncertainty
3. uncertainty about the number and size of historical (in-sample) breaks
4. uncertainty about future (out-of-sample) breaks

Our empirical results suggest, first, that the parameters of standard forecasting models appear to be highly unstable and subject to multiple shifts, many of which coincide with important historical events. Second, we find that, once such breaks are accounted for, the possibility of future breaks has a large impact on the optimal asset allocation of a Bayesian investor endowed with reasonable priors over the distribution from which new parameters are drawn following future breaks.

References

Aït-Sahalia, Y., and M. Brandt, 2001, Variable selection for portfolio choice, *Journal of Finance* 56, 1297–1351.

Ang, A., and G. Bekaert, 2002, International asset allocation with regime shifts, *Review of Financial Studies* 15, 1137–1187.

Ang, A., and J. Chen, 2002, Asymmetric correlations of equity portfolios, *Journal of Financial Economics* 63, 443–494.

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

Bai, J., R.L. Lumsdaine, and J.H. Stock, 1998, Testing for and dating common breaks in multivariate time series, *Review of Economic Studies* 65, 394–432.

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

Brandt, M., 1999, Estimating portfolio and consumption choice: A conditional euler equations approach, *Journal of Finance* 54, 1609–1645.

———, A. Goyal, P. Santa-Clara, and J. Stroud, 2002, A simulation approach to dynamic portfolio choice with an application to learning about predictability, *Review of Financial Studies* forthcoming.

Brennan, M., and Y. Xia, 2002, Dynamic asset allocation under inflation, *Journal of Finance* 41, 1201–1238.

Brennan, Michael J., Eduardo S. Schwartz, and Ronald Lagnado, 1997, Strategic asset allocation, *Journal of Economic Dynamics and Control* 21, 1377–1403.

Campbell, J., 1987, Stock returns and the term structure, *Journal of Financial Economics* 18, 373–399.

——— , Y. Chan, and L. Viceira, 2003, A multivariate model of strategic asset allocation, *Journal of Financial Economics* 67, 41–80.

Campbell, J., and R. Shiller, 1988, Stock prices, earnings, and expected dividends, *Journal of Finance* 43, 661–676.

Campbell, J., and L. Viceira, 1999, Consumption and portfolio decisions when expected returns are time varying, *Quarterly Journal of Economics* 114, 433–495.

——— , 2001, Who should buy long-term bonds?, *American Economic Review* 91, 99–127.

Chib, S., 1995, Marginal likelihood from the gibbs output, *Journal of the American Statistical Association* 90, 1313–1321.

——— , 1998, Estimation and comparison of multiple change point models, *Journal of Econometrics* 86, 221–241.

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

Fama, E. F., and K. R. French, 1988, Dividend yields and expected stock returns, *Journal of Financial Economics* 22, 3–25.

Farnsworth, H., W. Ferson, D. Jackson, and S. Todd, 2002, Performance evaluation with stochastic discount factors, *Journal of Business* pp. 473–504.

Ferson, W.E., and R. W. Schadt, 1996, Measuring fund strategy and performance in changing economic conditions, *Journal of Finance* 51, 425–461.

Ferson, Wayne E., and Campbell R. Harvey, 1991, The variation of economic risk premiums, *Journal of Political Economy* 99, 385–415.

George, E. I., U. E. Makov, and A. F. M. Smith, 1993, Conjugate likelihood distributions, *Scandinavian Journal of Statistics* 20, 147–156.

Guidolin, M., and A. Timmermann, 2004, Size and value anomalies under regime switching, Unpublished manuscript, Virginia and UCSD.

Harvey, Campbell, 1989, The world price of covariance risk, *Journal Finance* 46, 111–157.

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

Keim, D., and R. Stambaugh, 1986, Predicting returns in the stock and bond markets, *Journal of Financial Economics* 17, 357–390.

Lettau, Martin, and Sydney Ludvigsson, 2001, Consumption, aggregate wealth and expected stock returns, *Journal of Finance* 56, 815–849.

Lettau, M., S. Ludvigsson, and J. Wachter, 2004, The declining equity premium: What role does macroeconomic risk play?, mimeo, New York University.

Liechty, J. C., M. W. Liechty, and P. Müller, 2004, Bayesian correlation estimation, *Biometrika* 91-1, 1–14.

Pastor, L., and R. Stambaugh, 2001, The equity premium and structural breaks, *Journal of Finance* 56, 1207–1245.

Perez-Quiros, G., and A. Timmermann, 2000, Firm size and cyclical variations in stock returns, *Journal of Finance* 55, 1229–1262.

Pesaran, M. H., D. Pettenuzzo, and A. Timmermann, 2004, Forecasting time series subject to multiple break points, working paper University of Cambridge, Bocconi University and UCSD.

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

Stock, J.H., and M. W. Watson, 1996, Evidence on structural instability in macroeconomic time series relations, *Journal of Business and Economic Statistics* 14, 11–30.

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

Appendix: Gibbs Sampler for the Return Prediction Model with Multiple Breaks

This appendix extends results in Pesaran, Pettenuzzo, and Timmermann (2004) to cover multivariate dynamic models. We are interested in drawing from the posterior distribution $\pi\left(\Theta, H, p, \mathcal{S}_T \middle| \mathcal{Z}_T\right)$, where

$$\Theta = \left(vec(B)_1, \psi_1, R_1, ..., vec(B)_{K+1}, , \psi_{K+1}, \Lambda_{K+1}\right)$$

are the $K+1$ sets of regime-specific parameters (regression coefficients, error term variances and correlations) and

$$H = \left(b_0, V_0, v_{0,1}, d_{0,1}, ..., v_{0,m}, d_{0,m}, \mu_{\rho,12}, \sigma^2_{\rho,12}, ..., \mu_{\rho,m-1m}, \sigma^2_{\rho,m-1m}\right)$$

are the hyperparameters of the meta distribution that characterizes how much the parameters of the return model are allowed to vary across regimes. We also use the notation $\mathcal{S}_T = (s_1, ..., s_T)$ for the collection of values of the latent state variable and $\mathcal{Z}_T = (z_1, ..., z_T)'$ for the time-series of returns and predictor variables. Finally, $p = (p_{11}, p_{22,}..., p_{KK})'$ summarizes the unknown parameters of the transition probability matrix in (5).

The Gibbs sampler applied to our set up works as follows: First, states, $\mathcal{S}_T$, are simulated conditional on the data, $\mathcal{Z}_T$, the parameters, Θ, and meta hyperparameters, H; next, the parameters and hyperparameters of the meta distribution are simulated conditional on the data and $\mathcal{S}_T$. Specifically, the Gibbs sampler is implemented by simulating the following set of conditional distributions:

1. $\pi\left(\mathcal{S}_T \middle| \Theta, H, p, \mathcal{Z}_T\right)$
2. $\pi\left(\Theta, H \middle| p, \mathcal{S}_T, \mathcal{Z}_T\right)$
3. $\pi\left(p \middle| \mathcal{S}_T\right)$.

Here we used the identity $\pi\left(\Theta, H, p \middle| \mathcal{S}_T, \mathcal{Z}_T\right) = \pi\left(\Theta, H \middle| p, \mathcal{S}_T, \mathcal{Z}_T\right)\pi\left(p \middle| \mathcal{S}_T\right)$ and note that under our assumptions, $\pi\left(p \middle| \Theta, H, \mathcal{S}_T, \mathcal{Z}_T\right) = \pi\left(p \middle| \mathcal{S}_T\right)$.

Simulation of the states $\mathcal{S}_T$ requires 'forward' and 'backward' passes through the data. Define $\mathcal{S}_t = (s_1, ..., s_t)$ and $\mathcal{S}^{t+1} = (s_{t+1}, ..., s_T)$ as the state history up to time t and from time t to T, respectively. We partition the joint density of the states as follows:

$$p(s_{T-1}| s_T, \Theta, H, p, \mathcal{Z}_T) \times \cdots \times p(s_t| \mathcal{S}^{t+1}, \Theta, H, p, \mathcal{Z}_T) \times \cdots \times p(s_1| \mathcal{S}^2, \Theta, H, p, \mathcal{Z}_T). \tag{32}$$

Chib (1995) shows that the generic element of (32) can be decomposed as follows

$$p(s_t| \mathcal{S}^{t+1}, \Theta, H, p, \mathcal{Z}_T) \propto p(s_t| \Theta, H, p, \mathcal{Z}_T) p(s_t| s_{t-1}, \Theta, H, p), \tag{33}$$

where the normalizing constant is easily obtained since s_t takes only two values conditional on the value taken by s_{t+1}. The second term in (33) is simply the transition probability from the Markov chain. The first term can be computed by a recursive calculation (the forward pass through the data) where, for given $p(\,s_{t-1}|\,\Theta, H, p,\mathcal{Z}_{t-1})$, we obtain $p(\,s_t|\,\Theta, H, p,\mathcal{Z}_t)$ and $p(\,s_{t+1}|\,\Theta, H, p,\mathcal{Z}_{t+1})$, ..., $p(\,s_T|\,\Theta, H, p,\mathcal{Z}_t)$. Suppose $p(\,s_{t-1}|\,\Theta, H, p,\mathcal{Z}_{t-1})$ is available. Then

$$p(\,s_t = k|\,\mathcal{Z}_t, \Theta, H, p) = \frac{p(\,s_t = k|\,\mathcal{Z}_{t-1}, \Theta, H, p) \times f\,(\,z_t|\,vec(B)_k, \Sigma_k, \mathcal{Z}_{t-1})}{\sum\limits_{l=k-1}^{k} p(\,s_t = l|\,\Theta, H, p,\mathcal{Z}_{t-1}) \times f\,(\,z_t|\,vec(B)_l, \Sigma_l, \mathcal{Z}_{t-1})},$$

where, for $k = 1, 2, ..., K+1-$and recalling that p_{lk} is the Markov transition probability$-$

$$p(\,s_t = k|\,\Theta, H, p,\mathcal{Z}_{t-1}) = \sum_{l=k-1}^{k} p_{lk} \times p(\,s_{t-1} = l|\,\Theta, H, p,\mathcal{Z}_{t-1}).$$

For a given set of simulated states, $\mathcal{S}_T$, the data is partitioned into $K+1$ groups. Let $Z_j = \left(z'_{\tau_{j-1}+1}, ..., z'_{\tau_j}\right)'$ and $X_j = \left(z'_{\tau_{j-1}}, ..., z'_{\tau_j - 1}\right)'$ be the values of the dependent and independent variables within the jth regime. To obtain the conditional distributions for the regression parameters and hyperparameters, note that the conditional distributions of $vec(B)_j$ are independent across regimes with[15]

$$vec(B)_j|\,\Theta_{-vec(B)_j}, H, p, \mathcal{S}_T, \mathcal{Z}_T \sim N\left(\overline{vec(B)}_j, \overline{V}_j\right),$$

where

$$\begin{aligned}\overline{V}_j &= \left(X_j'\Sigma_j^{-1}X_j + V_0\right)^{-1}\\ \overline{vec(B)}_j &= \overline{V}_j\left(X_j'\Sigma_j^{-1}Z_j + V_0 b_0\right).\end{aligned}$$

The densities of the location and scale parameters of the meta distribution for the regression parameter, b_0 and V_0, take the form

$$\begin{aligned}b_0|\,\Theta, H_{-b_0}, p, \mathcal{S}_T, \mathcal{Z}_T &\sim N\left(\overline{\mu}_\beta, \overline{\Sigma}_\beta\right)\\ V_0|\,\Theta, H_{-V_0}, p, \mathcal{S}_T, \mathcal{Z}_T &\sim W\left(\overline{v}_\beta, \overline{V}_\beta^{-1}\right),\end{aligned}$$

where

$$\begin{aligned}\overline{\Sigma}_\beta &= \left(\underline{\Sigma}_\beta^{-1} + (K+1)\,V_0\right)^{-1}\\ \overline{\mu}_\beta &= \overline{\Sigma}_\beta\left(V_0\sum_{j=1}^{J} vec(B)_j + \underline{\Sigma}_\beta^{-1}\underline{\mu}_\beta\right),\end{aligned}$$

and

$$\begin{aligned}\overline{v}_\beta &= \underline{v}_\beta + (K+1)\\ \overline{V}_\beta &= \sum_{j=1}^{J}\left(vec(B)_j - b_0\right)\left(vec(B)_j - b_0\right)' + \underline{V}_\beta.\end{aligned}$$

[15] Using standard set notation we define A_{-b} as the complementary set of b in A, i.e. $A_{-b} = \{x \in A : x \neq b\}$.

Moving to the posterior for the precision parameters within each regime j and for each equation i, let $\Xi = (Z_j - X_j B_j)' (Z_j - X_j B_j)$ with Ξ_{ij} being its i-th row and j-th column element. Note that

$$s_{j,i}^{-2} \Big| \Theta_{-S_j}, H, p, \mathcal{S}_T, \mathcal{Z}_T \sim G\left(\frac{v_{0,i} + \Xi_{ii}}{2}, \frac{d_{0,i} + n_j}{2}\right),$$

where n_j is the number of observations assigned to regime j.

The location and scale parameters for the error term precision of each equation are then updated as follows:

$$v_{0,i} | \Theta, H_{-v_{0,i}}, p, \mathcal{S}_T, \mathcal{Z}_T \propto \prod_{j=1}^{K+1} G\left(s_{j,i}^{-2} \Big| v_{0,i}, d_{0,i}\right) \exp\left(v_{0,i} | \underline{\rho_{0,i}}\right) \tag{34}$$

$$d_{0,i} | \Theta, H_{-d_{0,i}}, p, \mathcal{S}_T, \mathcal{Z}_T \sim G\left(v_{0,i}(K+1) + \underline{c_{0,i}}, \sum_{j=1}^{K+1} s_{j,i}^{-2} + \underline{d_{0,i}}\right).$$

Drawing $v_{0,i}$ from (34) is complicated since we cannot make use of standard distributions. We therefore introduce a Metropolis-Hastings step in the Gibbs sampling algorithm. At each loop of the Gibbs sampling we draw a value $v_{0,i}^*$ from a Gamma distributed candidate generating density,

$$q\left(v_{0,i}^* | v_{0,i}^{g-1}\right) \sim Gamma\left(\varsigma, \varsigma / v_{0,i}^{g-1}\right).$$

This candidate generating density is centered on the last accepted value of $v_{0,i}$ in the chain, $v_{0,i}^{g-1}$, while the parameter ς defines the variance of the density and the rejection in the Metropolis-Hastings step. Higher values of ς mean a smaller variance for the candidate generating density and thus a smaller rejection rate. The acceptance probability is given by

$$\xi\left(v_{0,i}^* | v_{0,i}^{g-1}\right) = \min\left[\frac{\pi\left(v_{0,i}^* \Big| \Theta, H_{-v_{0,i}}, p, \mathcal{S}_T, \mathcal{Z}_T\right) / q\left(v_{0,i}^* \Big| v_{0,i}^{g-1}\right)}{\pi\left(v_0^{g-1} \Big| \Theta, H_{-v_{0,i}}, p, \mathcal{S}_T, \mathcal{Z}_T\right) / q\left(v_{0,i}^{g-1} \Big| v_{0,i}^*\right)}, 1\right]. \tag{35}$$

With probability $\xi\left(v_{0,i}^* \Big| v_{0,i}^{g-1}\right)$, the candidate value $v_{0,i}^*$ is accepted as the next value in the chain. Conversely, with probability $\left(1 - \xi\left(v_{0,i}^* \Big| v_{0,i}^{g-1}\right)\right)$ the chain remains at $v_{0,i}^{g-1}$. The acceptance ratio penalizes and rejects values of $v_{0,i}$ drawn from low posterior density areas.

Moving to the matrix of correlations within each regime, Λ_j, each element of these, $\lambda_{j,ic}$, is sampled independently from the other elements in Λ_j. Liechty, Liechty, and Müller (2004) show that−up to a proportionality constant−its distribution is

$$\begin{aligned} f\left(\lambda_{j,ic} | \Theta_{-R_j}, H, p, \mathcal{S}_T, \mathcal{Z}_T\right) \quad &\propto \quad |\Lambda_j|^{-m/2} \exp\left\{-tr\left(\Lambda_j^{-1} C_j\right)/2\right\} \\ &\quad \times \exp\left\{-\left(\lambda_{j,ic} - \mu_{\rho,ic}\right) / \left(2\sigma_{\rho,ic}^2\right)\right\} I\left\{R \in \mathcal{R}^m\right\}, \end{aligned} \tag{36}$$

where $I\{\cdot\}$ is the indicator function and C_j is the correlation matrix in regime j. The indicator function $I\{\Lambda \in \mathcal{R}^m\}$ ensures that the correlation matrix is positive definite and introduces dependence among the $\lambda_{j,ic}$-values.

The full conditional densities for $\mu_{\rho,ic}$ and $\sigma^2_{\rho,ic}$ are similar to the conjugate densities with an additional factor due to the constraint requiring Λ_j to be positive definite:

$$f\left(\mu_{\rho,ic}\middle|\,\Theta, H_{-\mu_{\rho,ic}}, p, \mathcal{S}_T, \mathcal{Z}_T\right) \propto \prod_{j=1}^{K+1} \exp\left\{-\left(\lambda_{j,ic}-\mu_{\rho,ic}\right)^2/\left(2\sigma^2_{\rho,ic}\right)\right\} \tag{37}$$
$$\times \exp\left\{-\left(\mu_{\rho,ic}-\underline{\mu_{\mu,ic}}\right)/\left(2\underline{\tau^2_{ic}}\right)\right\} I\left\{\Lambda \in \mathcal{R}^m\right\}$$

$$f\left(\sigma^2_{\rho,ic}\middle|\,\Theta, H_{-\sigma^2_{\rho,ic}}, p, \mathcal{S}_T, \mathcal{Z}_T\right) \propto \prod_{j=1}^{K+1} \exp\left\{-\left(r_{j,ic}-\mu_{\rho,ic}\right)^2/\left(2\sigma^2_{\rho,ic}\right)\right\} \tag{38}$$
$$\sigma_{\rho,ic}^{2\left(1-\underline{a_{\rho,ic}}\right)} \exp\left(-\underline{b_{\rho,ic}}/\sigma^2_{\rho,ic}\right) I\left\{R \in \mathcal{R}^m\right\}$$

The distributions of the correlation coefficients within each regime, $\lambda_{j,ic}$, and of the hyperparameters $\mu_{\rho,ic}$ and $\sigma^2_{\rho,ic}$ are not conjugate so sampling is accomplished using a Griddy Gibbs sampling step inside the main Gibbs sampling algorithm.

Finally, p is simulated from the conditional beta posterior

$$p_{jj}|\,\mathcal{S}_T \sim Beta(a + l_j, b + 1),$$

where $l_j = \tau_j - \tau_{j-1} - 1$ is the duration of regime j.

The distribution for the hyperparameters a and b is not conjugate so sampling is accomplished using a Metropolis-Hastings step. The conditional posterior distribution for a is

$$\pi\left(a|\,\Theta, H_{-b}, \mathcal{S}_T, p, \mathcal{Z}_T\right) \propto \prod_{j=1}^{K} Beta\left(p_{jj}|\,a, b\right) Gamma\left(a|\,\underline{a_0}, \underline{b_0}\right),$$

and similarly for b. To draw candidate values, we use a Gamma proposal distribution with shape parameter ς, mean equal to the previous draw a^g

$$q\left(a^*|\,a^g\right) \sim Gamma\left(\varsigma, \varsigma/a^g\right),$$

and acceptance probability

$$\xi\left(a^*|\,a^g\right) = \min\left[\frac{\pi\left(a^*|\,\Theta, H_{-b}, \mathcal{S}_T, p, \mathcal{Z}_T\right)/q\left(a^*|\,a^g\right)}{\pi\left(a^g|\,\Theta, H_{-b}, \mathcal{S}_T, p, \mathcal{Z}_T\right)/q\left(a^g|\,a^*\right)}, 1\right].$$

If no new breaks occur out-of-sample, we obtain a draw from $\pi\left(B_{K+1}, \Sigma_{K+1}|\,H, p, \mathcal{S}_T, \mathcal{Z}_T\right)$. Then, conditional on these parameters, we draw R_{T+h} from the posterior predictive density,

$$R_{T+h} \sim p\left(R_{T+h}|\,B_{K+1}, \Sigma_{K+1}, S_{T+h} = K+1, S_T = K+1, \mathcal{Z}_T\right). \tag{39}$$

With a single future break, we update the posterior distributions of b_0, V_0, $v_{0,1}$,$d_{0,1}$,$v_{0,2}$,$d_{0,2}$,$\mu_{\rho,12}$ and $\sigma^2_{\rho,12}$ as follows:

Draw b_0 from

$$b_0 \sim \pi\left(b_0|\,\Theta, H_{-b_0}, P, S_T, \mathcal{Z}_T\right),$$

and V_0 from

$$V_0 \sim \pi\left(V_0^{-1}\middle|\,\Theta, H_{-V_0}, P, S_T, \mathcal{Z}_T\right).$$

Draw $v_{0,1}$ and $v_{0,2}$ from

$$v_{0,i} \sim \pi\left(v_{0,i}\middle|\,\Theta, H_{-v_{0,i}}, P, \mathcal{S}_T, \mathcal{Z}_T\right),$$

and $d_{0,1}$ and $d_{0,2}$ from

$$d_{0,i} \sim \pi\left(d_{0,i}\middle|\,\Theta, H_{-d_{0,i}}, P, \mathcal{S}_T, \mathcal{Z}_T\right).$$

Draw $\mu_{\rho,12}$ from

$$\mu_{\rho,12} \sim \pi\left(\mu_{\rho,12}\middle|\,\Theta, H_{-\mu_{\rho,ic}}, P, \mathcal{S}_T, \mathcal{Z}_T\right),$$

and $\sigma^2_{\rho,12}$ from

$$\sigma^2_{\rho,12} \sim \pi\left(\sigma^2_{\rho,12}\middle|\,\Theta, H_{-\sigma^2_{\rho,ic}}, P, \mathcal{S}_T, \mathcal{Z}_T\right).$$

For a fixed set of hyperparameters, draw B_{K+2} and Σ_{K+2} from their respective priors given by $\pi\left(B_{K+2}\middle|\,b_0, V_0\right)$ and

$\pi\left(\Sigma_{K+2}\middle|\,v_{0,1}, d_{0,1}, v_{0,m}, d_{0,m}, \mu_{\rho,12}, \sigma^2_{\rho,12}\mathcal{Z}_T\right)$, respectively.

Draw R_{T+h} from the posterior predictive density,

$$R_{T+h} \sim p\left(R_{T+h}\middle|\,S_{T+h} = K+2, \tau_{K+1} = T+j, S_T = K+1, \mathcal{Z}_T\right). \tag{40}$$

To obtain the estimate of $p_{K+1,K+1}$ needed in equation (**??**), we combine information from the last regime with prior information, assuming the prior $p_{K+1,K+1} \sim Beta(a,b)$, so

$$p_{K+1,K+1}\middle|\,\mathcal{Z}_T \sim Beta(a + n_{K+1,K+1}, b+1) \tag{41}$$

where $n_{K+1,K+1}$ is the number of observations from regime $K+1$.

Turning to the case with an arbitrary number of future breaks, to draw from the distribution of the parameters a, b that characterize the break probability, we use that the conditional posterior distributions for a and b are

$$\begin{aligned}
\pi\left(a\middle|\,\Theta, H_{-b}, \mathcal{S}_T, P, \mathcal{Z}_T\right) &\propto \prod_{j=1}^{K} Beta\left(p_{jj}\middle|\,a,b\right) Gamma\left(a\middle|\,\underline{a_0}, \underline{b_0}\right) \\
\pi\left(b\middle|\,\Theta, H_{-b}, \mathcal{S}_T, P, \mathcal{Z}_T\right) &\propto \prod_{j=1}^{K} Beta\left(p_{jj}\middle|\,a,b\right) Gamma\left(b\middle|\,\underline{a_0}, \underline{b_0}\right).
\end{aligned}$$

Using these new posterior distributions, we generate draws for $p_{K+2,K+2}$ using the prior distribution for the p_{ii}'s and the resulting posterior densities for a and b,[16]

$$p_{K+2,K+2}\middle|\,a,b \sim Beta(a,b).$$

[16] Since we do not have any information about the length of regime $K+2$ from the estimation sample, we rely on prior information to get an estimate for $p_{K+2,K+2}$.

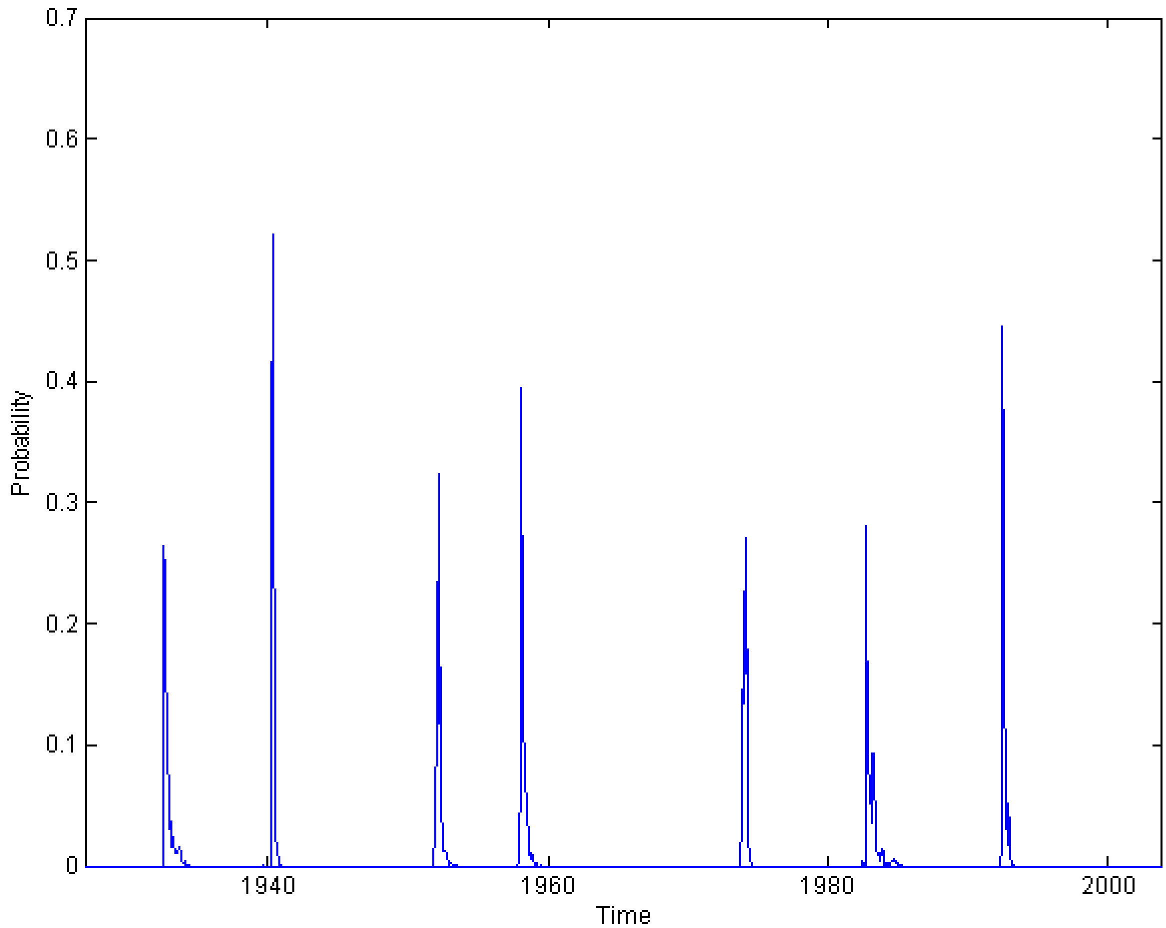

Figure 1: Posterior probabilities of breakpoint locations for the return prediction model with seven breaks based on the dividend yield. The estimation sample is 1926:12 - 2003:12.

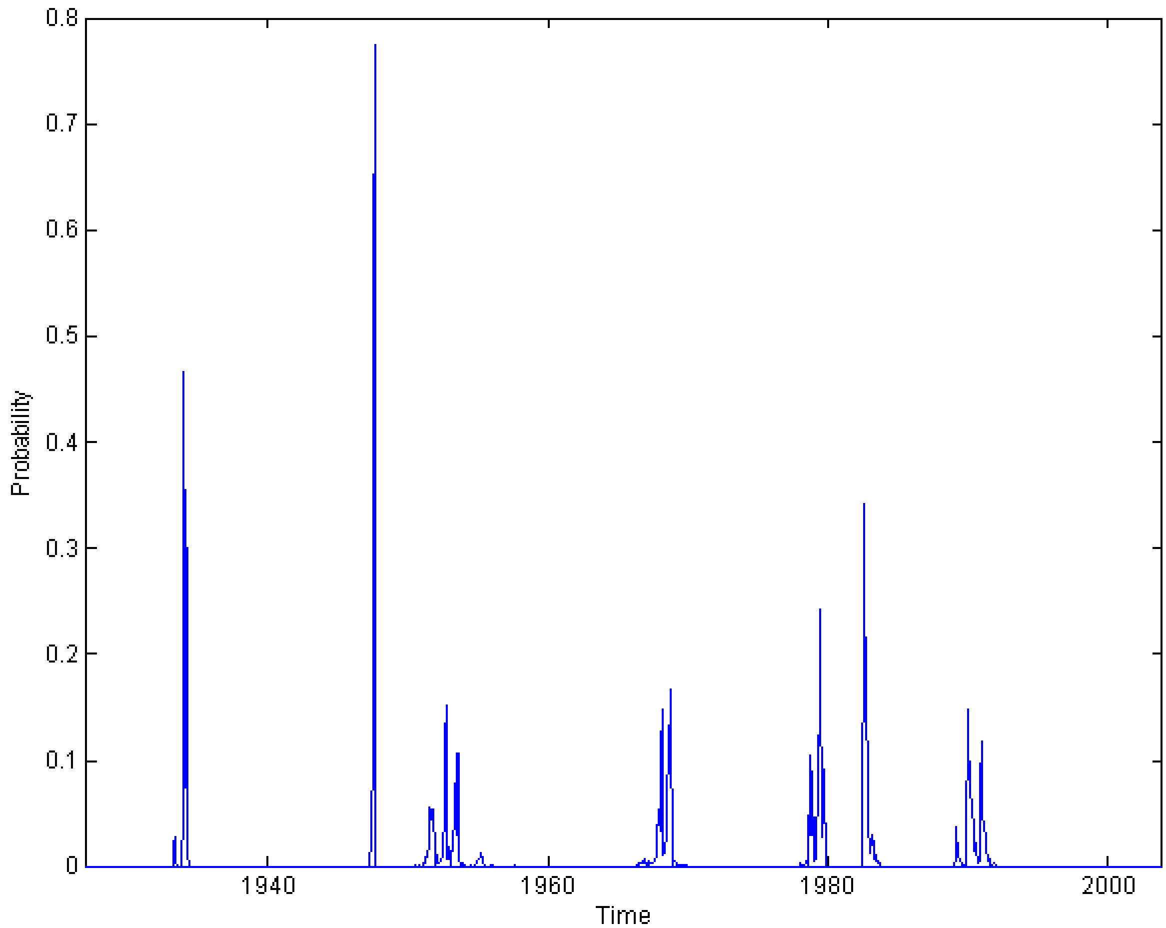

Figure 2: Posterior probabilities of breakpoint locations for the return prediction model with seven breaks based on the T-bill rate. The estimation sample is 1926:12 - 2003:12.

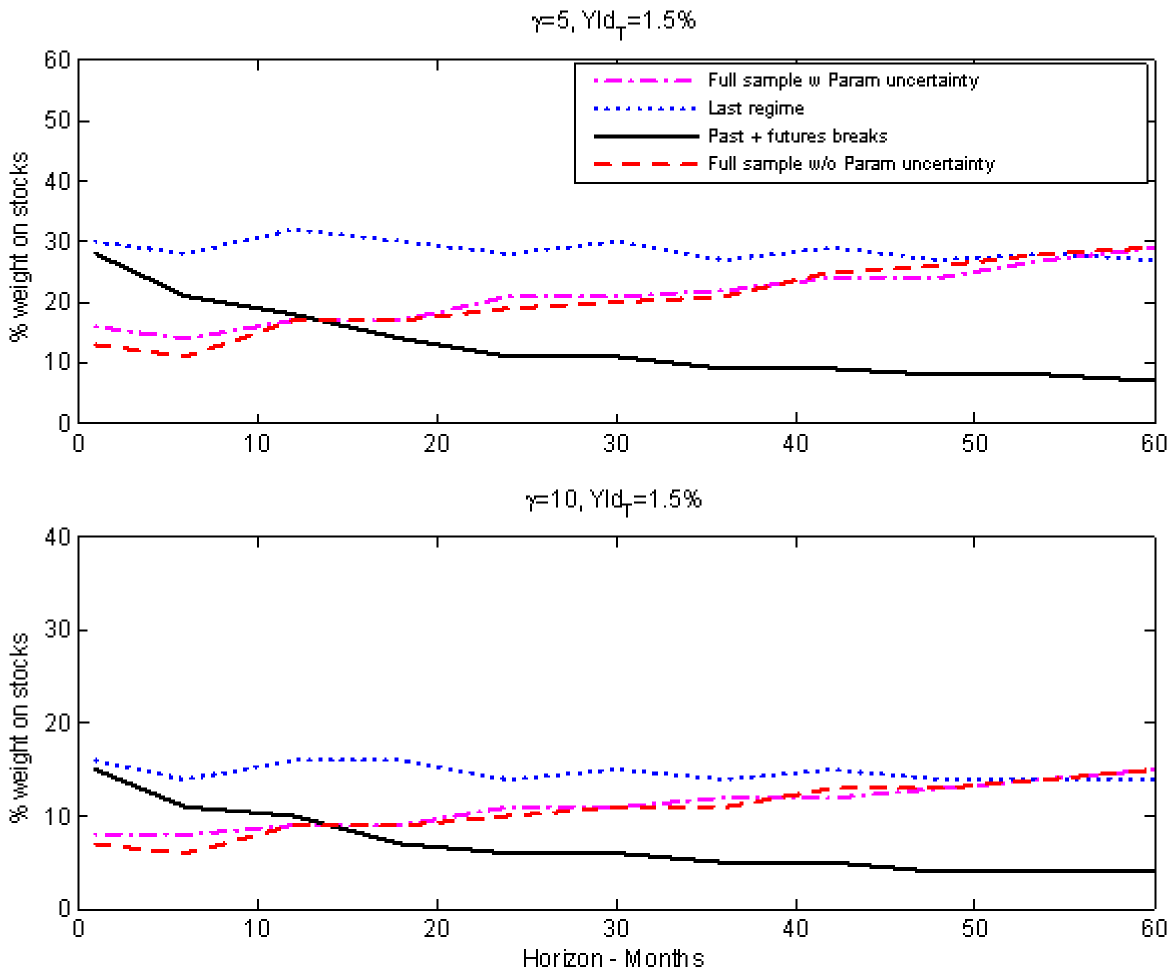

Figure 3: Optimal Asset Allocation as a function of the investment horizon for a buy-and-hold investor with power utility over terminal wealth, $U(W_{T+h}) = \frac{1}{(1-\gamma)} W_{T+h}^{1-\gamma}$, where h is the forecast horizon and γ is the coefficient of relative risk aversion. The panels show allocations to stocks under the assumption that the dividend yield is set at its value at the end of the sample, $Yld_T = 1.5\%$. The dotted line shows allocations starting from the regime at the end of the sample (2003:12). The dashed line shows the full-sample allocation ignoring breaks and parameter estimation uncertainty. The dashed/dotted line shows allocations based on full-sample parameter values (no breaks) but accounting for parameter uncertainty. Finally, the solid line shows allocations under the model that accounts for both past and future breaks.

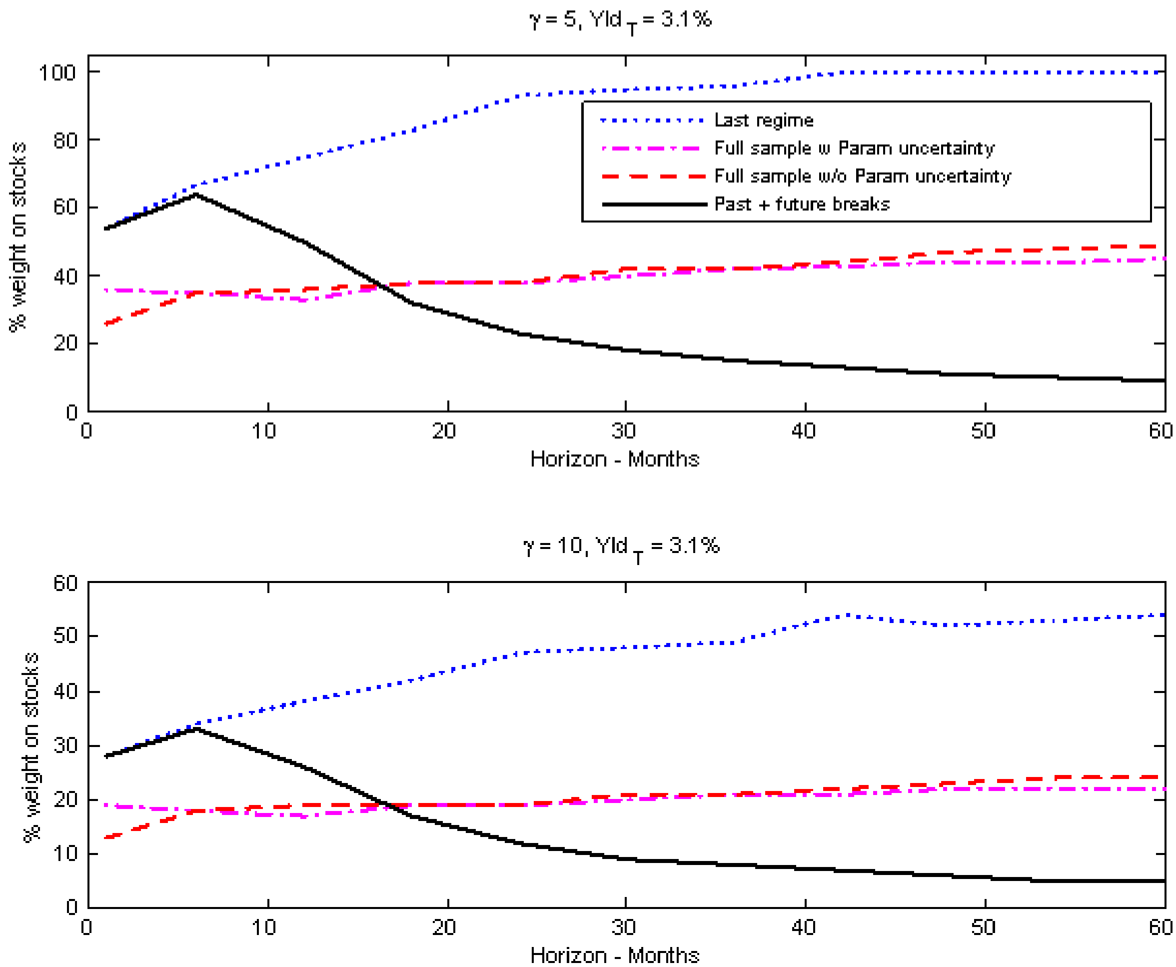

Figure 4: Optimal Asset Allocation as a function of the investment horizon for a buy-and-hold investor with power utility over terminal wealth, $U(W_{T+h}) = \frac{1}{(1-\gamma)} W_{T+h}^{1-\gamma}$, where h is the forecast horizon and γ is the coefficient of relative risk aversion. The panels show allocations to stocks under the assumption that the current regime parameters and the dividend yield are set at the values from the regime prevailing during 1958-1974. The dotted line shows allocations starting from the end of the 1952-1974 regime. The dashed line shows the full-sample allocation ignoring breaks and parameter estimation uncertainty. The dashed/dotted line shows allocations based on full-sample parameter values (no breaks) but accounting for parameter uncertainty. Finally, the solid line shows allocations under the model which accounts for both past and future breaks.

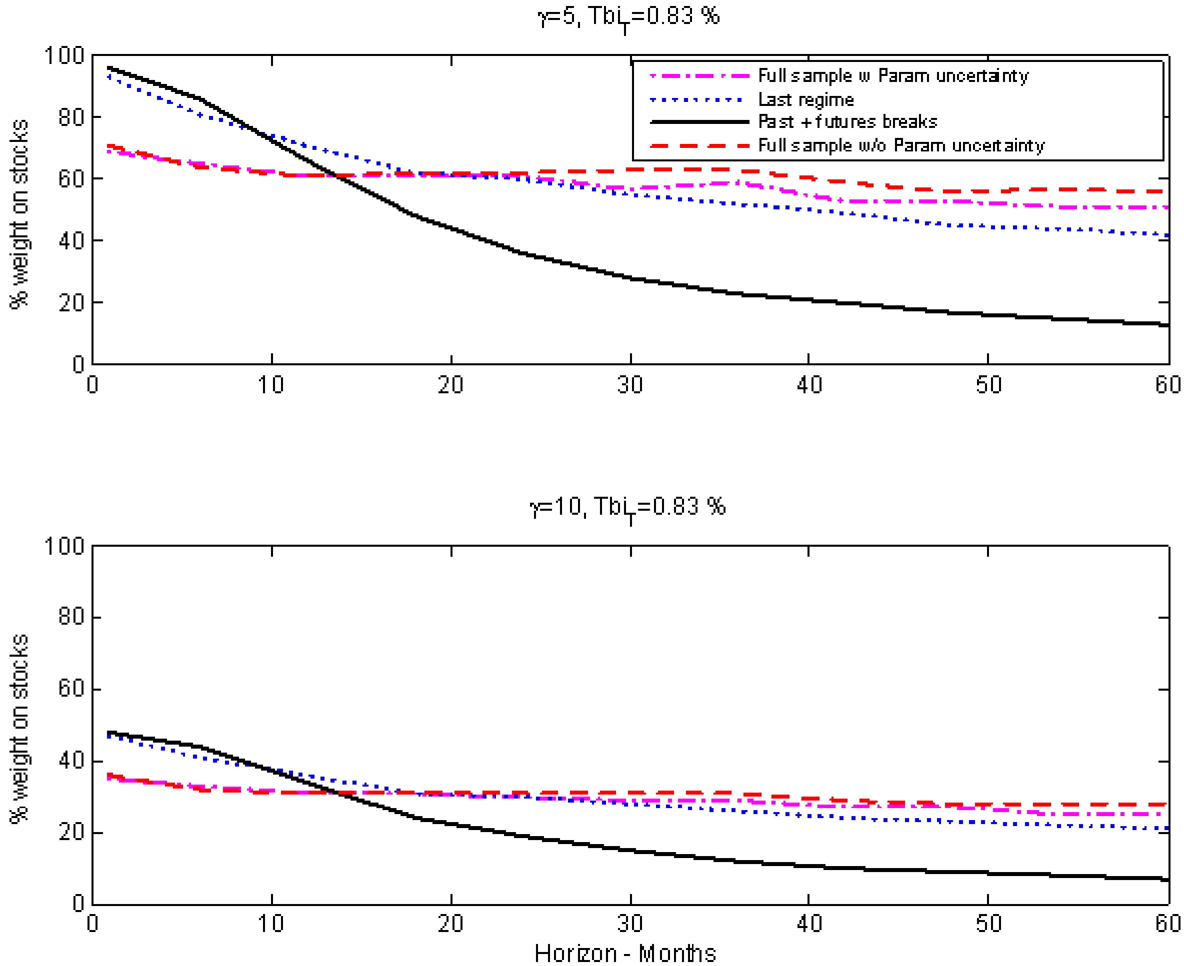

Figure 5: Optimal Asset Allocation as a function of the investment horizon for a buy-and-hold investor with power utility over terminal wealth, $U(W_{T+h}) = \frac{1}{(1-\gamma)} W_{T+h}^{1-\gamma}$, where h is the forecast horizon and γ is the coefficient of relative risk aversion. The panels show allocations to stocks under the assumption that the T-bill rate is set at its value at the end of the sample, $TB_T = 0.83\%$. The dotted line shows allocations starting from the regime at the end of the sample (2003:12). The dashed line shows the full-sample allocation ignoring breaks and parameter estimation uncertainty. The dashed/dotted line shows allocations based on full-sample parameter values (no breaks) but accounting for parameter uncertainty. Finally, the solid line shows allocations under the model that accounts for both past and future breaks.

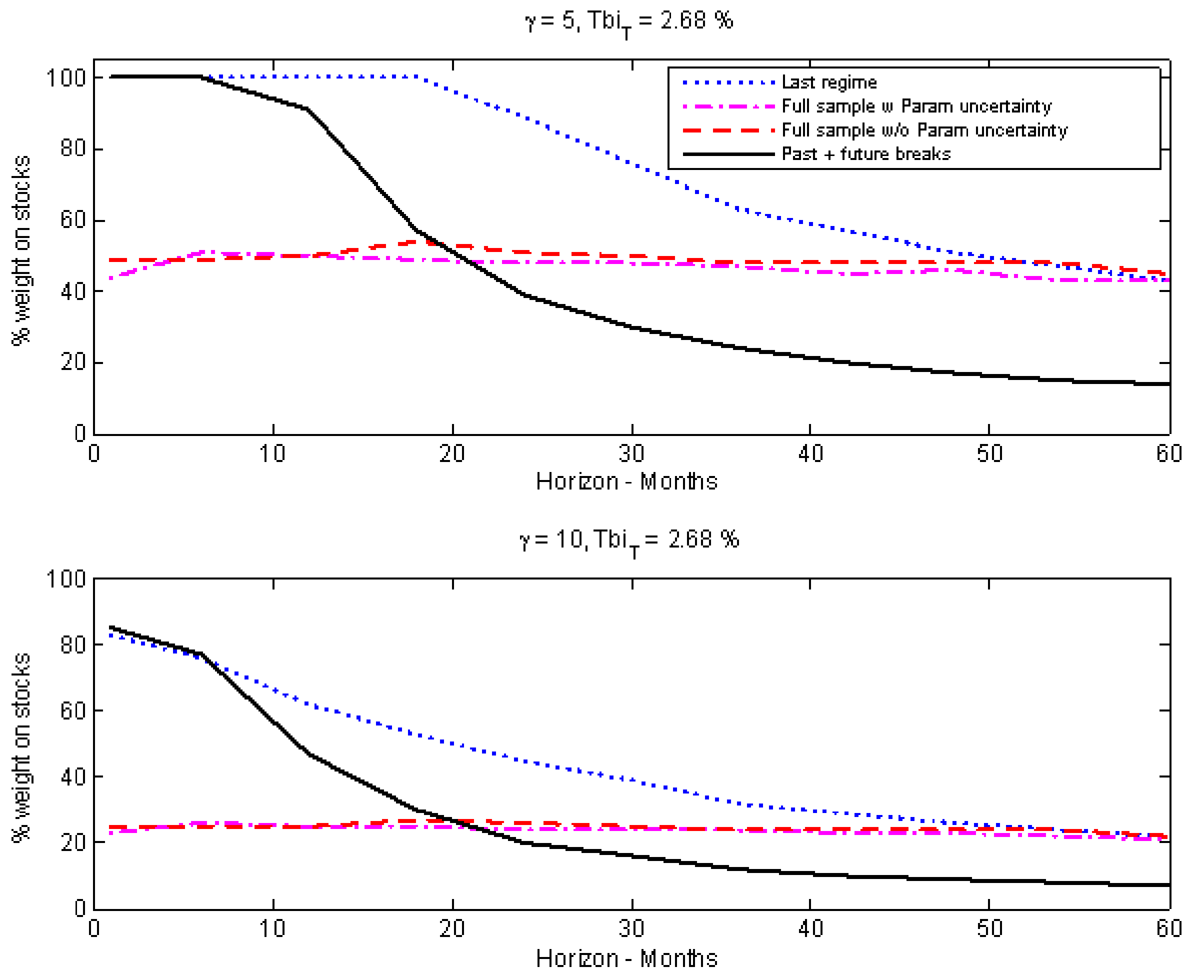

Figure 6: Optimal Asset Allocation as a function of the investment horizon for a buy-and-hold investor with power utility over terminal wealth, $U(W_{T+h}) = \frac{1}{(1-\gamma)}W_{T+h}^{1-\gamma}$, where h is the forecast horizon and γ is the coefficient of relative risk aversion. The panels show allocations to stocks under the assumption that the current regime parameters and the T-bill rate are set at the values from the regime prevailing during 1952-1968. The dotted line shows allocations starting from the end of the 1952-1968 regime. The dashed line shows the full-sample allocation ignoring breaks and parameter estimation uncertainty. The dashed/dotted line shows allocations based on full-sample parameter values (no breaks) but accounting for parameter uncertainty. Finally, the solid line shows allocations under the model that accounts for both past and future breaks.

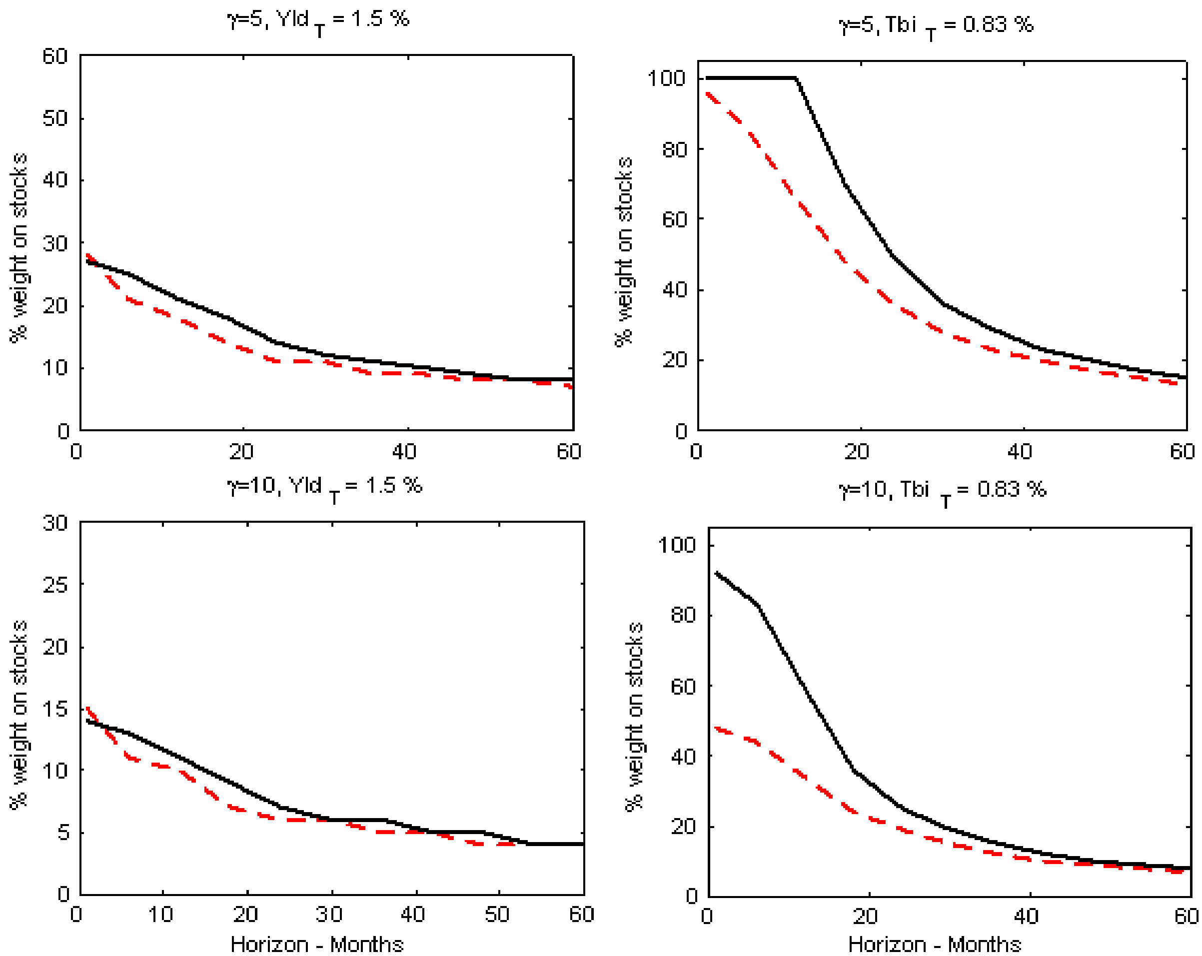

Figure 7: Stock allocations accounting for uncertainty about the number of in-sample breaks. The left panels show stock holdings that average across models based on the dividend yield as a predictor variable, using Bayesian Model Averaging to average across models with between zero and eight breaks. The right panels plot the allocations when averaging across models with between zero and eight breaks based on the T-bill rate as the predictor variable. In all panels, the dotted lines show stock holdings under the model with seven historical breaks, while the solid lines show allocation under Bayesian Model Averaging. In all cases the allocations allow for the possibility of future (out-of-sample) breaks.

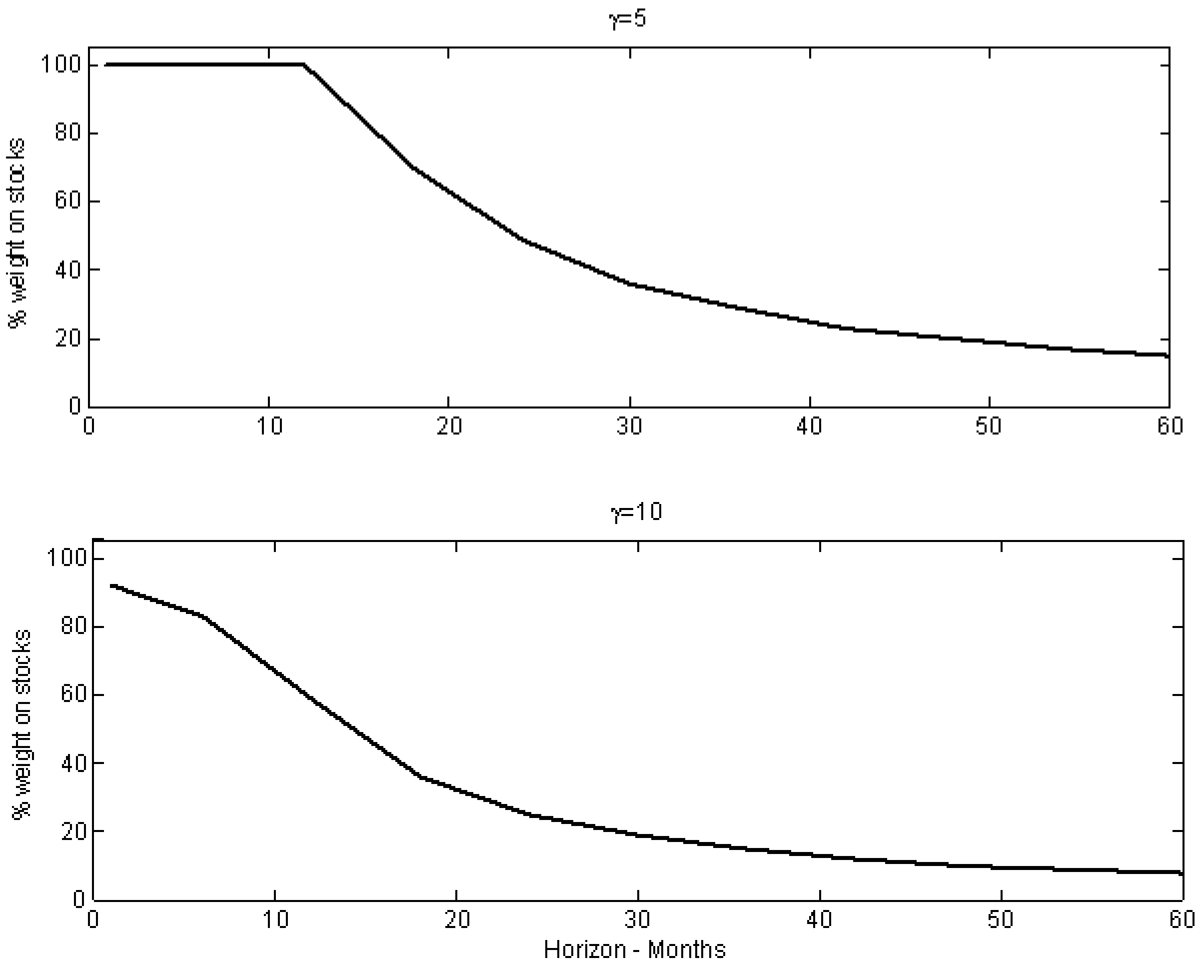

Figure 8: Stock allocations accounting for model uncertainty and uncertainty about the number of historical breaks. The panels show stock holdings that use Bayesian Model Averaging to integrate across models based either on the dividend yield or on the T-bill rate as predictor variables. Allocations are based on the composite model that accounts for both past and future breaks.

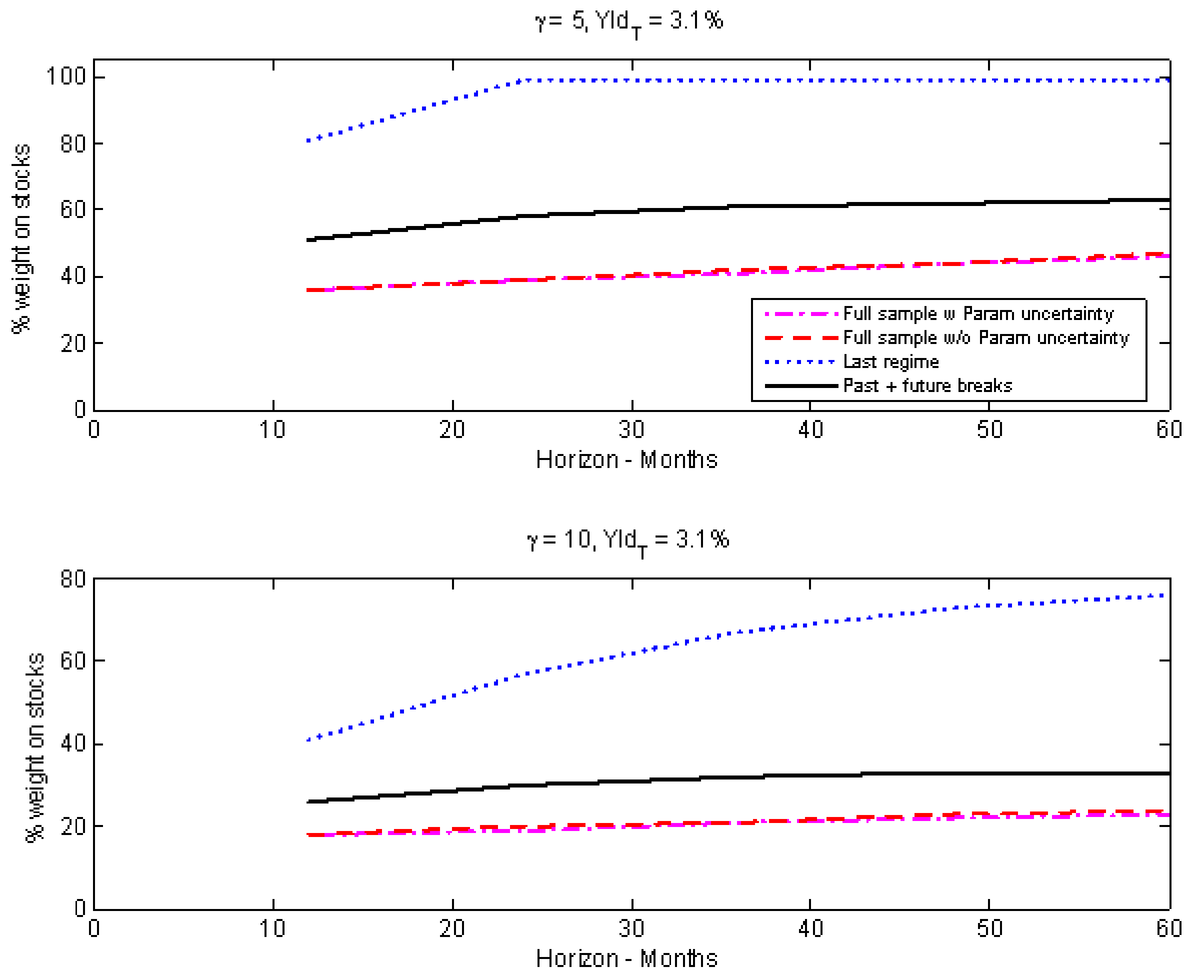

Figure 9: Optimal Asset Allocation as a function of the investment horizon for an investor who optimally rebalances every 12 months and has power utility over terminal wealth, $U(W_{T+h}) = \frac{1}{(1-\gamma)} W_{T+h}^{1-\gamma}$, where h is the forecast horizon and γ is the coefficient of relative risk aversion. The two panels show allocations to stocks under the assumption that the dividend yield is set at the average during the period 1958-1974, $Yld_T = 3.1\%$. The dotted line shows allocations starting from the end of the 1958-1974 regime. The dashed line shows the full-sample allocation ignoring breaks and parameter estimation uncertainty. The dashed/dotted line shows allocations based on full-sample parameter values (no breaks) but accounting for parameter uncertainty. Finally, the solid line shows allocations under the composite model which accounts for past and future breaks.

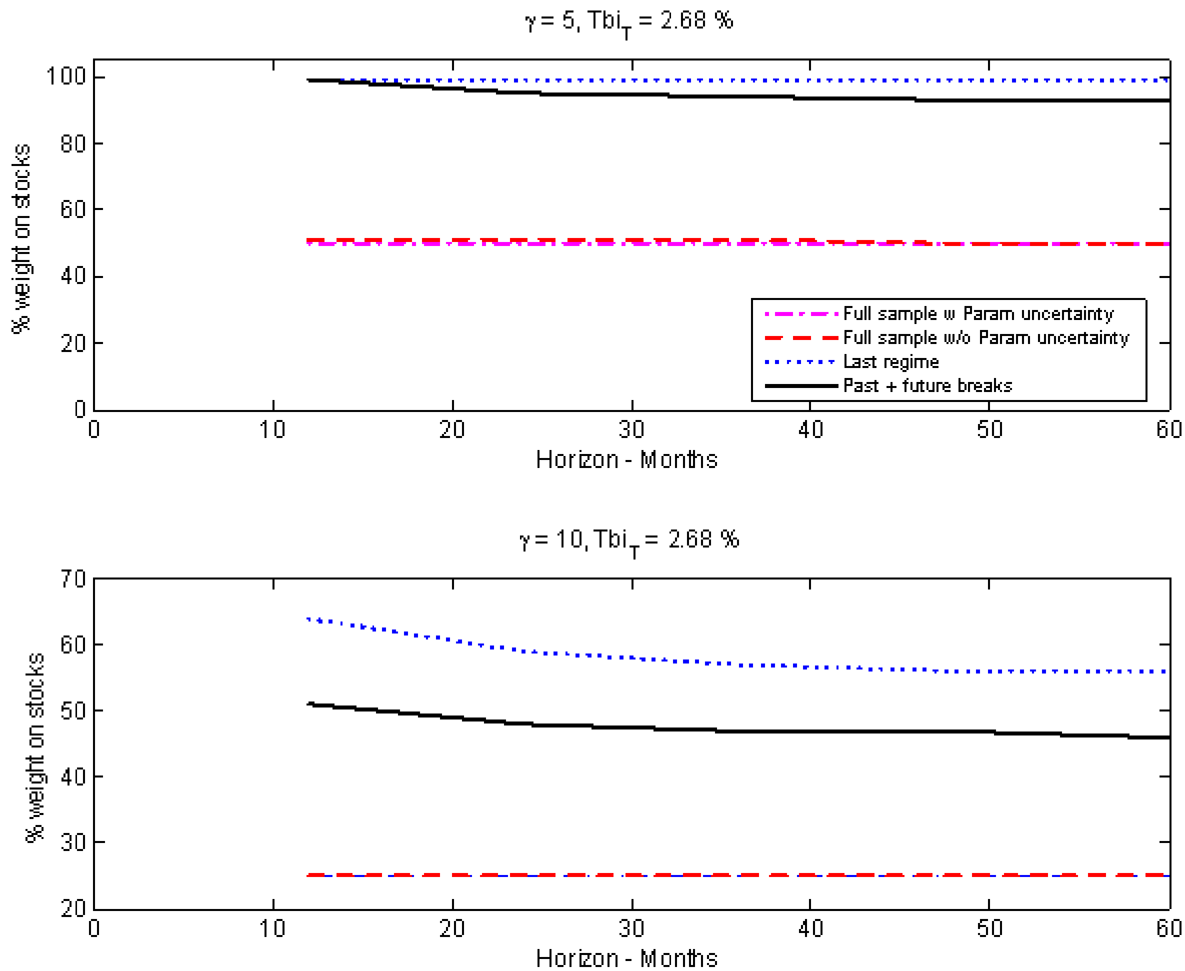

Figure 10: Optimal Asset Allocation as a function of the investment horizon for an investor who optimally rebalances every 12 months and has power utility over terminal wealth, $U(W_{T+h}) = \frac{1}{(1-\gamma)}W_{T+h}^{1-\gamma}$, where h is the forecast horizon and γ is the coefficient of relative risk aversion. The two panels show allocations to stocks under the assumption that the T-bill rate is set at the average during the period 1952-1968, $Tbi_T = 2.68\%$. The dotted line shows allocations starting from the end of the 1952-1968 regime. The dashed line shows the full-sample allocation ignoring breaks and parameter estimation uncertainty. The dashed/dotted line shows allocations based on full-sample parameter values (no breaks) but accounting for parameter uncertainty. Finally, the solid line shows allocations under the model which accounts for both past and future breaks.

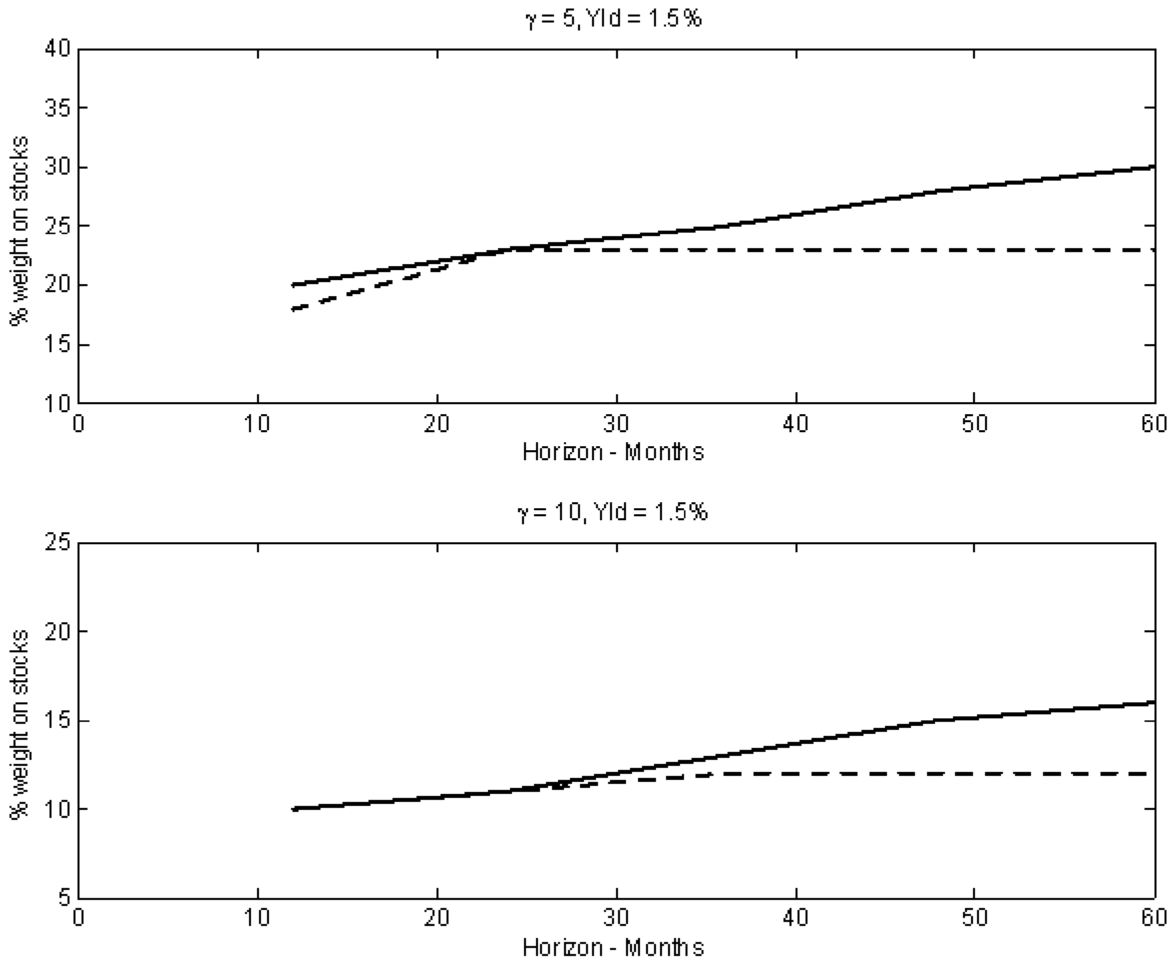

Figure 11: Optimal Asset Allocation as a function of the investment horizon for an investor who optimally rebalances every 12 months and learns about the probability of future break points. The investor is endowed with power utility over terminal wealth, $U(W_{T+h}) = \frac{1}{(1-A)} W_{T+h}^{1-A}$, where h is the forecast horizon and A is the coefficient of relative risk aversion. The panels show allocations to stocks under the assumption that the dividend yield is set at its value at the end of the sample, $Yld_T = 1.5\%$. The solid line shows allocations under the model which accounts for past and future breaks but ignores learning, while the dashed line shows allocations under this model and under learning.

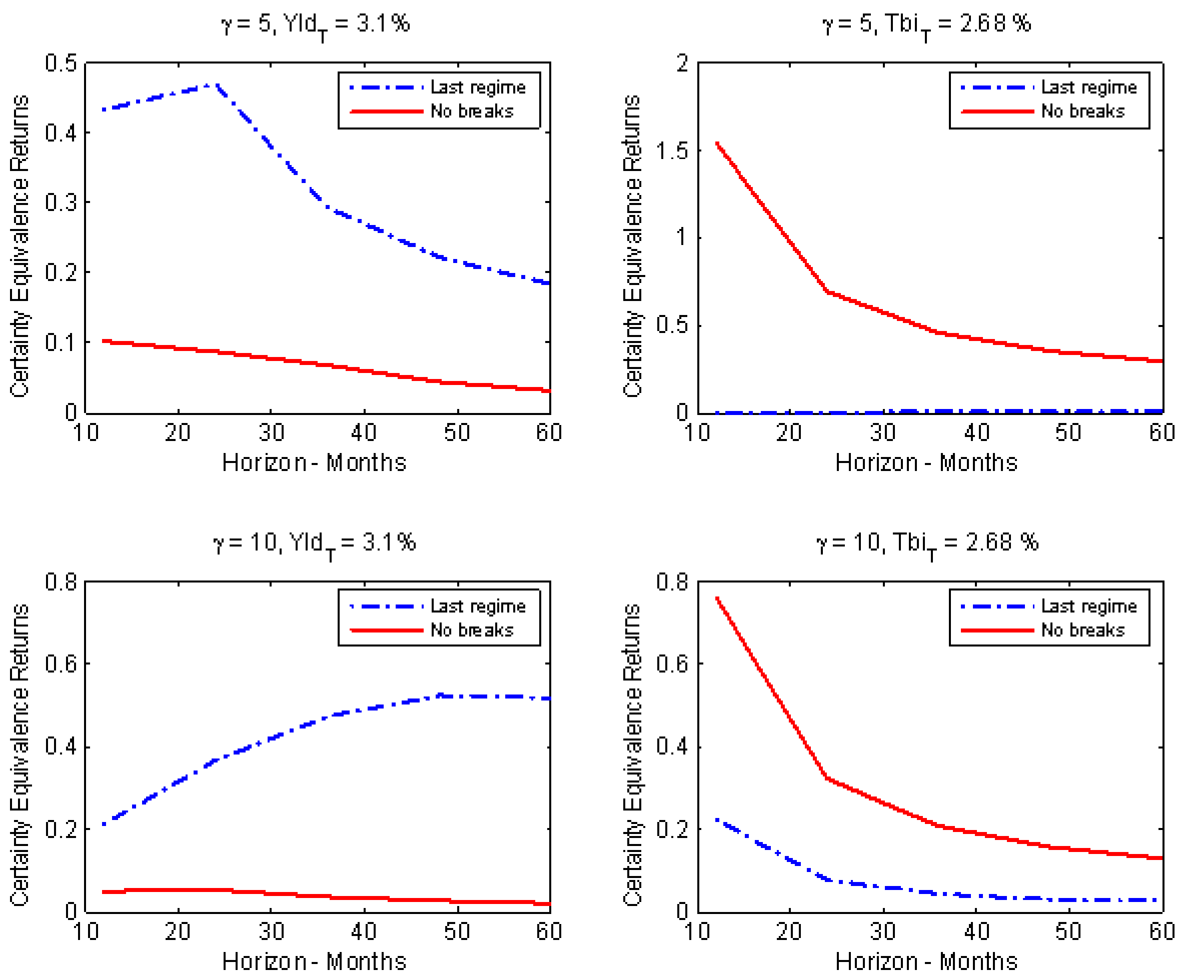

Figure 12: Certainty equivalence returns (in annualized percentage terms) under rebalancing as a function of the investment horizon for different levels of risk aversion, γ. For each panel, the solid line shows the difference in certainty equivalence returns between a model that allows for past and future breaks and a model that ignores breaks. The dashed/dotted line shows the difference in certainty equivalence returns between a model that allows for past and future breaks and a model based on the last regime that considers past breaks but ignores future breaks.

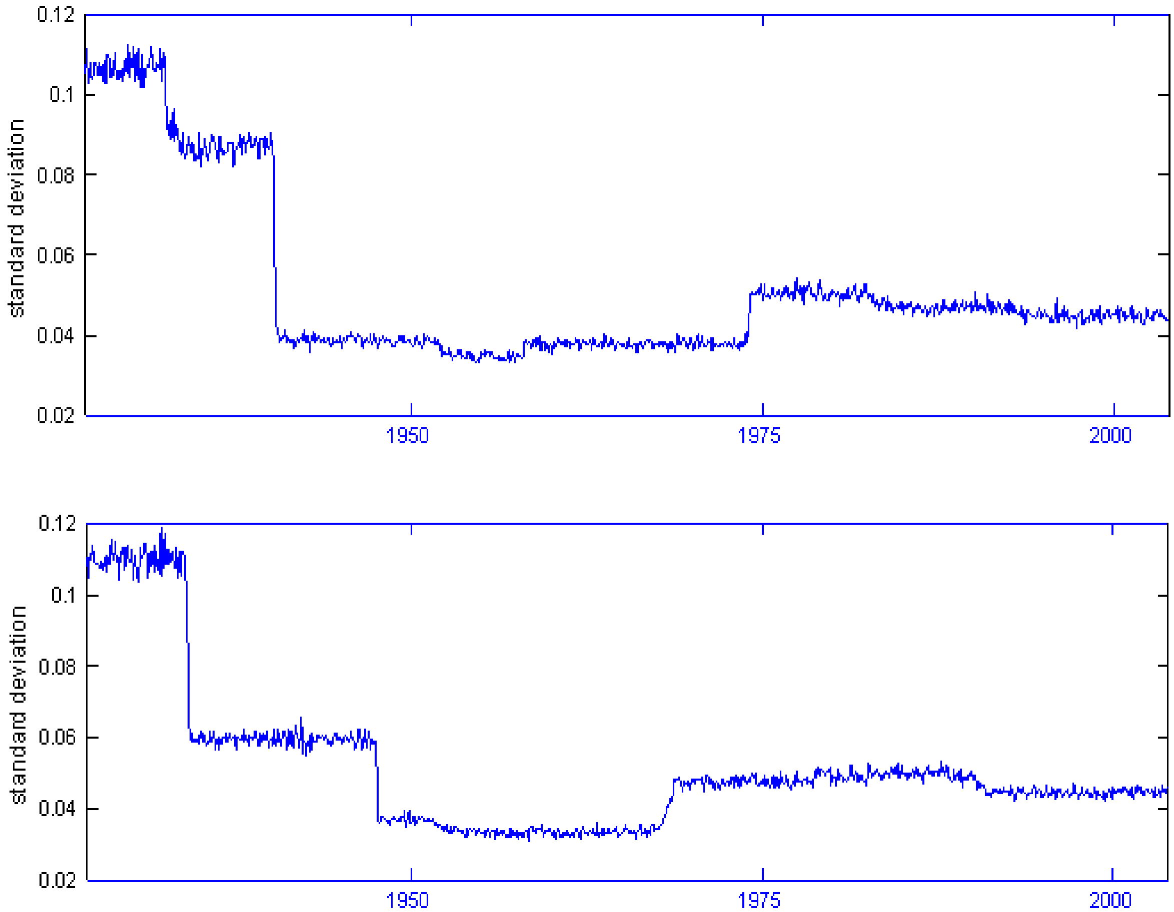

Figure 13: Standard deviations of the predictive distribution of excess returns when the predictor variable is the dividend yield (top panel) or the T-bill rate (bottom panel) under a model with seven breaks.

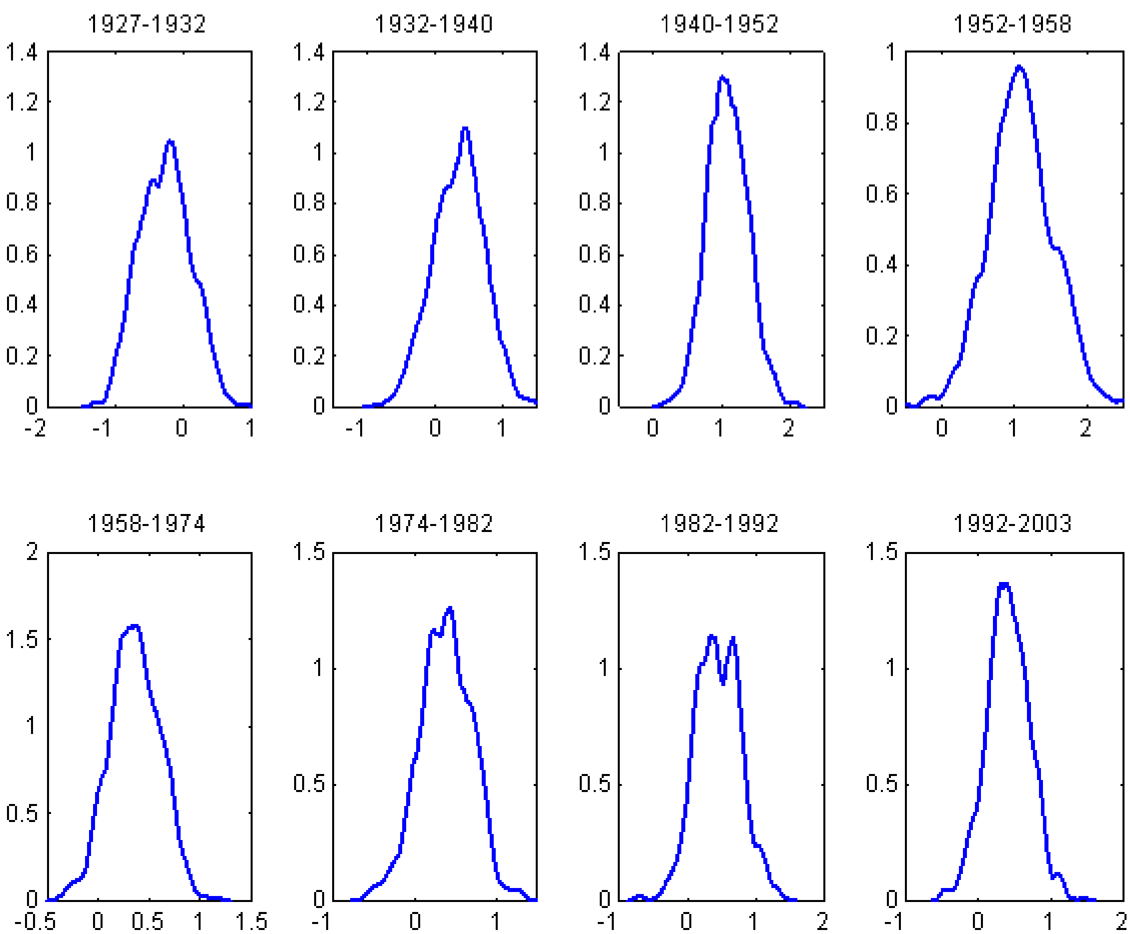

Figure 14: Annualized Sharpe ratios of the (posterior) predictive return distribution for the eight regimes based on a model with predictability from the dividend yield and seven breaks.

I. Excess returns - Dividend Yield							
# breaks	log lik.	marg. log lik.	Post. Prob.	Break locations			
0	1368.18	1319.34	0.00				
1	1433.09	1369.33	0.00	Apr-52			
2	1445.83	1383.28	0.00	Feb-52	Sep-54		
3	1454.38	1387.49	0.00	May-40	Apr-52	Jul-92	
4	1459.03	1390.04	0.00	May-40	Feb-58	Jan-74	Jul-92
5	1462.11	1391.51	0.00	May-40	Apr-52	Feb-58	Mar-74
				Jul-92			
6	1476.60	1404.75	0.01	Aug-32	May-40	Feb-58	Mar-74
				Oct-82	Jul-92		
7	1482.55	1408.74	0.69	Aug-32	May-40	Apr-52	Feb-58
				Mar-74	Oct-82	Jul-92	
8	1482.64	1407.89	0.30	Jun-32	May-40	Apr-52	Feb-58
				Mar-74	Oct-82	Apr-92	Jul-92
II. Excess returns - Treasury Bill Rate							
# breaks	log lik.	marg. log lik.	Post. Prob.	Break locations			
0	1368.52	1323.33	0.00				
1	1378.69	1318.98	0.00	Sep-57			
2	1448.08	1389.49	0.00	Nov-37	Aug-47		
3	1466.27	1405.79	0.00	Jan-34	Aug-47	Sep-57	
4	1471.20	1407.12	0.00	Jan-34	Aug-47	May-73	Jan-90
5	1476.13	1410.67	0.13	Jan-34	Aug-47	Oct-68	Jul-79
				Oct-82			
6	1474.18	1406.79	0.00	Jan-34	Aug-47	Sep-57	Jul-79
				Aug-82	Jan-90		
7	1481.13	1412.29	0.64	Jan-34	Aug-47	Oct-52	Oct-68
				Jul-79	Aug-82	Jan-90	
8	1481.41	1411.26	0.23	Jan-34	Aug-47	Oct-57	Oct-62
				Jul-66	Jul-79	Aug-82	Jan-90

Table 1: Model comparison and selection of the number of breaks in the return forecasting models. The table shows estimates of the log-likelihood for stock returns and the predictor variable (either the dividend yield or the T-bill rate), marginal log-likelihood estimates for returns and posterior probabilities for models with different numbers of breaks along with the time of the break points for the different models. The top and bottom panels display results when the predictor for the excess return is the lagged dividend yield (panel I) and the lagged T-Bill rate (panel II), respectively. The data sample is 1926:12 - 2003:12.

	Regimes								
	Full sample	27-32	32-40	40-52	52-58	58-74	74-82	82-92	92-03
				μ_r					
mean	-0.003	-0.028	-0.026	-0.016	-0.028	-0.062	-0.086	-0.056	-0.020
s.d.	0.005	0.029	0.033	0.015	0.024	0.021	0.033	0.028	0.011
				β_r					
mean	0.197	0.392	0.737	0.479	0.825	2.090	1.999	1.747	1.408
s.d.	0.120	0.489	0.701	0.262	0.508	0.680	0.724	0.782	0.576
				σ_r					
mean	0.055	0.105	0.086	0.038	0.034	0.038	0.050	0.046	0.045
s.d.	0.001	0.009	0.007	0.002	0.003	0.002	0.004	0.003	0.003
				μ_x					
mean	0.001	0.005	0.004	0.002	0.002	0.003	0.004	0.002	3.0E-04
s.d.	0.001	0.002	0.002	0.001	0.001	0.001	0.002	0.001	2.0E-04
				β_x					
mean	0.983	0.916	0.901	0.967	0.951	0.919	0.908	0.940	0.979
s.d.	0.006	0.033	0.035	0.016	0.022	0.021	0.034	0.026	0.009
				$\sigma_x \times 100$					
mean	0.286	0.714	0.443	0.235	0.148	0.111	0.226	0.149	0.065
s.d.	0.007	0.063	0.036	0.014	0.013	0.006	0.016	0.010	0.004
				ρ_{rx}					
mean	-0.872	-0.930	-0.936	-0.858	-0.928	-0.955	-0.963	-0.941	-0.946
s.d.	0.008	0.021	0.022	0.023	0.023	0.020	0.017	0.021	0.020
				p					
mean		0.982	0.986	0.990	0.983	0.992	0.987	0.988	N.A.
s.d.		0.013	0.010	0.008	0.013	0.006	0.010	0.010	N.A.

Table 2: Parameter estimates for the return (r_t) forecasting model with seven break points, based on the lagged dividend yield (x_{t-1}) as a predictor variable: $r_t = \mu_{r_j} + \beta_{r_j} x_{t-1} + \epsilon_{rt},\ \epsilon_{rt} \sim N\left(0, \sigma^2_{r_j}\right)$, $x_t = \mu_{x_j} + \beta_{x_j} x_{t-1} + \epsilon_{xt},\ \epsilon_{xt} \sim N\left(0, \sigma^2_{x_j}\right),\ \Pr\left(s_t = j \mid s_{t-1} = j\right) = p_{jj},\ corr\left(\epsilon_{rt}, \epsilon_{xt}\right) = \rho_{rx_j}$, $\tau_{j-1} + 1 \leq t \leq \tau_j$. The sample period is 1926:12-2003:12.

Hyperparameters of Meta distributions				
I Return equation				
Mean Parameters				
	mean	s.d.	95% conf interval	
$b_0(\mu_r)$	-0.042	0.038	-0.123	0.033
$b_0(\beta_r)$	1.218	0.508	0.225	2.209
II Dividend Yield equation				
Mean Parameters				
	mean	s.d.	95% conf interval	
$b_0(\mu_x)$	0.003	0.002	1.0E-04	0.008
$b_0(\beta_x)$	0.918	0.033	0.839	0.972
Correlation parameters				
	mean	s.d.	95% conf interval	
μ_ρ	-0.920	0.042	-0.984	-0.831
Transition Probability parameters				
	mean	s.d.	95% conf interval	
a_0	33.258	15.015	10.498	73.091
b_0	0.806	0.308	0.357	1.462

Table 3: Estimates of the parameters of the meta distribution that characterizes variation in the parameters of the return model across different regimes. The estimates are from a model with predictability of returns from the dividend yield and assume seven historical breaks. Within the jth regime the model is: $z_t = B_j' x_{t-1} + u_t$, where $z_t = (r_t, x_t)'$ is the vector of stock returns and the predictor variable, and $vec(B)_j \sim N(b_0, V_0)$. $\rho_j \sim N(\mu_\rho, \sigma_\rho^2)$ is the correlation between shocks to the dividend yield and shocks to returns in the jth regime, while $p_{jj} \sim Beta(a_0, b_0)$ is the probability of remaining in the jth regime. The sample period is 1926:12-2003:12.

	Regimes								
	Full sample	27-34	34-47	47-52	52-68	68-79	79-82	82-90	90-03
				μ_r					
mean	0.009	-0.001	0.007	0.016	0.027	0.037	0.092	0.021	0.009
s.d.	0.003	0.018	0.005	0.011	0.007	0.019	0.034	0.024	0.010
				β_r					
mean	-1.402	-0.439	-3.343	-4.353	-8.215	-8.203	-9.976	-2.452	-0.942
s.d.	0.731	6.886	11.343	10.542	3.015	4.074	3.656	4.086	2.724
				σ_r					
mean	0.055	0.109	0.059	0.036	0.033	0.047	0.047	0.049	0.044
s.d.	0.001	0.008	0.003	0.004	0.002	0.003	0.006	0.004	0.002
				$\mu_x \times 100$					
mean	0.004	0.006	0.001	0.017	0.011	0.032	0.132	0.049	0.006
s.d.	0.002	0.004	0.001	0.006	0.005	0.014	0.069	0.021	0.004
				β_x					
mean	0.986	0.958	0.924	0.844	0.959	0.937	0.832	0.914	0.974
s.d.	0.005	0.021	0.027	0.065	0.020	0.030	0.075	0.037	0.012
				$\sigma_x \times 100$					
mean	0.040	0.033	0.004	0.011	0.027	0.045	0.129	0.047	0.025
s.d.	0.001	0.003	2.7E-04	0.002	0.001	0.003	0.017	0.004	0.001
				ρ_{rx}					
mean	0.005	0.067	0.003	0.170	-0.011	-0.189	-0.311	0.302	0.067
s.d.	0.032	0.026	0.026	0.046	0.033	0.034	0.033	0.037	0.030
				p					
mean		0.985	0.991	0.981	0.992	0.989	0.976	0.987	N.A.
s.d.		0.012	0.007	0.015	0.006	0.009	0.019	0.010	N.A.

Table 4: Parameter estimates for the return (r_t) forecasting model with seven break points, based on the lagged T-bill rate (x_{t-1}) as a predictor variable: $r_t = \mu_{r_j} + \beta_{r_j} x_{t-1} + \epsilon_{rt}$, $\epsilon_{rt} \sim N\left(0, \sigma^2_{r_j}\right)$, $x_t = \mu_{x_j} + \beta_{x_j} x_{t-1} + \epsilon_{xt}$, $\epsilon_{xt} \sim N\left(0, \sigma^2_{x_j}\right)$, $\Pr\left(s_t = j \mid s_{t-1} = j\right) = p_{jj}$, $corr\left(\epsilon_{rt}, \epsilon_{xt}\right) = \rho_{rx_j}$, $\tau_{j-1} + 1 \leq t \leq \tau_j$. The sample period is 1926:12-2003:12.

Hyperparameters of Meta Distributions				
I Return equation				
Mean Parameters				
	mean	s.d.	95% conf interval	
$b_0(\mu_r)$	0.026	0.040	-0.062	0.101
$b_0(\beta_r)$	-4.645	4.840	-14.833	4.430
II T bill equation				
Mean Parameters				
	mean	s.d.	95% conf interval	
$b_0(\mu_x)$	4.6E-04	3.4E-04	4.8E-05	0.001
$b_0(\beta_x)$	0.897	0.042	0.793	0.961
Correlation parameters				
	mean	s.d.	95% conf interval	
μ_ρ	0.013	0.110	-0.205	0.229
Transition Probability parameters				
	mean	s.d.	95% conf interval	
a_0	26.625	13.103	7.656	54.678
b_0	0.660	0.288	0.237	1.271

Table 5: Estimates of the parameters of the meta distribution that characterizes variation in the parameters of the return model across different regimes. The estimates are from a model with predictability of returns from the T-bill rate and assume seven historical breaks. Within the jth regime the model is: $z_t = B_j' x_{t-1} + u_t$, where $z_t = (r_t, x_t)'$ is the vector of stock returns and the predictor variable, and $vec(B)_j \sim N(b_0, V_0)$. $\rho_j \sim N(\mu_\rho, \sigma_\rho^2)$ is the correlation between shocks to the T-bill and shocks to returns in the jth regime, while $p_{jj} \sim Beta(a_0, b_0)$ is the probability of remaining in the jth regime. The sample period is 1926:12-2003:12.

Hyperparameters of Meta distributions				
I Return equation				
Mean Parameters				
	mean	s.d.	95% conf interval	
μ_{μ_r}	-0.019	0.104	-0.188	0.134
μ_{β_r}	0.770	0.457	0.086	1.573
ρ_{μ_r}	0.490	0.286	0.048	0.945
ρ_{β_r}	0.465	0.275	0.043	0.924
II Dividend Yield equation				
Mean Parameters				
	mean	s.d.	95% conf interval	
μ_{μ_x}	1.5E-04	0.0986	-0.157	0.163
μ_{β_x}	0.490	0.288	0.037	0.931
ρ_{μ_x}	0.497	0.288	0.045	0.942
ρ_{β_x}	0.489	0.290	0.045	0.942

Table 6: Estimates of the hyperparameters of the meta distribution for the return forecasting model with seven break points, based on the dividend yield as the predictor variable under dependence of the regression parameters across regimes: $z_t = B_j' x_{t-1} + u_t$ where $z_t = (r_t, x_t)'$ is the excess return and the predictor variable, and $vec(B_j)_i \sim N(\mu_i + \rho_i vec(B_{j-1})_i, \sigma^2_{\eta,i})$, $j = 1, ..., K+1$ and $i = 1, .., m^2$. The sample period is 1926:12-2003:12.