Predictability of Stock Returns and Asset Allocation under Structural Breaks*

Davide Pettenuzzo
Bates White, LLC

Allan Timmermann
University of California, San Diego

March 1, 2010

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 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 constant. This paper adopts a new approach that accounts for breaks to return prediction models both in the historical estimation period and at future points. Empirically we find evidence of multiple breaks in return prediction models based on the dividend yield or a short interest rate. Our analysis suggests that model instability is a very important source of investment risk for investors with long horizons and that breaks can lead to a negative slope in the relationship between the investment horizon and the proportion of wealth that a buy-and-hold investor allocates to stocks.

*Three anonymous referees made constructive and helpful suggestions on an earlier version of the paper. We thank Jun Liu, Ross Valkanov, Jessica Wachter and Mark Watson as well as seminar participants at the Rio Forecasting conference, University of Aarhus (CAF), University of Arizona, New York University (Stern), Erasmus University at Rotterdam, Princeton, UC Riverside, UCSD, Tilburg, and Studiencenter Gerzensee for helpful comments on the paper. Alberto Rossi provided excellent research assistance. Timmermann acknowledges support from CREATES, funded by the Danish National Research Foundation.

1. **Introduction**

Stock market investors face a daunting array of risks. First and foremost is the innovation component of stock returns that cannot be predicted in the context of any model for the return generating process. This source of uncertainty is substantial, given the low predictive power of return forecasting models. Second, even conditional on a particular forecasting model, investors are confronted with parameter uncertainty, i.e. the effect of not knowing the true model parameters (Kandel and Stambaugh (1996) and Barberis (2000)). Third, investors do not know the state variables or functional form of the true return process and so face model uncertainty (Avramov (2002) and Cremers (2002)). This paper deals with a fourth source of uncertainty that is of particular importance to long-run investors, namely model instability, i.e. random changes or "breaks" to the parameters of the return generating process.

Conventional practice in economics and finance is to compute forecasts conditional upon a maintained model whose parameters are assumed to be constant both throughout the historical sample and during the future periods to which the forecasts apply. This procedure ignores that, over estimation samples that often span several decades, the relation between economic variables is likely to change. Instability in economic models may reflect institutional, legislative and technological change, financial innovation, changes in stock market participation, large macroeconomic (oil price) shocks and changes in monetary targets or tax policy.[1] In the context of financial return prediction models, Merton's intertemporal CAPM suggests that time-variations in aggregate risk aversion may lead to changes in the relationship between expected returns and predictor variables tracking movements in market risk or investment opportunities.[2]

Instability in the relation between stock returns and predictor variables such as the dividend yield and short-term interest rates has been documented empirically in several studies. Pesaran and Timmermann (1995), Bossaerts and Hillion (1999), Lettau and Ludvigson (2001), Paye and Timmermann (2006), Ang and Bekaert (2007) and Goyal and Welch (2008) find substantial variation across subsamples in the coefficients of return prediction models and in the degree of return predictability.[3] Building on this evidence, recent studies such as Dangl and Halling (2008) and Johannes, Korteweg, and Polson (2009) have proposed capturing time-variation in return prediction models by assuming that some of the model parameters follow a random walk and thus change every period.

In this paper we focus instead on the effect of rare but large structural breaks as opposed to small parameter changes occurring every period. The distinction between rare, large breaks versus

[1] For example, the introduction of SEC rule 10b-18 in November 1982 changed firms' ability to repurchase shares and thus may have changed firms' payout policy, in turn affecting the relation between stock returns and dividend yields. Examples of changes in the dynamics and predictive content of short-term interest rates include the Accord of 1951 and the monetarist experiment from 1979 to 1982.

[2] Menzly and Veronesi (2004) provide theoretical reasons for expecting time-variation in the relation between expected stock returns and predictor variables such as the dividend yield.

[3] Studies such as Barsky (1989), Dimson, Marsh, and Staunton (2002), McQueen and Roley (1993) and Boyd and Jagannathan (2005) have found evidence of time-variations in the correlation between stock and bond returns or stock returns and economic news variables.

frequent, small breaks can be difficult to make in practice (Elliott and Mueller (2006)). However, our analysis allows us to pinpoint the most important times where the return prediction model undergoes relatively sharp changes, which provides insights into the interpretation of the economic sources of model instability. Sudden, sharp changes in model parameters are consistent with empirical findings by both Dangl and Halling (2008) and Johannes, Korteweg, and Polson (2009) that the change in the parameters of return predictability models at times can be large. By considering few, large breaks, our approach is close in spirit to Pastor and Stambaugh (2001) who consider breaks in the risk-return trade-off and Lettau and van Nieuwerburgh (2008) who consider a discrete break to the steady state value of a single predictor variable (the dividend yield).

Our approach builds on Chib (1998), Pastor and Stambaugh (2001) and Pesaran, Pettenuzzo, and Timmermann (2006) in adopting a changepoint model driven by an unobserved discrete state variable. Specifically, we generalize the univariate model in Pesaran, Pettenuzzo, and Timmermann (2006) to a multivariate setting so instability can arise either in the conditional model used to forecast returns, in the marginal process generating the predictor variable(s) or in the correlation between innovations to the two equations. Forecasting returns in this model requires accounting for the probability and magnitude of future breaks. To this end, we introduce a meta distribution that straddles the parameters drawn for the individual regimes and characterizes how the parameters 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 in the distribution of 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 meta distribution.

The proposed model is very general and allows for uncertainty about the timing (dates) of historical breaks as well as uncertainty about the number of breaks and their magnitude. We also extend our setup to allow for uncertainty about the identity of the predictor variables (model uncertainty) using Bayesian model averaging techniques. Hence, investors are not assumed to know the true model or its parameter values, nor are they assumed to know the number, timing and magnitude 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 as new data is observed.

Instability in model parameters is particularly important to investors' long-run asset allocation decisions which crucially rely on forecasts of future returns. Long investment horizons make it more likely that breaks to model parameters will occur and some of these breaks could adversely affect the investment opportunity set, thereby significantly increasing investment risks. Asset allocation exercises mostly 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 remained constant through time (e.g., Barberis (2000), Pastor and Stambaugh (2009)). Studies that have allowed for time-varying model parameters such as Dangl and Halling (2008) and Johannes, Korteweg, and Polson (2009) only consider mean-variance investors with single-period investment horizons. Our focus is instead on the effect of model instability on the risks faced by investors with a long investment horizon.

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 multiple breaks in return models based on either predictor variable in data covering the period 1926-2005. Many of the break dates coincide with major events such as changes in the Fed's operating procedures (1979, 1982), the Great Depression, the Treasury-Fed Accord (1951) and the growth slowdown following the oil price shocks in the early 1970s. Variation in model parameters is found to be extensive. For example, the predictive coefficient of the dividend yield varies between zero and 2.6, while the coefficient of the T-bill rate varies even more, between -9.4 and 3.3, across break segments.

Structural breaks are found to have a large effect on investors' optimal asset allocations. For example, in the model with predictability from the dividend yield but no breaks, the allocation to stocks rises from 40% at short horizons to 60% at the five-year horizon. Once past and future breaks are considered, the allocation to stocks declines from close to 100% at short horizons to 10% at the five-year horizon. Our analysis suggests that model instability is a more important source of investment risk than parameter estimation uncertainty for investors with long horizons and that breaks 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.[4]

Our portfolio allocation results lend further credence to the finding in Pastor and Stambaugh (2009) that the long-run risks of stocks can be very high. In a model that allows for imperfect predictors and unknown, but stable parameters of the data generating process, Pastor and Stambaugh find that the true per-period predictive variance of stock returns can be increasing in the investment horizon due to the compound effect of uncertainties about current and future expected returns (and their relationship to observed predictor variables) and estimation risk. While this finding is similar to ours, the mechanism is very different: Pastor and Stambaugh (2009) derive their results from investors' imperfect knowledge of current and future expected returns and model parameters, whereas model instability is the key driver behind our results.

The paper is organized as follows. Section 2 introduces the breakpoint methodology and Section 3 presents empirical estimates for return prediction models based on the dividend yield or the short interest rate. Section 4 shows how investors' optimal asset allocation can be computed while accounting for past and future breaks. Section 5 considers asset allocations empirically for a buy-and-hold investor. Section 6 proposes various extensions to our approach and Section 7 concludes. Technical details are provided in appendices at the end of the paper.

[4] Consistent with our results, Johannes et al. (2009) also find that parameter estimation uncertainty has a smaller effect on the asset allocation than uncertainty about changes to model parameters. While Dangl and Halling (2008) find that estimation uncertainty plays a dominant role, they also report that uncertainty about time-variation in coefficients is important, particularly during periods with turmoil such as the early seventies. Barberis (2000) finds that estimation risk significantly affects investors' long-run asset allocations, but this finding is based on a relatively short data sample.

2. **Methodology**

Studies of asset allocation under 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. 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. Empirically, the coefficients on the lagged returns are usually found to be small, so we follow common practice and restrict them to be zero. 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 is the predictor variable and $u_t \sim IIDN(0, \Sigma)$, where $\Sigma = E[u_t u_t']$ 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:

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

2.1. *Predictive Distributions of Returns under Breaks*

Asset allocation decisions require the ability to evaluate expected utility associated with the realization of future payoffs on risky assets. This, in turn, requires computing expectations over the predictive distribution of cumulated 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 while allowing for breaks, we need to make assumptions about the probability that future breaks occur, their likely timing as well as the size of such breaks. 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 equation (2), we build on the multiple change point model proposed by Chib (1998). Shifts to the parameters of the return prediction 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, $s_t = k$ indicates that z_t has been drawn from $f\left(z_t | \mathcal{Z}_{t-1}, \Theta_k\right)$, where $\mathcal{Z}_{t-1} = \{z_1, ..., z_{t-1}\}$ is the information set at time $t-1$, while a change from $s_t = k$ to $s_{t+1} = k+1$ shows that a break has occurred at time $t+1$. Location and scale parameters in regime k are collected in $\Theta_k = (B_k, \Sigma_k)$. Allowing for K breaks or, equivalently, $K+1$ break segments, between $t=1$ and $t=T$, our model takes the form

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

Here $\Upsilon_K = \{\tau_0, ..., \tau_K\}$ is the collection of break points with $\tau_0 = 1$, and the innovations u_t are assumed to be multivariate Gaussian with zero mean. Within each regime we decompose the covariance matrix, Σ_k, into the product of a diagonal matrix representing the standard deviations of the variables, $diag(\psi_k)$, and a correlation matrix, Λ_k:

$$\Sigma_k = diag(\psi_k) \times \Lambda_k \times diag(\psi_k). \tag{4}$$

This specification allows both mean parameters, volatilities and correlations to vary across regimes.[5] We collect the regression coefficients, error term variances and correlation parameters in $\Theta = (vec(B)_k, \psi_k, \Lambda_k)_{k=1}^{K+1}$.

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.[6] The one-step-ahead transition probability matrix therefore takes the form

$$P = \left(\begin{array}{ccccc|c} p_{1,1} & p_{1,2} & 0 & \cdots & 0 & \\ 0 & p_{2,2} & p_{2,3} & \cdots & 0 & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \\ 0 & \cdots & 0 & p_{K,K} & p_{K,K+1} & \\ 0 & 0 & \dots & 0 & p_{K+1,K+1} & p_{K+1,K+2} \\ \hline 0 & 0 & \dots & & 0 & p_{K+2,K+2} \\ & & & & & \ddots \end{array} \right). \tag{5}$$

Here $p_{k,k+1} = Pr\left(s_t = k+1 | s_{t-1} = k\right)$ is the probability of moving to regime $k+1$ at time t given that we are in state k at time $t-1$ so $p_{k,k+1} = 1 - p_{k,k}$. K is the number of breaks in the historical

[5] Allowing for time-variations in both first and second moments could be important in practice. In a model that allows for stochastic volatility, Johannes, Korteweg, and Polson (2009) find that the level of return volatility affects the signal-to-noise ratio of the return equation and therefore also affects investors' ability to infer the underlying state and compute expected returns.

[6] Some studies assume that the parameters of the return equation are driven by a Markov switching process with two or three states, e.g., Ang and Bekaert (2002), Ang and Chen (2002), Guidolin and Timmermann (2008) and Perez-Quiros and Timmermann (2000). The assumption of a fixed number of states amounts to imposing a restriction that 'history repeats'. This approach is well suited to identify patterns in returns 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.

sample up to time T so the $(K+1) \times (K+1)$ sub-matrix in the upper left corner of P, denoted $p = (p_{1,1}, p_{2,2}, ..., p_{K+1,K+1})'$, 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 period from T to $T+h$.[7] The special case without breaks corresponds to $K = 0$ and $p_{1,1} = 1$.

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

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

This break model is quite different from the drifting coefficients (random walk) models studied by Dangl and Halling (2008) and Johannes, Korteweg, and Polson (2009). The latter are designed to obtain a good local approximation to parameter values at any given point in time, whereas our break model attempts to capture rare, but large shifts in parameter values that affect the return distribution, particularly at longer horizons.

2.2. *Meta Distributions*

Since we are interested in forecasting future returns, we follow Pastor and Stambaugh (2001) and Pesaran, Pettenuzzo, and Timmermann (2006) and adopt a hierarchical prior formulation, but extend those studies to allow for structural breaks in a multivariate setting.[8] To this end we assume that the location and scale parameters within each regime, (B_k, Σ_k), are drawn from common meta distributions which 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. This will show up in the form of a wide dispersion in the meta distribution for 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. By using meta distributions that pool information from different regimes, our approach makes sure that historical 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), $\{B_k, diag(\psi_k), \Lambda_k\}$. We assume that there is a single return series and, for generality, $m-1$ predictor variables for a total

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

[8] 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.

of m equations in the prediction model (3) and further assume that the m^2 location parameters are independent draws from a normal distribution, $vec(B)_k \sim N(b_0, V_0)$, $k = 1, ..., K+1$, while the m error term precision terms $\psi_{k,i}^{-2}$ are independent and identical draws (IID) from a Gamma distribution, $\psi_{k,i}^{-2} \sim Gamma\left(\frac{v_{0,i}}{2}, \frac{v_{0,i}d_{0,i}}{2}\right)$, $i = 1, ..., m$. Finally, the $m(m-1)/2$ correlations, $\lambda_{k,i,c}$, are IID draws from a normal distribution, $\lambda_{k,i,c} \sim N(\mu_{\rho,i,c}, \sigma^2_{\rho,i,c})$, $i, c = 1, ..., m$, $i < c$, truncated so the correlation matrix is positive definite which in the two-equation model means that $\lambda_{k,i,c} \in (-1, 1)$. Hence $b_0, v_{0,i}$ and $\mu_{\rho,i,c}$ represent location parameters, while V_0, $d_{0,i}$ and $\sigma^2_{\rho,i,c}$ are scale parameters of the three meta distributions.

The pooled scenario in which all parameters are identical across regimes and the case where the parameters of each regime are virtually unrelated can be seen as special cases 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 $\sigma^2_{\rho,i,c}$.

To characterize the parameters of the meta distribution, we assume that[9]

$$\begin{aligned} b_0 &\sim N\left(\underline{\mu}_\beta, \underline{\Sigma}_\beta\right), \\ V_0^{-1} &\sim W\left(\underline{V}_\beta^{-1}, \underline{\nu}_\beta\right), \end{aligned} \tag{7}$$

where $W(.)$ is a Wishart distribution and $\underline{\mu}_\beta$, $\underline{\Sigma}_\beta$, $\underline{v}_\beta$ and $\underline{V}_\beta$ 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 (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:

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

$$\sigma^{-2}_{\rho,i,c} \sim Gamma\left(\underline{a}_{\rho,i,c}, \underline{b}_{\rho,i,c}\right), \tag{11}$$

where again $\underline{\mu}_{\mu,i,c}$, $\underline{\tau}^2_{i,c}$, $\underline{a}_{\rho,i,c}$ and $\underline{b}_{\rho,i,c}$ are prior hyperparameters for each element of the correlation matrix. Finally, we specify a prior distribution for the hyperparameters a and b of the transition probabilities,

$$a \sim Gamma\left(\underline{a}_0, \underline{b}_0\right), \tag{12}$$

$$b \sim Gamma\left(\underline{a}_0, \underline{b}_0\right). \tag{13}$$

These are all standard choices of distributions. We collect the hyperparameters of the meta distribution in $H = \left(b_0, V_0, v_{0,1}, d_{0,1}, ..., v_{0,m}, d_{0,m}, \mu_{\rho,1,2}, \sigma^2_{\rho,1,2}, ..., \mu_{\rho,m-1,m}, \sigma^2_{\rho,m-1,m}, a, b\right)$.

[9] Throughout the paper we use underscore bars (e.g. $\underline{a}_0$) to denote prior hyperparameters.

2.3. *Likelihood function and approximate marginal likelihood*

The likelihood function is obtained by extending to the hierarchical setting the approach proposed by Chip (1998). The likelihood function, evaluated at the posterior means of the regime specific parameters, Θ^*, hyperparameters, H^*, and transition probabilities, p^*, is obtained from the decomposition

$$\ln f\left(\mathcal{Z}_T \middle| \Theta^*, H^*, p^*\right) = \sum_{t=1}^{T} \ln f\left(z_t \middle| \mathcal{Z}_{t-1}, \Theta^*, H^*, p^*\right), \tag{14}$$

where

$$f\left(z_t \middle| \mathcal{Z}_{t-1}, \Theta^*, H^*, p^*\right) = \sum_{k=1}^{K+1} f\left(z_t \middle| \mathcal{Z}_{t-1}, \Theta^*, H^*, p^*, s_t = k\right) p\left(s_t = k \middle| \Theta^*, H^*, p^*, \mathcal{Z}_{t-1}\right), \tag{15}$$

and $f\left(z_t \middle| \mathcal{Z}_{t-1}, \Theta^*, H^*, p^*, s_t = k\right)$ is the conditional density of z_t given the regime $s_t = k$, while

$$p\left(s_t = k \middle| \Theta^*, H^*, p^*, \mathcal{Z}_{t-1}\right) = \sum_{l=k-1}^{k} p^*_{l,k} \times p\left(s_{t-1} = l \middle| \Theta^*, H^*, p^*, \mathcal{Z}_{t-1}\right), \tag{16}$$

(see Appendix 1.) The following expression, which is proportional to the SIC, is used to compute an asymptotic approximation to the marginal likelihood for a model with K breaks, M_K:

$$p\left(M_K \middle| \mathcal{Z}_T\right) \propto \ln f\left(\mathcal{Z}_T \middle| \Theta^*, H^*, p^*\right) - \frac{N_K \times \ln(T)}{2},$$

where N_K is the number of parameters for model M_K. Approximate posterior model probabilities for models with up to $\bar{K}$ breaks can be computed by exponentiating the approximate marginal likelihood for a model with K breaks divided by the sum of the corresponding terms across models with $K = 0, ..., \bar{K}$ breaks.

2.4. *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 6 we conduct a sensitivity analysis to shed light on the importance of these choices.

We impose two constraints on the parameters in the return prediction model, (3). First, to rule out explosive behavior in the predictor variable (and consequently in stock returns), we impose that $\beta_x < 1$. Second, we require the unconditional mean of the predictor variable in each state to be non-negative, i.e. $\mu_x/(1-\beta_x) \geq 0$. This has a very limited impact on the posterior in-sample distributions of the individual regime parameters. However, this restriction is important when generating out-of-sample forecasts and helps eliminate economically non-sensible trajectories for the predictor variables.

Starting with the prior hyperparameters for the mean of the regression coefficient, b_0, we set $\underline{\mu}_\beta = [0, 0, 0, 0.9]'$ and $\underline{\Sigma}_\beta = diag(sc, sc, sc, 1)$, where sc is a scale factor set to 1,000 to reflect uninformative priors. Both predictor variables that we shall consider (the dividend yield and the T-bill rate) are highly persistent so we specify a more informative prior for the autoregressive coefficient, β_x, and center it at 0.9. The hyperparameters for the prior variance of the regression coefficient, V_0, are set at $\underline{\nu}_\beta = 2m + 2$, $\underline{V}_\beta = diag(0.5, 5, 0.00001, 0.1)$ for the dividend yield specification and $\underline{V}_\beta = diag(0.1, 1000, 0.00001, 0.1)$ for the T-bill specification. This is sufficient to preserve the variation in the regression coefficients across regimes and ensures that the mean of the inverse Wishart distribution exists. The small variation in μ_x and the somewhat larger variation in μ_r reflect the high persistence of the predictor variables, i.e., β_x is close to one.

Moving to the variance hyperparameters, we maintain uninformative priors and set $\underline{c}_{0,i} = 1$, $\underline{d}_{0,i} = \epsilon$ (the smallest number that matlab can interpret), and $\underline{\rho}_{0,i} = 1/\epsilon$ in all equations, hence specifying a very large variance. For the correlation coefficient, $\lambda_{j,1,2}$, we use an uninformative prior, i.e. $\underline{\mu}_{\mu,1,2} = 0$, $\underline{\tau}^2_{1,2} = 100$, $\underline{a}_{\rho,1,2} = 1$ and $\underline{b}_{\rho,1,2} = 0.01$.

Finally, we specify uninformative priors for the hyperparameters a and b of the transition probabilities $p_{k,k}$ in (5), namely $\underline{a}_0 = 1$ and $\underline{b}_0 = \epsilon$. By using uninformative priors for the hyperparameters governing the diagonal elements of the transition probability matrix, we allow the data to dictate the frequency of breaks.

3. 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.

3.1. *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 1926:12-2005:12. 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 the asset allocation implications of return predictability (Kandel and Stambaugh (1996) and Barberis (2000).) Due to its persistence and the large negative 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 predict stock returns (Campbell (1987) and Ang and Bekaert (2002).) Table 1 reports descriptive statistics for the three variables.

Before turning to the empirical results we briefly summarize the estimation setup. Both the dividend yield and T-bill rate models were estimated using a Gibbs sampler with 2,500 draws and the first 500 draws discarded to allow the sampler to achieve convergence.[10] We performed a variety of MCMC convergence diagnostics, ranging from autocorrelation estimates, Raftery and Lewis (1992a), Raftery and Lewis (1992b) and Raftery and Lewis (1995) MCMC diagnostics, Geweke (1992) numerical standard errors and relative numerical efficiency estimates, and the Geweke chi-squared test comparing the means from the first and last part of the sample. We found very little evidence of autocorrelation in the Gibbs sampler draws. This is further confirmed by the thinning ratio estimates obtained from the Raftery and Lewis (1995) diagnostics which were very close to unity. Finally, the Geweke chi-squared test of the means from the first 20% of the sample versus the last 50% confirmed that the Gibbs sampler has achieved an equilibrium state. Appendix 1 provides details of the Gibbs sampler used to estimate the return prediction model with multiple breaks.

3.2. *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 number of breaks, K, we get a new model, M_K, with its own set of parameters. For all values of K, the models are estimated by maximum likelihood with states based on the posterior modes of the break point probabilities. Table 2 provides a comparison of models with different numbers of breaks by reporting various measures of model fit such as the log-likelihood and the approximate marginal likelihood described earlier.

We find strong support for structural break in the return prediction model based on the dividend yield. The approximate posterior odds ratios for the models with multiple breaks are all very high relative to a model with no breaks. Among models with up to ten breaks, an eight-break specification obtains a posterior probability weight of nearly one. Although eight breaks may appear to be a large number, it is consistent with the evidence reported by Pastor and Stambaugh (2001) of 15 breaks in the equity premium over a sample (1834-1999) twice the period covered here.

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 overfit the data. This is not an issue here, however, since we select the break point specification by the SIC, which approximates the marginal likelihood. Marginal likelihood captures the models' out-of-sample prediction record and so penalizes for an increase in the numbers of estimated parameters.

For the model with eight breaks, Figure 1 shows the time of the associated breaks. More precisely, for an interval around the posterior modes of the eight break dates, each diagram shows the posterior probability of there being one break. Most break dates are reasonably precisely identified in the form of either single spikes with probabilities exceeding one-half or narrow spans covering a few months. There are some exceptions to this, however, notably the break dated 1940,

[10] We used a Windows XP based server with 8 Xeon x5355 2.66 GHz processors and 32 Gigabytes of DRAM. The Gibbs sampler for the dividend yield model based on eight breaks finished in 35 minutes, while the Gibbs sampler for the T-bill model based on five breaks ran for 33 minutes.

where the alternative date of 1943 also achieves a high posterior probability, and the break dated 1951, for which 1954 is an equally plausible date.[11]

Most of the break locations are associated with major events and occurred around the Great Depression (1933), World War II (1940), the Treasury-Fed Accord (1951), and the major oil price shocks of the early seventies and the resulting growth slowdown (1974). Some breaks are also associated with changes in price dynamics, e.g., the interval spanning the October 1987 stock market crash (1986 and 1988) and, more recently, the take-off accompanying the bull market of the nineties (1996).

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 caused, e.g., by major supply shocks.

Parameter estimates for the model with eight breaks (nine regimes) as well as the no-break model are reported in Table 3. Consistent with results in the empirical literature, full-sample estimates of the parameters in the return equation (2) with no breaks (shown in the first column) reveal a mean coefficient on the dividend yield that is positive but slightly less than two standard errors away from zero.

Turning to the estimates of the break model, the mean of the dividend yield coefficient in the return equation ranges from a low of zero in the earliest sample (1926-1933) to 2.6 during the final period (1996-2005). The substantial time variation in the coefficient of the dividend yield is consistent with the sub-sample estimates reported by Ang and Bekaert (2007). It is also consistent with the finding in Lettau and van Nieuwerburgh (2008) that uncertainty over the magnitude of breaks is very large. The standard deviation parameter of the return equation also varies considerably across regimes, from a high of 10% per month during the Great Depression to a low of only 3.1% per month from 1988-1996.

The parameter estimates for the dividend yield equation show that this process is highly persistent in all regimes with a mean autoregressive parameter that varies from 0.94 to 0.97. The variance of the dividend yield is again highest in the first regime and becomes much lower after the final break in 1996. Correlation estimates for the innovations to stock returns and the lagged dividend yield are large and negative in all regimes with mean values ranging from -0.97 to -0.92. Transition probabilities are high with mean values that always exceed 0.97 and go as high as 0.992, corresponding to mean durations ranging from 40 to 140 months.

One of the questions we set out to address in our paper was how similar the parameters of the return equation are across regimes. To address this question, information on the posterior estimates of the hyperparameters of the meta distribution is provided in Table 4. To preserve space we only report the values of the parameters that are easiest to interpret. The parameter tracking the grand mean of the slope of the dividend yield in the return equation is centered on 0.92 with a standard deviation centered at 0.50, giving rise to a 95% confidence interval of $[0.0, 2.0]$. The autoregressive

[11] Lettau and van Nieuwerburgh (2008) find breaks in the mean of the dividend yield in 1954 and 1994. These are very similar to two of our break dates, namely 1951 and 1996 with differences likely to be attributed to uncertainty in the determination of the break dates (1954) and differences in the number of breaks allowed.

slope β_x in the dividend yield equation is centered on a value of 0.95 with a much smaller standard deviation of only 0.03 and a 95% confidence interval of $[0.88, 0.99]$. Similarly, the hyperparameter tracking the correlation between shocks to returns and shocks to the dividend yield is centered on -0.94 with a modest standard deviation of 0.03. 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 coefficient of the lagged dividend yield in the return equation, the volatility of stock returns and the duration of the regimes. There is considerably less uncertainty about the persistence of the dividend yield or the correlation between shocks to returns and shocks to the dividend yield.

3.3. *Predictability from the Short Interest Rate*

Turning to the return model based on the short interest rate, Table 2 shows that a model with five breaks is strongly supported by the data. These breaks again appear around the time of major events such as the Great Depression (1934), the Fed-Treasury Accord (1951), the Vietnam War (1969) and the beginning and end of the change to the Fed's operating procedures (1979 and 1982).

Figure 2 shows the posterior probabilities surrounding the modes of the five breakpoints. The break dates are quite precisely estimated as, in each case, the posterior probabilities define narrow ranges. The break dated 1969 is surrounded by the greatest uncertainty.

Parameter estimates for the return model with five breaks are displayed in Table 5. The mean of the coefficient on the lagged T-bill rate in the return equation varies significantly over time, ranging from -9.4 during the very volatile "monetarist policy experiment" from 1979 to 1982 to 3.3 during 1934-1951. Furthermore, the estimates of the slope on the T-bill rate within each regime are surrounded by large standard errors, particularly prior to 1951.

The process for the short interest rate is highly persistent with the mean of the autoregressive coefficient ranging from a low of 0.94 to a high of 0.996. The correlation between shocks to returns and shocks to the T-bill rate varies much more across regimes than in the dividend yield model, ranging from a low of -0.47 during 1979-1982 to a high of 0.08 during 1926-34. 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 0.973 to 0.993, resulting in state durations between 40 and more than 160 months.

Turning finally to the meta distribution parameters for the T-bill rate model shown in Table 6, once again the chief source of uncertainty is the slope coefficient of the interest rate in the return equation. For example, $b_0(\beta_r)$ has a mean of -2.8 and a standard deviation of 5.9, giving a very long 95% confidence interval that ranges from -14.6 to 9.1. 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 and the wide 95% confidence interval from -0.45 to 0.24.

4. **Asset Allocation under Structural Breaks**

Investors are concerned with instability in the return model because this affects 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 with a horizon of h periods who at time T has power utility over terminal wealth, W_{T+h}, and coefficient of relative risk aversion, γ:

$$u(W_{T+h}) = \frac{W_{T+h}^{1-\gamma}}{1-\gamma}, \quad \gamma > 0. \tag{17}$$

Following Kandel and Stambaugh (1996) and Barberis (2000), we assume that the investor has access to a risk-free asset whose single-period return is denoted $r_{f,T+1}$, and a risky stock market portfolio whose return, measured in excess of the (single-period) risk-free rate, is denoted r_{T+1}. All returns are continuously compounded. The risk-free rate is allowed to change every period.

4.1. *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

$$W_{T+h} = (1-\omega)\exp(\sum_{\tau=1}^{h} r_{f,T+\tau}) + \omega \exp(\sum_{\tau=1}^{h}(r_{T+\tau} + r_{f,T+\tau})). \tag{18}$$

Subject to the no short-sale constraint $0 \leq \omega \leq 0.99$,[12] the buy-and-hold investor solves the following problem

$$\max_{\omega} E_T \left(\frac{((1-\omega)\exp(R_{f,T+h}) + \omega \exp(R_{s,T+h}))^{1-\gamma}}{1-\gamma} \right), \tag{19}$$

where the cumulative $h-$period returns on stocks and the corresponding return from rolling over one-period T-bills are given by $R_{s,T+h} = \sum_{\tau=1}^{h}(r_{T+\tau} + r_{f,T+\tau})$ and $R_{f,T+h} = \sum_{\tau=1}^{h} r_{f,T+\tau}$ and 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.

4.2. *No Breaks, no Parameter Uncertainty*

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 $\Theta = (B, \Sigma)$ can be estimated and the model can be iterated forward conditional on these parameter estimates.

[12] We use an upper bound on stock holdings of $\omega = 0.99$ in order to ensure that the expected utility is bounded which might otherwise be a problem, see (Geweke (2001)), Kandel and Stambaugh (1996) and Barberis (2000).

Collecting cumulative stock and T-bill returns in the vector $R_{T+h} = (R_{s,T+h}, R_{f,T+h})$, we can generate a distribution for future asset 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

$$\max_{\omega} \int u(W_{T+h})p(R_{T+h}|\widehat{\Theta}, S_{T+h} = 1, \mathcal{Z}_T)dR_{T+h}. \tag{20}$$

Here we used that, from (18), 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.[13]

4.3. *No Breaks with Parameter Uncertainty*

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.[14] 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 $T + h$:

$$p(R_{T+h}|S_{T+h} = 1, \mathcal{Z}_T) = \int p(R_{T+h}|\Theta, S_{T+h} = 1, \mathcal{Z}_T)\pi(\Theta|S_{T+h} = 1, \mathcal{Z}_T)d\Theta. \tag{21}$$

This investor therefore solves the asset allocation problem

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

Comparing stock holdings in (20) and (22) gives a measure of the economic importance of parameter estimation uncertainty. Both solutions ignore model instability, however.

4.4. *Past and Future Breaks*

Both past and future breaks matter for the investor's estimates of the future return distribution. The predictive density of returns conditional on $K + 1$ regimes having emerged up to time T can be computed by integrating over the parameters, $\pi\left(\Theta, H, p|\, S_T = K + 1, \mathcal{Z}_T\right)$:

$$\begin{aligned} & p(R_{T+h}|S_T = K + 1, \mathcal{Z}_T) \\ = & \int\int\int p\left(R_{T+h}|\, \Theta, H, p, S_T = K + 1, \mathcal{Z}_T\right) \pi\left(\Theta, H, p|\, S_T = K + 1, \mathcal{Z}_T\right) d\Theta dH dp. \end{aligned} \tag{23}$$

Appendix 2 explains the steps involved in obtaining draws from the predictive distribution of cumulative returns that account for possible future breaks. An investor who considers the uncertainty about out-of-sample breaks but conditions on K historical (in-sample) breaks therefore solves

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

[13] 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, with $K = 0$ here. The importance of M_{k_x} will become clear when we integrate out uncertainty about the number of breaks and the predictor variables.

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

This expression does not restrict the number of future breaks, nor does it take the parameters as known. It does, however, take the number of historical breaks as fixed and also ignores uncertainty about the forecasting model itself. We next relax these assumptions.

4.5. *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 historical 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 number of breaks, K, and predictor variable, x, we get a model M_{K_x} with predictive density $p_{K_x}(R_{T+h}|S_T = K+1, X = x, \mathcal{Z}_T)$. Integrating over the number of breaks (but keeping the choice of predictor variables, x, fixed), the predictive density under the Bayesian model average is

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

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

$$p\left(M_{K_x}\middle| \mathcal{Z}_T\right) \propto f\left(\mathcal{Z}_T\middle| M_{K_x}\right) p\left(M_{K_x}\right). \tag{26}$$

4.6. *Model uncertainty*

In addition to not knowing the parameters of a given return forecasting model and not knowing the number of historical breaks, investors are unlikely to know the true identity of the predictor variables. 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 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:

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

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

5. **Empirical Asset Allocation Results**

We next use 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 appendices.

Before moving to the results, it is worth recalling two important effects for asset allocation under return 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 hedging demand for stocks, see Campbell, Chan, and Viceira (2003). Negative shocks to stock prices 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. Second, parameter estimation uncertainty 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 similar 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, breaks to the parameters tend to lower the precision of current parameter estimates and thus increase 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.

Our analysis focuses on in-sample predictability of stock returns. An alternative way to evaluate the importance of return predictability is to conduct a recursive out-of-sample analysis as is done by Dangl and Halling (2008), Johannes, Korteweg, and Polson (2009) and Wachter and Warusawitharana (2009). In fact, our finding of large breaks to the return prediction model is likely to be an important reason for the widely-reported poor out-of-sample forecasting performance of return prediction models. For example, Lettau and van Nieuwerburgh (2008) find that a prediction model that allows for breaks to the steady state of the dividend yield produces better in-sample forecasts, although they also report that it is difficult to exploit such breaks in real time due to the uncertainty surrounding the magnitude of the shift in the mean dividend yield.

5.1. *Results Based on the Dividend Yield*

Figures 3 and 4 plot the allocation to stocks under the three scenarios discussed in Section 4, namely (i) no breaks, no parameter uncertainty; (ii) no breaks with parameter uncertainty; (iii) past and future breaks. The first two scenarios ignore breaks and so use full-sample parameter estimates. We compute the optimal weight on stocks under two values for the coefficient of relative risk aversion, namely $\gamma = 5$ and $\gamma = 10$. To be consistent with the results for the T-bill rate model reported in Table 5, the asset allocation results assume that the short rate follows a persistent process that is subject to breaks.

Figure 3 starts the dividend yield off from its value at the end of sample (2005:12) which is 1.8%. Under the models that assume no breaks, and setting $\gamma = 5$, 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 (1926-2005) 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.

Allowing for past and future breaks, the weight on stocks starts out at 70% at the 1-month horizon and declines to a level below 10% at the five-year horizon. Parameter instability clearly dominates the increased demand for stocks induced by return predictability from the dividend yield.

When risk aversion is increased to $\gamma = 10$, the weight on stocks declines uniformly to levels close to half their values for $\gamma = 5$. The resulting stock allocations may seem low but are also affected by the assumed initial value of the dividend yield which, at 1.8%, is close to its historical minimum.

To demonstrate this point, Figure 4 shows the allocation to stocks when the initial value of the dividend yield is set at its sample mean of 4.1%. Comparing Figures 3 and 4, the level of the optimal stock holding appears to be quite sensitive to the initial value of the dividend yield. The allocation to stocks under the no-break models now starts close to 40% at short investment horizons and increases to nearly 60% at the five-year horizon. Stock holdings under the model that accounts for breaks start just below 100% at the 1-month horizon but rapidly decline as the investment horizon is expanded to reach a level close to 10% at the five-year horizon. Stock holdings are reduced considerably if γ is raised from five to ten, but the qualitative findings remain the same.

These findings suggest that the allocation to stocks is generally increasing in the horizon if breaks in the return prediction model based on the dividend yield are ignored. If past and future breaks are considered, we generally see a strongly declining allocation to stocks, the longer the investment horizon.

Parameter instability in our model has a much 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 80 years of data.

This conclusion is related to the analysis by Pastor and Stambaugh (2009) who find that the variance of stock returns can increase more than in proportion with the forecast horizon due to a combination of uncertainty about current and future expected returns and estimation risk. We also find that the per-period variance of stock returns increases with the horizon in the dividend-yield model. A closely related property is that, in the standard model with no return predictability, the Sharpe ratio divided by the square root of the horizon should be constant. Introducing return predictability from the dividend yield, but ignoring breaks, means that the Sharpe ratio becomes higher, the longer the investment horizon. For the no-break dividend yield model, we find that the mean Sharpe ratio is nine percent higher at the 36 month horizon and 17 percent higher at the 60-month horizon, compared with the mean 12-month Sharpe ratio. Conversely, allowing for breaks to the parameters of this model, we find that the mean of the 36-month and 60-month Sharpe ratios are ten percent and 20 percent *lower*, respectively, than the mean of the 12-month Sharpe ratio. These numbers help explain why long-run investors find it less attractive to hold stocks under the model that allows for breaks.

5.2. *Results based on the Short Interest Rate*

Optimal stock holdings under the return prediction model based on the T-bill rate, set at its value at the end of the sample of 3.8%, are shown in Figure 5. When $\gamma = 5$, the allocation to stocks is flat at 40% under the no-break model irrespective of whether parameter estimation uncertainty is considered. To see why, note that while shocks to the dividend yield and stock returns are strongly negatively correlated and thus give rise to a hedging demand for stocks that grows with the investment horizon, shocks to the short rate and stock returns are−on average−largely uncorrelated, see Table 5. This means that the long-run risk of stocks is perceived to be higher under the T-bill rate model and so helps explain the absence of an increase in the stock allocation, the longer the investment horizon. Raising the risk aversion from $\gamma = 5$ to $\gamma = 10$, the allocation to stocks is reduced to roughly half its previous level.

When past and future breaks are considered, the allocation to stocks declines from a level near 70% at short horizons to only 10% at the five-year horizon. Thus, we continue to find that the level and slope of stock holdings as a function of the investment horizon are highly sensitive to assumptions about model stability.

5.3. *Uncertainty about the number of historical breaks*

We next follow the analysis in Section 4 and integrate out uncertainty about the number of historical (in-sample) breaks. The effect of using this approach is illustrated in Figure 6. This figure compares plots of optimal stock holdings under a forecasting model that assumes eight historical breaks for the dividend yield model (or five breaks for the T-bill rate model) against holdings computed under Bayesian Model Averaging which considers all the models displayed in Table 2 with identical prior weights on each of these models. For both predictor variables the allocations calculated under the models with the highest posterior probabilities versus those calculated under the model averaging approach are virtually identical. This is to be expected given the very high posterior probability assigned to the top models shown in Table 2.

5.4. *Model Uncertainty*

Figure 6 also shows the effect of accounting for model uncertainty in a simple experiment that combines the return forecasting models listed in Table 2 and includes either the dividend yield or the T-bill rate as a predictor variable, weighting the models according to equation (27). In total, predictive densities across 20 different models are thus considered.

Optimal stock holdings most resemble the allocation under the five-break forecasting model based on the T-bill rate. This happens because this model achieves a better fit than the best model based on the dividend yield and hence gets a greater weight in the forecast combination. If we considered a larger universe−in particular models with more than one predictor variable−it is less likely that any particular model would dominate in the way we find here.

This analysis is only meant to illustrate how our approach can be extended to account for model uncertainty. In reality, model uncertainty is far greater than shown here due to the typically large dimension of the set of possible predictor variables. For example, Avramov (2002) considers 2^{14} different models.

6. **Sensitivity Analysis and Extensions**

6.1. *Robustness to Priors*

We next investigate the sensitivity of our empirical results with regard to the assumed priors. 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. If we make $\underline{V}_{\beta}$ larger than the value assumed in the empirical analysis, the posterior distribution of the parameter estimates within each regime gets more dispersed, whereas the location of the breaks is less affected. Conversely, for smaller values of $\underline{V}_{\beta}$, the parameter estimates in the various regimes become more similar than suggested by the empirical results since this reduces the variation in the posterior mean of the parameter estimates across regimes. The choice of $\underline{V}_{\beta}$ in the empirical analysis 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 the constraint that the predictor variable, x_t, is stationary ($0 < \beta_x < 1$) does not have much effect on the results. An additional parameter constraint which requires the unconditional mean excess stock return within each regime to fall between zero and 1% per month ($0 \leq \mu_r + \frac{\beta_r \mu_x}{1-\beta_x} \leq 0.01$) also does not affect the results in any major way.

Following the analysis in Stambaugh (1999), we consider using an informative prior that centers the correlation between innovations in the return and the dividend yield equations, $\lambda_{j,1,2}$ on -0.9, with $\underline{\mu}_{\mu,1,2} = -0.9$, $\underline{\tau}^2_{1,2} = 0.00001$, $\underline{a}_{\rho,1,2} = 1000$ and $\underline{b}_{\rho,1,2} = 100$. This again makes little difference to the results, although at short horizons of up to six months, the allocation to stocks is reduced slightly compared to the case with uninformative priors on this parameter.

Our benchmark analysis uses an informative prior for the autoregressive parameter that is centered at 0.9. We have also analyzed a model with an uninformative prior for this coefficient and find that this has little effect on the posterior parameter estimates and asset allocation results.

We finally estimate a model that imposes identical transition probability parameters across states by setting $p_{k,k} = p$ for all k. Such a restriction is natural to consider since we effectively only have one observation for each $p_{k,k}$. Notice that this restriction changes the structure of the prior for the transition probability matrix, P, since we no longer have a hierarchical prior on a, b. This modification has little effect on the results. For example, for the dividend yield model with eight breaks, two of the estimated break dates (1940 and 1958) change a little (to 1943 and 1954, respectively) but the estimated slope coefficients of the dividend yield in the return equation are very similar to the baseline case in Table 3 and asset allocations are basically unchanged. Similar results prevail for the return prediction model based on the T-bill rate, for which the parameter estimates and break dates in a five-break model change very little when imposing that $p_{k,k} = p$.

6.2. *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 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 7 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 volatility of returns varies significantly from close to 10% per month around the Great Depression to 3-4% per month in the middle of the sample. This finding shows 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 returns.

6.3. *Welfare Costs from Ignoring Breaks*

To quantify the economic costs of ignoring breaks, we undertook the following exercise. For each of the return prediction models we simulated returns under the assumption that the true data generating process corresponds to the breakpoint models reported in Tables 3-4 and 5-6, respectively. We then computed the optimal asset allocation under the three different model specifications considered earlier, i.e., (i) no breaks without parameter uncertainty; (ii) no breaks with parameter uncertainty; and (iii) past and future breaks. This exercise thus addresses what loss an investor would occur if the true return process experiences breaks, but the investor is forced to hold a portfolio that is optimized under a model that ignores breaks.

Table 7 reports results in the form of annualized certainty equivalent returns, a common measure of performance that allows us to measure the economic loss from using a misspecified return prediction model, in this case one that ignores breaks. Once again we compute results for risk aversion coefficients of $\gamma = 5$ and $\gamma = 10$ and we set the values of the predictor variables close to their sample means.

First consider the model based on the dividend yield. When $\gamma = 5$, the loss in certainty equivalent return is quite modest at horizons up to one year. However, the loss quickly grows as the horizon expands. In fact, the models that ignore breaks generate negative certainty equivalent returns at horizons of 30 months or longer, compared with a certainty equivalent return around 5% for the model that accounts for breaks.

The finding that the cost of ignoring breaks increases with the investment horizon can be explained as follows. First, the probability of a break increases, the longer the investment horizon. Second, the difference between the allocation to stocks under the no-break and break models increases as the horizon grows, with the no-break models allocating far too much to stocks. As a result, under the assumption that breaks do in fact affect the return generating process, the no-break models lead investors to take on too much risk and result in far lower certainty equivalent returns. This finding is particularly strong under the dividend yield prediction model since this assumes strong mean reversion in returns and so leads investors to underestimate the long-run risks from holding stocks. Accounting for parameter uncertainty helps investors a bit, although it only reduces the loss in certainty equivalent returns by around 2% compared to the case where parameter uncertainty is ignored. Notice also that losses from ignoring breaks are much smaller when investors are more risk averse ($\gamma = 10$) since this leads them to naturally dampen their allocation to stocks and, and as a consequence, their over-exposure to stocks is greatly reduced.

Turning to the return prediction model based on the T-bill rate, Table 7 shows much lower losses in certainty equivalent returns. Losses are now only a few basis points for investment horizons shorter than one year. Although losses due to ignoring breaks grow as the investment horizon expands, they remain modest at just above one percent per annum for $\gamma = 5$ even at the five-year horizon. Once again losses are lower when the risk aversion is increased from $\gamma = 5$ go $\gamma = 10$.

7. Conclusion

This paper provided an analysis of the stability of return prediction models and the asset allocation implication of breaks to model parameters. Our analysis accounts for several sources of uncertainty, namely (i) parameter uncertainty; (ii) model uncertainty; (iii) uncertainty about the number, location and size of historical breaks to model parameters; and (iv) uncertainty about future (out-of-sample) breaks.

Our empirical results suggest that the parameters of standard forecasting models are highly unstable and subject to multiple breaks, many of which coincide with important historical events. Such breaks compound the effect of parameter estimation uncertainty. Moreover, we find that the possibility of past and future breaks has a large impact on investors' optimal asset allocation. Overall, we conclude that instabilities in standard return prediction models can imply long-run risk estimates that are far greater than usually perceived and hence make stocks less attractive to long-run investors.

References

Ang, A., and G. Bekaert (2002): "Regime Switches in Interest Rates," *Journal of Business and Economic Statistics*, 20, 163–182.

——— (2007): "Stock return predictability: Is it there?," *Review of Financial Studies*, 20(3), 651.

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. Lumsdaine, and J. Stock (1998): "Testing For and Dating Common Breaks in Multivariate Time Series," *Review of Economic Studies*, 65, 394–432.

Barberis, N. (2000): "Investing for the Long Run when Returns are Predictable," *Journal of Finance*, 55(2), 225 – 264.

Barsky, R. (1989): "Why don't the prices of stocks and bonds move together?," *American Economic Review*, 79, 1132–1145.

Bossaerts, P., and P. Hillion (1999): "Implementing statistical criteria to select return forecasting models: what do we learn?," *Review of Financial Studiess*, 12(2), 405–428.

Boyd, J.H., J. H., and R. Jagannathan (2005): "The stock market's reaction to unemployment news: Why bad news is usually good for stocks," *Journal of Finance*, 60, 649–672.

Campbell, J., Y. Chan, and L. Viceira (2003): "A Multivariate Model of Strategic Asset Allocation," *Journal of Financial Economics*, 67, 41–80.

Campbell, J., and L. Viceira (2001): "Who Should Buy Long-Term Bonds?," *American Economic Review*, 91, 99–127.

Campbell, J. Y. (1987): "Stock Returns and the Term Structure," *Journal of Financial Economics*, 18, 373–399.

Campbell, J. Y., and R. J. Shiller (1988): "Stock Prices, Earnings, and Expected Dividends," *Journal of Finance*, 43(3), 661–676.

Chib, S. (1996): "Calculating Posterior Distribution and Modal Estimates in Markov Mixture Models," *Journal of Econometrics*, 75, 79–97.

——— (1998): "Estimation and Comparison of Multiple Change Point Models," *Journal of Econometrics*, 86, 221–241.

Cremers, K. (2002): "Stock Return Predictability: A Bayesian Model Selection Perspective," *The Review of Financial Studies*, 15, 1223–1249.

Dangl, T., and M. Halling (2008): "Predictive regressions with time-varying coefficients," Working Paper, Vienna University of Technology.

Dimson, E., P. Marsh, and M. Staunton (2002): *Triumph of the optimists: 101 years of global investment returns*. Princeton Univ Press.

Elliott, G., and U. Mueller (2006): "Efficient Tests for General Persistent Time Variation in Regression Coefficients," *Review of Economic Studies*, 73, 907–940.

Fama, E. F., and K. R. French (1988): "Dividend Yields and Expected Stock Returns," *Journal of Financial Economics*, 22(1), 3–25.

George, E. I., U. E. Makov, and A. F. M. Smith (1993): "Conjugate Likelihood Distributions," *Scandinavian Journal of Statistics*, 20, 147–156.

Geweke, J. (1992): "Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments (Disc: P189-193)," *Bayesian Statistics*, 4, 169–188.

——— (2001): "A note on some limitations of CRRA utility," *Economics Letters*, 71(3), 341–345.

Goyal, A., and I. Welch (2008): "A comprehensive look at the empirical performance of equity premium prediction," *Review of Financial Studies*, 21(4), 1455–1508.

Guidolin, M., and A. Timmermann (2008): "Size and Value Anomalies under Regime Switching," *Journal of Financial Econometrics*, (6), 1–48.

Johannes, M., A. Korteweg, and N. Polson (2009): "Sequential learning, predictive regressions, and optimal portfolio returns," Mimeo, Columbia University.

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, M., and S. Ludvigson (2001): "Consumption, aggregate wealth and expected stock returns," *Journal of Finance*, 56, 815–849.

Lettau, M., and S. van Nieuwerburgh (2008): "Reconciling the return predictability evidence," *Review of Financial Studies*, 21(4), 1607.

Liechty, J. C., M. W. Liechty, and P. Müller (2004): "Bayesian Correlation Estimation," *Biometrika*, 91-1, 1–14.

McQueen, G., and V. Roley (1993): "Stock prices, news, and business conditions," *The Review of Financial Studies*, 6(3), 683–707.

MENZLY, L., T. S., AND P. VERONESI (2004): "Understanding predictability," *Journal of Political Economy*, 112, 1–47.

PASTOR, L., AND R. STAMBAUGH (2001): "The Equity Premium and Structural Breaks," *Journal of Finance*, 56, 1207–1245.

PASTOR, L., AND R. STAMBAUGH (2009): "Are Stocks Really Less Volatile in the Long Run?," *Mimeo, University of Chicago*.

PAYE, B., AND A. TIMMERMANN (2006): "Instability of Return Prediction Models," *Journal of Empirical Finance*, 13, 274–315.

PEREZ-QUIROS, G., AND A. TIMMERMANN (2000): "Firm Size and Cyclical Variations in Stock Returns," *Journal of Finance*, 55, 1229–1262.

PESARAN, M., D. PETTENUZZO, AND A. TIMMERMANN (2006): "Forecasting Time Series Subject to Multiple Break Points," *Review of Economic Studies*, (73), 1057–1084.

PESARAN, M. H., AND A. TIMMERMANN (1995): "Predictability of Stock Returns: Robustness and Economic Significance," *Journal of Finance*, 50(4), 1201–1228.

RAFTERY, A., AND S. LEWIS (1992a): "How many iterations in the Gibbs sampler," *Bayesian statistics*, 4(2), 763–773.

RAFTERY, A., AND S. LEWIS (1992b): "One Long Run with Diagnostics: Implementation Strategies for Markov Chain Monte Carlo," *Statistical Science*, 7(4), 493–497.

——— (1995): "The number of iterations, convergence diagnostics and generic Metropolis algorithms," *Practical Markov Chain Monte Carlo*.

STAMBAUGH, R. (1999): "Predictive Regressions," *Journal of Financial Economics*, 54, 375–421.

WACHTER, J., AND M. WARUSAWITHARANA (2009): "Predictable returns and asset allocation: Should a skeptical investor time the market?," *Journal of Econometrics*, 148(2), 162–178.

Technical Appendices

Appendix 1. Gibbs sampler for the return prediction model with multiple breaks

This appendix extends results in Pesaran, Pettenuzzo, and Timmermann (2006) 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

$$\Theta = \left(vec(B)_1, \psi_1, \Lambda_1, ..., vec(B)_{K+1}, \psi_{K+1}, \Lambda_{K+1}\right)$$

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

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

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, $p = (p_{1,1}, p_{2,2}, ..., p_{K+1,K+1})'$ 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, the meta hyperparameters, H, and the elements of the transition probability matrix, P; next, the parameters and hyperparameters of the meta distributions are simulated conditional on the data and $\mathcal{S}_T$. Specifically, the Gibbs sampler is implemented by simulating the following set of conditional distributions: $\pi\left(\mathcal{S}_T \middle| \Theta, H, p, \mathcal{Z}_T\right)$, $\pi\left(\Theta, H \middle| p, \mathcal{S}_T, \mathcal{Z}_T\right)$, $\pi\left(p \middle| \mathcal{S}_T\right)$.[15]

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 $t+1$ to T, respectively. We partition the joint density of the states as follows:

$$p(s_{T-1}| s_T, \Theta, H, p, \mathcal{Z}_T) \times \cdots \times p(s_t| \mathcal{S}^{t+1}, \Theta, H, p, \mathcal{Z}_T) \times \cdots \times p(s_1| \mathcal{S}^2, \Theta, H, p, \mathcal{Z}_T). \tag{28}$$

Chib (1996) shows that the generic element of (28) can be decomposed as follows

$$p(s_t| \mathcal{S}^{t+1}, \Theta, H, p, \mathcal{Z}_T) \propto p(s_t| \Theta, H, p, \mathcal{Z}_t) p(s_{t+1}| s_t, p), \tag{29}$$

where the normalizing constant is easily obtained since s_t takes only two values conditional on the value taken by s_{t+1}. The last term in (29) 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 a given $p(s_{t-1}| \Theta, H, p, \mathcal{Z}_{t-1})$, we obtain $p(s_t| \Theta, H, p, \mathcal{Z}_t)$ and $p(s_{t+1}| \Theta, H, p, \mathcal{Z}_{t+1})$, and so on until $p(s_T| \Theta, H, p, \mathcal{Z}_T)$. Suppose $p(s_{t-1}| \Theta, H, p, \mathcal{Z}_{t-1})$ is available. From Chib (1998),

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

[15] We use 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, $\pi\left(p \middle| \Theta, H, \mathcal{S}_T, \mathcal{Z}_T\right) = \pi\left(p \middle| \mathcal{S}_T\right)$.

for $k = 1, 2, ..., K + 1$, and where $p_{l,k}$ is the Markov transition probability. In addition,

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

For a given set of simulated states, $\mathcal{S}_T$, the data is partitioned into $K + 1$ groups. Let $Z_k = \left(z'_{\tau_{k-1}+1}, ..., z'_{\tau_k}\right)'$ and $X_k = \left(z'_{\tau_{k-1}}, ..., z'_{\tau_k - 1}\right)'$ be the values of the dependent and independent variables within the kth regime. To obtain the conditional distributions for the regression parameters and hyperparameters, note that the conditional distributions of $vec(B)_k$ are independent across regimes with[16]

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

where

$$\begin{aligned} \overline{V}_k &= \left(X'_k \Sigma_k^{-1} X_k + V_0^{-1}\right)^{-1}, \\ \overline{vec(B)}_k &= \overline{V}_k \left(X'_k \Sigma_k^{-1} Z_k + V_0^{-1} b_0\right). \end{aligned}$$

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

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

where

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

and

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

Moving to the posterior for the precision parameters within each regime k and for each equation i, let $\Xi = (Z_k - X_k B_k)'(Z_k - X_k B_k)$ with $\Xi_{i,j}$ being its i-th row and j-th column element. Note that

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

where n_k is the number of observations assigned to regime k.

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

Location and scale parameters for the error term precision of each equation are then updated as follows:[17]

$$v_{0,i}|\,\Theta, H_{-v_{0,i}}, p, \mathcal{S}_T, \mathcal{Z}_T \propto \prod_{k=1}^{K+1} G\left(s_{k,i}^{-2}\middle|\, v_{0,i}, d_{0,i}\right)\exp\left(-\underline{\rho}_{0,i}\right), \tag{30}$$

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

The full conditional densities for $\mu_{\rho,i,c}$ and $\sigma^2_{\rho,i,c}$ are similar to conjugate densities with an additional factor due to the constraint requiring Λ_k to be positive definite (we write $\mathcal{R}^m$ to identify the space of all correlation matrices of dimension m):

$$\begin{aligned} f\left(\mu_{\rho,i,c}\middle|\,\Theta, H_{-\mu_{\rho,i,c}}, p, \mathcal{S}_T, \mathcal{Z}_T\right) \quad \propto \quad & \prod_{k=1}^{K+1}\exp\left\{-\left(\lambda_{k,i,c}-\mu_{\rho,i,c}\right)^2/\left(2\sigma^2_{\rho,i,c}\right)\right\} \\ & \times\exp\left\{-\left(\mu_{\rho,i,c}-\underline{\mu}_{\mu,i,c}\right)^2/\left(2\underline{\tau}^2_{i,c}\right)\right\} I\left\{\Lambda_k\in\mathcal{R}^m\right\}, \end{aligned} \tag{31}$$

$$\begin{aligned} f\left(\sigma^2_{\rho,i,c}\middle|\,\Theta, H_{-\sigma^2_{\rho,i,c}}, p, \mathcal{S}_T, \mathcal{Z}_T\right) \quad \propto \quad & \prod_{k=1}^{K+1}\exp\left\{-\left(\lambda_{k,i,c}-\mu_{\rho,i,c}\right)^2/\left(2\sigma^2_{\rho,i,c}\right)\right\} \\ & \sigma_{\rho,i,c}^{2\left(1-\underline{a}_{\rho,i,c}\right)}\exp\left(-\underline{b}_{\rho,i,c}/\sigma^2_{\rho,i,c}\right) I\left\{\Lambda_k\in\mathcal{R}^m\right\}. \end{aligned} \tag{32}$$

The posterior distributions of the correlation coefficients within each regime, $\lambda_{k,i,c}$, and of the hyperparameters $\mu_{\rho,i,c}$ and $\sigma^2_{\rho,i,c}$ are non standard so sampling is accomplished using a Griddy Gibbs sampling step inside the main Gibbs sampling algorithm.

Finally, $p_{k,k}$ is simulated from the conditional beta posterior

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

where $l_k = \tau_k - \tau_{k-1} - 1$ is the duration of regime k. The posterior distribution for the hyperparameters a and b in 12 and 13 is not conjugate so sampling is accomplished using a Metropolis-Hastings step.

Appendix 2. Algorithm for generating draws of returns under breaks

This appendix describes the steps used to obtain draws from the predictive distribution of returns that account for uncertainty about past and future breaks.

For each draw j from the Gibbs sampler ($j = 1, ..., J$),

1. First, obtain a draw for $p^j_{K+1,K+1}$ from its posterior distribution. This is achieved by combining the information from the last regime with the prior information for $p_{k,k}$ in (6)

$$p^j_{K+1,K+1}\Big|\,\mathcal{S}^j_T \sim Beta(a+l^j_{K+1}, b+1),$$

[17] See George, Makov, and Smith (1993), pp. 154-155. Drawing $v_{0,i}$ from (30) is complicated since we cannot make use of standard distributions so we use an adaptive rejection sampling step in the Gibbs sampling algorithm.

where $l^j_{K+1} = T - \tau^j_K - 1$ is the number of observations in regime $K+1$ in round j of the Gibbs sampler. Hence the probability that at time $T+s$ $(1 \leq s \leq h)$ the jth draw of the Gibbs sampler remains in regime $K+1$, conditional on $\mathcal{S}^j_{T+s-1}$, is $p^j_{K+1,K+1}$, while the probability of moving to a new regime is $1 - p^j_{K+1,K+1}$.

Next, for each period $T+s$ $(1 \leq s \leq h)$ proceed as follows:

2. Draw a realization, U^j_{T+s}, from a uniform distribution, $U^j_{T+s} \sim U[0,1]$. If $U^j_{T+s} \leq p^j_{K+1,K+1}$, the sampler remains in the current regime; if $U^j_{T+s} > p^j_{K+1,K+1}$ the sampler moves to the next regime.

2a. If $U^j_{T+s} \leq p^j_{K+1,K+1}$, stay in the current regime and draw returns from

$$r^j_{T+s} \sim p\left(r_{T+s} \middle| \Theta^j_{K+1}, H^j, p^j, S^j_{T+s} = K+1, \mathcal{Z}_T\right).$$

Then go back to step 2 after incrementing the time indicator, s, by one.

2b. If $U^j_{T+s} > p^j_{K+1,K+1}$, start by drawing a new set of hyperparameters H^j from their meta distributions in equations (7)-(11). Next, draw B^j_{K+2} and Σ^j_{K+2} from $\pi\left(B^j_{K+2} \middle| H^j, \mathcal{Z}_T\right)$ and $\pi\left(\Sigma^j_{K+2} \middle| H^j, \mathcal{Z}_T\right)$, respectively. Finally, draw returns from the posterior predictive density,

$$r^j_{T+s} \sim p\left(r_{T+s} \middle| \Theta^j_{K+2}, H^j, S^j_{T+s} = K+2, \mathcal{Z}_T\right),$$

3. If $U^j_{T+s} > p^j_{K+1,K+1}$, so a break occurred, draw a^j and b^j from their conditional posterior distributions in (12) and (13). Generate a draw for $p^j_{K+2,K+2}$ using the prior distribution for $p_{k,k}$ in (6) and a^j and b^j,[18]

$$p^j_{K+2,K+2}\Big| a^j, b^j \sim Beta(a^j, b^j).$$

Then go back to step 2 of the algorithm after incrementing the time indicator, s, by one and increasing the regime indicator, currently set to its last in-sample value, $K+1$, by one.

4. When $s = h$, add returns from periods $T+1, ..., T+h$ to get the cumulated return over the investment horizon. Repeating across $j = 1, .., J$ we obtain $p(R_{T+h}|S_T = K+1, \mathcal{Z}_T)$.

[18] 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 $p^j_{K+2,K+2}$.

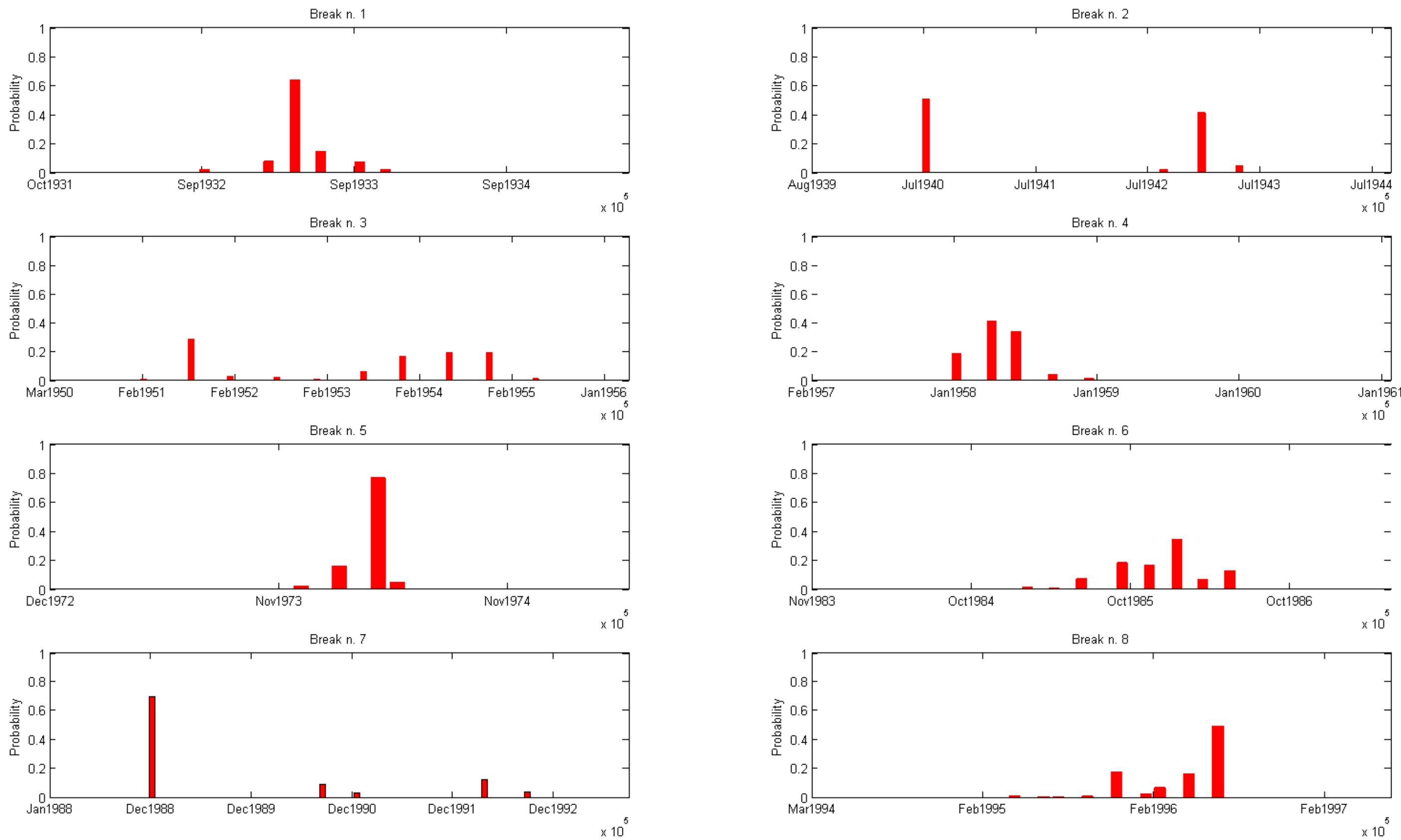


Figure 1: Posterior probabilities for breakpoint locations in the return prediction model with eight breaks based on the dividend yield.

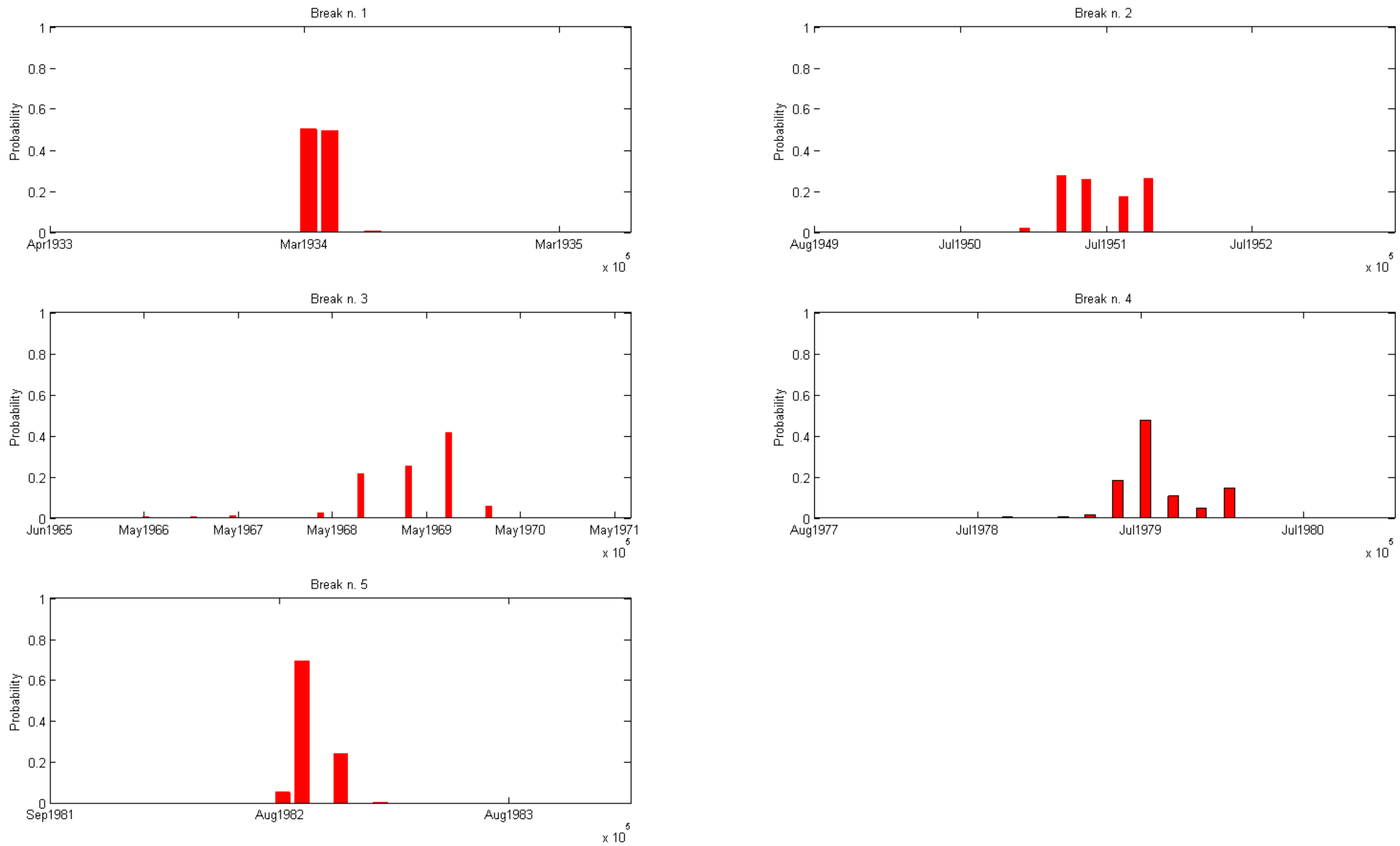


Figure 2: Posterior probabilities for breakpoint locations in the return prediction model with five breaks based on the T-bill rate.

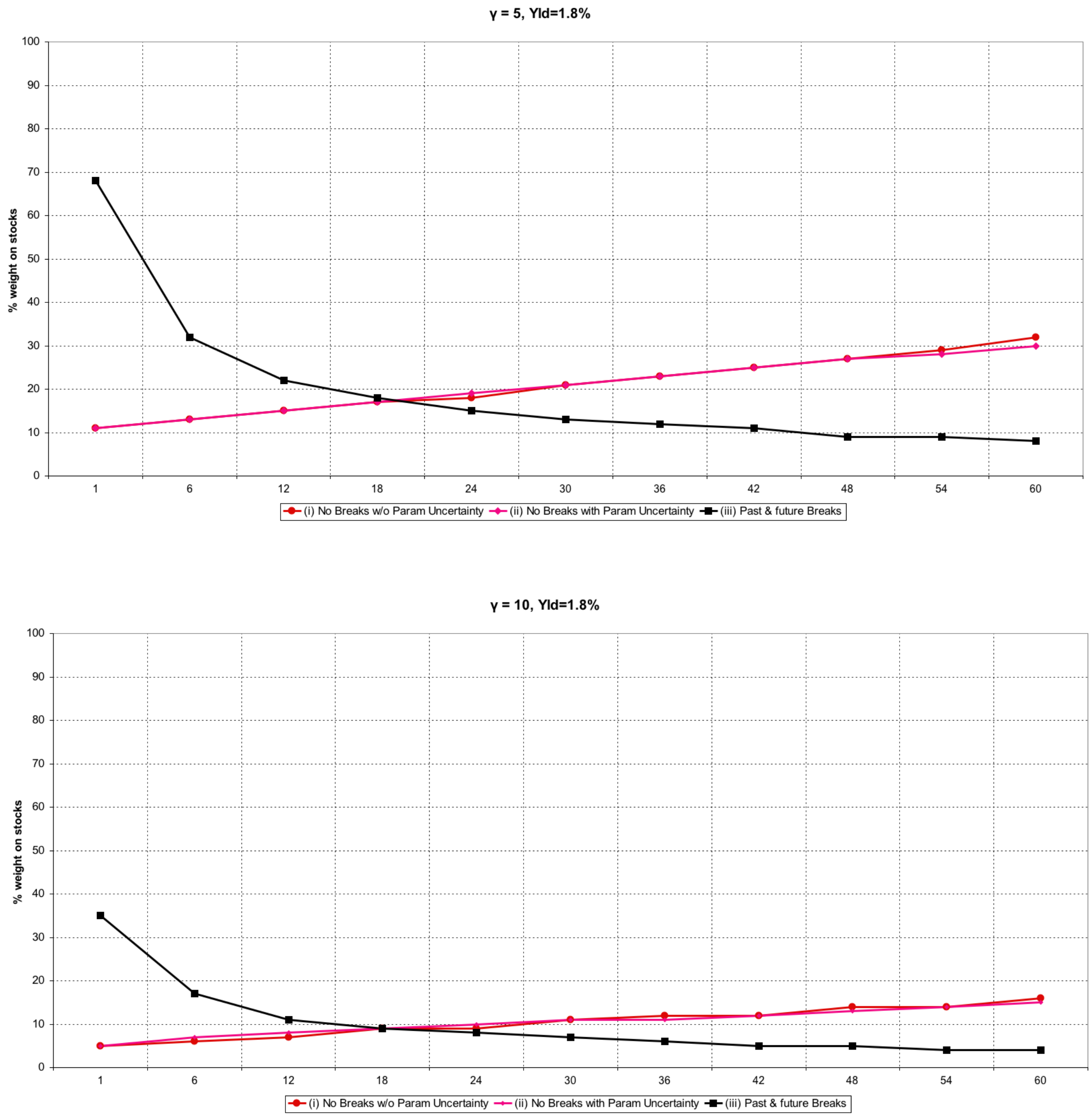


Figure 3: Optimal Asset Allocation as a function of the investment horizon for a buy-and-hold investor with power utility over terminal wealth, $U(W_{T+h}) = \frac{1}{1-\gamma}W_{T+h}^{1-\gamma}$, where h is the forecast horizon and γ is the coefficient of relative risk aversion. The calculations use the dividend yield as a predictor variable. The panels show percentage allocations to stocks plotted against the investment horizon measured in months under the assumption that the dividend yield is set at its value at the end of the sample, i.e., 1.8%. The upward sloping curves track stock allocations under no breaks, while the downward sloping curves allow for past and future breaks.

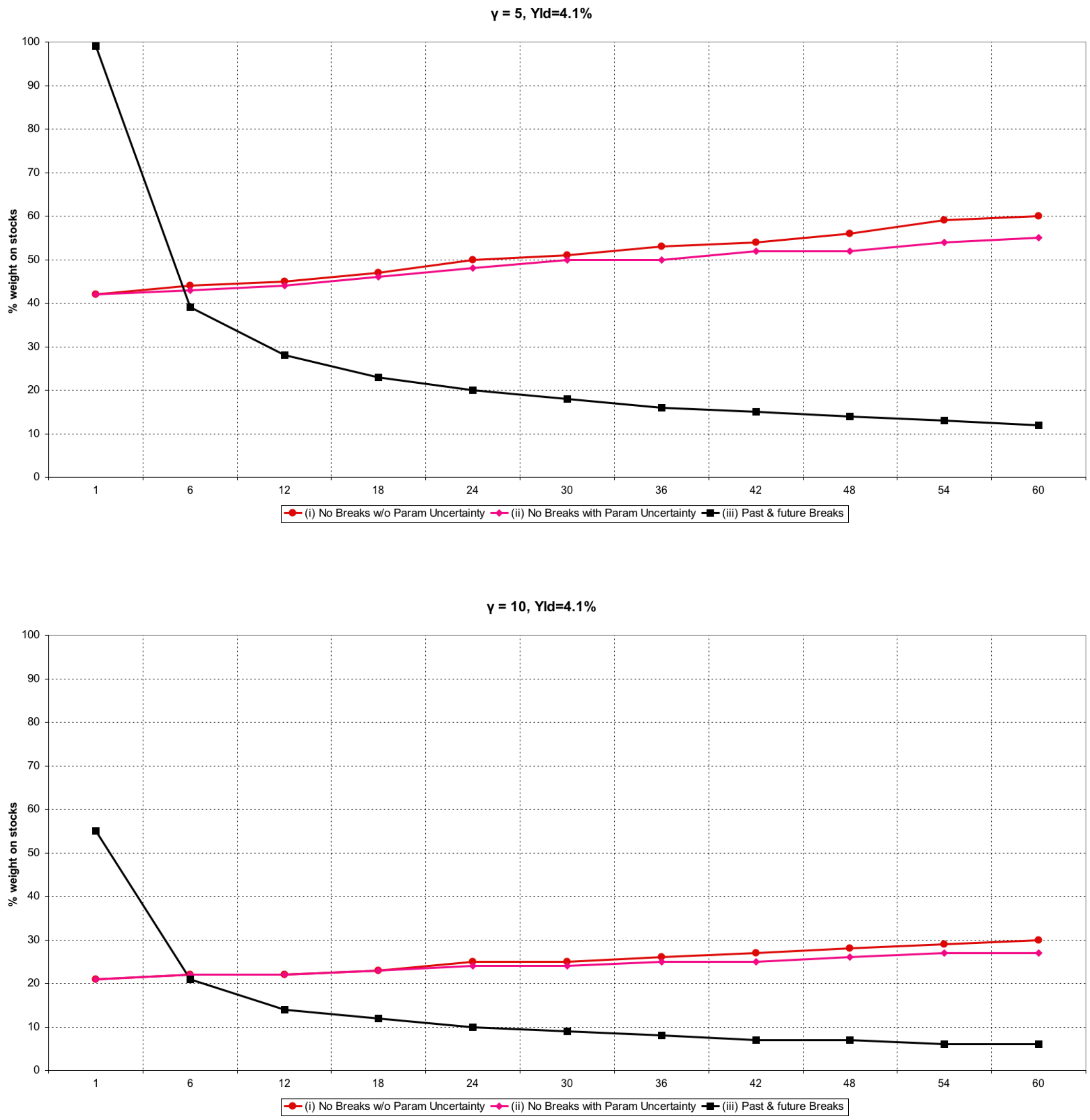
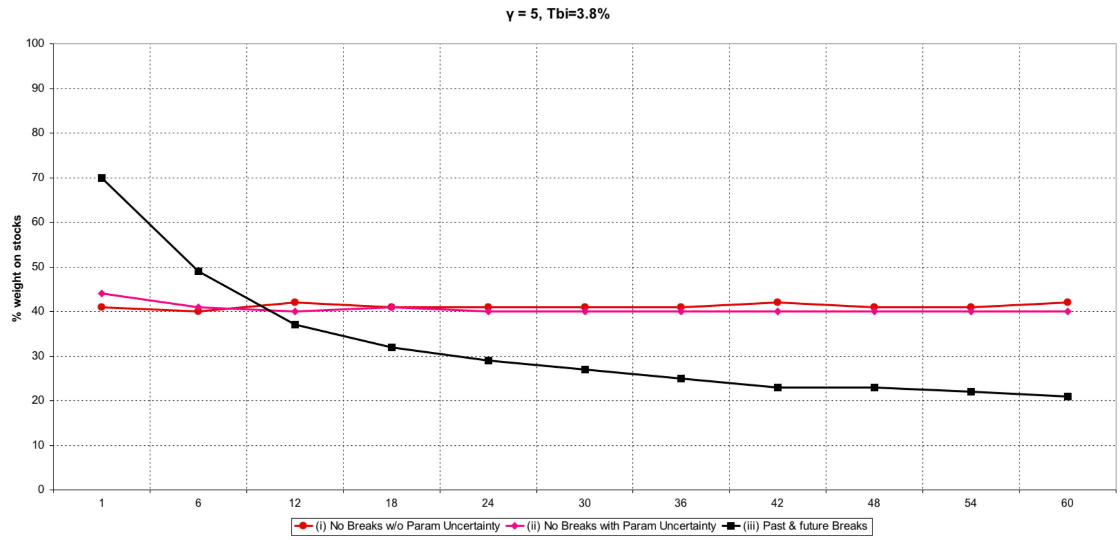


Figure 4: Optimal Asset Allocation as a function of the investment horizon for a buy-and-hold investor with power utility over terminal wealth, $U(W_{T+h}) = \frac{1}{1-\gamma} W_{T+h}^{1-\gamma}$, where h is the forecast horizon and γ is the coefficient of relative risk aversion. The calculations use the dividend yield as a predictor variable. The panels show percentage allocations to stocks plotted against the investment horizon measured in months under the assumption that the dividend yield is set at its mean value, i.e., 4.1%. The upward sloping curves track stock allocations under no breaks, while the downward sloping curves allow for past and future breaks.



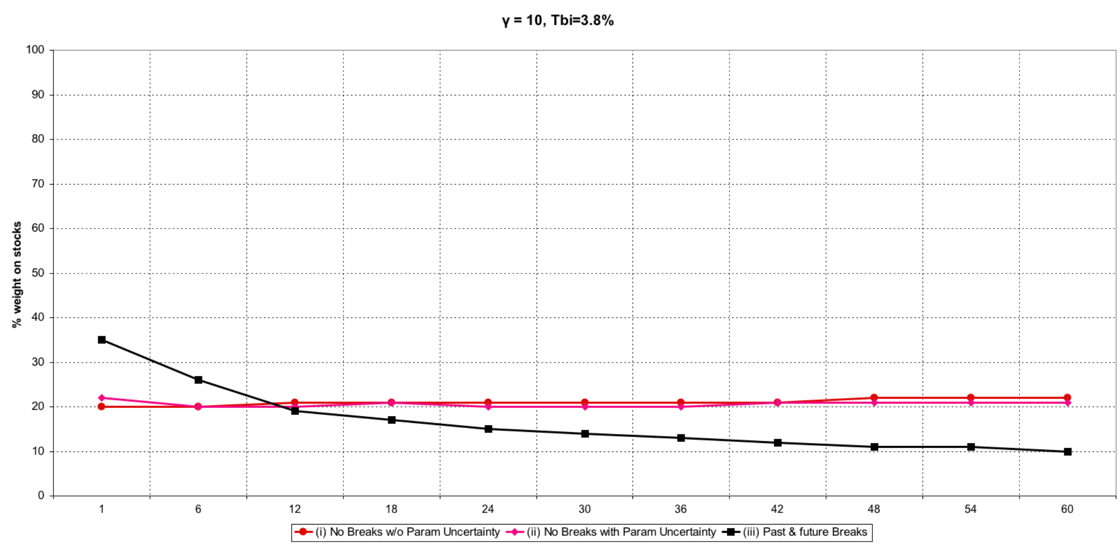
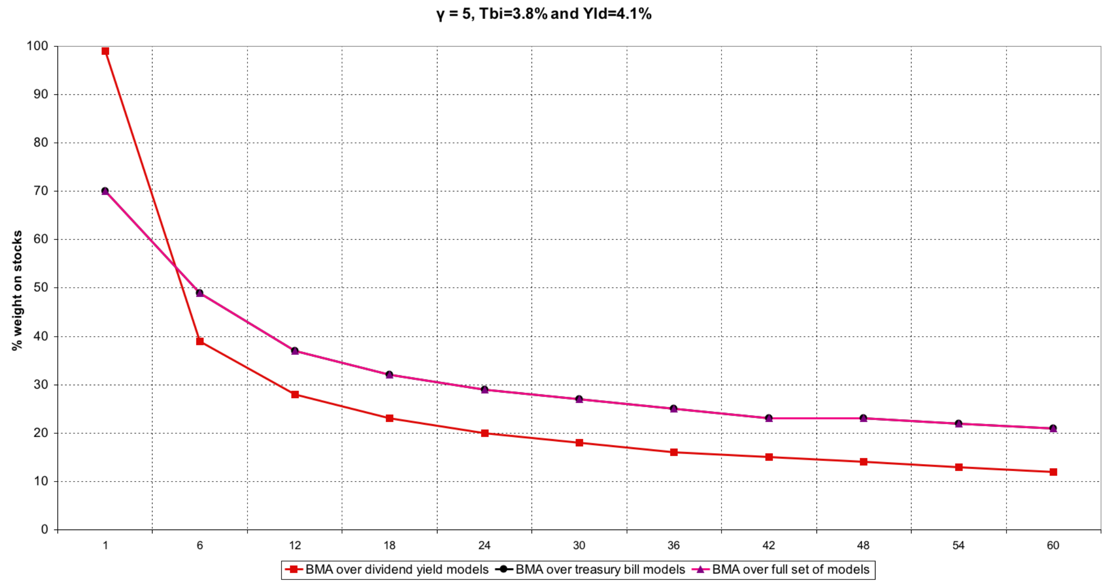


Figure 5: Optimal Asset Allocation as a function of the investment horizon for a buy-and-hold investor with power utility over terminal wealth, $U(W_{T+h}) = \frac{1}{1-\gamma} W_{T+h}^{1-\gamma}$, where h is the forecast horizon and γ is the coefficient of relative risk aversion. The calculations use the T-bill rate as a predictor variable. The panels show percentage allocations to stocks plotted against the investment horizon measured in months under the assumption that the T-bill rate is set at its value at the end of the sample, i.e., 3.8% per annum (0.03% per month). Flat curves track stock allocations under no breaks, while the downward sloping curves allow for past and future breaks.



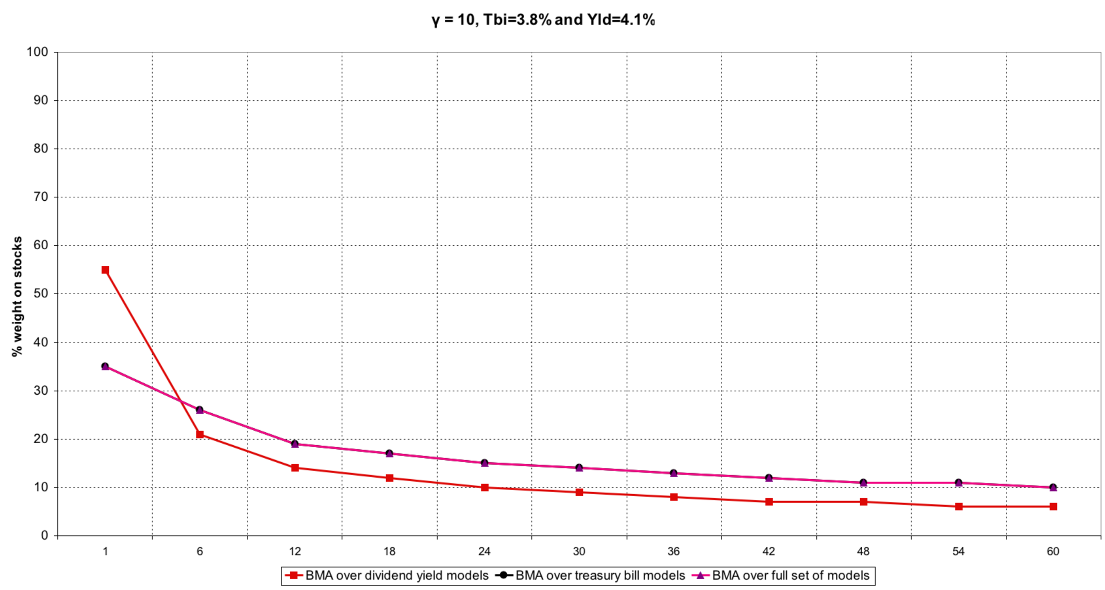
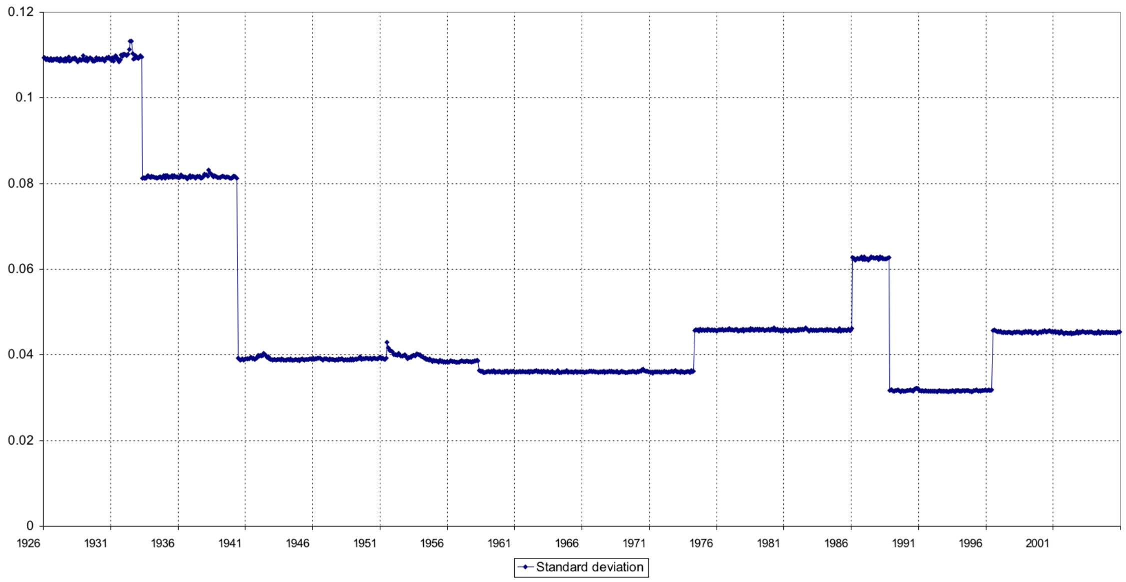


Figure 6: Optimal asset allocation computed under Bayesian model averaging, considering the dividend yield specification with up to 10 breaks (BMA over dividend yield modes), the T-bill rate specification with up to 8 breaks (BMA over treasury bill models) and the combined set of all dividend yield and T-bill rate return prediction models.



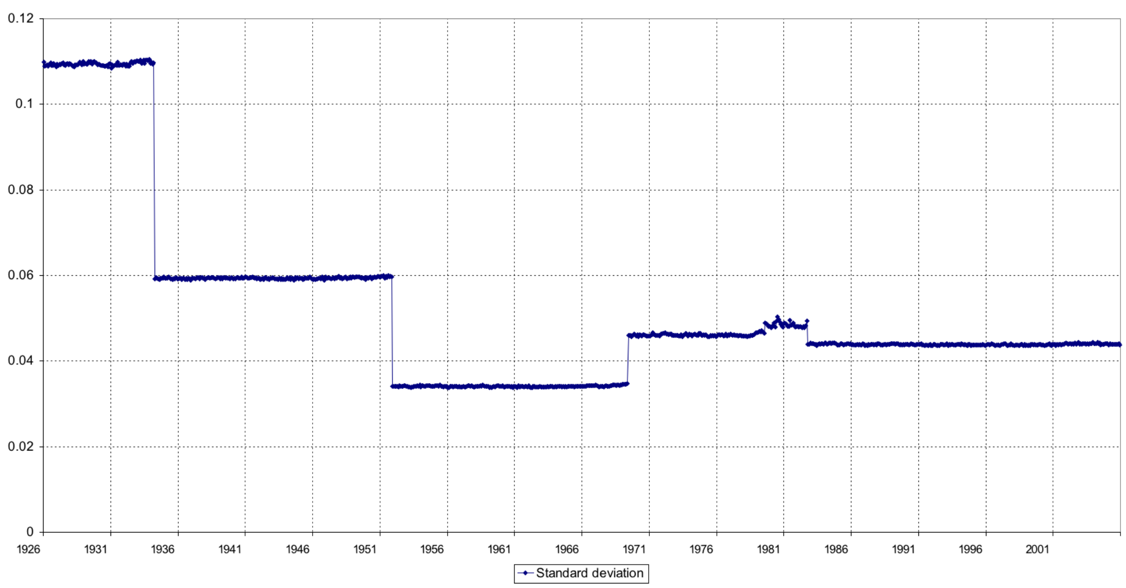


Figure 7: Standard deviations of the predictive distribution of stock returns (in decimals per month) when the predictor variable is the dividend yield (top panel) or the T-bill rate (bottom panel) under models with eight and five breaks, respectively.

Statistic	Excess returns	Dividend yield	Treasury bill
Mean	0.0050	0.0408	0.0030
St. Deviation	0.0558	0.0170	0.0024
Kurtosis	7.9662	4.1017	0.9237
Skewness	-0.3788	1.1463	0.9317
Minimum	-0.3391	0.0108	8.3E-06
Maximum	0.3478	0.1536	0.0126

Table 1: Summary statistics for the excess return, dividend yield and T-bill rate data used throughout the paper. Mean, standard deviation, coefficient of kurtosis, coefficient of skewness, minimum and maximum values are reported for each series.

I. Excess Returns - Dividend Yield

Number of breaks	Joint log lik.	Approx. marg. lik.	Posterior probability	Break locations				
0	6188.4	6164.4	0.00000					
1	6866.7	6739.7	0.00000	Feb-43				
2	6892.1	6737.6	0.00000	May-42	Jun-45			
3	7223.9	7041.9	0.00000	Jun-40	May-54	Mar-95		
4	7321.4	7112.0	0.00000	Jun-40	May-58	May-74	Mar-95	
5	7403.8	7166.8	0.00000	Jun-40	May-58	May-74	Nov-85	Mar-95
6	7434.9	7170.5	0.00008	May-33	Feb-43	May-58	May-74	Nov-85
				Mar-95				
7	7455.6	7163.7	0.00000	May-33	Jun-40	Jul-51	May-58	May-74
				Nov-85	Mar-95			
8	7499.3	7180.0	0.99984	May-33	Jun-40	Jul-51	May-58	May-74
				Feb-86	Nov-88	Jul-96		
9	7516.6	7169.8	0.00004	May-33	Jul-40	Feb-43	Jul-58	Aug-66
				May-74	Feb-86	Nov-88	Jul-96	
10	7544.3	7170.0	0.00005	May-33	Jul-40	Apr-43	Jul-51	May-58
				May-74	Nov-82	Feb-86	Nov-88	Jul-96

II. Excess Returns - Treasury Bill Rate

Number of breaks	Joint log lik.	Approx. marg. lik.	Posterior probability	Break locations				
0	7858.2	7834.1	0.00000					
1	8046.6	7919.6	0.00000	Jun-69				
2	8306.2	8151.7	0.00000	Apr-34	Dec-52			
3	8337.2	8155.2	0.00000	Jun-40	Jun-69	Jul-85		
4	8550.8	8341.4	0.00029	Apr-34	Dec-51	Aug-79	Oct-82	
5	8586.5	8349.5	0.99971	Apr-34	Dec-51	Jun-69	Aug-79	Oct-82
6	8598.2	8333.8	0.00000	Apr-34	Dec-51	Jun-69	Aug-79	Nov-82
				Jul-85				
7	8620.9	8329.0	0.00000	Apr-34	Jul-47	Dec-51	Jun-69	Aug-79
				Nov-82	Jul-85			
8	8653.0	8333.7	0.00000	Apr-34	Jul-40	Dec-51	Nov-60	Jul-66
				Aug-79	Nov-82	Jul-85		

Table 2: Model comparison and selection of the number of breaks in the return forecasting models. The table shows estimates of the joint log-likelihood for stock returns and the predictor variable (either the dividend yield or the T-bill rate), approximate marginal likelihood values and approximate posterior probabilities for models with different numbers of breaks along with the posterior modes for the time of the break points. 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.

	Regimes									
	Full sample	**26-33**	**33-40**	**40-51**	**51-58**	**58-74**	**74-86**	**86-88**	**88-96**	**96-05**
					$\boldsymbol{\mu_r}$					
mean	-0.0033	0.0037	0.0052	-0.0289	-0.0236	-0.0355	-0.0284	-0.0291	-0.0263	-0.0377
s.d.	0.0046	0.0139	0.0162	0.0163	0.0208	0.0159	0.0147	0.0282	0.0176	0.0168
					$\boldsymbol{\beta_r}$					
mean	0.2018	-0.0153	0.0512	0.7042	0.8321	1.2070	0.6760	1.0804	1.1649	2.6278
s.d.	0.1039	0.2683	0.3289	0.3030	0.5281	0.5009	0.3112	0.8222	0.5895	1.0458
					$\boldsymbol{\sigma_r}$					
mean	0.0558	0.1082	0.0806	0.0386	0.0378	0.0359	0.0455	0.0609	0.0308	0.0449
s.d.	0.0013	0.0084	0.0087	0.0026	0.0041	0.0019	0.0028	0.0090	0.0050	0.0031
					$\boldsymbol{\mu_x \times 100}$					
mean	0.0830	0.2028	0.2063	0.1605	0.1889	0.1815	0.1903	0.1816	0.1382	0.0503
s.d.	0.0290	0.0818	0.0791	0.0637	0.0678	0.0545	0.0684	0.0711	0.0532	0.0265
					$\boldsymbol{\beta_x}$					
mean	0.9790	0.9567	0.9571	0.9722	0.9484	0.9429	0.9591	0.9416	0.9472	0.9663
s.d.	0.0066	0.0189	0.0185	0.0117	0.0172	0.0172	0.0146	0.0231	0.0191	0.0165
					$\boldsymbol{\sigma_x \times 100}$					
mean	0.3533	0.9016	0.4976	0.2305	0.1627	0.1208	0.2256	0.2114	0.0991	0.0717
s.d.	0.0082	0.0707	0.0476	0.0182	0.0180	0.0061	0.0136	0.0327	0.0188	0.0047
					$\boldsymbol{\rho_{rx}}$					
mean	-0.8807	-0.9359	-0.9399	-0.9290	-0.9529	-0.9732	-0.9734	-0.9324	-0.9666	-0.9547
s.d.	0.0074	0.0222	0.0207	0.0248	0.0217	0.0131	0.0131	0.0232	0.0160	0.0193
					$\boldsymbol{p}$					
mean		0.9857	0.9866	0.9895	0.9802	0.9920	0.9895	0.9763	0.9837	—
s.d.		0.0110	0.0102	0.0081	0.0150	0.0061	0.0081	0.0189	0.0130	—

Table 3: Parameter estimates for the return (r_t) forecasting model with eight break points, based on the lagged dividend yield (x_{t-1}) as a predictor variable: $r_t = \mu_{r_k} + \beta_{r_k} x_{t-1} + \epsilon_{rt}$, $\epsilon_{rt} \sim N\left(0, \sigma_{r_k}^2\right)$, $x_t = \mu_{x_k} + \beta_{x_k} x_{t-1} + \epsilon_{xt}$, $\epsilon_{xt} \sim N\left(0, \sigma_{x_k}^2\right)$, $\Pr\left(s_t = k \middle| s_{t-1} = k\right) = p_{k,k}$, $corr\left(\epsilon_{rt}, \epsilon_{xt}\right) = \rho_{rx_k}$, $\tau_{k-1} + 1 \leq t \leq \tau_k$, $k = 1, ..., 9$.

Hyperparameters of meta distributions

I Return equation

Mean parameters

	mean	s.d.	95% conf.	interval
$b_0(\mu_r)$	-0.0248	0.0791	-0.1780	0.1331
$b_0(\beta_r)$	0.9264	0.5009	0.0029	2.0354
$\sqrt{(V_0(\mu_r))}$	0.2296	0.0545	0.1497	0.3621
$\sqrt{(V_0(\beta_r))}$	1.1151	0.3246	0.6512	1.8838

II Dividend Yield equation

Mean parameters

	mean	s.d.	95% conf.	interval
$b_0(\mu_x) \times 100$	0.1673	0.0508	0.0761	0.2752
$b_0(\beta_x)$	0.9478	0.0309	0.8806	0.9953
$\sqrt{(V_0(\mu_x))} \times 100$	0.0688	0.0232	0.0372	0.1256
$\sqrt{(V_0(\beta_x))}$	0.1020	0.0234	0.0674	0.1577

Correlation parameters

	mean	s.d.	95% conf.	interval
μ_ρ	-0.9377	0.0341	-0.9878	-0.8606

Transition probability parameters

	mean	s.d.	95% conf.	interval
a_0	34.7707	18.4461	9.7272	75.7877
b_0	0.8126	0.3663	0.2995	1.7041

Table 4: 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 eight historical breaks. Within the kth regime the model is: $z_t = B_k' x_{t-1} + u_t$, where $z_t = (r_t, x_t)'$ is the vector of stock returns and the predictor variable, and $vec(B)_k \sim N(b_0, V_0)$. $\rho_k \sim N(\mu_\rho, \sigma_\rho^2)$ is the correlation between shocks to the dividend yield and shocks to returns in the kth regime, while $p_{k,k} \sim Beta(a_0, b_0)$ is the probability of remaining in the kth regime.

	Full sample	26-34	34-51	51-69	69-79	79-82	82-05
			Regimes				
			$\boldsymbol{\mu}_r$				
mean	0.0081	0.0000	0.0065	0.0250	0.0166	0.0902	0.0042
s.d.	0.0030	0.0174	0.0052	0.0061	0.0169	0.0384	0.0066
			$\boldsymbol{\beta}_r$				
mean	-1.0334	0.1414	3.2724	-7.0505	-3.6461	-9.3972	0.5025
s.d.	0.7537	7.0858	8.7202	2.4093	3.3379	3.9141	1.4710
			$\boldsymbol{\sigma}_r$				
mean	0.0559	0.1083	0.0591	0.0339	0.0456	0.0470	0.0437
s.d.	0.0013	0.0078	0.0029	0.0017	0.0029	0.0057	0.0019
			$\boldsymbol{\mu}_x \times \mathbf{100}$				
mean	0.0025	0.0047	0.0007	0.0045	0.0151	0.0484	0.0031
s.d.	0.0013	0.0033	0.0003	0.0022	0.0087	0.0319	0.0020
			$\boldsymbol{\beta}_x$				
mean	0.9922	0.9668	0.9962	0.9886	0.9746	0.9405	0.9902
s.d.	0.0034	0.0163	0.0034	0.0083	0.0176	0.0349	0.0048
			$\boldsymbol{\sigma}_x \times \mathbf{100}$				
mean	0.0295	0.0285	0.0038	0.0169	0.0328	0.1112	0.0181
s.d.	0.0007	0.0021	0.0002	0.0009	0.0023	0.0140	0.0008
			$\boldsymbol{\rho}_{rx}$				
mean	-0.0793	0.0746	0.0181	-0.0336	-0.4010	-0.4736	0.0131
s.d.	0.0321	0.0252	0.0269	0.0280	0.0267	0.0352	0.0239
			$\boldsymbol{p}$				
mean		0.9866	0.9930	0.9929	0.9891	0.9732	—
s.d.		0.0109	0.0054	0.0059	0.0085	0.0225	—

Table 5: Parameter estimates for the return (r_t) forecasting model with five break points, based on the lagged T-bill rate (x_{t-1}) as a predictor variable: $r_t = \mu_{r_k} + \beta_{r_k} x_{t-1} + \epsilon_{rt}$, $\epsilon_{rt} \sim N\left(0, \sigma^2_{r_k}\right)$, $x_t = \mu_{x_k} + \beta_{x_k} x_{t-1} + \epsilon_{xt}$, $\epsilon_{xt} \sim N\left(0, \sigma^2_{x_k}\right)$, $\Pr\left(s_t = k \vert\, s_{t-1} = k\right) = p_{k,k}$, $corr\left(\epsilon_{rt}, \epsilon_{xt}\right) = \rho_{rx_k}$, $\tau_{k-1} + 1 \leq t \leq \tau_k$, $k = 1, ..., 9$.

Hyperparameters of meta distributions

I Return equation

Mean parameters

	mean	s.d.	95% conf.	interval
$b_0(\mu_r)$	0.0252	0.0524	-0.0783	0.1292
$b_0(\beta_r)$	-2.7676	5.8697	-14.6019	9.1074
$\sqrt{(V_0(\mu_r))}$	0.1287	0.0361	0.0787	0.2198
$\sqrt{(V_0(\beta_r))}$	13.6488	4.1496	8.3880	24.0129

II T-bill equation

Mean parameters

	mean	s.d.	95% conf.	interval
$b_0(\mu_x) \times 100$	0.0201	0.0155	0.0009	0.0592
$b_0(\beta_x)$	0.9497	0.0377	0.8571	0.9978
$\sqrt{(V_0(\mu_x))} \times 100$	0.0434	0.0139	0.0257	0.0783
$\sqrt{(V_0(\beta_x))}$	0.1239	0.0363	0.0764	0.2171

Correlation parameters

	mean	s.d.	95% conf.	interval
μ_ρ	-0.1250	0.1714	-0.4541	0.2360

Transition probability parameters

	mean	s.d.	95% conf.	interval
a_0	26.8935	14.4843	6.4051	60.5744
b_0	0.6732	0.3085	0.2348	1.3780

Table 6: 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 five historical breaks. Within the kth regime the model is: $z_t = B_k' x_{t-1} + u_t$, where $z_t = (r_t, x_t)'$ is the vector of stock returns and the predictor variable, and $vec(B)_k \sim N(b_0, V_0)$. $\rho_k \sim N(\mu_\rho, \sigma_\rho^2)$ is the correlation between shocks to the T-bill and shocks to returns in the kth regime, while $p_{k,k} \sim Beta(a_0, b_0)$ is the probability of remaining in the kth regime.

Certainty Equivalent Return Estimates

I. Dividend yield return prediction model

Horizon	No Breaks w/o Param. Uncertainty	No Breaks with Param. Uncertainty	Past & Future Breaks	Horizon	No Breaks w/o Param. Uncertainty	No Breaks with Param. Uncertainty	Past & Future Breaks
	Yld=4.1%; $\gamma = 5$				**Yld=4.1%; $\gamma = 10$**		
1	0.0907	0.0907	0.1226	1	0.0649	0.0649	0.0809
6	0.0712	0.0716	0.0724	6	0.0556	0.0556	0.0558
12	0.0462	0.0483	0.0628	12	0.0451	0.0451	0.0507
18	0.0239	0.0273	0.0591	18	0.0360	0.0360	0.0485
24	0.0004	0.0091	0.0568	24	0.0254	0.0283	0.0472
30	-0.0105	-0.0058	0.0553	30	0.0222	0.0252	0.0462
36	-0.0225	-0.0088	0.0542	36	0.0174	0.0204	0.0455
42	-0.0275	-0.0185	0.0533	42	0.0136	0.0193	0.0448
48	-0.0350	-0.0178	0.0527	48	0.0108	0.0164	0.0443
54	-0.0459	-0.0245	0.0522	54	0.0085	0.0138	0.0437
60	-0.0468	-0.0264	0.0515	60	0.0063	0.0138	0.0430

II. T-billl rate return prediction model

Horizon	No Breaks w/o Param. Uncertainty	No Breaks with Param. Uncertainty	Past & Future Breaks	Horizon	No Breaks w/o Param. Uncertainty	No Breaks with Param. Uncertainty	Past & Future Breaks
	Tbi=3.8%; $\gamma = 5$				**Tbi=3.8%; $\gamma = 10$**		
1	0.0646	0.0656	0.0696	1	0.0516	0.0523	0.0541
6	0.0622	0.0624	0.0630	6	0.0506	0.0506	0.0509
12	0.0571	0.0574	0.0577	12	0.0480	0.0482	0.0482
18	0.0539	0.0539	0.0561	18	0.0463	0.0463	0.0473
24	0.0505	0.0512	0.0547	24	0.0445	0.0451	0.0465
30	0.0478	0.0486	0.0539	30	0.0430	0.0437	0.0458
36	0.0458	0.0467	0.0534	36	0.0417	0.0425	0.0454
42	0.0420	0.0443	0.0525	42	0.0401	0.0401	0.0446
48	0.0422	0.0434	0.0524	48	0.0382	0.0392	0.0444
54	0.0408	0.0420	0.0519	54	0.0370	0.0381	0.0437
60	0.0381	0.0406	0.0514	60	0.0358	0.0369	0.0431

Table 7: Annualized certainty equivalent return estimates computed under the assumption that the break point model is the return generating process. Asset allocations are computed when the investor (i) ignores breaks and parameter uncertainty; (ii) ignores breaks but accounts for parameter uncertainty; and (iii) accounts for breaks and parameter uncertainty. The horizon is measured in months.