Optimal Asset Allocation under Structural Breaks
Job Market Paper

Davide Pettenuzzo       Allan Timmermann
Bocconi University     University of California, San Diego

November 3, 2004

Abstract

An extensive literature in finance has found that return predictability has important effects on optimal asset allocations. While some papers have also considered the portfolio effects of parameter and model uncertainty in the context of Bayesian models, model instability (‘breaks’) has received far less attention. This poses an important concern when the parameters of return prediction models are estimated using data samples spanning several decades where parameters are unlikely to remain stable. In this paper we adopt a new approach that accounts for breaks to multivariate return prediction models both in the historical sample and at future (out-of-sample) points. Our empirical findings suggest that model instability has a large effect on optimal stock holdings both in absolute terms and when compared to the effect of parameter estimation and model uncertainty. The effect of future breaks is, unsurprisingly, largest at long investment horizons, whereas historical (in-sample) breaks affect current parameter estimates and therefore have a large effect also at short investment horizons. Our analysis covers optimal asset allocation under parameter uncertainty, model uncertainty and uncertainty about the stability of the return forecasting model.
1. Introduction

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

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

One aspect of return predictability that has received far less attention is model instability. Asset allocation exercises invariably assume that although the parameters of the return prediction model or the identity of the “true” model may not be known to investors, the parameters of the data generating process remain constant through time.

The model stability assumption is important since asset allocation decisions require forecasting future returns. Suppose that it is found that there are structural breaks in the parameters of the return prediction model over the historical sample. If such breaks occurred in the past it would seem plausible to assume that they could also appear in the future. This introduces an extra source of risk concerning when the next break(s) will occur, how long the new ‘regime’ will last and how large any shifts in the parameters of the return equation will be.

There are many reasons for questioning this stability assumption. Structural instability is known to affect a large majority of time-series models for economic and financial variables, c.f. Stock and Watson (1996). Natural candidates for explanations of structural shifts such as institutional and technological change, large macroeconomic (oil price) shocks or changes in tax policy are known to occur in samples spanning long periods of time. This is important since predictability in stock returns is generally rather weak, necessitating the use of long spans of data in order to obtain reasonably precise estimates of the underlying regression coefficients. For example, Barberis (2000) uses monthly data from 1927 to 1995 to estimate the coefficient of the dividend yield in a return forecasting model. However, it is perhaps difficult to believe that this coefficient remained constant through a sample spanning the Great Depression, World War II, the stagflation period of the seventies and the run-up in stock prices during the 1990s.


Unsurprisingly, the assumption of a stable relationship between stock returns and standard predictor variables has increasingly come under criticism both on theoretical and empirical grounds. Most notably, despite the positive historical correlation between stock returns and the lagged dividend yield, the late 1990s saw an unprecedented bull market with large positive mean stock returns and historically low values of the dividend yield. This brought many to question the stability of the relation between the dividend yield and stock returns. Lettau and Ludvigson (2001) found evidence of a breakdown in their prediction model based on their $cay$ variable in the mid-nineties.

This paper adopts an approach that accounts for structural breaks in return forecasting models. Our approach builds on Chib (1998), Pastor and Stambaugh (2001) and Pesaran, Pettenuzzo, and Timmermann (2004) in using a changepoint model driven by a hierarchical Hidden Markov Chain (HMC). This allows us to characterize structural breaks in the historical data sample. Furthermore, the model nests as special cases both a pooled scenario where the commonality between the parameters in the different ‘regimes’ is very strong (corresponding to a narrow dispersion of the meta distribution for these parameters) as well as a more idiosyncratic case 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 of the meta distribution.

To forecast future returns under breaks, we follow Pesaran, Pettenuzzo, and Timmermann (2004) and extend the Chib (1998) model for transitions between states. We do so by utilizing a meta distribution characterizing the parameters of the return forecasting model across the break segments. This allows us to answer questions such as “when will the next break occur?” and “what happens to the parameters of the return model after a future break?”.

The approach is very general and allows both for uncertainty about the timing (dates) of the breaks as well as uncertainty about the number of breaks. We also extend our setup to allow for uncertainty about the identity of the predictor variables (model uncertainty) using Bayesian model averaging techniques as proposed by Avramov (2002) and Cremers (2002). Hence, investors are neither assumed to know the true parameter values or the true model nor are they assumed to know the number and timing of past or future breaks. Instead, they come with prior beliefs about the “meta” distribution from which current and future values of the parameters of the return model are drawn.

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

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

Another literature that is related to our paper assumes that the parameters of the return equation are driven by a Markov switching process with a small number of states, c.f. Ang and Bekaert (2002), Ang and Chen (2002), Guidolin and Timmermann (2004) and Perez-Quiros and Timmermann (2000). The assumption of a fixed number of states amounts to imposing a restriction that ‘history repeats’. For example, the majority of papers on Markov switching in stock returns assume only two states so the mean and variance of returns can either be high or low depending on which state the model is in. This approach is well suited to identify patterns in returns linked to repeated events such as recessions and expansions. It is not clear that it is able to capture equally well the effects of institutional and technological changes that are more likely to lead to genuinely new and historically unique regimes.

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

Structural breaks are found to have a large effect on optimal asset allocations. Using a return model based on the dividend yield, Barberis (2000) found that parameter estimation uncertainty leads risk averse investors to pursue a less aggressive allocation to stocks. Parameter estimation uncertainty can also alter how the investment horizon maps into the optimal stock allocation which can turn from upward-sloping to downward-sloping once parameter estimation uncertainty is accounted for. We find that model instability can have an even larger effect on the asset allocation than parameter estimation uncertainty and lead to an even steeper negative slope in the relationship between the investment horizon and the proportion of wealth that a buy-and-hold investor allocates to stocks. This reflects the extent to which the coefficients of the predictor variables varied in the return equation across our sample.

The paper is organized as follows. Section 2 presents the basic breakpoint methodology and introduces the hierarchical hidden Markov chain approach. Section 3 presents empirical estimates for return prediction models based on the dividend yield or the short interest rate. Section 4 shows how models for optimal asset allocation are affected by structural breaks while section 5 considers their effect empirically. Section 6 proposes various extensions to the basic approach and also conducts a robustness analysis. Finally Section 7 concludes. Technical details are provided in appendices at the end of the paper.
2. Methodology

Studies of asset allocation in the presence of return predictability (e.g., Barberis (2000), Campbell and Viceira (2001), Campbell, Chan, and Viceira (2003) and Kandel and Stambaugh (1996)) have mostly used vector autoregressions (VARs) to capture the dynamic relations 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 a bivariate model relating returns (or excess returns) on the risky asset to a predictor variable, $x_t$. Constraints that the coefficients on lagged values of returns are zero are commonly imposed. This reflects the common finding that stock returns are not strongly serially correlated and reduces the number of parameters to be estimated. The constrained model takes the form

$$z_t = B' x_{t-1} + u_t$$

where $z_t = (r_t, x_t)'$, $x_{t-1} = (1, x_{t-1})'$, $r_t$ is the stock return at time $t$ in excess of a short risk-free rate, while $x_{t-1}$ is the lagged predictor variable and $E[u_t u_t'] = \Sigma$ is the covariance matrix. The first and second columns of $B'$ contain intercept and lagged predictor variable coefficients, respectively.

For future reference, we define $m = 2$ as the number of equations in (1). Furthermore, we let $\mu_r$ and $\mu_x$ be the intercepts in the equation for the return and predictor variable, respectively, while $\beta_r$ and $\beta_x$ are the coefficients on the predictor variable in the two equations:

$$r_t = \mu_r + \beta_r x_{t-1} + \varepsilon_{rt}$$
$$x_t = \mu_x + \beta_x x_{t-1} + \varepsilon_{xt}.$$  \hspace{1cm} (2)

To capture breaks, we build on the multiple change point model proposed by Chib (1998). Breaks in the parameters of the return model are captured through an integer-valued state variable, $s_t$, that tracks the regime from which a particular observation of returns and the predictor variable, $x_t$, are drawn. For example, $s_t = l$ indicates that $z_t$ has been drawn from $f(z_t | \mathcal{Z}_{t-1}, \Theta_l)$, where $\mathcal{Z}_t = \{z_1, ..., z_t\}$ is the current information set, while a change from $s_t = l$ to $s_{t+1} = l+1$ shows that a break has occurred at time $t+1$. Allowing for $K$ breaks or $K+1$ break segments, our breakpoint model takes the form

$$z_t = B'_1 x_{t-1} + u_t, \quad E[u_t u_t'] = \Sigma_1 \quad \text{for } \tau_0 \leq t \leq \tau_1 \quad (s_t = 1)$$
$$z_t = B'_2 x_{t-1} + u_t, \quad E[u_t u_t'] = \Sigma_2 \quad \text{for } \tau_1 + 1 \leq t \leq \tau_2 \quad (s_t = 2)$$
$$\vdots$$
$$z_t = B'_l x_{t-1} + u_t, \quad E[u_t u_t'] = \Sigma_l \quad \text{for } \tau_{l-1} + 1 \leq t \leq \tau_l \quad (s_t = l)$$
$$\vdots$$
$$z_t = B'_{K+1} x_{t-1} + u_t, \quad E[u_t u_t'] = \Sigma_{K+1} \quad \text{for } \tau_K + 1 \leq t \leq T \quad (s_t = K+1)$$

Here $\mathcal{T}_K = \{\tau_0, ..., \tau_K\}$ is the collection of break points with $\tau_0 = 1$. $\Theta_l = (B_l, \Sigma_l)$ can be interpreted as the location and scale parameters in regime $l$.

The state variable $s_t$ is assumed to be driven by a first order hidden Markov chain (HMC) whose transition probability matrix is constrained to reflect a multiple change point model. At each point in time, $s_t$ can either remain in the current state or jump to the subsequent state. Assuming $K$
breaks in the historical sample \((1 \leq t \leq T)\), the one-step-ahead transition probability matrix takes the form

\[
P = \begin{pmatrix}
p_{11} & p_{12} & 0 & \cdots & 0 \\
0 & p_{22} & p_{23} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & \cdots & 0 & p_{K,K} & p_{K,K+1} \\
0 & 0 & \cdots & 0 & 1
\end{pmatrix}.
\]

(4)

Here \(p_{j-1,j} = Pr( s_t = j | s_{t-1} = j - 1) \) is the probability of moving to regime \(j\) at time \(t\) given that we are in state \(j - 1\) at time \(t - 1\) so \(p_{j,j} + p_{j,j+1} = 1\). \(p_{K+1,K+1} = 1\) due to the assumption of \(K\) breaks up to period \(T\). This means that - conditional on \(K\) breaks occurring in the historical sample - the process terminates in state \(K + 1\). The special case without breaks corresponds to \(K = 0\).

The diagonal elements of (4), \(p_{j,j}\), are assumed to be independent of \(p_{i,i}, j \neq i\), and are drawn from a beta distribution,

\[
p_{j,j} \sim \text{Beta} (a,b), \text{ for } j = 1, 2, ..., K.
\]

(5)

The joint density of \(p = (p_{11}, ..., p_{K,K})'\) is then given by

\[
\pi (p) = c_K \prod_{j=1}^{K} p_{j,j}^{(a-1)} (1 - p_{j,j})^{(b-1)},
\]

(6)

where \(c_K = \Gamma (a + b) / \Gamma (a) \Gamma (b) \)^{K}.

2.1. Predictive Distributions of Returns under Breaks

At the heart of an investor’s asset allocation problem lies the predictive distribution of returns conditional on the information set at time \(T\), \(Z_T\). This conditional distribution can be computed under a range of scenarios depending on the investor’s assumptions concerning the possibility of breaks over the investment horizon \([T, T + h]\).

Since we are interested in forecasting future returns, we follow Pastor and Stambaugh (2001) and Pesaran, Pettenuzzo, and Timmermann (2004) and adopt a hierarchical break point formulation that makes use of meta distributions for the unknown parameters. To this end we assume that the coefficient vector, \(B_j\), and variance covariance matrix of the error terms, \(\Sigma_j\), in each regime are drawn from common (meta) distributions.

An assumption that the investor knows that no new breaks occur between \(T\) and \(T + h\) appears implausible, particularly at long investment horizons. If future breaks can occur, forecasts of \(r_{T+h}\) based solely on the posterior distribution of the parameters \(B_{K+1}\) and \(\Sigma_{K+1}\) will clearly be biased, with a bias that grows the further away from the average (meta) distribution the parameters of the last regime are.

To compute the predictive distribution of returns, we need to make assumptions about the probability that future breaks occur as well as the size of such breaks. Most obviously, we need an estimate of the probability of staying in the current regime, \(p_{K+1,K+1}\). If more than one break can
occur over the course of the investment horizon, we also need to model the distribution from which future ‘stayer’ probabilities \( p_{K+2,K+2}, p_{K+3,K+3} \ldots \) are drawn. We next explain how this is done.

Suppose we know that we are in regime \( K + 1 \) at time \( T \). The probability that a single break occurs at a future point in time, \( T + j (1 \leq j \leq h) \), is

\[
Pr(\tau_{K+1} = T + j | s_{T+h} = K + 2, s_T = K + 1) \propto (1 - p_{K+1,K+1}) p_{K+1,K+1}^{j-1}.
\]

Here \( \tau_{K+1} \in [T + 1; T + h] \) tracks when the next break \((K + 1)\) happens and \( p_{K+1,K+1} \) is the probability of remaining in state \( K + 1 \). To forecast the likelihood of future breaks therefore requires extending the transition probability matrix (4) which conditions on \( K \) in-sample breaks and therefore assumes that \( p_{K+1,K+1} = 1 \). We follow Pesaran, Pettenuzzo, and Timmermann (2004) by modifying (4) by a transition probability matrix capturing the dynamics also after period \( T \):

\[
\tilde{P} = \begin{pmatrix}
p_{11} & p_{12} & 0 & \cdots & 0 \\
0 & p_{22} & p_{23} & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & \cdots & 0 & p_{K,K} & p_{K,K+1} \\
0 & 0 & \cdots & 0 & p_{K+1,K+1} \\
0 & 0 & \cdots & 0 & p_{K+2,K+2} \\
\end{pmatrix}.
\]

Here the \((K + 1) \times (K + 1)\) sub-matrix in the upper left corner describes possible breaks in the historical data sample, \( \{z_1, \ldots, z_T\} \) and is identical to (4) except for the final element, \( p_{K+1,K+1} \) which in general is different from unity. The remaining part of \( \tilde{P} \) describes the breakpoint dynamics over the future out-of-sample investment period from \( T \) to \( T + h \).

To model the parameters after a break, we need to make assumptions about the distribution from which such new parameters are drawn. We do this by adopting a hierarchical setup that assumes meta distributions for the location and scale parameters. Our setup extends the general idea of Pesaran, Pettenuzzo, and Timmermann (2004) of forecasting univariate time series subject to structural breaks to a multivariate setting. This is an important exercise that, as we show below, yields important new insights into commonly used return prediction models.\footnote{Within each regime we decompose the covariance matrix, \( \Sigma_j \), into the product of a diagonal matrix representing the standard deviations of the variables, \( \text{diag}(\psi_j) \), and a correlation matrix, \( R_j \), each of which is modeled separately: \( \Sigma_j = \text{diag}(\psi_j) \times R_j \times \text{diag}(\psi_j) \).}

\[\Sigma_j = \text{diag}(\psi_j) \times R_j \times \text{diag}(\psi_j).\]
2.2. Meta Distributions

After a break the values of the parameters in the new regime are drawn from a set of meta distributions. As pointed out by Pastor and Stambaugh (2001), p. 1207 “Finance practitioners and academics often elect to rely on more recent data, however, motivated in part by concerns that the probability distribution of excess returns changes over time, experiencing shifts known as “structural breaks”. Discarding pre-break data tends to lead to imprecise estimates and also goes against the intuition that pre-break data contains some information about the parameters in the new regime. Using a meta distribution from which the parameters within each regime are drawn gets around this problem.

We next describe the meta distributions. The \( m^2 \times 1 \) vector of coefficients are independent draws from a normal distribution, \( \text{vec}(B)_j \sim N(b_0, V_0) \), \( j = 1, \ldots, K + 1 \), while the \( m \) error term precisions \( \psi_{j,i}^{-2} \) are independently and identical (IID) draws from a Gamma distribution, \( \psi_{j,i}^{-2} \sim \text{Gamma}(v_{0,i}, d_{0,i}) \), \( i = 1, \ldots, m \). Finally the \( (m-1)/2 \) correlation elements, \( r_{j,ic} \), are IID draws from a normal distribution, \( r_{j,ic} \sim N(\mu_{p,ic}, \sigma_{p,ic}^2) \), \( i, c = 1, \ldots, m, i < c \). Notice that \( b_0, v_{0,i} \) and \( \mu_{p,ic} \) represent the location parameters, while \( V_0, d_{0,i} \) and \( \sigma_{p,ic}^2 \) are the scale parameters of the three meta distributions.

The assumption that the parameters are drawn from a meta distribution implies that data from previous regimes carry information relevant for current data and for the new parameters after a future break.

The pooled scenario (all parameters are identical across regimes) and the regime-specific scenario (the parameters of each regime are unrelated) can be seen as extreme special cases. Which scenario most closely represents the data can be inferred from the estimates of the hyperparameters \( V_0, d_{0,i} \) and \( \sigma_{p,ic}^2 \). We use a random coefficient model to introduce a hierarchical prior for the regime coefficients \( \{B_j, \text{diag}(\psi_j), R_j\} \).

For the meta distribution parameters we assume that

\[
\begin{align*}
b_0 & \sim N(\mu_\beta, \Sigma_\beta) \\
V_0^{-1} & \sim W(\nu_\beta, V_\beta^{-1})
\end{align*}
\]

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

\[
\begin{align*}
v_{0,i} & \sim \text{Exp}(\rho_{0,i}) \\
d_{0,i} & \sim \text{Gamma}(c_{0,i}, d_{0,i})
\end{align*}
\]

\( \rho_{0,i}, c_{0,i} \) and \( d_{0,i} \) are prior hyperparameters. Following Liechty, Liechty, and Müller (2004), we specify

---

\( ^6 \) Given the symmetry of the correlation matrix \( R_j \), we only model the element above the main diagonal, i.e. \( r_{j,ic} \), \( i, c = 1, \ldots, m \) such that \( i < c \).

\( ^7 \) Throughout the paper we use underscore bars (e.g. \( a_\beta \)) to denote prior hyperparameters.
the following distributions for the hyperparameters of the elements of the correlation matrix:

\[
\mu_{\rho,ic} \sim N \left( \mu_{\rho,ic}, \tau_{\rho,ic}^2 \right) \tag{13}
\]

\[
\sigma_{\rho,ic}^{-2} \sim \text{Gamma} \left( a_{\rho,ic}, b_{\rho,ic} \right). \tag{14}
\]

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

\[
a \sim \text{Gamma} \left( a_0, b_0 \right),
\]

\[
b \sim \text{Gamma} \left( a_0, b_0 \right).
\]

2.3. 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 of these choices.

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

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

Moving to the variance hyperparameters, we maintain uninformative priors and set \( c_{0,i} = 1, \)

\( d_{0,i} = 1/\infty \) and \( \rho_{0,i} = \infty \) in all equations. This corresponds to centering these distributions at \( \infty \) and specifying a very large variance. We also use uninformative priors for the correlation coefficient, \( r_{j,12} \), i.e. \( \mu_{\rho,12} = 0, \tau_{12}^2 = \infty, a_{\rho,12} = 1 \) and \( b_{\rho,12} = 0.01 \). Finally, we assume a more informative

\footnote{We could allow for a unit root (\( \beta_x = 1 \)), but this turns out not to affect our results greatly.}

\footnote{If the upper constraint on the mean of the predictor variable is ignored while negative values are ruled out, the mean of the predictive density tends to increase too much at very long investment horizons. We impose these constraints by including an accept-reject step in the section of the Gibbs sampler that relates to \( \mu_x \) and \( \beta_x \). This discards draws of the parameters violating the constrains.}
prior for the diagonal elements, \( p_{ii} \) in (4), centered at 0.98, namely \( a_0 = 1 \) and \( b_0 = 0.02 \). This ensures that breaks do not occur too frequently.

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 that use the dividend yield or the short interest rate as predictor variables. We provide two sets of results. First, we use a model with uninformative (diffuse) priors for the meta hyperparameters. Second, we use economically motivated constraints on the priors of these parameters.

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 1926:12-2003:12. All data is obtained from the Center for Research in Security Prices (CRSP).

As forecasting variables we include a constant and either the dividend-price ratio \( (Yld) \) — 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 \( (Tb) \). The dividend yield has been found to predict stock returns by many authors including Campbell (1987), Campbell and Shiller (1988), Keim and Stambaugh (1986), Fama and French (1988). It has played a special role in the literature on asset allocation implications of return predictability, c.f. Kandel and Stambaugh (1996) and Barberis (2000). Furthermore, due to its persistence and the large negative correlation between shocks to the 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 rate has also been found to reliably predict stock returns, c.f. Ang and Bekaert (2002).

Table 1 shows posterior parameter estimates when the constrained VAR model (2) is fitted to the full sample. These results use an uninformative prior and do not impose any constraints on the parameters. The estimates are in line with what has been reported in the literature. The (mean) coefficient on the dividend yield in the return equation is positive and slightly less than two standard errors away from zero. The dividend yield is highly persistent with an autoregressive parameter close to 0.98. Importantly for asset allocation purposes, there is a large negative correlation between innovations to the dividend yield and innovations to returns.

Full sample estimates for the return model based on the short interest rate indicate a large negative coefficient on the T-bill rate about two standard errors away from zero. Once again this regressor is highly persistent with an autoregressive coefficient exceeding 0.98. In contrast to the model that includes the dividend yield, however, the correlation between shocks to the T-bill rate and shocks to returns is close to zero.
3.2. Return Forecasting Model Based on 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. To this end, Table 2 provides a comparison of models with different numbers of breaks by reporting estimates of the conditional log-likelihood as well as the marginal log-likelihood for different numbers of breaks in the return forecasting model. For a given choice of number of breaks, $K$, we get a new model, $M_K$, with its own set of parameters.

First consider the return model based on the dividend yield. For this model there is strong support for structural breaks as the marginal log-likelihood function for the return model with no breaks is much smaller than the values observed for the models that include breaks. Overall, the evidence supports the presence of seven breaks. When additional breaks were included the algorithm essentially chose the same break date twice suggesting that further breaks are not well defined. Although seven break points may appear to be a large number, it is consistent with the evidence reported by Pastor and Stambaugh (2001) of 15 break points in the equity premium over a sample (1834-1999) a bit longer than twice the period covered here.

Among all models with up to seven breaks, the fourth column of Table 2 shows that the model with seven breaks obtains a posterior probability weight of 98%. Although we cannot be sure that there are seven breaks in the sample, nevertheless the data strongly supports this specification. Table 2 also shows for each model the time of the associated breaks. More precisely, these are the posterior modes identified for the break dates since our model only provides probability estimates of the break dates. To emphasize this, Figure 1 shows posterior probabilities of the break locations for the models with seven breaks. The break dates are very well identified in the form of single spikes with probabilities ranging from 0.27 to over 0.50. This suggests that there is not much uncertainty about the locations of the break dates for this model.

Five of the break locations are associated with major events and occurred around the Great Depression (1932), the beginning of World War II (1940), the oil price shock (1974), the end to the change in the Fed’s operating procedures (1982) and more recently at the beginning of the bull market of the nineties (1992). The two remaining break dates (1952 and 1958) are harder to interpret, but the break in the late fifties appears to match a similar increase in the posterior probability of a break in the unconditional equity premium model estimated by Pastor and Stambaugh (2001) (Figure 1, top panel).

Parameter estimates for the model with seven breaks (eight regimes) are reported in Table 3. The mean of the dividend yield coefficients in the return equation range from a low of 0.39 prior to the Great Depression to values around two during the regimes from 1958-1974 and again from 1974-1982 before declining to levels near 1.74 and 1.40 in the last regimes. The mean value of the standard deviation parameters of the return equation also vary considerably over the sample, from a high near 10% per month around the Great Depression to a low of only 3.4% per month during the 1950s. Since the mid-seventies, return volatility appears to have been quite stable at 4-5% per month.

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

Information on the posterior estimates of the hyperparameters of the meta distribution is provided in Table 4. The hyperparameter tracking the mean slope of the dividend yield in the return equation is centered on 1.2 with a large standard deviation centered on 0.50. Notice also the large mean value of the variance parameter \( V_0(\beta_y) \) which means that the meta distribution allows for considerable heterogeneity in the slope of the dividend yield across regimes. The autoregressive slope \( \beta_x \) in the dividend yield equation is centered on a value of 0.92 with a standard deviation of only 0.033. Its variance parameter \( V_0(\beta_x) \) is centered on 0.014 and is hence surrounded by far less uncertainty. The same holds for the hyperparameter tracking the correlation between shocks to returns and to the dividend yield. This is centered on -0.92 with a modest standard deviation of 0.04. The posterior distributions of the hyperparameters of the transition probability, \( a_0 \) and \( b_0 \), are surrounded by greater uncertainty as indicated by their relatively large standard deviations.

These findings suggest that the greatest uncertainty in our model is associated with the effect of the dividend yield on 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 the dividend yield. Both of these parameters are important determinants of the optimal asset allocation.

3.3. Return Forecasting Model Based on the Short Interest Rate

Turning to the return model based on the short interest rate, Table 2 also suggests the presence of seven breaks. Some of these breaks again appear around the time of major historical events such as the Great Depression (1934), the end of World War II (1947), the beginning and end of the change to the Fed’s operating procedures (1979 and 1982) and the beginning of protracted bull market of the 1990s (1990). The two remaining breaks are estimated to have occurred around 1952 and 1968.

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

Parameter estimates for the model with seven breaks are displayed in Table 5. The mean of the

---

10 The conditional log-likelihood function of the return equation declines as we move from five to six breaks. This can happen even as the number of breaks increases as long as the likelihood of the joint distribution of stock returns and the short interest rate increases, which indeed it does uniformly as the number of breaks (and hence the number of parameters) goes up.
coefficient on the lagged T-bill rate in the return equation is always negative but varies significantly over time, ranging from only -0.44 in the first regime to -9.98 in the very volatile regime from 1979 to 1982 when the Fed changed its monetary policy. Furthermore, the estimates of the slope on the T-bill rate within each regime are surrounded by large standard errors, particularly in the regimes from 1934-1947 and 1947-1952.

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

Turning finally to the meta distribution parameters for the short 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 -4.6 and a standard deviation of 4.8, giving a very long 95% confidence interval that ranges from -14.8 to 4.4. Compared with the model based on the dividend yield, there is now also greater uncertainty about the correlation between shocks to returns and shocks to the T-bill rate. The variance of this parameter is more than four times higher than was found in Table 4. In contrast, the meta distribution of the persistence parameter in the interest rate equation is quite precisely determined with a mean of 0.90 and a standard deviation of 0.042.

4. Asset Allocation under Structural Breaks

Investors are interested in instability in the return model because breaks affect future asset payoffs and therefore may alter the optimal asset allocation. To study the economic importance of structural breaks in the return model, this section considers the optimal asset allocation under a range of alternative modeling assumptions for a buy-and-hold investor with power utility over terminal wealth:

$$u(W_{T+h}) = \frac{W_T^{1-\gamma}}{1-\gamma},$$

(16)

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

Without loss of generality we set initial wealth at one, \( W_T = 1 \), and let \( \omega \) be the allocation to stocks. Terminal wealth is given by

\[
W_{T+h} = (1 - \omega) \exp(r_f h) + \omega \exp(r_f h + r_{T+1} + \ldots + r_{T+h}).
\]

(17)

Defining the cumulative excess returns over \( h \) periods as

\[
R_{T+h} = r_{T+1} + r_{T+2} + \ldots + r_{T+h},
\]

the buy-and-hold investor solves the program

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

(18)

where \( E_T \) is the conditional expectation given information at time \( T \). This reflects the modeling assumptions made by the investor.

The predictive density for the \( h \)-period cumulative returns, \( R_{T+h} \), can be constructed using the iterative scheme of Barberis (2000) that takes advantage of the assumed VAR structure. To see how this works, rewrite (2) as

\[
z_t = \mu_t + \beta_{0t} z_{t-1} + u_t \quad \text{where} \quad \mu_t = (\mu_{rt} \, \mu_{xt})'
\]

and

\[
\beta_{0t} = \begin{bmatrix}
0 & \beta_{rt} \\
0 & \beta_{xt}
\end{bmatrix}.
\]

The \( t \)-subscripts on the parameters reflects the possibility of breaks. Iterating forward on this model, we have

\[
\begin{align*}
z_{T+1} &= \mu_{T+1} + \beta_{0T+1} z_T + u_{T+1} \\
z_{T+2} &= \mu_{T+2} + \beta_{0T+2} \mu_{T+1} + \beta_{0T+2} \beta_{0T+1} z_T + \mu_{T+2} + \beta_{0T+2} u_{T+1} \\
z_{T+3} &= \mu_{T+3} + \beta_{0T+3} \mu_{T+2} + \beta_{0T+3} \beta_{0T+2} \mu_{T+1} + \beta_{0T+3} \beta_{0T+2} \beta_{0T+1} z_T \\
&\quad + \mu_{T+3} + \beta_{0T+3} u_{T+2} + \beta_{0T+3} \beta_{0T+2} u_{T+1} \\
&\quad \quad \vdots
\end{align*}
\]

(19)

\[
z_{T+h} = \mu_{T+h} + u_{T+h} + \sum_{j=1}^{h-1} \left( \prod_{i=j+1}^{h} \beta_{0T+i} \right) (\mu_{T+j} + u_{T+j}) + \left( \prod_{i=j+1}^{h} \beta_{0T+i} \right) z_T.
\]

Assuming normality of \( u_T \), in the special case where \( \mu \) and \( \beta_0 \) do not vary over time, we get the results reported by Barberis (2000), namely \( z_{T+h} \sim N \left( \mu_{z_{T+h}}, \Sigma_{z_{T+h}} \right) \), where

\[
\mu_{z_{T+h}} = \sum_{s=0}^{h-1} \beta_0^s \mu + \beta_0^h z_T,
\]

\[
\Sigma_{z_{T+h}} = \left( \sum_{s=0}^{h-1} \beta_0^s \right) \Sigma \left( \sum_{s=0}^{h-1} \beta_0^s \right)'.
\]
Furthermore, the sum  \( z_{T+1} = z_{T+1} + z_{T+2} + \ldots + z_{T+h} \) is distributed as a multivariate normal variable with mean vector  \( \mu_{\text{sum}} \) and variance-covariance matrix  \( \Sigma_{\text{sum}} \):

\[
\mu_{\text{sum}} = h\mu + (h - 1)\beta_0\mu + (h - 2)\beta_0^2\mu + \ldots + \beta_0^{h-1}\mu + (\beta_0 + \beta_0^2 + \ldots + \beta_0^h)z_T
\]

(20)

\[
\Sigma_{\text{sum}} = \Sigma + (I + \beta_0)\Sigma(I + \beta_0)'
\]
\[
+ (I + \beta_0 + \beta_0^2)\Sigma(I + \beta_0 + \beta_0^2)'
\]
\[
+ (I + \beta_0 + \ldots + \beta_0^{h-1})\Sigma(I + \beta_0 + \ldots + \beta_0^{h-1})'.
\]

(21)

Comparing (19) to (20) and (21), in the general case with time-varying parameters, it is far more complicated to evaluate the predictive return distribution and we must rely on numerical methods.

We next consider a range of assumptions about how investors deal with their limited knowledge of the parameters of the return model.

4.2. No Breaks

First consider the asset allocation problem for an investor who ignores parameter estimation uncertainty and breaks. Once the predictors have been specified, the VAR parameters  \( \Theta = (B, \Sigma) \) can be estimated and, using (20) and (21), the VAR model can be iterated forward conditional on these parameter estimates. This generates a distribution for future stock returns,  \( p(R_{T+h}|\tilde{\Theta}, S_{T+h} = 1, Z_T) \) where  \( S_{T+h} = 1 \) shows that past and future breaks are ignored. This investor therefore solves the problem

\[
\max_{\omega} \int u(W_{T+h})p(R_{T+h}|\tilde{\Theta}, S_{T+h} = 1, Z_T)dR_{T+h}.
\]

(22)

This of course ignores that  \( \Theta \) is not known precisely but typically estimated with considerable uncertainty. To be more precise, we could condition also on  \( M_K \) i.e. the return prediction model based on the predictor variable  \( x \) and conditional on  \( K \) in sample breaks. The importance of  \( M_K \) will become clear when we integrate out uncertainty about the number of in-sample breaks and uncertainty about the predictor variables.

Next consider the decision of an investor who accounts for parameter estimation errors but ignores both past and future breaks, i.e., assumes that  \( S_{T+h} = 1 \). In the absence of breaks the posterior distribution\(^{11} \pi(\Theta|S_{T+h} = 1, Z_T) \) summarizes the uncertainty about the parameters given the historical data sample. Integrating over this distribution leads to the predictive distribution conditioned only on the observed sample (and not on any fixed  \( \Theta \)) and the assumption of no breaks:

\[
p(R_{T+h}|S_{T+h} = 1, Z_T) = \int p(R_{T+h}|\Theta, S_{T+h} = 1, Z_T)\pi(\Theta|S_{T+h} = 1, Z_T)d\Theta.
\]

(23)

This investor therefore solves the problem

\[
\max_{\omega} \int u(W_{T+h})p(R_{T+h}|S_{T+h} = 1, Z_T)dR_{T+h}.
\]

(24)

\(^{11}\)Throughout the paper,  \( \pi(\cdot|Z_T) \) refers to a parameter posterior distribution conditioned on information contained in  \( Z_T \).
Comparing stock holdings in (22) and (24) gives a measure of the economic importance of parameter estimation uncertainty.

Both solutions in (22) and (24) ignore model instability, however. To illustrate the effect of breaks in the parameters of the return prediction model, we separately consider scenarios with past breaks but no future breaks as well as scenarios allowing for both past and future breaks.

4.3. Past Breaks Only

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

$$
\int p(R_{T+h}|\Theta_{K+1}, S_{T+h} = K + 1, S_T = K + 1, Z_T) \pi(\Theta_{K+1}|H, p, S_T, Z_T) d\Theta_{K+1},
$$

where $S_T = (s_1, \ldots, s_T)$ is the collection of values of the state variable underlying the breakpoint process up to period $T$, $\Theta_{K+1} = (\text{vec}(B)_{K+1}, \psi_{K+1}, R_{K+1})$ are the $K + 1$–th regime-specific parameters (regression coefficients, error term variances and correlations), and

$$
H = (b_0, V_0, v_{0,1}, d_{0,1}, \ldots, v_{0,m}, d_{0,m}; \mu_{p,12}, \sigma_{p,12}^2, \ldots, \mu_{p,m-1m}, \sigma_{p,m-1m}^2, a, b)
$$

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

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

4.4. Past and Future Breaks

Assuming again that there are $K$ in-sample breaks, after a future break the process is generated under the parameters from regime number $K + 2$. First, suppose that we limit ourselves to a single future break during the investor’s holding period so $S_T = K + 1$ and $S_{T+h} = K + 2$. For a given future break date, $T + j$ ($1 \leq j \leq h$), we obtain $p(R_{T+h}|S_{T+h} = K + 2, \tau_{K+1} = T + j, S_T = K + 1, Z_T)$ from

$$
\int\int p(R_{T+h}|\Theta_{K+2}, H, S_{T+h} = K + 2, \tau_{K+1} = T + j, S_T = K + 1, Z_T) \times \pi(\Theta_{K+2}, H|Z_T) d\Theta_{K+2}dH.
$$

12This is an over-simplification since these parameters depend on other regimes through the meta distribution. Furthermore, it should also be recalled that we allow for uncertainty about the time of the most recent break, $\tau_K$. Later we show how to integrate out uncertainty about the number of in-sample breaks, $K$.

13We have simplified the notation here since anytime we are drawing from $\pi(\Theta_{K+1}|H, p, S_T, Z_T)$, we are actually drawing from $\pi(B_{K+1}, \psi_{K+1}, R_{K+1}|H, p, S_T, Z_T)$ where $\psi_{K+1}$ and $R_{K+1}$ contain the variances and correlations of $Z_T$, respectively.
Summing across all possible breakpoint locations, \( T + j, (1 \leq j \leq h) \), we get the predictive density of returns conditional on a single break occurring between period \( T \) and period \( T + h \):

\[
p(\{R_{T+h}\mid S_{T+h} = K + 2, S_T = K + 1, Z_T\})
= \sum_{j=1}^{h-1} \int \int p(\{R_{T+h}\mid \Theta_{K+2}, H, S_{T+h} = K + 2, \tau_{K+1} = T + j, S_T = K + 1, Z_T\})
\times \pi(\Theta_{K+2}, H \mid Z_T) \times \pi(\tau_{K+1} = T + j \mid S_{T+h} = K + 2, S_T = K + 1, Z_T) d\Theta_{K+2} dH.
\]

This can readily be extended to allow for multiple breaks over the investment horizon \([T, T + h]\) in cases where \( h \geq 2 \). For such cases we first need to know the distribution of the probability of staying in future regimes, \( K + j, p_{K+j,K+j}, j \geq 2 \). Using the hierarchical specification for \( p_{j,j} \), \( p_{K+j,K+j} \) values are drawn from the conditional beta posterior

\[
p_{K+j,K+j} \mid S_T \sim Beta(a + l_j, b + 1),
\]

where \( l_j = \tau_j - \tau_{j-1} \) is the duration of the \( l \)th break segment.

The number of possible out-of-sample break point scenarios becomes very large as the number of possible breaks increases. To see this, consider the case with only two breakpoints in the out-of-sample period. This gives \( \sum_{i=1}^{h-1} (h - i) = h(h - 1)/2 \) possible break point locations. For a five-year investment horizon \((h = 60)\) this gives 1,770 possible break locations. Conditioning on two breaks, the probability that the first break occurs at time \( T + j_1 \), followed by a second break occurring at time \( T + j_2 \) is (assuming \( j_1 < j_2 \leq h \))

\[
\Pr(\tau_{K+2} = T + j_2 \mid S_{T+h} = K + 3, \tau_{K+1} = T + j_1, S_T = K + 1) 
\propto p_{j_1-1}^{j_1-1} (1 - p_{K+1,K+1}) p_{j_2-j_1-1}^{j_2-j_1-1} (1 - p_{K+2,K+2}).
\]

Using this, we can readily compute the predictive density conditional on two breaks occurring between period \( T \) and period \( T + h \) by accounting for all possible locations of the break points:

\[
p(\{R_{T+h}\mid S_{T+h} = K + 3, S_T = K + 1, Z_T\})
= \sum_{j=1}^{h-1} \sum_{j_2=j_1+1}^{h} \int \int \int \int \int p(\{R_{T+h}\mid \Theta_{K+2}, \Theta_{K+3}, H, S_{T+h} = K + 3, \tau_{K+1} = T + j_1, \tau_{K+2} = T + j_2, S_T = K + 1, Z_T\})
\times \pi(\tau_{K+1} = T + j_1, \tau_{K+2} = T + j_2 \mid S_{T+h} = K + 3, S_T = K + 1, Z_T) 
\times \pi(\Theta_{K+2}, \Theta_{K+3}, H \mid Z_T) d\Theta_{K+2} d\Theta_{K+3} dH.
\]

The extension to \( n_h > 2 \) break points is tedious but conceptually straightforward. The transition probability for this case takes the form

\[
\Pr(\tau_{K+n_h} = T + j_{n_h} \mid S_{T+h} = K + n_h + 1, \ldots, \tau_{K+2} = T + j_2, \tau_{K+1} = T + j_1, S_T = K + 1) 
\propto \prod_{j=1}^{n_h} p_{K+j,K+j}^{d_j} (1 - p_{K+j,K+j}),
\]

\footnote{We are explicitly using the independence of the break location \( \tau_{K+1} \) from the regime parameters, \( \Theta_{K+2} \) and the meta distribution hyperparameters \( H \), i.e. \( \pi(\tau_{K+1} = T + j \mid S_{T+h} = K + 2, S_T = K + 1, \Theta_{K+2}, H, Z_T) = \pi(\tau_{K+1} = T + j \mid S_{T+h} = K + 2, S_T = K + 1, Z_T) \). We will make use of this property extensively.}
where $d_j \geq 1$ is the duration of regime $K + j$. Similarly, integrating over all possible combinations of $n_b$ breaks, we have

$$p(R_{T+h}|S_{T+h} = K + n_b + 1, S_T = K + 1, Z_T) = \sum_{j_1=1}^{h-n_b+1} \ldots \sum_{j_{n_b}=j_{n_b-1}+1}^{h} \int \ldots \int p(R_{T+h}|\Theta_{K+2}, \ldots, \Theta_{K+n_b+1}, H, S_{T+h} = K + n_b + 1, \tau_{K+1} = T + j_1, \ldots, \tau_{K+n_b} = T + j_{n_b}, S_T = K + 1, Z_T)$$

$$\times \pi(\tau_{K+1} = T + j_1, \ldots, \tau_{K+n_b} = T + j_{n_b} | S_{T+h} = K + n_b + 1, S_T = K + 1)$$

$$\times \pi(\Theta_{K+2}, \ldots, \Theta_{K+n_b+1}, H | Z_T) d\Theta_{K+2} \ldots d\Theta_{K+n_b+1} dH. \quad (27)$$

To integrate out the state variable $S_{T+h}$, we finally need to combine all the possible scenarios defined above. We do this by first conditioning on the maximum number of future breaks between $T$ and $T + h$. Conditional on this we then compute the probability of the break locations. We finally sum over both the number of breaks and break locations.

For example, conditioning on at most $n_b$ break out of sample, we have that

$$p(S_{T+h} = K + 1 | S_T = K + 1, Z_T) = p^h_{K+1,K+1}$$

$$p(S_{T+h} = K + 2 | S_T = K + 1, Z_T) = \sum_{j_1=1}^{h} (1 - p_{K+1,K+1}) p^{j_1-1}_{K+1,K+1}$$

$$p(S_{T+h} = K + 3 | S_T = K + 1, Z_T) = \sum_{j_1=1}^{h-1} \sum_{j_2=j_1+1}^{h} p^{j_2-1}_{K+1,K+1} (1 - p_{K+1,K+1}) p^{j_2-1}_{K+2,K+2} (1 - p_{K+2,K+2})$$

$$\vdots$$

$$p(S_{T+h} = K + n_b + 1 | S_T = K + 1, Z_T) = \sum_{j_1=1}^{h-n_b+1} \ldots \sum_{j_{n_b}=j_{n_b-1}+1}^{h} \left( \prod_{j=1}^{n_b} p^{d_j}_{K+j,K+j} (1 - p_{K+j,K+j}) \right).$$

Integrating out the uncertainty about the future number of breaks, we have

$$p(R_{T+h} | S_T = K + 1, Z_T) = \sum_{j=1}^{n_b+1} p(R_{T+h} | S_{T+h} = K + j, S_T = K + 1, Z_T) \times p(S_{T+h} = K + j | S_T = K + 1, Z_T) \quad (28)$$

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

$$\max_{\hat{K}} \int u(W_{T+h}) p(R_{T+h} | S_T = K + 1, Z_T) dR_{T+h}. \quad (29)$$

4.5. **Uncertainty about the number of past breaks**

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

\[
p_x(R_{T+h}|Z_T) = \sum_{k_x=0}^{K_x} p_x(R_{T+h}|S_T = k_x + 1, Z_T) p(M_{k_x}|Z_T),
\]

where \( K_x \) is an upper limit on the largest number of breaks that is entertained. The weights used in the average are proportional to the posterior of model \( M_{k_x} \) and are hence given by the product of the prior for model \( M_{k_x} \), \( p(M_{k_x}) \), and the marginal likelihood, \( f(Z_T|M_{k_x}) \),

\[
p(M_{k_x}|Z_T) \propto f(Z_T|M_{k_x}) p(M_{k_x}).
\]

4.6. Model uncertainty

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

We follow this analysis by integrating across the two return prediction models considered in this study based on the dividend yield and the short interest rate. This is simply an illustration of how to handle model uncertainty and our analysis could of course be extended to a much larger set of variables and could also be extended beyond bivariate regressions. 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. Hence we get a predictive density of returns that accounts for

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

\[
p(R_{T+h}|Z_T) = \sum_{x=1}^{\bar{X}} \sum_{k=0}^{K_x} p_x(R_{T+h}|S_T = k_x + 1, Z_T) p(M_{k_x}|Z_T).
\]

Here \( p(M_{k_x}|Z_T) \) is the posterior probability of the model with \( k \) breaks using \( x \) as the predictor variable and \( \bar{X} \) is the number of different sets of predictor variables used to forecast stock returns.
5. **Empirical Asset Allocation Results**

This section uses the methods from Section 4 to assess empirically the effect of structural breaks on investors’ 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 this distribution are provided in Appendix B.

### 5.1. Results Based on the Dividend Yield

Figure 3 plots the allocation to stocks under four scenarios discussed in Section 4, namely (i) no breaks, ignoring parameter estimation uncertainty; (ii) no breaks, accounting for parameter estimation uncertainty; (iii) breaks in the historical sample, but no future breaks (so the last regime remains in effect over the course of the investment horizon); (iv) past and future breaks allowed. The first two scenarios ignore breaks and so use full-sample parameter estimates. They correspond to the cases covered by Barberis (2000). We compute the optimal weight on stocks under two values for the coefficient of relative risk aversion, namely $\gamma = 5$ and $\gamma = 10$. Figure 3 sets the initial value of the dividend yield at its terminal value at the end of sample (2003:12) which is 1.5%.

Before interpreting the results, it is worth recalling two important effects on asset allocation under predictability from the yield. First, the dividend yield identifies a mean-reverting component in stock returns which means that the risk of stock returns grows more slowly than in the absence of predictability, creating a large hedging demand for stocks, c.f. Lynch (2001), Campbell, Chan, and Viceira (2003). Negative shocks to returns are bad news in the period when they occur but tend to increase subsequent values of the dividend yield and thus become associated with higher future expected stock returns. This effect is particularly important at long investment horizons. Second, parameter estimation uncertainty 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 one period after the investment decision is made, this will affect returns along the entire investment horizon in a similar way to a permanent negative shock.

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

Under the models that assume no (past or future) breaks, Figure 3 shows that the weight on stocks rises from a level near 10% at short investment horizons to 30% at the five-year horizon. The assumed absence of a break means that a very long data sample (1926-2003) is available for parameter estimation. This reduces parameter estimation uncertainty and leads to an increasing...
weight on stocks, the longer the investment horizon. This interpretation is confirmed by the finding that stock holdings are very similar irrespective of whether parameter estimation uncertainty is accounted for. In contrast, the model estimated under the parameters of the last regime—which acknowledges past breaks but ignores future breaks—implies a short-run allocation to stocks around 30% that is essentially independent of the investment horizon. In this case parameter estimation uncertainty is much larger due to the shortness of the data sample after the most recent break whose most likely location is 1992.\textsuperscript{15} In Figure 3 the greater risk due to parameter estimation uncertainty largely cancels out against the mean reversion in returns identified by the dividend yield.

Turning to the allocation to stocks under the composite model that allows for both past and future breaks—with the latter weighted by their probabilities computed under the assumed change-point model—the weight on stocks is 20% at the short horizon and declines to a level below 10% at the five-year horizon. Parameter estimation uncertainty now dominates the hedging demand for stocks induced by return predictability from the dividend yield.

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

Figure 4 shows the allocation to stocks when the initial value of the dividend yield is set at its historical average of 4%. Comparing Figures 3 and 4 it is clear that the level of the optimal stock holding is very different depending on the initial value of the dividend yield. The allocation to stocks under the no-break, model now starts at a level close to 40% at short investment horizons and increases to nearly 45% at the five-year horizon - or 50% if parameter estimation uncertainty is ignored. Stock holdings under the composite model that accounts for past and future breaks start at a level near 100% but declines quite steeply after one year to a level near 20% at the five-year horizon.

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

They also suggest that breaks have an even larger effect on the optimal asset allocation than parameter estimation error. This can be seen by comparing the full sample (no break) plots in Figures 3 and 4 with and without estimation error. In both cases these are very similar. This is to be expected since investors have access to 75 years of data. In fact, the large effect of parameter estimation error documented by Barberis (2000) was found in a sample where the parameters were estimated using a much shorter data set from 1986 to 1995.

\textsuperscript{15}The difference observed between the stock holding under this scenario and that under no breaks and no parameter estimation uncertainty is similar to the differences in asset allocation shown by Barberis (2000) for a short data sample (1985-1996).
5.2. Results based on the Short Interest Rate

Optimal stock holdings under the return prediction model based on the short rate are shown in Figures 5 and 6, again corresponding to initial values of the short rate that prevailed at the end of the sample, 2003:12, (0.83%) and the sample mean (3.6%).

First consider the results when the short rate is set at its terminal value. These results are quite different from those based on the dividend yield model. The allocation to stocks is downward sloping as a function of the investment horizon irrespective of the assumed breakpoint scenario.

There are two reasons for this. First, while innovations to the dividend yield and stock returns are strongly negatively correlated and thus give rise to a strong hedging demand for stocks, shocks to the short rate and stock returns are—on average—largely uncorrelated, c.f. Table 1. Second, as was the case for the dividend yield, the T-bill rate ended up far below its historical average in 2003:12. However, since the coefficient on the T-bill rate in the return prediction model is negative, this raises the expected stock return, whereas the low terminal value of the dividend yield reduces the expected stock return.

We continue to find that the level and slope of the stock holdings as a function of the investment horizon are sensitive to assumptions about breaks. Under the assumption of no breaks (but accounting for parameter estimation uncertainty), the weight on stocks starts from a level near 60% and declines to around 50% at the five year horizon. This allocation is only marginally higher if parameter estimation uncertainty is ignored.

Suppose next that we consider historical breaks but ignore future breaks. The far greater uncertainty about current parameter values under the model based on the last regime (which bases estimation on data after 1990 and hence uses only 13 years of data) means that the allocation to stocks drops more precipitously from a level near 80% at the shortest investment horizon to 40% at the longest horizon. The drop is even sharper under the composite model which allows for past and future breaks. This model sees the weight on stocks decline from 85% to a level around 10% at the five year horizon, suggesting that the possibility of breaks has an even greater impact on the optimal allocation in the return prediction model based on the T-bill rate than in the model based on the dividend yield. This is a reflection of the greater dispersion of parameter estimates found for this model, as can be seen by comparing Tables 5 and 6 to Tables 3 and 4.

Once again we computed allocations under a higher coefficient of risk aversion ($\gamma = 10$). Stock holdings continue to slope downwards as a function of the investment horizon and are about half their level when $\gamma = 5$. Raising the initial value of the short rate to its historical average (4%) has a similar effect of reducing the allocation to stocks, c.f. Figure 6. Under the full sample (no break) scenarios with and without parameter estimation uncertainty, stock holdings are somewhat lower and much flatter as a function of the investment horizon while the allocation to stocks continues to be downward sloping under the model that allows for past and future breaks.
5.3. **Uncertainty about the number of breaks**

The effect of integrating out uncertainty about the number of historical (in-sample) breaks is shown in Figure 7 which plots optimal stock holdings for different initial levels of the dividend yield (left panel) or the short interest rate (right panel). These results assume identical prior weights on each of the models in Table 2. Stock holdings are very similar to those found when we conditioned on the presence of seven breaks in the historical sample (shown in Figures 3 and 4 for the dividend yield model and in Figures 5 and 6 for the model based on the T-bill rate.) This comes as no surprise, since Table 2 shows that the model with seven historical breaks gets a weight of 98% in the combined model (30) - (31) based on the dividend yield and a weight of 83% in the combined model based on the T-bill rate.

5.4. **Model Uncertainty**

Figure 8 shows the results of accounting for model uncertainty in a simple experiment that combines the return forecasting models with up to seven breaks and includes either the dividend yield or the T-bill rate as a predictor variable, weighting these according to (31) and (32). Predictive densities across 16 different models are considered here (namely two predictor variables, each with between zero and seven breaks). Optimal stock holdings most resemble the allocation under the forecasting model based on the T-bill rate. It is easy to understand why: the return forecasting model based on the T-bill rate gives a better fit than the model based on the dividend yield irrespective of the number of breaks and hence gets a greater weight in the forecast combination.

6. **Sensitivity Analysis**

6.1. **Robustness to Priors**

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

Insights into the effect of imposing constraints on the parameters of the return forecasting model can be gained from Figure 9 which shows the mean and standard deviations of the predictive densities of excess returns within each of the individual regimes, the full sample (no break) and under the meta distribution. The two left windows impose the constraints that the driving variable, $x_t$, is stationary ($0 < \beta_x < 1$) and that the unconditional mean of this variable is non-negative with symmetric constraints around its historical average ($0 \leq \mu_x/(1 - \beta_x) \leq 0.08$ in the case of the dividend yield). This figure assumes that the dividend yield is set at its mean value of 4%. 

22
Reassuringly the mean of the full sample and meta distributions lie well within the highest and lowest regimes as they should since they can be viewed as weighted averages of the predictive densities of returns within the individual regimes.

Turning to the plot of the (mean value of the) standard deviation, once again the full sample estimate is centered in the middle, surrounded by plots for individual regimes above and below. The standard deviation of the predictive density under the meta distribution is generally higher, however, particularly at the longer horizon. This reflects the extra uncertainty due to the possibility of future breaks and means that the standard deviation of returns under the composite model is no longer a weighted average of the standard deviations within each regime since these do not account for shifts between regimes.

Imposing an additional parameter constraint which requires that the unconditional mean excess stock return within each regime lies between zero and 1% per month ($0 \leq \mu_r + \frac{\beta_t\mu_x}{\beta_x} \leq 0.01$) has little effect on the mean of the predictive densities as shown in the third window of Figure 9. However, the standard deviation of the predictive density under the meta distribution is now reduced significantly and lies below that of the full sample (no break) model. The effect of imposing a tight constraint on the mean excess stock return within each regime is to reduce the uncertainty of the meta distribution very significantly—more so than the reduction seen under the full sample model. This is because the restriction is imposed on each regime and hence is a much stronger restriction on the meta distribution than on the full-sample distribution which assumes only a single regime.

6.2. Time-variations in predictive density of returns

The return model that allows for breaks includes a larger number of parameters than the conventional full-sample model so one might be concerned that it ‘overfits’ the data. We do not believe that this is a particular cause for concern since the number of breaks is selected using a criterion (marginal log-likelihood) that is known to penalize large models heavily.

Some insights can be gained, however, by studying the time-series of the first two moments of the predictive density for stock returns shown in Figures 11 and 12. To save space we only present results for the model based on the dividend yield. The top left window shows the mean return series under the full-sample estimates. Mean returns vary between -1/2% and 2% per month and are highest around the Great Depression. The model that allows for multiple breaks (top right window) experiences a greater variation in the mean of stock returns that is now mostly confined to an interval between -2% and 2%. Considering that these are in-sample fitted values, such estimates do not appear implausible.

An even more interesting distinction comes from comparing the plots of the standard deviation of the predictive density of returns under the two models. In the full-sample, no-break model the standard deviation of returns is centered at 5.5% per month and only varies very little. In contrast, since the standard deviation of returns (and of the yield) is allowed to vary across regimes in the break model, the plot for this model follows a step function that tracks the various regimes. In fact, the (mean value of the) standard deviation of returns varies significantly from a level around 10%
around the Great Depression to a level near 3-4% in the middle of the sample. This means that the model with breaks is capable of accounting for heteroskedasticity in returns in as far as this coincides with the identified regimes. This is an important consideration since stock market returns were clearly far more volatile during periods such as the Great Depression.

6.3. **Sharpe Ratios Within Each of the Regimes**

The time-series plots of the expected excess return and the standard deviation of excess returns take on very sensible values in the presence of structural breaks. Even so, it can be difficult to assess whether the priors assumed for the hyperparameters of the meta distribution in our model are economically sensible. One way to go about this is to see whether the Sharpe ratio—defined as the mean excess return over the standard deviation of the excess return—is sensible within each of the eight regimes identified in the empirical analysis.

To this end we plot in Figure 13 the predictive distributions of the Sharpe ratios in each regime. These plots assume that the predictive variable (in this case the dividend yield) equals the mean value within a given regime. This ensures that the plots provide the typical distribution of Sharpe ratios emerging within a given regime. The Sharpe ratios are generally centered on a positive value around 0.10 with a 95% spread close to 0.6. For comparison, the predictive density of the full-sample Sharpe ratio is centered on 0.1 and largely contained in the interval from zero to 0.2. Unsurprisingly, therefore, the presence of breaks results in a greater degree of parameter estimation uncertainty and a wider spread for the Sharpe ratio, but again the values appear quite sensible.

7. **Conclusion**

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

We found, first, that the parameters of standard forecasting models appear to be highly unstable and subject to multiple breaks, many of which coincide with important historical events. Second, we found that, once such breaks are accounted for, the possibility of future breaks has a large impact on the optimal asset allocation of a Bayesian investor endowed with a reasonable meta distribution determining draws of new parameters following future breaks.

Our analysis can be extended in several directions. First, we considered only the asset allocation for a buy-and-hold investor and it would be interesting to explore how the results change if the investor is allowed to rebalance portfolio weights as new information becomes available. We
anticipate that interesting effects will emerge due to the uncertainty concerning the location and size of future breaks. This means that investors constantly have to assess if a break has occurred recently. Realizations of the predictor variable and the stock return in the middle of the distribution of the current regime will not lead to large revisions in investor beliefs and will reduce parameter estimation uncertainty since more data can be used to estimate the parameters of the current regime. Outliers, on the other hand, will lead investors to infer that there is a higher chance that a break has occurred in the return model, thereby increasing the uncertainty and leading investors to put more weight on the meta distribution. We intend to pursue how these shifts in investors’ beliefs affect optimal asset allocations in future work.

References


Cocco, J., F. Gomes, and P. Maenhout, 2001, Consumption and portfolio choice over the life-cycle, mimeo, London Business School and INSEAD.


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


———, J. Wachter, and W. Boudry, 2002, Does mutual fund performance vary over the business cycle?, Discussion paper NYU/NBER.


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


Appendix A: Multivariate Breakpoint model

Our generalization of the breakpoint model to a multivariate setting involves some modeling choices which are worth explaining in some detail. The main complication that arises when going to a multivariate setting is how to parameterize the variance-covariance matrix within each regime and also how to estimate its meta distribution parameters. Several approaches have been proposed for the choice of prior probability models in the literature. The most commonly used prior is the conjugate inverse Wishart (Bernardo and Smith (1994)). However, this distribution rules out using different degrees of prior informativeness for each of the included time series since the degree of freedom parameter $\nu$ is the only ‘tuning’ parameter available to express uncertainty in this model. The factor setup proposed by West (2003) and Aguilar and West (2000) also has the problem that it is difficult to interpret the factors and factor loadings. This is an important consideration since it restricts our ability to suggest informative prior distributions.

Several noninformative priors have been proposed for the covariance matrices, such as Jeffreys’ prior $p_j(\Sigma) = 1/|\Sigma|^{(m+1)/2}$, (where $m$ is the number of time series in the VAR) or the reference prior, $p_R(\Sigma) \propto 1/\left| \Sigma \right| \prod_{i<j} (d_i - d_j)$ (where $d_i$ are the eigenvalues of $\Sigma$) proposed by Yang and Berger (1994). The lack of intuition for the relationship between the eigenvalues and correlation parameters makes it difficult to work with this model and impose economically motivated constraints. Daniels (1999) proposes a uniform shrinkage prior that obtains the posterior mean as a linear combination of the prior mean and the sample average. The log matrix prior introduced by Leonard and Hsu (1992) uses a logarithmic transformation of the eigenvalue/eigenvector decomposition of $\Sigma$ and allows for hierarchical shrinkage based on the eigenvalues. This achieves dimensionality reduction, but once again it is difficult to interpret the relationship between the log of the eigenvalues and the correlations and standard deviations.

Barnard, McCulloch, and Meng (2000) propose to decompose $\Sigma$ by assuming independent priors for the standard deviations $\text{diag}(\psi)$—where $\psi$ is an $m \times 1$ vector of standard deviations—and the $m \times m$ correlation matrix $R$: $\Sigma = \text{diag}(\psi) \times R \times \text{diag}(\psi)$. They propose a variety of prior models for $R$ including a model where the marginal prior for each element of $R$, $r_{ij}$, is a modified beta distribution over $[-1, 1]$. With an appropriate choice of the beta parameters this yields a uniform marginal prior distribution. Wong, Carter, and Kohn (2003) propose a prior probability model for the precision matrix, $P = \Sigma^{-1}$, that is similar to the approach we adopt. Daniels and Kass (1999) extend this setting and allow for a hierarchical formulation. Their prior specifications include a normal prior for a transformation of the correlation coefficients. The constraint of positive definiteness amounts to appropriate truncation of the normal prior.

Since we are interested in drawing new values of the covariance matrix parameters following a break, we follow Daniels and Kass (1999) and introduce additional hierarchical structure on the correlation matrix. We use the decomposition of Barnard, McCulloch, and Meng (2000) and thus model the standard deviations $\text{diag}(\psi)$ and the correlation matrix $R$ separately. This makes economic interpretation of the priors much easier - an essential requirement in financial applications.
Posterior inference in our model relies on Markov chain Monte Carlo simulation methods (Tierney (1994)) but is computationally challenging because we need to sample from truncated distributions due to the constraint that the covariance matrix is positive definite. The truncated distributions involve analytically intractable normalizing constants that are functions of the parameters. To overcome this problem we use Griddy Gibbs Sampling. The Gibbs sampler is explained in Appendix B.

**Appendix B: Gibbs Sampler for the VAR Model with Multiple Breaks**

This appendix builds on Pesaran, Pettenuzzo, and Timmermann (2004) but extends results in this paper to multivariate dynamic models. We are interested in drawing from the posterior distribution 

\[ \pi (\Theta, H, p, S_T | \mathcal{Z}_T) \]

where 

\[ \Theta = \left( \text{vec}(B)_1, \psi_1, R_1, \ldots, \text{vec}(B)_{K+1}, \psi_{K+1}, R_{K+1} \right) \]

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

\[ H = (b_0, V_0, v_{0,1}, d_{0,1}, \ldots, v_{0,m}, d_{0,m}, \mu_{\rho,12}, \sigma_{\rho,12}^2, \ldots, \mu_{\rho,m-1m}, \sigma_{\rho,m-1m}^2) \]

are the hyperparameters of the meta distribution. We also use the notation \( S_T = (s_1, \ldots, s_T) \) for the collection of values of the latent state variable, \( \mathcal{Z}_T = (z_1, \ldots, z_T)' \), and \( p = (p_{11}, p_{22}, \ldots, p_{KK})' \). These summarize the unknown parameters of the transition probability matrix in (4).

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

1. \( \pi (S_T | \Theta, H, p, \mathcal{Z}_T) \)
2. \( \pi (\Theta, H | p, S_T, \mathcal{Z}_T) \)
3. \( \pi (p | S_T) \).

Here we used the identity 

\[ \pi (\Theta, H, p | S_T, \mathcal{Z}_T) = \pi (\Theta, H | p, S_T, \mathcal{Z}_T) \pi (p | S_T) \]

and note that under our assumptions 

\[ \pi (p | \Theta, H, S_T, \mathcal{Z}_T) = \pi (p | S_T) \]

Simulation of the states \( S_T \) requires ‘forward’ and ‘backward’ passes through the data. Define \( S_t = (s_1, \ldots, s_t) \) and \( S_{t+1} = (s_{t+1}, \ldots, s_T) \) as the state history up to time \( t \) and from time \( t \) to \( T \), respectively. We partition the states’ joint density as follows:

\[ p(s_{T-1} | s_T, \Theta, H, p, \mathcal{Z}_T) \times \cdots \times p(s_t | S_{t+1}, \Theta, H, p, \mathcal{Z}_T) \times \cdots \times p(s_1 | S^2, \Theta, H, p, \mathcal{Z}_T). \] (33)

Chib (1995) shows that the generic element of (33) can be decomposed as follows

\[ p(s_t | S_{t+1}, \Theta, H, p, \mathcal{Z}_T) \propto p(s_t | \Theta, H, p, \mathcal{Z}_T)p(s_t | s_{t-1}, \Theta, H, p), \] (34)

30
where the normalizing constant is easily obtained since \( s_t \) takes only two values conditional on the value taken by \( s_{t+1} \). The second term in (34) is simply the transition probability from the Markov chain. The first term can be computed by a recursive calculation (the forward pass through the data) where, for given \( p(s_{t-1} | \Theta, H, p, Z_{t-1}) \), we obtain \( p(s_t | \Theta, H, p, Z_t) \) and \( p(s_{t+1} | \Theta, H, p, Z_{t+1}) \), ..., \( p(s_T | \Theta, H, p, Z_T) \). Suppose \( p(s_{t-1} | \Theta, H, p, Z_{t-1}) \) is available. Then

\[
p(s_t = k | Z_t, \Theta, H, p) = \frac{p(s_t = k | Z_{t-1}, \Theta, H, p) \times f(z_t | vec(B)_k, \Sigma_k, Z_{t-1})}{\sum_{l=0}^{K} p(s_t = l | \Theta, H, p, Z_{t-1}) \times f(z_t | vec(B)_l, \Sigma_l, Z_{t-1})}
\]

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

\[
p(s_t = k | \Theta, H, p, Z_{t-1}) = \sum_{l=0}^{K} p_{lk} \times p(s_{t-1} = l | \Theta, H, p, Z_{t-1}).
\]

For a given set of simulated states, \( S_T \), the data is partitioned into \( K + 1 \) groups. Let \( Z_j = (z'_{t_{j-1}+1}, ..., z'_{t_j}) \) and \( X_j = (z'_{t_{j-1}+1}, ..., z'_{t_{j-1}}) \) be the values of the dependent and independent variables within the \( j \)th regime. To obtain the conditional distributions for the regression parameters and meta hyperparameters, note that the conditional distributions of \( vec(B)_j \) are independent across regimes with\(^{16}\)

\[
vec(B)_j | \Theta - vec(B)_j, H, p, S_T, Z_T \sim N \left( vec(B)_j, \bar{V}_j \right),
\]

where

\[
\bar{V}_j = \left( X_j' \Sigma_j^{-1} X_j + V_0 \right)^{-1}
\]

\[
vec(B)_j = \bar{V}_j \left( X_j' \Sigma_j^{-1} Z_j + V_0 b_0 \right).
\]

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

\[
b_0 | \Theta, H_{-b_0}, p, S_T, Z_T \sim N \left( \bar{\mu}_\beta, \Sigma_\beta \right)
\]

\[
V_0 | \Theta, H_{-V_0}, p, S_T, Z_T \sim W \left( \bar{v}_\beta, V_{\bar{\beta}}^{-1} \right),
\]

where

\[
\Sigma_\beta = \left( \Sigma_\beta^{-1} + (K + 1) V_0 \right)^{-1}
\]

\[
\bar{\mu}_\beta = \Sigma_\beta \left( V_0 \sum_{j=1}^{J} vec(B)_j + \Sigma_\beta^{-1} \mu_\beta \right),
\]

and

\[
\bar{v}_\beta = v_\beta + (K + 1)
\]

\[
V_{\bar{\beta}} = \sum_{j=1}^{J} (vec(B)_j - b_0) (vec(B)_j - b_0)' + V_\beta.
\]

\(^{16}\)Using standard set notation we define \( A_{-b} \) as the complementary set of \( b \) in \( A \), i.e. \( A_{-b} = \{ x \in A : x \neq b \} \).
Moving to the posterior for the precision parameters within each regime $j$ and for each equation $i$, let $\Lambda = (Z_j - X_j B_j)' (Z_j - X_j B_j)$ with $\Lambda_{ij}$ being its $i$-th row and $j$-th column element. Note that 

$$s_{j,i}^{-2} \Theta_{-S_j, H, p, S_T, Z_T} \sim G \left( \frac{v_{0,i} + \Lambda_{ii}}{2}, \frac{d_0,i + n_j}{2} \right),$$

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

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

$$v_{0,i} \Theta, H_{-v_{0,i}}, p, S_T, Z_T \propto \prod_{j=1}^{K+1} G \left( s_{j,i}^{-2} | v_{0,i}, d_0,i \right) \exp \left( v_{0,i} | \rho_{0,i} \right)$$

where

$$d_0,i \Theta, H_{-d_0,i}, p, S_T, Z_T \sim G \left( v_{0,i} (K + 1) + c_{0,i}, \sum_{j=1}^{K+1} s_{j,i}^{-2} + d_0,i \right).$$

Drawing $v_{0,i}$ from (35) is complicated since we cannot make use of any standard distributions. We therefore introduce a Metropolis-Hastings (M-H) step in the Gibbs sampling algorithm. At each loop of the Gibbs sampling we draw a value $v_{0,i}^*$ from a Gamma distributed candidate generating density,

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

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

$$\xi \left( v_{0,i}^* | v_{0,i}^{g-1} \right) = \min \left[ \frac{\pi \left( v_{0,i}^* | \Theta, H_{-v_{0,i}}, p, S_T, Z_T \right)}{\pi \left( v_{0,i}^{g-1} | \Theta, H_{-v_{0,i}}, p, S_T, Z_T \right)} \right] = \frac{1}{\xi \left( v_{0,i}^* | v_{0,i}^{g-1} \right)}.$$ (36)

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

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

$$f \left( r_{j,ic} | \Theta_{-R_j}, H, p, S_T, Z_T \right) \propto |R_j|^{-m/2} \exp \left\{ -tr \left( R_j^{-1} C_j \right) / 2 \right\} \times \exp \left\{ -\left( r_{j,ic} - \mu_{r,ic} \right) / (2 \sigma_{r,ic}^2) \right\} I \{ R \in \mathbb{R}^m \}$$

where $I \{ \cdot \}$ is the indicator function and $C_j$ is the correlation matrix in regime $j$. The indicator function $I \{ R \in \mathbb{R}^m \}$ ensures that the correlation matrix is positive definite and introduces dependence among the $r_{j,ic}$-values.
The full conditional densities for $\mu_{p,ic}$ and $\sigma^2_{p,ic}$ are similar to the conjugate densities with an additional factor due to the constraint requiring $R_j$ to be positive definite:

$$
f \left( \mu_{p,ic} \mid \Theta, H_{-\mu_{p,ic}}, S_T, Z_T \right) \propto \prod_{j=1}^{K+1} \exp \left\{ -\left( r_{j,ic} - \mu_{p,ic} \right)^2 / \left( 2\sigma^2_{p,ic} \right) \right\} \exp \left\{ -\left( \mu_{p,ic} - \mu_{\mu,ic} \right) / \left( 2\sigma^2_{\mu,ic} \right) \right\} I \{ R \in \mathcal{R}^m \} \tag{38}
$$

$$
f \left( \sigma^2_{p,ic} \mid \Theta, H_{-\sigma^2_{p,ic}}, p, S_T, Z_T \right) \propto \prod_{j=1}^{K+1} \exp \left\{ -\left( r_{j,ic} - \mu_{p,ic} \right)^2 / \left( 2\sigma^2_{p,ic} \right) \right\} \sigma^2_{p,ic} \left( 1 - a_{p,ic} \right) \exp \left( -b_{p,ic} / \sigma^2_{p,ic} \right) I \{ R \in \mathcal{R}^m \} \tag{39}
$$

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

Finally, $p$ is simulated from the conditional beta posterior

$$p_{ij} \mid S_T \sim Beta(a + l_j, b + 1),$$

where $l_j = \tau - \tau - 1$ is the duration of regime $j$.

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

$$\pi(a \mid \Theta, H_{-b}, S_T, p, Z_T) \propto \prod_{j=1}^{K} Beta(p_{ij} \mid a, b) Gamma(a \mid a_0, b_0),$$

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

$$q(a^* \mid a^g) \sim G(\zeta, \zeta / a^g),$$

and acceptance probability

$$\xi(a^* \mid a^g) = \min \left[ \frac{\pi(a^* \mid \Theta, H_{-b}, S_T, p, Z_T) / q(a^* \mid a^g)}{\pi(a^g \mid \Theta, H_{-b}, S_T, p, Z_T) / q(a^g \mid a^* \mid a^g), 1} \right].$$

If there are no new breaks in the out-of-sample period, we obtain a draw from $\pi(B_{K+1}, \Sigma_{K+1} \mid H, p, S_T, Z_T)$ and then, conditional on these parameters, draw $R_{T+h}$ from the posterior predictive density,

$$R_{T+h} \sim p(R_{T+h} \mid B_{K+1}, \Sigma_{K+1}, S_{T+h} = K + 1, S_T = K + 1, Z_T) \tag{40}.$$

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

Draw $b_0$ from

$$b_0 \sim \pi(b_0 \mid \Theta, H_{-b_0}, P, S_T, Z_T),$$
and $V_0$ from
\[ V_0 \sim \pi \left( V_0^{-1} \mid \Theta, H^{-V_0}, P, S_T, Z_T \right). \]

Draw $v_{0,1}$ and $v_{0,2}$ from
\[ v_{0,i} \sim \pi \left( v_{0,i} \mid \Theta, H^{-v_{0,i}}, P, S_T, Z_T \right), \]
and $d_{0,1}$ and $d_{0,2}$ from
\[ d_{0,i} \sim \pi \left( d_{0,i} \mid \Theta, H^{-d_{0,i}}, P, S_T, Z_T \right). \]

Draw $\mu_{\rho,12}$ from
\[ \mu_{\rho,12} \sim \pi \left( \mu_{\rho,12} \mid \Theta, H^{-\mu_{\rho,12}}, P, S_T, Z_T \right), \]
and $\sigma^2_{\rho,12}$ from
\[ \sigma^2_{\rho,12} \sim \pi \left( \sigma^2_{\rho,12} \mid \Theta, H^{-\sigma^2_{\rho,12}}, P, S_T, Z_T \right). \]

Draw $B_{K+2}$ and $\Sigma_{K+2}$ from their respective priors given by $\pi \left( B_{K+2} \mid b_0, V_0 \right)$ and
\[ \pi \left( \Sigma_{K+2} \mid v_{0,1}, d_{0,1}, v_{0,m}, d_{0,m}, \mu_{\rho,12}, \sigma^2_{\rho,12}, Z_T \right), \]
respectively, for a fixed set of hyperparameters.

Draw $R_{T+h}$ from the posterior predictive density,
\[ R_{T+h} \sim p \left( R_{T+h} \mid S_{T+h} = K + 2, \tau_{K+1} = T + j, S_T = K + 1, Z_T \right). \tag{41} \]

To obtain the estimate of $p_{K+1,K+1}$ needed in (7), we combine information from the last regime with prior information, assuming the prior $p_{K+1,K+1} \sim \text{Beta}(a, b)$, so
\[ p_{K+1,K+1} \mid Z_T \sim \text{Beta}(a + n_{K+1,K+1}, b + 1) \tag{42} \]
where $n_{K+1,K+1}$ is the number of observations from regime $K + 1$.

Turning to the case with an arbitrary number of future breaks, to draw from the distribution of the parameters $a, b$ that characterize the break point probability, we use that the conditional posterior distribution for $a$ and $b$ are
\[ \pi \left( a \mid \Theta, H^{-b}, S_T, P, Z_T \right) \propto \prod_{j=1}^{K} \text{Beta} \left( p_{jj} \mid a \right) \text{Gamma} \left( a \mid a_0, b_0 \right) \]
\[ \pi \left( b \mid \Theta, H^{-b}, S_T, P, Z_T \right) \propto \prod_{j=1}^{K} \text{Beta} \left( p_{jj} \mid a \right) \text{Gamma} \left( b \mid a_0, b_0 \right). \]

Using these new posterior distributions, we generate draws for $p_{K+2,K+2}$ using the convolution of the prior distribution for the $p_{ii}$’s and the resulting posterior densities for $a$ and $b$,\(^\dagger\)
\[ p_{K+2,K+2} \mid a, b \sim \text{Beta}(a, b). \]

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

34
Table 1: Posterior parameter estimates for the full sample (1926:12 - 2003:12) VAR model with uninformative priors and no breaks. The model in the left panel includes excess returns ($r$) and the dividend yield ($Yld$), while the model in the right panel includes excess returns and the T-bill rate ($TB$).
|   | $f(Z_T|\Theta_i, M_i)$ | $f(Z_T|M_i)$ | $P(M_i|Z_T)$ | Break locations |
|---|------------------|-------------|-------------|-----------------|
| 0 | 1368.18          | 1319.34     | 0.00        |                 |
| 1 | 1433.09          | 1369.33     | 0.00        | Apr-52          |
| 2 | 1445.83          | 1383.28     | 0.00        | Feb-52          |
| 3 | 1454.38          | 1387.49     | 0.00        | May-40 Oct-52   |
| 4 | 1459.03          | 1390.04     | 0.00        | May-40 Feb-58   |
| 5 | 1462.11          | 1391.51     | 0.00        | May-40 Feb-58   |
| 6 | 1476.60          | 1404.75     | 0.02        | Aug-32 May-40   |
| 7 | 1482.55          | 1408.74     | 0.98        | Aug-32 May-40   |

|   | $f(Z_T|\Theta_i, M_i)$ | $f(Z_T|M_i)$ | $P(M_i|Z_T)$ | Break locations |
|---|------------------|-------------|-------------|-----------------|
| 0 | 1368.52          | 1323.33     | 0.00        |                 |
| 1 | 1378.69          | 1318.98     | 0.00        | Sep-57          |
| 2 | 1448.08          | 1389.49     | 0.00        | Nov-37 Aug-47   |
| 3 | 1466.27          | 1405.79     | 0.00        | Jan-34 Aug-47   |
| 4 | 1471.20          | 1407.12     | 0.00        | Jan-34 Aug-47   |
| 5 | 1476.13          | 1410.67     | 0.16        | Jan-34 Aug-47   |
| 6 | 1474.18          | 1406.79     | 0.00        | Jan-34 Aug-47   |
| 7 | 1481.13          | 1412.29     | 0.83        | Jan-34 Aug-47   |

Table 2: Model comparison and selection of the number of breaks in the return forecasting model. The table shows estimates of the conditional log-likelihood for stock returns and the predictor variable (either the dividend yield or the T-bill rate), $Z_T$, along with marginal log-likelihood estimates for returns and posterior probabilities for different numbers of breaks along with the time of the break points for the different models. $M_i$ denotes the return forecasting model with $i$ breaks. The top and bottom panels display results when the predictor for the excess return is the lagged dividend yield and the lagged T-Bill rate, respectively. The data sample is 1926:12 - 2003:12.
<table>
<thead>
<tr>
<th>Regimes</th>
<th>27-32</th>
<th>32-40</th>
<th>40-52</th>
<th>52-58</th>
<th>58-74</th>
<th>74-82</th>
<th>82-92</th>
<th>92-03</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_r$</td>
<td>0.016</td>
<td>-0.028</td>
<td>-0.026</td>
<td>-0.006</td>
<td>-0.086</td>
<td>-0.056</td>
<td>-0.020</td>
<td></td>
</tr>
<tr>
<td>s.d.</td>
<td>0.015</td>
<td>-0.024</td>
<td>0.033</td>
<td>0.021</td>
<td>0.033</td>
<td>0.028</td>
<td>0.011</td>
<td></td>
</tr>
<tr>
<td>$\beta_r$</td>
<td>0.825</td>
<td>0.737</td>
<td>0.479</td>
<td>2.090</td>
<td>1.999</td>
<td>1.747</td>
<td>1.408</td>
<td></td>
</tr>
<tr>
<td>s.d.</td>
<td>0.026</td>
<td>0.033</td>
<td>0.015</td>
<td>0.024</td>
<td>0.021</td>
<td>0.033</td>
<td>0.028</td>
<td></td>
</tr>
<tr>
<td>$\sigma_r$</td>
<td>0.105</td>
<td>0.086</td>
<td>0.038</td>
<td>0.034</td>
<td>0.038</td>
<td>0.050</td>
<td>0.046</td>
<td>0.045</td>
</tr>
<tr>
<td>s.d.</td>
<td>0.009</td>
<td>0.007</td>
<td>0.002</td>
<td>0.003</td>
<td>0.002</td>
<td>0.004</td>
<td>0.003</td>
<td>0.003</td>
</tr>
<tr>
<td>$\mu_x$</td>
<td>0.002</td>
<td>0.002</td>
<td>0.001</td>
<td>0.001</td>
<td>0.002</td>
<td>0.001</td>
<td>2.0E-04</td>
<td></td>
</tr>
<tr>
<td>s.d.</td>
<td>0.002</td>
<td>0.002</td>
<td>0.001</td>
<td>0.001</td>
<td>0.002</td>
<td>0.001</td>
<td>2.0E-04</td>
<td></td>
</tr>
<tr>
<td>$\beta_x$</td>
<td>0.967</td>
<td>0.916</td>
<td>0.967</td>
<td>0.951</td>
<td>0.919</td>
<td>0.908</td>
<td>0.940</td>
<td>0.979</td>
</tr>
<tr>
<td>s.d.</td>
<td>0.016</td>
<td>0.035</td>
<td>0.016</td>
<td>0.022</td>
<td>0.021</td>
<td>0.034</td>
<td>0.026</td>
<td>0.009</td>
</tr>
<tr>
<td>$\sigma_x$</td>
<td>0.007</td>
<td>0.004</td>
<td>0.002</td>
<td>0.001</td>
<td>0.001</td>
<td>0.002</td>
<td>0.001</td>
<td>6.4E-04</td>
</tr>
<tr>
<td>s.d.</td>
<td>6.3E-04</td>
<td>3.6E-04</td>
<td>1.4E-04</td>
<td>1.3E-04</td>
<td>5.8E-05</td>
<td>1.6E-04</td>
<td>1.0E-04</td>
<td>3.8E-05</td>
</tr>
<tr>
<td>$\rho_{rx}$</td>
<td>-0.930</td>
<td>-0.936</td>
<td>-0.858</td>
<td>-0.928</td>
<td>-0.955</td>
<td>-0.963</td>
<td>-0.941</td>
<td>-0.946</td>
</tr>
<tr>
<td>s.d.</td>
<td>0.021</td>
<td>0.022</td>
<td>0.023</td>
<td>0.023</td>
<td>0.020</td>
<td>0.017</td>
<td>0.021</td>
<td>0.020</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.982</td>
<td>0.986</td>
<td>0.990</td>
<td>0.983</td>
<td>0.992</td>
<td>0.987</td>
<td>0.988</td>
<td>1</td>
</tr>
<tr>
<td>s.d.</td>
<td>0.013</td>
<td>0.010</td>
<td>0.008</td>
<td>0.013</td>
<td>0.006</td>
<td>0.010</td>
<td>0.010</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 3: Hierarchical hidden Markov chain estimates for the return ($r_t$) forecasting model with seven break points, based on the lagged dividend yield ($x_{t-1}$) as a predictor variable: $r_t = \mu_{r_j} + \beta_{r_j} x_{t-1} + \epsilon_{rt}, \epsilon_{rt} \sim N(0, \sigma_r^2), x_t = \mu_{x_j} + \beta_{x_j} x_{t-1} + \epsilon_{xt}, \epsilon_{xt} \sim N(0, \sigma_x^2), P(r_s = j | s_{t-1} = j) = p_{jj}, \text{corr}(\epsilon_{rt}, \epsilon_{xt}) = \rho_{rx_j}, \tau_{j-1} + 1 \leq t \leq \tau_j$. 

37
### Hyperparameters of Meta distributions

#### I Return equation

<table>
<thead>
<tr>
<th>Mean Parameters</th>
<th>mean</th>
<th>s.d.</th>
<th>95% conf interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b_0(\mu_r)$</td>
<td>-0.042</td>
<td>0.038</td>
<td>-0.123 0.033</td>
</tr>
<tr>
<td>$b_0(\beta_r)$</td>
<td>1.218</td>
<td>0.508</td>
<td>0.225 2.209</td>
</tr>
</tbody>
</table>

**Variance Parameters**

<table>
<thead>
<tr>
<th>mean</th>
<th>s.d.</th>
<th>95% conf interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_0(\mu_r)$</td>
<td>0.013</td>
<td>0.007</td>
</tr>
<tr>
<td>$V_0(\beta_r)$</td>
<td>1.984</td>
<td>1.321</td>
</tr>
</tbody>
</table>

**Error term precision**

<table>
<thead>
<tr>
<th>mean</th>
<th>s.d.</th>
<th>95% conf interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>$v_{01}$</td>
<td>3.277</td>
<td>1.118 1.479 6.258</td>
</tr>
<tr>
<td>$d_{01}$</td>
<td>0.007</td>
<td>0.003 0.002 0.013</td>
</tr>
</tbody>
</table>

#### II Dividend Yield equation

<table>
<thead>
<tr>
<th>Mean Parameters</th>
<th>mean</th>
<th>s.d.</th>
<th>95% conf interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b_0(\mu_y)$</td>
<td>0.003</td>
<td>0.002</td>
<td>1.0E-04 0.008</td>
</tr>
<tr>
<td>$b_0(\beta_y)$</td>
<td>0.918</td>
<td>0.033</td>
<td>0.839 0.972</td>
</tr>
</tbody>
</table>

**Variance Parameters**

<table>
<thead>
<tr>
<th>mean</th>
<th>s.d.</th>
<th>95% conf interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_0(\mu_y)$</td>
<td>1.0E-04</td>
<td>1.0E-04</td>
</tr>
<tr>
<td>$V_0(\beta_y)$</td>
<td>0.014</td>
<td>0.008</td>
</tr>
</tbody>
</table>

**Error term precision**

<table>
<thead>
<tr>
<th>mean</th>
<th>s.d.</th>
<th>95% conf interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>$v_{02}$</td>
<td>0.913</td>
<td>0.388 0.341 2.054</td>
</tr>
<tr>
<td>$d_{02}$</td>
<td>1.9E-06</td>
<td>9.6E-07 5.9E-07 3.8E-06</td>
</tr>
</tbody>
</table>

**Correlation parameters**

<table>
<thead>
<tr>
<th>mean</th>
<th>s.d.</th>
<th>95% conf interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_{p}$</td>
<td>-0.920</td>
<td>0.042 -0.984 0.831</td>
</tr>
<tr>
<td>$\sigma_{p}^2$</td>
<td>0.020</td>
<td>0.016 0.010 0.058</td>
</tr>
</tbody>
</table>

**Transition Probability parameters**

<table>
<thead>
<tr>
<th>mean</th>
<th>s.d.</th>
<th>95% conf interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_0$</td>
<td>33.258</td>
<td>15.015 10.498 73.091</td>
</tr>
<tr>
<td>$b_0$</td>
<td>0.806</td>
<td>0.308 0.357 1.462</td>
</tr>
</tbody>
</table>

Table 4: Hierarchical hidden Markov chain prior hyperparameter estimates for return forecasting model with seven break points, based on the dividend yield as the predictor variable. $z_t = B'_j x_{t-1} + u_t$ where $z_t = (r_t, x_t)'$, the excess return and the predictor variable, and $E[u_t u_t'] = \Sigma_j = diag(\psi_j) \times R_j \times diag(\psi_j)$, $\tau_{j-1} + 1 \leq t \leq \tau_j$. The meta distributions are specified as: $vec(B)_j \sim N(b_0, V_0)$, $j = 1, ..., K + 1$, $\psi_{j,i}^{-2} \sim Gamma (v_{0,i}, d_{0,i})$, $i = 1, 2$, $\rho_j \sim N(\mu_p, \sigma_p)$, $p_{jj} \sim Beta (a_0, b_0)$. 38
Table 5: Hierarchical hidden Markov chain estimates for the return ($r_t$) forecasting model with seven break points, based on the lagged T-bill ($x_{t-1}$) as a predictor variable: $r_t = \mu_{r_j} + \beta_{r_j} x_{t-1} + \epsilon_{rt}$, $\epsilon_{rt} \sim N\left(0, \sigma_{r_j}^2\right)$, $x_t = \mu_{x_j} + \beta_{x_j} x_{t-1} + \epsilon_{xt}$, $\epsilon_{xt} \sim N\left(0, \sigma_{x_j}^2\right)$, $\Pr(s_t = j | s_{t-1} = j) = p_{jj}$, $corr(\epsilon_{rt}, \epsilon_{xt}) = \rho_{rx_j}$, $\tau_{j-1} + 1 \leq t \leq \tau_j$. 

<table>
<thead>
<tr>
<th>Regimes</th>
<th>27-34</th>
<th>34-47</th>
<th>47-52</th>
<th>52-68</th>
<th>68-79</th>
<th>79-82</th>
<th>82-90</th>
<th>90-03</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_r$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td>-0.001</td>
<td>0.007</td>
<td>0.016</td>
<td>0.027</td>
<td>0.037</td>
<td>0.092</td>
<td>0.021</td>
<td>0.009</td>
</tr>
<tr>
<td>s.d.</td>
<td>0.018</td>
<td>0.005</td>
<td>0.011</td>
<td>0.007</td>
<td>0.019</td>
<td>0.034</td>
<td>0.024</td>
<td>0.010</td>
</tr>
<tr>
<td>$\beta_r$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>s.d.</td>
<td>6.886</td>
<td>11.343</td>
<td>10.542</td>
<td>3.015</td>
<td>4.074</td>
<td>3.656</td>
<td>4.086</td>
<td>2.724</td>
</tr>
<tr>
<td>$\sigma_r$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td>0.109</td>
<td>0.059</td>
<td>0.036</td>
<td>0.033</td>
<td>0.047</td>
<td>0.047</td>
<td>0.049</td>
<td>0.044</td>
</tr>
<tr>
<td>s.d.</td>
<td>0.008</td>
<td>0.003</td>
<td>0.004</td>
<td>0.002</td>
<td>0.003</td>
<td>0.006</td>
<td>0.004</td>
<td>0.002</td>
</tr>
</tbody>
</table>

| $\mu_x$ |       |       |       |       |       |       |       |       |
| mean    | 6.0E-05| 1.3E-05| 1.7E-04| 1.1E-04| 3.2E-04| 1.3E-03| 4.9E-04| 6.0E-05|
| s.d.    | 4.2E-05| 5.3E-06| 6.1E-05| 4.8E-05| 1.4E-04| 6.9E-04| 2.1E-04| 4.1E-05|
| $\beta_x$ |       |       |       |       |       |       |       |       |
| mean    | 0.958 | 0.924 | 0.844 | 0.959 | 0.937 | 0.832 | 0.914 | 0.974 |
| s.d.    | 0.021 | 0.027 | 0.065 | 0.020 | 0.030 | 0.075 | 0.037 | 0.012 |
| $\sigma_x$ |       |       |       |       |       |       |       |       |
| mean    | 3.4E-04| 4.4E-05| 1.1E-04| 2.7E-04| 4.5E-04| 0.0013| 4.7E-04| 2.5E-04|
| s.d.    | 2.7E-05| 2.7E-06| 1.6E-05| 1.5E-05| 2.9E-05| 1.7E-04| 3.6E-05| 1.5E-05|
| $\rho_{rx}$ |       |       |       |       |       |       |       |       |
| mean    | 0.067 | 0.003 | 0.170 | -0.011 | -0.189 | -0.311 | 0.302 | 0.067 |
| s.d.    | 0.026 | 0.026 | 0.046 | 0.033 | 0.034 | 0.033 | 0.037 | 0.030 |
| $p$     |       |       |       |       |       |       |       |       |
| mean    | 0.985 | 0.991 | 0.981 | 0.992 | 0.989 | 0.976 | 0.987 | 1    |
| s.d.    | 0.012 | 0.007 | 0.015 | 0.006 | 0.009 | 0.019 | 0.010 | 0    |
### Hyperparameters of Meta Distributions

#### I Return equation

<table>
<thead>
<tr>
<th>Mean Parameters</th>
<th>mean</th>
<th>s.d.</th>
<th>95% conf interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>(b_0(\mu_r))</td>
<td>0.026</td>
<td>0.040</td>
<td>-0.062 0.101</td>
</tr>
<tr>
<td>(b_0(\beta_r))</td>
<td>-4.645</td>
<td>4.840</td>
<td>-14.833 4.430</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variance Parameters</th>
<th>mean</th>
<th>s.d.</th>
<th>95% conf interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>(V_0(\mu_r))</td>
<td>0.013</td>
<td>0.009</td>
<td></td>
</tr>
<tr>
<td>(V_0(\beta_r))</td>
<td>167.497</td>
<td>104.511</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Error term precision</th>
<th>mean</th>
<th>s.d.</th>
<th>95% conf interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>(v_{01})</td>
<td>4.103</td>
<td>1.818</td>
<td>1.303 8.093</td>
</tr>
<tr>
<td>(d_{01})</td>
<td>0.009</td>
<td>0.004</td>
<td>0.002 0.018</td>
</tr>
</tbody>
</table>

#### II T-bill equation

<table>
<thead>
<tr>
<th>Mean Parameters</th>
<th>mean</th>
<th>s.d.</th>
<th>95% conf interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>(b_0(\mu_t))</td>
<td>4.6E-04</td>
<td>3.4E-04</td>
<td>4.8E-05 0.001</td>
</tr>
<tr>
<td>(b_0(\beta_t))</td>
<td>0.897</td>
<td>0.042</td>
<td>0.793 0.961</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variance Parameters</th>
<th>mean</th>
<th>s.d.</th>
<th>95% conf interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>(V_0(\mu_t))</td>
<td>0.011</td>
<td>0.007</td>
<td></td>
</tr>
<tr>
<td>(V_0(\beta_t))</td>
<td>0.016</td>
<td>0.012</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Error term precision</th>
<th>mean</th>
<th>s.d.</th>
<th>95% conf interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>(v_{02})</td>
<td>0.492</td>
<td>0.174</td>
<td>0.207 0.902</td>
</tr>
<tr>
<td>(d_{02})</td>
<td>7.5E-09</td>
<td>4.2E-09</td>
<td>2.1E-09 1.6E-08</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Correlation parameters</th>
<th>mean</th>
<th>s.d.</th>
<th>95% conf interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\mu_\rho)</td>
<td>0.013</td>
<td>0.110</td>
<td>-0.205 0.229</td>
</tr>
<tr>
<td>(\sigma^2_\rho)</td>
<td>0.095</td>
<td>0.091</td>
<td>0.020 0.336</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Transition Probability parameters</th>
<th>mean</th>
<th>s.d.</th>
<th>95% conf interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a_0)</td>
<td>26.625</td>
<td>13.103</td>
<td>7.656 54.678</td>
</tr>
<tr>
<td>(b_0)</td>
<td>0.660</td>
<td>0.288</td>
<td>0.237 1.271</td>
</tr>
</tbody>
</table>

Table 6: Hierarchical hidden Markov chain prior hyperparameter estimates for return forecasting model with seven break points, based on the tT-bill as the predictor variable. \(z_t = B_j' x_{t-1} + u_t\), where \(z_t = (r_t, x_t)'\), the excess return and the predictor variable, and \(E[u_t u'_t] = \Sigma_j = diag(\psi_j) \times R_j \times diag(\psi_j), \tau_{j-1} + 1 \leq t \leq \tau_j\). The meta distributions are specified as: \(\text{vec}(B)_j \sim N(b_0, V_0)\), \(j = 1, ..., K + 1, \psi_{ji}^2 \sim \Gamma(v_{0,i}, d_{0,i}), i = 1, 2, \rho_j \sim N(\mu_\rho, \sigma_\rho), p_{ji} \sim \text{Beta}(a_0, b_0)\).
Figure 1: Posterior probabilities of breakpoint locations for the return prediction model based on the dividend yield with seven breaks. The estimation sample is 1926:12 - 2003:12.
Figure 2: Posterior probabilities of breakpoint locations for the return prediction model based on the T-bill rate with seven breaks. The estimation sample is 1926:12 - 2003:12.
Figure 3: Optimal Asset Allocation as a function of the investment horizon for an investor with power utility over terminal wealth, $U(W_{T+h}) = \frac{1}{(1-A)} W_{T+h}^{1-A}$ where $h$ is the forecast horizon and $A$ is the coefficient of relative risk aversion. The panels show allocations to stocks under the assumption that the dividend yield is set at its value at the end of the sample, $Yld_T = 1.5\%$. The dotted line shows allocations starting from the regime at the end of the sample (2003:12). The dashed line shows the full-sample allocation ignoring breaks and parameter estimation uncertainty. The dashed/dotted line shows allocations based on full-sample parameter values (no breaks) but accounting for parameter uncertainty. Finally, the solid line shows allocations under the composite model which accounts for historical breaks and considers the possibility of future breaks.
Figure 4: Optimal Asset Allocation as a function of the investment horizon for an investor with power utility over terminal wealth, $U(W_{T+h}) = \frac{1}{1-A} W_{T+h}^{1-A}$ where $h$ is the forecast horizon and $A$ is the coefficient of relative risk aversion. The panels show allocations to stocks under the assumption that the dividend yield is set at its sample mean, $Yld_T = 4\%$. The dashed line shows the full-sample allocation ignoring breaks and parameter estimation uncertainty. The dashed/dotted line shows allocations based on full-sample parameter values (no breaks) but accounting for parameter uncertainty. Finally, the solid line shows allocations under the composite model which accounts for historical breaks and considers the possibility of future breaks.
Figure 5: Optimal Asset Allocation as a function of the investment horizon for an investor with power utility over terminal wealth, $U(W_{T+h}) = \frac{1}{(1-A)} W_{T+h}^{1-A}$ where $h$ is the forecast horizon and $A$ is the coefficient of relative risk aversion. The panels show allocations to stocks under the assumption that the T-bill rate is set at its value at the end of the sample, $TB_T = 0.83\%$. The dotted line shows allocations starting from the regime at the end of the sample (2003:12). The dashed line shows the full-sample allocation ignoring breaks and parameter estimation uncertainty. The dashed/dotted line shows allocations based on full-sample parameter values (no breaks) but accounting for parameter uncertainty. Finally, the solid line shows allocations under the composite model which accounts for historical breaks and considers the possibility of future breaks.
Figure 6: Optimal Asset Allocation as a function of the investment horizon for an investor with power utility over terminal wealth, $U(W_{T+h}) = \frac{1}{(1-A)} W_{T+h}^{1-A}$ where $h$ is the forecast horizon and $A$ is the coefficient of relative risk aversion. The panels show allocations to stocks under the assumption that the T-bill rate is set at its sample average, $TB_T = 3.63\%$. The dashed line shows the full-sample allocation ignoring breaks and parameter estimation uncertainty. The dashed/dotted line shows allocations based on full-sample parameter values (no breaks) but accounting for parameter uncertainty. Finally, the solid line shows allocations under the composite model which accounts for historical breaks and considers the possibility of future breaks.
Figure 7: Stock allocations accounting for uncertainty about the number of in-sample breaks. The left panels show stock holdings that average across models based on the dividend yield as a predictor variable, but using Bayesian Model Averaging to average across models with different numbers of breaks. The right panels plot the allocations when averaging across models with different numbers of breaks based on the T-bill rate as the predictor variable. In all cases the allocations are based on the composite model and hence allows for the possibility of future (out-of-sample) breaks.
Figure 8: Stock allocations accounting for model uncertainty and uncertainty about the number of in-sample breaks. The panels show stock holdings that use Bayesian Model Averaging to integrate across models based either on the dividend yield or on the T-bill rate as predictor variables. We also integrate over the number of assumed in-sample breaks. Allocations are based on the composite model and hence allows for the possibility of future (out-of-sample) breaks.
Figure 9: Mean (upper panel) and standard deviation (lower panel) of the predictive density of excess returns in the eight regimes. Dotted lines show the means (upper panels) or standard deviations (lower panels) of the predictive density within each of the individual regimes. The dashed/dotted lines show the mean and standard deviations under the full sample (no break) case, while the solid lines show first and second moments under the meta distribution. All plots are based on the model that uses the dividend yield as the predictor variable. The left panel imposes the constraints $0 \leq \frac{\mu_x}{1 - \beta_x} \leq 0.08$, $\beta_x < 1$ while models in the right panel are estimated under the additional constraint $0 \leq \mu_x + \beta_x \frac{\mu_y}{1 - \beta_x} \leq 0.01$. 


Figure 10: Mean (upper panel) and standard deviation (lower panel) of the predictive density of excess returns in the eight regimes. Dotted lines show the means (upper panels) or standard deviations (lower panels) of the predictive density within each of the individual regimes. The dashed/dotted lines show the mean and standard deviations under the full sample (no break) case, while the solid lines show first and second moments under the meta distribution. All plots are based on the model that uses the T-bill rate as the predictor variable. The left panel imposes the constraints \(0 \leq \frac{\mu_x}{1 - \beta_x} \leq 0.009, \beta_x < 1\) while models in the right panel are estimated under the additional constraint \(0 \leq \mu_r + \beta_r \frac{\mu_x}{1 - \beta_x} \leq 0.01\).
Figure 11: In-sample mean and standard deviations of the predictive density for excess return model based on the dividend yield predictor. The left panels show in-sample means of the predictive densities when the model used to obtain the forecast is the full sample (no break) while the right panels allow for seven breaks. Mean excess returns are shown in the top panels while the bottom panels show standard deviations of the predictive return distribution.
Figure 12: In-sample mean and standard deviations of the predictive density for excess return model based on the T-bill rate predictor variable. The left panels show in-sample means of the predictive densities when the model used to obtain the forecast is the full sample (no break) while the right panels allow for seven breaks. Mean excess returns are shown in the top panels while the bottom panels show standard deviations of the predictive return distribution.
Figure 13: Sharpe ratio predictive densities for all regimes, where the forecast horizon $h = 1$. 