The Equity Premium and Structural Breaks

ĽUBOŠ PÁSTOR and ROBERT F. STAMBAUGH*

ABSTRACT

A long return history is useful in estimating the current equity premium even if the historical distribution has experienced structural breaks. The long series helps not only if the timing of breaks is uncertain but also if one believes that large shifts in the premium are unlikely or that the premium is associated, in part, with volatility. Our framework incorporates these features along with a belief that prices are likely to move opposite to contemporaneous shifts in the premium. The estimated premium since 1834 fluctuates between 4 and 6 percent and exhibits its sharpest drop in the last decade.

ONE OF THE MOST IMPORTANT but elusive quantities in finance is the equity premium, the expected rate of return on the aggregate stock market in excess of the riskless interest rate (the expected "excess return"). It is well known that estimates of the equity premium based on historical data can vary widely, depending on the methodology and the sample period, and the imprecision in such estimates can figure prominently in inference and decision making. Pástor and Stambaugh (1999) conclude, for example, that seven decades of data produce an equity premium estimate whose imprecision typically accounts for the largest fraction of uncertainty about a firm's cost of equity. Long histories offer the prospect of increased precision, and researchers have constructed and analyzed series of U.S. equity returns and interest rates that begin early in the 19th century (e.g., Schwert (1990), Siegel (1992), and Goetzmann and Ibbotson (1994)). 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." We incorporate various economic considerations in estimating the equity premium from a long series of returns whose distribution is subject to structural breaks.

In standard approaches to models that admit structural breaks, estimates of current parameters rely on data only since the most recent estimated

break. Discarding the earlier data reduces the risk of contaminating an estimate of the equity premium with data generated under a different mean. That practice seems prudent, but it contends with the reality that shorter histories typically yield less precision. Equity returns observed before a suspected break are likely to provide at least some information about the current premium. To take an extreme example, suppose one is confident that a shift in the equity premium occurred just a month ago. Discarding virtually all of the historical data on equity returns would certainly remove the risk of contamination by pre-break data, but it hardly seems sensible in estimating the current equity premium. Completely discarding the pre-break returns is appropriate only when one believes the premium might have shifted to such a degree that the pre-break returns are no more useful in estimating the current premium than, say, pre-break rainfall data, but such a view almost surely ignores economics.

A long return series also helps in estimating the current equity premium if one believes that, across subperiods separated by breaks, there is at least some positive association between the equity premium and volatility. Such an association might not be represented well by a specific parametric relation, but it can be represented as a flexible prior belief that, when combined with a long return history, provides information about the current equity premium. In essence, each earlier subperiod's ratio of equity premium to variance, its "price of risk," provides some information about the current price of risk and, given current estimated volatility, about the current equity premium as well. The strength of one's prior belief about the premium–volatility link is characterized simply by the dispersion in the prices of risk across subperiods. Although he implements the idea differently, using parametric relations, Merton (1980) also proposes that one impose a prior belief in a premium–volatility link when estimating the current equity premium.

Basic principles of discounting suggest that a shift in the equity premium is likely to be accompanied by a price change in the opposite direction. To incorporate this property, we assume that returns during a transition from one level of the premium to the next are drawn from a distribution whose mean is negatively related to the premium shift. The strength of that negative relation is specified using a prior distribution. This feature of our approach plays a significant role in making inferences about the timing of breaks and in estimating the equity premium.

Our estimates of the equity premium also incorporate the fact that the timing of structural breaks remains uncertain after examining the data. That is, the estimate of the equity premium on any given date reflects the uncertainty about where that date lies relative to breaks in the distribution. This feature of our approach, which applies the methodology of Chib (1998), stands in contrast to the commonly used maximum-likelihood procedure of estimating the dates of the breaks and then estimating parameters in each subperiod conditional on those dates.

The approach we develop and implement here is univariate, relying on a single time series of equity returns. As such, our approach is perhaps best viewed as an alternative to the popular method, also univariate, of estimat-

ing the equity premium by computing a sample average but using less of the available history. One can also view our approach as an alternative to modeling a time-varying equity premium as a function of observable state variables. Rather than specify those state variables and the function defining their roles, we simply augment the equity return series with the economically motivated beliefs that changes in the equity premium are unlikely to be extreme, are associated in part with shifts in volatility, and are likely to be accompanied by price changes in the opposite direction.

When these beliefs are incorporated, the estimated equity premium since 1834 fluctuates between roughly 4 and 6 percent (annualized). It rises through much of the 1800s, reaches its peak in the 1930s, and declines fairly steadily thereafter, except for a brief upward spike in the early 1970s. The sharpest decline in the premium occurs in the 1990s. The latter inference is influenced by the prior belief that the premium and the price tend to change in opposite directions. When that aspect of the model is omitted, the estimated premium instead increases during the last decade. The prior beliefs about shifts in the premium and about the premium's association with volatility are also shown to play important roles in estimating the premium. In our model with structural breaks but economically motivated prior beliefs, the precision associated with the estimate of the current equity premium is nearly as high as what one would attribute to an estimate based on the long-sample average when potential breaks are ignored.

The remainder of the article is organized as follows. Section I describes the stochastic setting and the priors used in our Bayesian approach. Section II presents the empirical results, and Section III briefly reviews the conclusions.

I. Methodology

This section describes our Bayesian framework for making inferences about the equity premium in the presence of structural change in the distribution of excess market returns. Although this framework is newly developed, our analysis shares some features with previous studies dealing with structural change.[1] This section first introduces the stochastic setting, and then discusses the prior distributions for the model's parameters. The general approach for obtaining posterior distributions is discussed briefly at the

---

[1] For surveys of early studies, too numerous to list, see Zacks (1983), Broemeling and Tsurumi (1987), Krishnaiah and Miao (1988), and Bhattacharya (1994). Some of the more recent studies in a frequentist setting include Andrews (1993), Andrews and Ploberger (1994), Bai (1995, 1997), Diebold and Chen (1996), Sowell (1996), Liu, Wu, and Zidek (1997), Bai, Lumsdaine, and Stock (1998), Bai and Perron (1998), Diebold and Chen (1996), Liu, Wu, and Zidek (1997), and Bai and Perron (1998). Perhaps the first Bayesian study on structural breaks is Chernoff and Zacks (1964), and more recent studies include Carlin, Gelfand, and Smith (1992), Stephens (1994), and Chib (1998). Markov switching models, proposed by Hamilton (1989), are studied in a Bayesian context by Albert and Chib (1993) and McCulloch and Tsay (1994). Recent studies that investigate structural breaks in some financial time series include Inclán (1993), Chen and Gupta (1997), Viceira (1997), and Ang and Bekaert (1998).

conclusion of this section, but the details of the computations are given in the Appendix.

A. Stochastic Framework

The data consist of T observations of excess market returns. Let x_t denote the excess return at time t, and $x = (x_1,\ldots,x_T)$. The sample period is split into $2K + 1$ regimes, K of which are transition regimes (TRs), during which the probability distribution of returns changes. In the $K + 1$ regimes separated by the TRs, referred to as stable regimes (SRs), the return distribution does not change. The SRs and TRs alternate in order, beginning and ending with an SR. The times at which the SRs change into the TRs and vice versa, the "changepoints," are unknown and denoted by $q_1,\ldots,q_{2K}$. Let $q_0 = 0$ and $q_{2K+1} = T$. The time spans for the ith SR and the jth TR can then be defined as

$$SR_i = \{q_{2i-2} + 1,\ldots,q_{2i-1}\}, \quad i = 1,\ldots,K + 1 \tag{1}$$

$$TR_j = \{q_{2j-1} + 1,\ldots,q_{2j}\}, \quad j = 1,\ldots,K. \tag{2}$$

We denote the duration of the kth regime as $l_k = q_k - q_{k-1}$, so the duration of SR_i is l_{2i-1} and the duration of TR_j is l_{2j}. Within each stable regime SR_i, the excess market returns are assumed to be normally distributed with mean μ_i and variance σ_i^2:

$$x_t \sim N(\mu_i,\sigma_i^2), \quad t \in SR_i, \quad i = 1,\ldots,K + 1. \tag{3}$$

We define $\Delta_i \equiv \mu_{i+1} - \mu_i$. Later in this section, we incorporate informative prior beliefs about the magnitude of Δ_i and about a positive relation between μ_i and σ_i^2. Within each transition regime TR_j, the excess market returns are assumed to be normally distributed with mean $[(\mu_j + \mu_{j+1})/2] + b_j\Delta_j$ and variance $\sigma_{j,j+1}^2$:

$$x_t \sim N\left(\frac{\mu_j + \mu_{j+1}}{2} + b_j\Delta_j,\sigma_{j,j+1}^2\right), \quad t \in TR_j, \quad j = 1,\ldots,K. \tag{4}$$

The mean excess return during the transition regime is conditioned on the shift in the premium. The first term, the average of the premiums in the neighboring stable regimes, is intended to represent the unconditional expected return during the transition. The second term, which reflects the conditioning on the premium shift Δ_j, allows us to impose a prior belief that b_j is negative. That is, we expect to see high returns during a TR in which the premium falls and low returns during a TR in which the premium rises.

Let $\mu = (\mu_1,\ldots,\mu_{K+1})$ denote the vector of equity premiums, let $\sigma_{SR} = (\sigma_1,\ldots,\sigma_{K+1})$ denote the vector of standard deviations ("volatilities") in the SRs, let $\sigma_{TR} = (\sigma_{1,2},\ldots,\sigma_{K,K+1})$ denote the vector of volatilities in the TRs, let $q = (q_1,\ldots,q_{2K})$ denote the set of changepoints, and let $b = (b_1,\ldots,b_K)$.

The likelihood function can be written as a product of $(2K + 1)$ normal densities:

$$p(x \mid \mu, \sigma_{SR}, \sigma_{TR}, b, q)$$

$$\propto \left(\prod_{i=1}^{K+1} \frac{1}{\sigma_i^{l_{2i-1}}} \right) \exp \left\{ -\frac{1}{2} \sum_{i=1}^{K+1} \sum_{t=q_{2i-2}+1}^{q_{2i-1}} \frac{(x_t - \mu_i)^2}{\sigma_i^2} \right\}$$

$$\times \left(\prod_{j=1}^{K} \frac{1}{\sigma_{j,j+1}^{l_{2j}}} \right) \exp \left\{ -\frac{1}{2} \sum_{j=1}^{K} \sum_{t=q_{2j-1}+1}^{q_{2j}} \frac{\left(x_t - \left(\frac{\mu_j + \mu_{j+1}}{2} + b_j \Delta_j \right) \right)^2}{\sigma_{j,j+1}^2} \right\},$$

(5)

where "$\propto$" denotes "proportional to" (up to a factor not involving μ, σ_{SR}, σ_{TR}, b, or q).

B. Prior Beliefs

Bayesian estimators combine the sample information contained in the likelihood function with prior information about the values of the model parameters and the relations among them. The prior beliefs used in this study are motivated by economic arguments. First, we impose a prior belief that the equity premium is positive. This prior reflects a simple argument that, in an equilibrium with risk-averse investors, the expected return on a value-weighted portfolio of all risky assets should exceed the risk-free rate of return. Merton (1980), for example, argues that the nonnegativity restriction on the expected excess market return should be imposed in estimating the equity premium. The rest of this section explains how our framework also incorporates informative prior beliefs about the relation between mean returns in the TRs and changes in the equity premium, about the duration of the TRs, about the premium's association with volatility, and about the magnitudes of the changes in the premium.

B.1. Beliefs About the Transition Regime Parameters and Duration

The TRs are relatively short periods during which the mean and the volatility of equity returns change. Recall that the expected return during TR_j is $(\mu_j + \mu_{j+1})/2 + b_j \Delta_j$. A drop in the equity premium during TR_j ($\Delta_j < 0$) is likely to accompany a drop in the rate at which future dividends are discounted (unless an increase in the interest rate accounts totally for the drop in the premium). With a drop in the discount rate, it seems reasonable to expect that the returns during the TR are high (unless there is an offsetting drop in expected dividends). Similarly, if the premium rises ($\Delta_j > 0$), the TR returns are likely to be low. This motivates our informative prior belief that $b_j < 0$. Therefore, the prior on b_j is assumed to be normal with a mean $\bar{b} < 0$ and variance σ_b^2:

$$p(b_j) \propto \exp \left\{ -\frac{(b_j - \bar{b})^2}{2\sigma_b^2} \right\}, \quad j = 1, \ldots, K.$$

(6)

The prior mean $\bar{b}$ is set equal to -15.13. This specification, explained in the Appendix, is based on Campbell's (1991) variance decomposition of aggregate equity returns. The prior standard deviation σ_b is set equal to a third of the absolute value of $\bar{b}$, so that virtually all of the prior mass of b_j is below zero.

The prior on the TR volatility, $\sigma_{j,j+1}$, is an inverted gamma distribution with $\eta = 10$ degrees of freedom:

$$p(\sigma_{j,j+1}) \propto \frac{1}{\sigma_{j,j+1}^{\eta+1}} \exp\left\{-\frac{(\eta-2)\alpha^2}{2\sigma_{j,j+1}^2}\right\}, \quad \sigma_{j,j+1} > 0, \quad j = 1,\ldots,K. \tag{7}$$

The parameter α^2, equal to the prior mean of $\sigma_{j,j+1}^2$, is set equal to 0.000634. This choice is also based on the results in Campbell (1991), as explained in the Appendix. The prior is informative about $\sigma_{j,j+1}$ to roughly the same extent as a sample of 10 observations of returns with a sample variance equal to α^2. Because some TRs can potentially be as short as one month, some prior information is needed to estimate the TR volatilities.

We explore a model with $K = 15$ transition regimes. The explanation for this choice is postponed until Section II.B. Our inference about the $2K = 30$ changepoints is based on Chib (1998). Chib formulates a multiple change-point model in terms of a latent state variable $s_t \in \{1,2,\ldots,2K+1\}$, whose value at time t indicates the regime from which the time-t observation has been drawn. This state variable follows a Markov process with a transition matrix P constrained such that, if $s_{t-1} = k$, then s_t can only take two values, k or $k+1$ (for $k = 1,\ldots,2K$). We need to specify a prior distribution for each diagonal element of P, $p_{k,k} = \text{Prob}(s_t = k|s_{t-1} = k)$, which denotes the probability of staying in regime k. By construction, $p_{2K+1,2K+1} = 1$. Following Chib, the prior of $p_{k,k}$ for $k = 1,\ldots,2K$ is specified as a beta distribution with parameters a_k and c_k:

$$p(p_{k,k}) = \frac{\Gamma(a_k + c_k)}{\Gamma(a_k)\Gamma(c_k)} p_{k,k}^{a_k-1}(1 - p_{k,k})^{c_k-1}. \tag{8}$$

Chib recommends specifying the prior parameters such that they correspond to prior beliefs about the mean duration of each regime. Given $p_{k,k}$, the prior density of the duration d_k of the regime k is $p(d_k|p_{k,k}) = p_{k,k}^{d_k-1}(1 - p_{k,k})$, and its prior mean is $E(d_k|p_{k,k}) = (1 - p_{k,k})^{-1}$. The duration's unconditional prior density and its moments can also be derived analytically. The unconditional prior mean is

$$E(d_k) = \frac{a_k + c_k - 1}{c_k - 1}. \tag{9}$$

For TRs (i.e., for even values of k), we set $a_k = 11$ and $c_k = 2$. These values imply that the prior distribution of the duration of each TR has a mean of 12 months, a median of 5 months, a mode of 1 month, and its 95th percentile is 39 months. This specification seems reasonable, in that most of the prior

mass is on very short TR durations, but the skewness implies some chance that a TR might instead last for several years.

For stable regimes (i.e., for odd values of k), we set $c_k = 2$ and compute a_k such that

$$E(d_{SR}) = \frac{T - KE(d_{TR})}{K + 1}.$$ (10)

Such a specification makes the prior internally consistent, in that the number of the TRs times their expected duration plus the number of the SRs times their expected duration equals the sample size, $T = 1{,}982$. The resulting a_k equals 223.25, the expected duration of each SR is 113 months, and the 95th percentile of each SR's duration is 708 months.

B.2. Beliefs About the Premium's Association with Volatility

In a study about estimating the equity premium, Merton (1980) proposes models in which the equity premium is linked positively to volatility. In motivating such models, Merton notes that, to preclude arbitrage, the equity premium must be zero if volatility is zero. Moreover, at positive levels of market volatility, risk-averse investors must in general be compensated by a positive equity premium. Thus, at least to this degree, a positive relation between the equity premium and volatility seems likely. Merton essentially proposes a positive relation as a reasonable prior belief, as opposed to a regularity that one might verify with the data. Attempts to do the latter, beginning with French, Schwert, and Stambaugh (1987), have produced mixed results, but such studies have generally investigated the presence of a relation at higher frequencies than envisioned in our setting.[2] One might believe that occasional changes in the equity premium during TRs, typically separated by a number of years, are associated to some degree with changes in volatility. At the same time, one might be less inclined to believe that the equity premium changes with higher-frequency fluctuations in volatility, which are essentially ignored in the present setting with returns assumed to be i.i.d. within each regime.[3] The prior link between the equity premium and volatility that we introduce below can take the form of a weak positive as-

---

[2] Some examples illustrate the range of the results. French, Schwert, and Stambaugh (1987), Harvey (1989), Turner, Startz, and Nelson (1989), and Tauchen and Hussey (1991) find a positive relation between the conditional market premium and conditional variance, and Scruggs (1998) finds a significant positive partial relation. Baillie and DeGennaro (1990) and Chan, Karolyi, and Stulz (1992) find that the conditional market premium is unrelated to its own conditional variance. Whitelaw (1994) finds a weak negative relation, and Campbell (1987) and Glosten, Jagannathan, and Runkle (1993) find a significant negative relation.

[3] In parameterized versions of equilibrium models in which moments of the aggregate endowment follow Markov-switching processes, Kandel and Stambaugh (1990) and Backus and Gregory (1993) show that the relation between the equity premium and volatility need not be positive. Campbell (1987) considers conditions under which the intertemporal CAPM implies an approximately proportional relation between the conditional mean and conditional variance of market returns.

sociation, as opposed to a strict parametric relation, and we suggest that such priors offer a sensible framework in which to explore the potential importance of volatility.

A prior association between the equity premium and volatility is introduced as follows. For a scalar parameter $\gamma > 0$, let

$$\mu_i = \gamma \psi_i \sigma_i^2, \qquad i = 1, \ldots, K + 1, \tag{11}$$

and let $\psi = (\psi_1, \ldots, \psi_{K+1})$. As explained below, the prior on each ψ_i is centered at one. Hence, γ can be viewed a priori as the average market "price of risk," defined as the ratio of the equity premium to the equity variance. The prior on γ is specified as a gamma distribution with parameters a_γ and b_γ:

$$p(\gamma) \propto \gamma^{a_\gamma - 1} \exp\left\{-\frac{\gamma}{b_\gamma}\right\}, \quad \gamma > 0. \tag{12}$$

The prior parameters are specified using an empirical Bayes approach as $a_\gamma = 18.7$ and $b_\gamma = 0.1$. These values equate the prior mean of γ to the unconditional sample estimate of the price of risk from the overall sample (1.98). The prior standard deviation of γ is equated to the sampling uncertainty in the price of risk estimate, 0.46, which is computed as the standard error of the sample mean return divided by the sample variance (this computation assumes for simplicity that the unconditional return variance is estimated without error). The 1st and the 99th percentiles of the prior for γ are 1.07 and 3.20.

In each stable regime SR_i, the price of risk is equal to $\psi_i \gamma$, with the ψ_is assumed to be independent across subperiods. The prior on each ψ_i is a gamma distribution, with parameters $\nu/2$ and $2/\nu$,

$$p(\psi_i) \propto \psi_i^{(\nu/2)-1} \exp\left\{-\frac{\psi_i \nu}{2}\right\}, \quad \psi_i > 0, \quad i = 1, \ldots, K + 1. \tag{13}$$

The prior on ψ_i implies that[4]

$$\mathrm{E}(\psi_i) = 1 \tag{14}$$

$$\mathrm{Var}(\psi_i) = \frac{2}{\nu}. \tag{15}$$

The desired degree of association between μ_i and σ_i^2 across the SRs is achieved by specifying the parameter ν. At one extreme, as $\nu \to 0$, the prior on ψ_i approaches a standard diffuse or noninformative prior, $p(\psi_i) \propto 1/\psi_i$. With a noninformative prior on ψ_i, no association between the elements of μ and σ_{SR} is imposed a priori. At the other extreme, as $\nu \to \infty$, it follows from equation (15) that $\mathrm{Var}(\psi_i) \to 0$, so $\psi_i = 1$ for all i, which imposes a perfect

---

[4] The two moment equations follow from standard results for the gamma density, such as in Zellner (1971). The moments exist for all $\nu > 0$, but the density has no mode for $\nu < 2$.

link between μ_i and σ_i^2 of the form $\mu_i = \gamma\sigma_i^2$. A positive but finite value of ν implies an intermediate degree of association between the equity premium and volatility: the higher the value of ν, the stronger the prior belief that the equity premium is linked positively to volatility. In the empirical analysis, a range of values for ν is entertained.

B.3. Beliefs About Magnitudes of Changes in the Premium

We use a "hierarchical" prior distribution on μ, given by

$$p(\mu|\bar{\mu}) \propto \exp\left\{-\frac{1}{2}(\mu - \bar{\mu}\iota)'V_\mu^{-1}(\mu - \bar{\mu}\iota)\right\}, \quad \mu > 0, \tag{16}$$

$$p(\bar{\mu}) \propto 1, \quad \bar{\mu} > 0, \tag{17}$$

where ι denotes a $(K + 1) \times 1$ vector of ones. The scalar $\bar{\mu}$ is a "hyperparameter" that can be interpreted roughly as a cross-period grand mean of the elements of μ.[5] The prior for μ conditional on $\bar{\mu}$ is a truncated normal distribution whose location depends on $\bar{\mu}$. The prior distribution of $\bar{\mu}$ is noninformative, except for the positivity restriction. As a result, the unconditional variance of each element of μ is large, and the marginal prior for each element of μ is noninformative.

The elements of V_μ can be specified such that equation (16) is informative about differences between the elements of μ. Recall that $\Delta_i = (\mu_{i+1} - \mu_i)$, $i = 1, \ldots, K$, and let $\Delta = (\Delta_1, \ldots, \Delta_K)$. The elements of Δ represent the magnitudes by which the market premium changes in the TRs. Note that equation (16) implies that the prior on each Δ_i is centered at zero, so the prior is noninformative about the direction of any shift in the premium. Some might find it reasonable to believe, as we do, that extremely large shifts in the equity premium are unlikely. For example, one could believe that the probability is only 5 percent that the annual equity premium can shift by more than 6 percent during any TR. This type of prior belief can be expressed by specifying a value for the standard deviation of the prior distribution of each Δ_i, denoted by σ_Δ.[6] In the preceding example, $\sigma_\Delta = 3\%$. At one extreme, setting $\sigma_\Delta = \infty$ assigns equal prior probabilities to fixed-width neighborhoods around all values of Δ_i, however large. One consequence of such a noninformative belief about Δ is that, in estimating the equity premium in SR_i, the data from all other regimes are discarded (in the absence of other informative prior beliefs). In other words, this prior results in a use of the data that corresponds to common practice. At the other extreme, setting $\sigma_\Delta = 0$ reflects a dogmatic belief that all $\Delta_i = 0$ and that there has never been a change

---

[5] In the absence of truncation in equation (16), $\bar{\mu}$ would be the mean of $p(\mu|\bar{\mu})$.

[6] In general, the literature treats the parameters before and after a structural break as independent of each other (see Carlin et al. (1992) and Barry and Hartigan (1993) for Bayesian examples and Liu et al. (1997) and Bai and Perron (1998) for frequentist examples). An exception is the early study by Chernoff and Zacks (1964), who, in a simpler setting, place an informative prior on the difference in subperiod means.

in the equity premium, in which case data from the entire sample are simply "pooled," roughly speaking, to estimate the premium. In intermediate cases, the smaller the value of σ_Δ, the more attention is paid to data from other SRs. In order to explore the effect of prior beliefs about Δ on the estimates of the equity premium, this study entertains a wide range of values of σ_Δ.

The value of σ_Δ is implied by the covariance matrix V_μ in equation (16). It is assumed that $V_\mu = \sigma_\mu^2 I_{K+1}$, where I_{K+1} denotes an identity matrix of size $K + 1$. Conditional on $\bar{\mu}$, and in the absence of truncation, the prior variance of each μ_i equals σ_μ^2. The unconditional prior variance of Δ_i for any i is equal to $\sigma_\Delta^2 = \mathrm{Var}(\mu_{i+1} - \mu_i)$. The value of σ_μ^2 that produces a desired value of σ_Δ is computed by simulation. In the resulting prior, the equity premium is believed to fluctuate independently across the stable regimes and thereby exhibit "immediate" mean reversion to a grand mean.

Let the vector θ contain the elements of b, σ_{TR}, P, γ, ψ, μ, and $\bar{\mu}$. It is assumed that all the elements of θ, except for μ and $\bar{\mu}$, are independent a priori, which implies that the joint prior on all the parameters in the model can be written as

$$
p(\theta) = \left(\prod_{j=1}^{K} p(b_j) \right) \left(\prod_{j=1}^{K} p(\sigma_{j,j+1}) \right) \left(\prod_{k=1}^{2K} p(p_{k,k}) \right)
$$
$$
\times \, p(\gamma) \left(\prod_{i=1}^{K+1} p(\psi_i) \right) p(\mu|\bar{\mu}) \, p(\bar{\mu}).
$$
(18)

The densities multiplied on the right-hand side are given in equations (6), (7), (8), (12), (13), (16), and (17).

C. Posterior Distribution

In a Bayesian setting, a posterior probability distribution for the unknown parameters is obtained by updating a prior distribution with the information in the data transmitted through the likelihood function. Substituting for the elements of σ_{SR} from equation (11), the reparameterized likelihood from equation (5) can be written as[7]

$$
p(x|\mu,\psi,\gamma,b,\sigma_{TR},q)
$$

$$
\propto \left(\prod_{i=1}^{K+1} \left(\frac{\psi_i \gamma}{\mu_i} \right)^{l_{2i-1}/2} \right) \exp\left\{ -\frac{1}{2} \sum_{i=1}^{K+1} \sum_{t=q_{2i-2}+1}^{q_{2i-1}} \frac{(x_t - \mu_i)^2}{\mu_i} \gamma\psi_i \right\}
$$
$$
\times \left(\prod_{j=1}^{K} \sigma_{j,j+1}^{-l_{2j}} \right) \exp\left\{ -\frac{1}{2} \sum_{j=1}^{K} \sum_{t=q_{2j-1}+1}^{q_{2j}} \frac{\left(x_t - \left(\frac{\mu_j + \mu_{j+1}}{2} + b_j\Delta_j\right)\right)^2}{\sigma_{j,j+1}^2} \right\}.
$$
(19)

---

[7] In the case with no premium–volatility link ($\nu = 0$), there is no need to substitute for σ_{SR} from equation (11) and reparameterize the likelihood, so we work with the likelihood in equation (5). Such an equivalent specification turns out to save some computation time.

Multiplying the prior in equation (18) by the likelihood in equation (19) gives the joint posterior distribution, $p(\theta|x)$. Posterior distributions for parameters of interest, such as the equity premium μ, are computed numerically using a Metropolis–Hastings algorithm combined with the algorithm of Chib (1998) for drawing the changepoints.

Note that frequentist analyses of structural breaks typically estimate the break locations and then estimate the remaining parameters of interest conditional on those break locations. Recent examples include Bai (1995, 1997), Bai and Perron (1998), and Liu et al. (1997). The usual argument in favor of such a two-step procedure is that, under certain assumptions, the resulting parameter estimates are consistent. Treating estimates of the breakpoints as true values ignores the potential error in those estimates ("estimation risk") and could thereby compromise inferences in finite samples. In contrast, a Bayesian approach can account for the uncertainty about the locations of the breakpoints.[8] Instead of conducting inference based on the posterior $p(\theta|x,\hat{q})$, which conditions on the break estimates, we obtain results based directly on $p(\theta|x)$, which incorporates the uncertainty. The algorithm of Chib (1998) allows us to generate the posterior distribution $p(q|x)$ of the locations of structural breaks. A Bayesian approach integrates over this posterior,

$$p(\theta|x) = \int_q p(\theta|x,q)p(q|x)\,dq, \tag{20}$$

thereby incorporating break uncertainty in estimating θ.

For the purpose of estimating and plotting a monthly series of the equity premium, we define the premium in month t as

$$\mu^t = \begin{cases} \mu_i & \text{if } t \in SR_i \\ \mu_{(j)} & \text{if } t \in TR_j \end{cases}, \tag{21}$$

where $\mu_{(j)} \equiv (\mu_j + \mu_{j+1})/2$. That is, we include only the unconditional mean of the excess return during the transition regime and omit the portion $b_j\Delta_j$ negatively associated with the premium shift. We estimate μ^t by its posterior mean, using iterated expectations:[9]

$$E(\mu^t|x) = \sum_{i=1}^{K+1} E(\mu_i|x)p(t \in SR_i|x) + \sum_{j=1}^{K} E(\mu_{(j)}|x)p(t \in TR_j|x). \tag{22}$$

Because the regime to which a particular month belongs is uncertain, the estimated equity premium in our framework generally fluctuates at a monthly frequency, as opposed to remaining constant between fixed estimates of break locations.

---

[8] Examples of earlier Bayesian studies that account for the uncertainty about breakpoints include Chernoff and Zacks (1964), Hsu (1982), and Broemeling and Tsurumi (1987).

[9] The posterior mean is the estimate that minimizes the expected value of a quadratic loss function. For additional details, as well as alternative posterior estimators, see Berger (1985).

The posterior variance of μ^t is calculated by decomposing it as the weighted average of the posterior variance in each regime plus the variance of the posterior mean across regimes:

$$\text{Var}(\mu^t|x) = \sum_{i=1}^{K+1} [\text{Var}(\mu_i|x) + (E(\mu_i|x) - E(\mu^t|x))^2]p(t \in SR_i|x)$$

$$+ \sum_{j=1}^{K} [\text{Var}(\mu_{(j)}|x) + (E(\mu_{(j)}|x) - E(\mu^t|x))^2]p(t \in TR_j|x).$$

(23)

The posterior moments and probabilities are calculated across a large number of parameter draws from the joint posterior for θ and the s_ts. The details are provided in the Appendix.

II. Empirical Analysis

A. *The Market Excess-Return Series*

The data used in this study consist of monthly returns on a broadly based equity portfolio in excess of returns on a short-term riskless instrument. The equity-return series and the risk-free return series, described in this subsection, cover the period from January 1834 to June 1999. The equity series from January 1926 to June 1999 consists of returns on the value-weighted portfolio of NYSE stocks, obtained from the Center for Research in Security Prices (CRSP). Equity returns before 1926 are taken from Schwert (1990), who relies on a variety of historical indexes to construct a series of U.S. monthly returns over the past two centuries.[10] Up through 1862, his index is based on the returns on financial firms and railroads from Smith and Cole (1935). For the period 1863 through 1870, Schwert uses the returns on the railroad index from Macaulay (1938), and for the 1871 through 1885 period, he uses returns on the value-weighted market index constructed by Cowles and Associates (1939). Finally, the 1885 to 1925 data consist of returns on the Dow Jones index of industrial and railroad stocks as sourced from Dow Jones (Farrell (1972)).[11] Schwert adjusts the series for the effects of time averaging present in the Cowles and Macaulay series. Also, he acknowledges that the returns on the original Smith and Cole and Macaulay indexes do not include dividend yield and adds the dividend yield back based on an estimate from the Cowles series.

Although the series constructed in Schwert (1990) begins in 1802, we use the series back only to 1834 because the earlier data do not appear to capture aggregate equity returns. Prior to 1834, the Smith and Cole index is

---

[10] We thank Bill Schwert for providing these data.

[11] The four observations for August through November of 1914 are missing because the stock markets were temporarily closed due to the beginning of World War I (see Schwert (1989)).

based only on financial firms, whose returns were much less volatile than returns on a typical industrial company. Through 1814, the Smith and Cole index is an equally weighted portfolio of only seven banks, and those seven were chosen in hindsight from a larger group. Also, in their careful historical account of the early years on Wall Street, Werner and Smith (1991, p. 38) note that " . . . in periods of speculative fever, such as 1824 and 1825, trading volume and share prices both rose sharply . . ." and "[l]ate in 1825, the securities market bubble burst." An unusual price increase is not evident in the Smith and Cole data, however, as the annualized mean excess return on the index between January 1824 and August 1825 is only 1 percent. Also, there is only a mild fall in the prices of the financial firms at the end of 1825. Thus, one might suspect that the returns on a small set of financial companies fail to convey much of the information about overall equity returns in that period. After 1834, the Smith and Cole data expand to include a portfolio of up to 27 railroad stocks, which were among the most important industrial companies during much of the 19th century. Noting other properties of the Smith and Cole index prior to 1834, Schwert (1989) also excludes the data up to that point.

The short-term risk-free return series is based on the data constructed by Siegel (1992).[12] From 1926 until 1999, the returns on a one-month Treasury security are obtained from CRSP. For the period 1920 through 1925, the rates on three-month Treasuries are taken from Homer (1963). Prior to 1920, short-term Treasury securities in their current form were nonexistent. As a result, most of the data on U.S. short-term interest rates prior to 1920 are based on commercial paper rates quoted in Macaulay (1938).[13] As Siegel demonstrates, however, commercial paper in the 19th century was subject to a high and variable risk premium, which appears to render a raw series of returns on commercial paper a poor proxy for a risk-free rate of return. In order to remove the risk premium on commercial paper, Siegel constructs a synthetic "riskless" short-term interest rate series by assuming that the average term premiums on long-term high-grade securities were the same in the United States as in the United Kingdom.[14] Monthly returns are derived from Siegel's annual series using linear interpolation, treating his values as corresponding to the last month of the year. Given that the volatility of the annual series over this period is substantially lower than that of annual equity returns, we suspect that the problems induced by this simplification are relatively unimportant in the empirical analyses we conduct.

[12] We thank Jeremy Siegel for providing these data.

[13] For the period 1857 through 1919, Macaulay uses prime two-month and three-month commercial paper. For the period 1831 through 1856, he uses data from Bigelow (1862) on commercial paper with maturity varying between three and six months.

[14] In the 19th century, the capital markets in the United Kingdom were far more developed than those in the United States. Siegel (1992) motivates his assumption about the equality of the average term premiums by noting that real returns on long-term bonds in the United Kingdom and in the United States have behaved similarly over the past two centuries.

B. Structural Breaks

Recall that the framework proposed here includes K transition regimes (TRs), separated from the stable regimes (SRs) by $2K$ changepoints, or structural breaks. This framework will sometimes be referred to as a framework "with TRs." For comparison, we also consider a case "with no TRs," which can be viewed as a special case of our framework in which all TRs have zero length. In this latter specification, the beginning and ending points of each TR collapse into one, so that there are only K structural breaks.

Figure 1 displays, for each month from January 1834 through June 1999, the posterior probability that one of $K = 15$ TRs begins in that month. We sometimes refer to this probability as the "posterior break probability." The figure contains four plots. The first plot corresponds to a specification with TRs, no mean-variance link ($\nu = 0$), and $\sigma_\Delta = 2\%$. The sample contains 20 months in which the posterior break probability exceeds 10 percent. There is an 83 percent probability of a TR beginning between May and October of 1940, and a 60 percent probability between July and November of 1873. More recently, there is a 74 percent probability of a TR beginning between December 1991 and April 1992, a 23 percent probability between November 1994 and February 1995, and a 20 percent probability between April and August of 1996. An almost identical plot is obtained in the framework with $\nu = 10$ and $\sigma_\Delta = 3\%$, which is later referred to as "benchmark prior beliefs."

The second plot corresponds to the same specification as the first plot, but with return volatility constrained to be equal across the SRs. Fewer TRs are identified than in the first plot, which implies, perhaps not surprisingly, that changes in volatility are an important source of structural changes in the return distribution. Nevertheless, even with constant volatility, the specification with TRs can clearly identify several changepoints. For example, June 1932 and March 1933 both receive a 100 percent posterior break probability. The cumulative equity excess return in the two months following June 1932 is 82 percent. Those months are followed by several more with negative or low returns, and this pattern leads to the identification of June 1932 as the probable beginning of a TR. Similarly, the cumulative equity excess return in the three months following March 1933 is 90 percent, followed by lower returns thereafter, so that March 1933 is identified as the beginning of a TR. In contrast, the specification with constant volatility and no TRs is unable to identify any structural changes: all posterior break probabilities in the third plot are below 1.8 percent. Not surprisingly, changes in the expected return are very hard to detect without additional stochastic structure. Finally, the fourth plot, which corresponds to the framework with TRs, a mean-variance link ($\nu = 10$), and $\sigma_\Delta = \infty$, shows that the presence of the link also helps identify the TRs.

All the results in the paper are reported for $K = 15$. This choice is motivated by comparing the plots of the posterior break probabilities for different Ks in specifications with $\sigma_\Delta = 2\%$ and no mean-variance link. We find
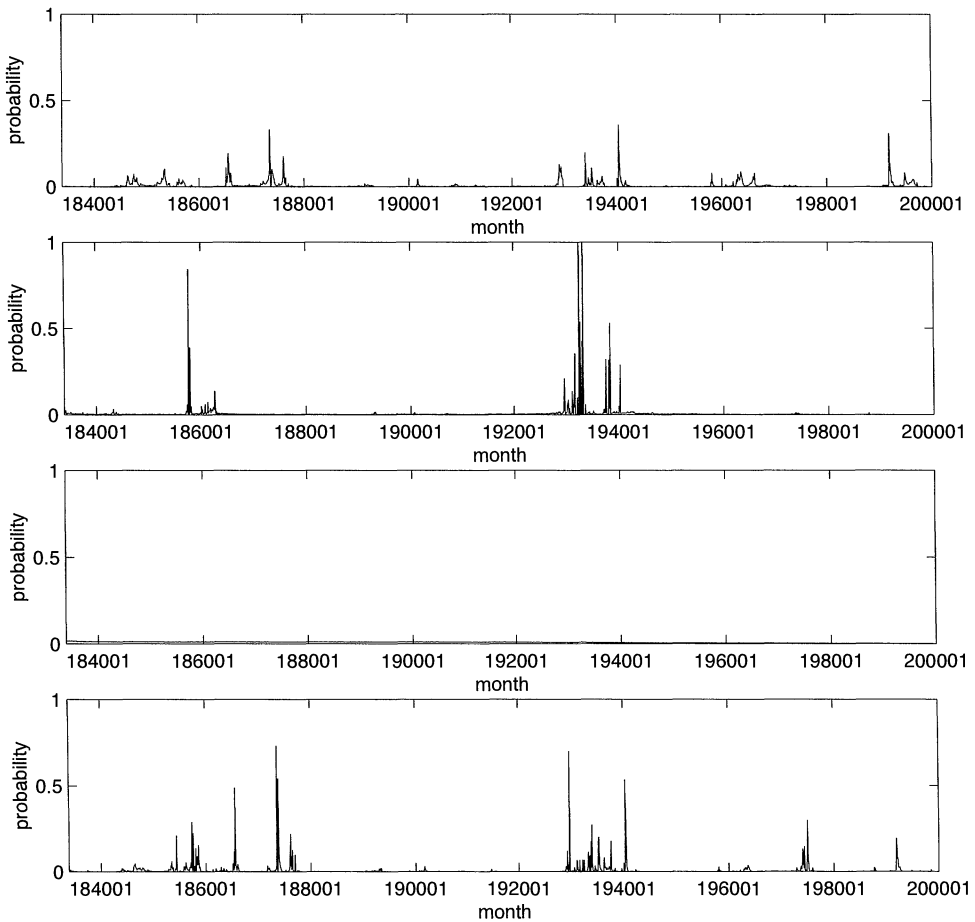
Figure 1. Posterior break probabilities. The figure plots, for each month, the posterior probability that one of 15 stable regimes in the return distribution ends in that month. (The 16th stable regime is assumed to continue through the last month of the sample.) The first plot corresponds to the specification with transition regimes (TRs), no mean-variance link, and $\sigma_\Delta = 2\%$, the second plot to the same specification as the first plot but with constant return volatility across stable regimes, the third plot to the same specification as the second plot but with no TRs, and the fourth plot to the specification with TRs, with the mean-variance link ($\nu = 10$), and $\sigma_\Delta = \infty$.

that the plot with $K = 20$ looks very similar to our first plot in Figure 1 with $K = 15$. The $K = 20$ framework identifies the same structural changes as the $K = 15$ framework, and the five additional TRs receive a low posterior probability. Therefore, increasing the number of TRs is unlikely to lead to substantial differences in the estimates of the premium. At the same time, the plot with $K = 10$ looks sufficiently different from the $K = 15$ plot that it cannot justify reducing the number of TRs from 15 to 10.

C. The Equity Premium Over Time

C.1. Benchmark Prior Beliefs

As argued earlier, it seems reasonable to believe that extremely large shifts in the equity premium are unlikely and that, to at least some extent, the premium is related to equity volatility. This subsection presents results that reflect moderately informative prior beliefs along these lines. The following subsections report the results for various other prior beliefs to demonstrate the role of economically motivated priors in estimating the premium.

Recall that the relation between the mean and the volatility of equity excess returns is established by specifying the parameter ν in the informative prior on ψ_i, defined in equation (11). Here we specify $\nu = 10$, which implies a plausible intermediate degree of the mean-variance link: there is a 10 percent prior probability that the price of risk in any SR is less than half its prior mean (the overall sample value), and there is a 10 percent probability that the price of risk is more than 1.6 times its prior mean. Also recall that prior beliefs about the magnitudes of changes in the premium during TRs are specified by choosing the value of σ_Δ, the prior standard deviation of the shift in the premium. Here we choose $\sigma_\Delta = 3\%$, whereby we assign only about 5 percent prior probability to the event that the annualized premium could shift by more than 6 percent during a TR. Note that, with a prior belief in a mean-variance link, one should not specify too small a value for σ_Δ. Because data suggest that equity volatility changes substantially over time, a belief that the premium is linked with volatility implies a belief that the premium changes over time as well. A prior specifying a nontrivial mean-variance association should not simultaneously be too restrictive about shifts in the premium.

Figure 2 plots the evolution of the posterior mean (solid) and posterior standard deviation (dotted) of the equity premium μ^t over time. The estimated equity premium has been fairly stable between January 1834 and June 1999. The annualized premium fluctuates between 3.9 percent in January 1849 and 6 percent in April 1934, with a downward trend since the mid-1930s. It is interesting that, although $\sigma_\Delta = 3\%$ allows for fairly large shifts in the premium, the evolution of the premium is rather smooth. Over the last decade, the premium decreased by 0.5 percent to its current level of about 4.8 percent. Although this decrease is perhaps not large in absolute terms, it is the most dramatic decrease in the premium over the last 165 years. The posterior standard deviation associated with the estimate of the current premium is about 1.4 percent.

Note that the decrease in the premium is not due to low recent returns, since the average equity excess return in the 1990s is about twice the long-run sample mean of 5.71 percent. Rather, our framework suggests that a significant portion of the recent run-up in stock prices occurred during TRs, so those high returns are consistent with a drop in the premium. Previous studies that identify a recent decrease in the premium rely on parametric
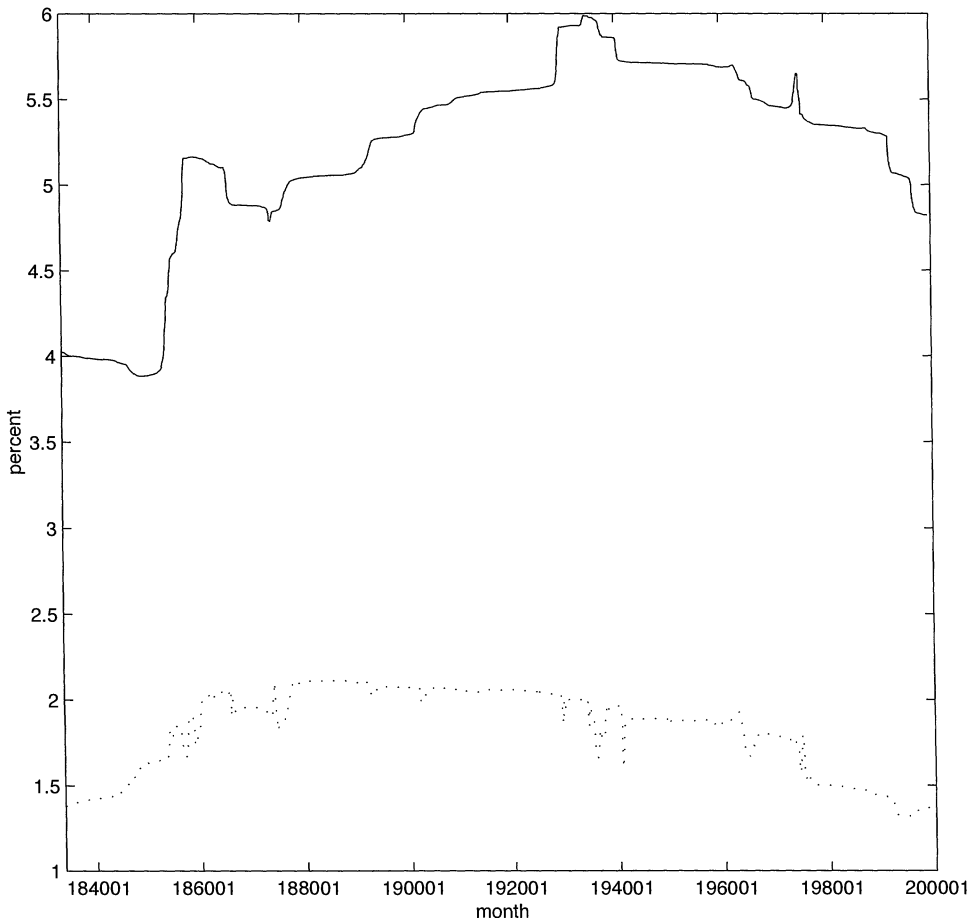
Figure 2. Equity premium with benchmark priors. The solid line plots, for each month, the posterior mean of the equity premium μ^t obtained using the specification with transition regimes, $\sigma_\Delta = 3\%$, and $\nu = 10$. The dotted line plots the posterior standard deviation of μ^t in each month.

relations between stock returns and other variables, such as dividend–price ratios, earnings–price ratios, and so forth. This paper identifies the decrease based only on a simple model for equity excess returns and economically motivated prior information.

C.2. The Effect of Beliefs About Magnitudes of Changes in the Premium

A simple example can illustrate the effect of informative prior beliefs about the magnitude of potential shifts in the premium. A common empirical tradition in finance is to estimate the equity premium using data beginning in January 1926, the starting date for widely used datasets produced by CRSP

Table I

Estimates of the Equity Premium with a Single Break Specified in December 1925

The table reports posterior moments of the equity premium (in percent per annum) in a model with a single structural break specified at December 1925. The equity premium is defined as the expected rate of return on the aggregate stock market portfolio in excess of the short-term interest rate, and μ_1 and μ_2 denote the equity premiums before and after the break. The break is associated with a shift in the equity premium given by $\Delta = \mu_2 - \mu_1$. The prior standard deviation of Δ is σ_Δ (annualized in the table).

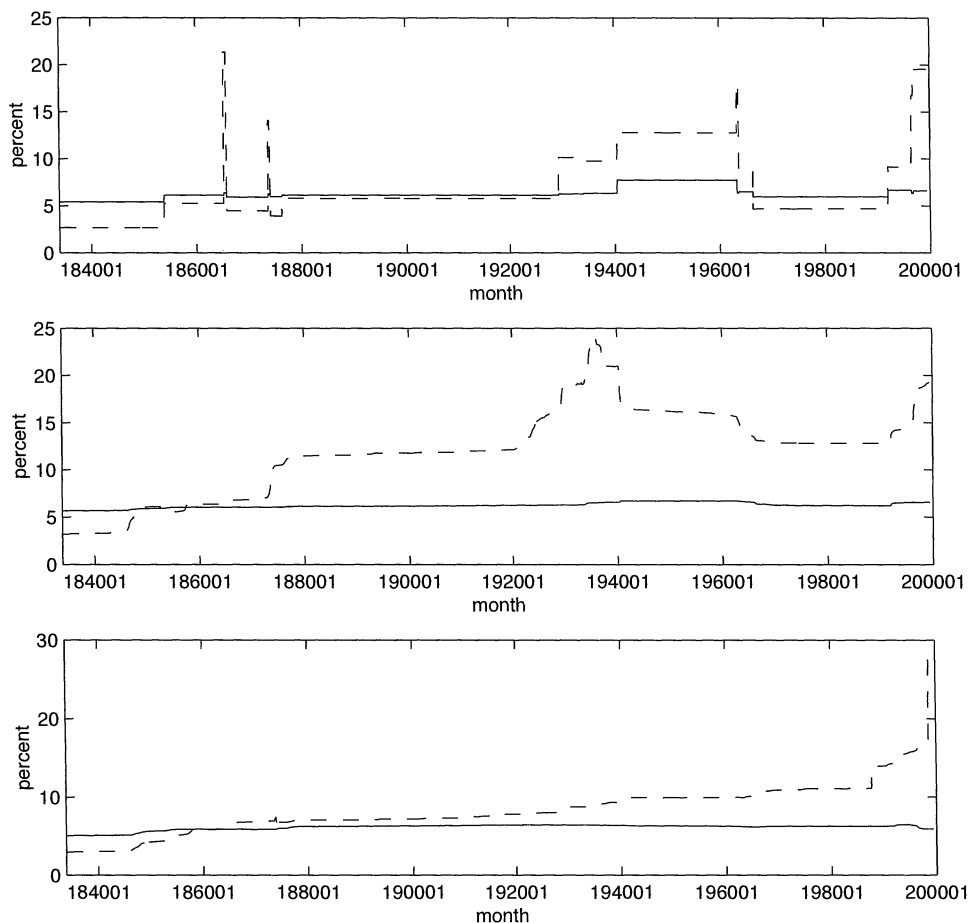|	Posterior Mean		Posterior Std. Dev.	
σ_Δ (%)	1834–1925	1926–1999	1834–1925	1926–1999
0	5.22	5.22	1.27	1.27
1	5.03	5.59	1.31	1.42
2	4.65	6.33	1.39	1.67
3	4.33	6.93	1.44	1.83
5	4.00	7.69	1.49	2.01
∞	3.64	8.36	1.53	2.19

and Ibbotson Associates. Given the availability of the earlier data, using just the post-1925 data is equivalent to specifying a structural break in December 1925 and having noninformative priors about Δ and ψ. Table I reports posterior moments obtained using the same 1834 to 1999 excess-return series as before, but in a model with only a single break, exogenously specified at December 1925. To isolate the effect of an informative prior on Δ, no association between the premium and the volatility is imposed (i.e., $\nu = 0$). With a noninformative prior ($\sigma_\Delta = \infty$), the data before the break are discarded, and the equity premium for the post-1925 period has a posterior mean of 8.36 percent, similar to standard textbook values.[15] With an informative prior on Δ, the data before the break are useful in estimating the post-break premium. Because the posterior mean of the premium in the 1834 to 1925 subperiod is only 3.64 percent with noninformative priors, specifying an informative prior for Δ lowers the mean of the post-1925 equity premium compared to the 8.36 percent produced with the noninformative prior. The mean equity premium is lower by 1 percent with $\sigma_\Delta = 4\%$ and by 2 percent with $\sigma_\Delta = 2\%$. Therefore, the common practice of using the average post-1925 excess return overstates the equity premium if one believes that large shifts in the premium are unlikely. At the same time, simply averaging the data beginning in 1834 produces too low an estimate, unless one believes that a shift in the premium did not occur. These two extreme approaches essentially correspond to the cases in Table I for $\sigma_\Delta = \infty$ and $\sigma_\Delta = 0$.

[15] For example, at several places in their popular text, Brealey and Myers (1996) use an equity premium of 8.4 percent, which they report is an estimate based on the 1926 to 1994 period.

The premium estimate with $\sigma_\Delta = 0$ is 5.22 percent, which is slightly below the simple arithmetic average of the excess returns over the entire sample (5.71 percent). The difference arises essentially because the sample averages from the two subperiods are weighted by the reciprocals of the subperiod volatilities, in addition to the lengths of the subperiods (the weights applied in computing the arithmetic average). The mean return is estimated with less precision in the more volatile second subperiod (the volatility is 18.93 percent versus 15.16 percent in the first subperiod), so the higher average return from that subperiod is given less weight (much as in weighted least squares). Also note that the premium's posterior standard deviation with $\sigma_\Delta = 0$ is 1.27 percent, less than the 1.53 percent and 2.19 percent obtained for the two subperiods when $\sigma_\Delta = \infty$. Naturally, posterior uncertainty about the equity premium is lower when inference is based on data from the entire 165-year sample as opposed to just a subperiod. To sum up, σ_Δ plays an important role in this simple example with one fixed structural break.

Prior beliefs about Δ are also important in our framework with uncertainty about the locations of the 15 TRs. Figure 3 displays the equity premium estimated with $\sigma_\Delta = 2\%$ (solid) and $\sigma_\Delta = \infty$ (dashed). As in the previous example, no association between the premium and the volatility is imposed in order to isolate the effect of an informative prior on Δ. The figure contains three plots. The first plot corresponds to the scenario in which the structural breaks are fixed at their estimates. The breaks are fixed at the highest posterior probability locations from the framework with no TRs and $\sigma_\Delta = 2\%$, except that clusters of adjacent months with high break probabilities are treated as one break. The second plot also corresponds to the scenario with no TRs, but the uncertainty about the locations of the 15 breaks is incorporated. The third plot incorporates both TRs and break uncertainty. In all three plots, there are substantial differences between the premium estimates for $\sigma_\Delta = 2\%$ and $\sigma_\Delta = \infty$. For example, the estimates for $\sigma_\Delta = 2\%$ are much less variable over time. Also, the estimates of the current premium for $\sigma_\Delta = \infty$ in all three plots are implausibly high, between 19 percent and 28 percent per year, mostly due to the high recent returns. In contrast, with $\sigma_\Delta = 2\%$, the current premium is estimated between 5.9 percent and 6.6 percent. Also note that a lower σ_Δ reduces the posterior uncertainty associated with the estimate of the current premium. Table II shows that the posterior standard deviation of the current premium with $\sigma_\Delta = \infty$ is huge, 10.18 percent, because the current premium is estimated based on the data only since the last structural break (whose location is uncertain). With $\sigma_\Delta = 2\%$, the posterior standard deviation is much smaller, 1.73 percent, as a result of incorporating the earlier data. These examples demonstrate that informative prior beliefs about magnitudes of changes in the premium have a big impact on the premium estimates.

Comparison of the first two plots in Figure 3 reveals the importance of the uncertainty about the locations of the 15 structural breaks. The second plot, which incorporates the uncertainty, produces a smoother pattern of the pre-

Figure 3. Equity premiums for different σ_Δs. The figure plots, for each month, the posterior mean of the equity premium μ^t obtained in specifications with no mean-variance link. The first plot arises from a specification with no transition regimes (TRs) and fixed breaks, the second from a specification that accounts for break uncertainty but has no TRs, and the third from a specification that includes TRs and accounts for break uncertainty. The solid line corresponds to $\sigma_\Delta = 2\%$ and the dashed line to $\sigma_\Delta = \infty$, where σ_Δ is the prior standard deviation of the shift in the equity premium at a structural break.

miums over time than the first plot, which ignores the uncertainty by conditioning on the break estimates.

C.3. The Effect of Beliefs About the Premium's Association with Volatility

As demonstrated in the previous subsection, the data before a structural break are relevant for estimating the current equity premium if one believes that extremely large shifts in the premium are unlikely. The data before a

Table II

Posterior Uncertainty About the Current Premium

The table reports posterior standard deviations of the current equity premium, as of June 1999, in percent per annum. The equity premium is defined as the expected rate of return on the aggregate stock market portfolio in excess of the short-term interest rate. The sample period is split into $2K + 1$ regimes, $K + 1$ of which are stable regimes (SRs), separated from each other by K transition regimes (TRs). Throughout, $K = 15$. The equity premium in the ith SR is denoted by μ_i. The jth TR is associated with a shift in the equity premium given by $\Delta_j = \mu_{j+1} - \mu_j$. In the framework "with no TRs," the beginning and ending points of TRs coincide. For $j = 1 \ldots, K$, the prior standard deviation of Δ_j is σ_Δ (annualized in the table). The standard deviation (volatility) of the excess stock return in the ith SR is denoted by σ_i. In each SR ($i = 1, \ldots, K + 1$), the relation between the equity premium and variance is given by $\mu_i = \psi_i \gamma \sigma_i^2$, where the prior for each ψ_i is a gamma distribution with parameters $\nu/2$ and $2/\nu$.

Panel A: No Prior Association Between μ_i and σ_i^2		
	$\sigma_\Delta = 2$	$\sigma_\Delta = \infty$
With TRs	1.73	10.18
No TRs	1.81	10.94
No TRs, fixed breaks	1.79	9.01

Panel B: With TRs and a Prior Association Between μ_i and σ_i^2		
	$\sigma_\Delta = 3$	$\sigma_\Delta = \infty$
$\nu = 10$ (moderate)	1.37	2.27
$\nu = \infty$ (perfect)	—	1.08

break can also be relevant if one believes that, across SRs, the equity premium has at least some degree of positive association with stock market volatility. For example, if the equity volatility in the last SR is low by historical standards, a prior belief in a link between the equity premium and volatility leads to an inference that the equity premium in the last SR is also low. Of course, such inference relies on data before the most recent break.

Figure 4 plots the equity premium estimated with different degrees of the mean-variance link. To isolate the effect of the link, the prior on Δ is noninformative ($\sigma_\Delta = \infty$). The solid line plots the premium estimated with a moderate link as in Section II.C.1 ($\nu = 10$), the dashed line plots the premium for a perfect link ($\nu = \infty$; the prior spiked at $\psi_i = 1$), and the dotted line plots the premium estimated with no mean-variance link ($\nu = 0$). The figure reveals that the effect of the link on the premium estimates is substantial. For example, in the period of high volatility in the 1930s, the premium estimates with a perfect link are more than double the estimates with no link. The current premium is estimated to be 4.3 percent with a perfect link, 6.4 percent with a moderate link ($\nu = 10$), and 27.7 percent with no link. The lower current premium in the presence of the link is due to the fact that the equity volatility in the 1990s has been 12.8 percent, less than the 17 percent volatility in the overall sample. These results clearly indicate that volatility can exert a strong effect on the estimate of the equity premium.
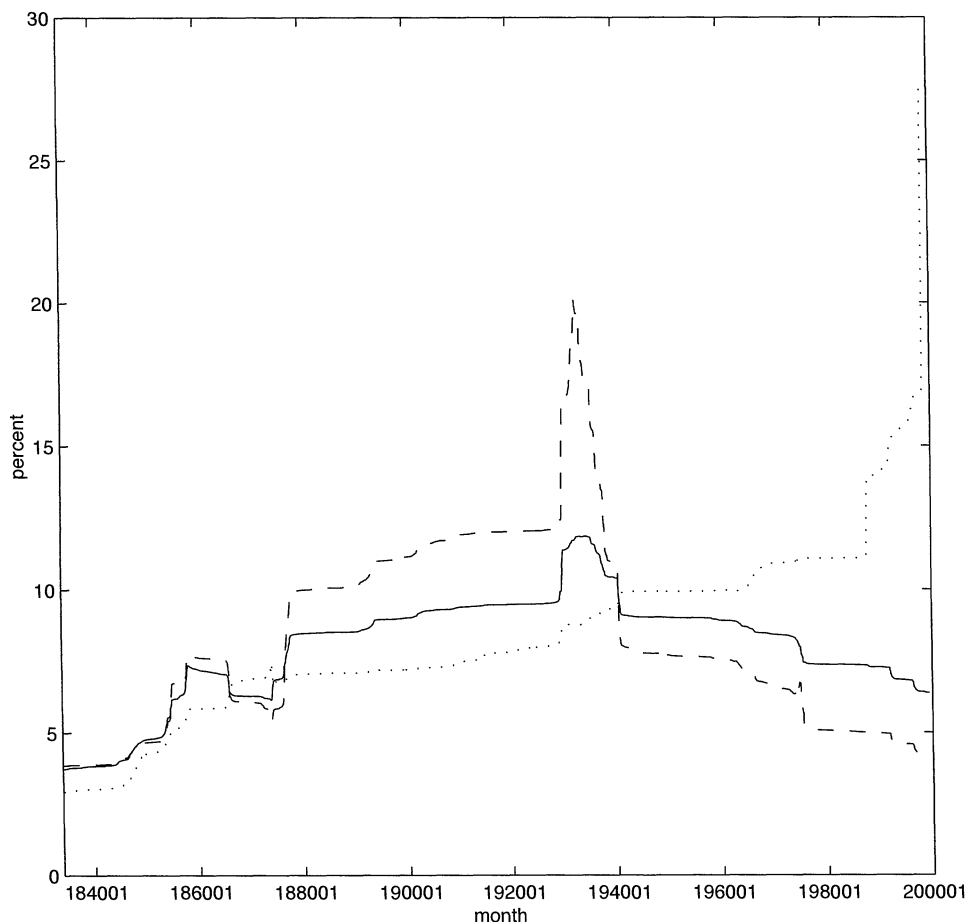
Figure 4. Equity premiums for different degrees of a mean-variance association. The figure plots, for each month, the posterior mean of the equity premium μ^t obtained in specifications with transition regimes and $\sigma_\Delta = \infty$. The solid line corresponds to a moderate prior mean-variance association ($\nu = 10$), the dashed line to a perfect link ($\nu = \infty$), and the dotted line to no link ($\nu = 0$).

Also note that the link reduces the posterior uncertainty associated with the current premium estimate, as shown in Table II. Whereas the posterior standard deviation of the current premium is 10.18 percent when $\sigma_\Delta = \infty$ and no link is imposed, it drops to 2.27 percent with the moderate link, and to 1.08 percent with a perfect link. Strengthening the link increases the precision of the current premium estimate for two reasons. First, it is well known that monthly data are more informative about volatilities than about means, especially for subperiods of modest length. A prior that links the mean to the volatility and is mildly informative about the price of risk allows us to learn about the mean from the volatility. The second reason is the

relatively low volatility in the 1990s. To see the basic point clearly, assume that σ_i^2 is known with certainty and that

$$\mu_i = \gamma\sigma_i^2, \quad i = 1,\ldots,K+1, \tag{24}$$

which is the perfect volatility link ($\nu = \infty$). Then

$$\text{Std}\{\mu_i|x,q\} = \sigma_i^2\,\text{Std}\{\gamma|x,q\}, \tag{25}$$

where "Std" denotes standard deviation. In this simplified setting, the posterior standard deviation of the equity premium in a given SR is proportional to that regime's equity variance, so a low-variance regime yields a lower standard deviation of the equity premium. To a rough approximation, the same reasoning applies to an imperfect but positive link.

C.4. *The Effect of the Transition Regimes*

Figure 5 plots the equity premium estimated in the specifications with TRs (solid) and without TRs (dashed). In both cases, there is no volatility link and $\sigma_\Delta = 2\%$. The effect of adding TRs is smaller than the previously observed effects of changing σ_Δ and ν, but it is still significant. For example, with no TRs, the equity premium rises from 6.2 percent to 6.6 percent in the 1990s, due to high recent returns. With TRs, however, the premium drops from 6.5 percent to 5.9 percent over the last five years, because much of the recent price increase is attributed to TRs during which the premium decreases. The inclusion of TRs in the statistical model is therefore not only theoretically appealing, but also empirically important.

III. Conclusions

A long history of aggregate stock returns contains information about the current equity premium even if the historical distribution of returns has experienced structural breaks. This study estimates the equity premium using a framework that combines the information in the entire return history with economically motivated prior beliefs. Our estimates also incorporate uncertainty about the timing of breaks.

Several economic considerations enter the specification of priors. First, changes in the equity premium are unlikely to be extreme. With an economically reasonable (i.e., finite) prior variance for shifts in the premium, equity returns before suspected breaks are still somewhat informative about the current premium. Second, across subperiods separated by structural breaks, it seems reasonable to believe that the equity premium is positively associated to at least some degree with equity volatility. We introduce a flexible prior that avoids specifying a parametric relation between the premium and volatility but allows information about the price of risk from earlier sub-
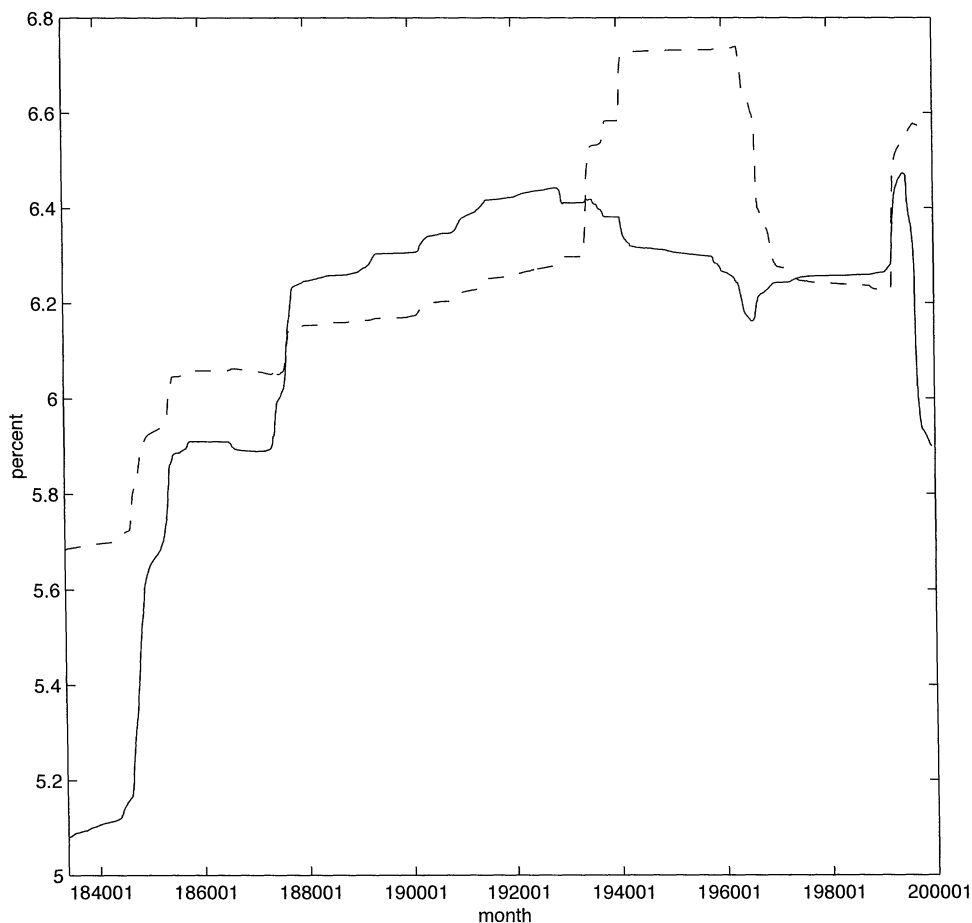
Figure 5. Equity premiums with and without transition regimes. The figure plots, for each month, the posterior mean of the equity premium μ^t obtained in specifications with $\sigma_\Delta = 2\%$ and no mean-variance link. The solid line corresponds to the specification with transition regimes (TRs), and the dashed line to the one without TRs.

periods to be used in estimating the current premium. Third, shifts in the equity premium are likely to be accompanied by contemporaneous price changes in the opposite direction. We incorporate this prior belief by introducing "transition" regimes between the "stable" regimes, where the latter have the usual interpretation associated with subperiods separated by structural breaks. Within a transition regime, our prior favors a negative relation between the equity return and the change in the premium between the previous and subsequent stable regimes.

Estimates of the equity premium based on reasonable priors fluctuate between 3.9 and 6.0 percent over the period from January 1834 through June 1999. The estimated premium rises through much of the 19th century and the first few decades of the 20th century, but it declines fairly steadily after

the 1930s except for a brief period in the mid-1970s. The estimated premium exhibits its sharpest decline of the entire period, to 4.8 percent, during the decade of the 1990s. We find that economically sensible priors are important in estimating the equity premium as well as in identifying the most likely dates at which breaks occurred. They also enable the current equity premium to be estimated with almost as much precision as what one would attribute to an estimate based on the long-sample average when potential breaks are ignored.

Appendix

Prior Means for the Transition-Regime Parameters

This part of the Appendix uses the results in Campbell (1991) in an empirical Bayes approach to specify sensible prior means for the TR parameters b_j and $\sigma^2_{j,j+1}$. (These prior means are denoted as $\bar{b}$ and α in equations (6) and (7).) Campbell's framework does not correspond exactly to ours. He models expected returns as linear functions of predetermined state variables, whereas no such dependence is modeled here. Also, his expected returns can change in any period, whereas ours can change only during TRs. Despite these differences, the underlying economics is the same in both frameworks, so we believe that his results are to some extent useful also in our framework. We use those results to center our mildly informative prior beliefs about b_j and $\sigma_{j,j+1}$.

Campbell (1991) decomposes unexpected aggregate stock returns $(x_t - E_{t-1}(x_t))$ into changes in the expectations of future dividend growth $(\eta_{d,t})$ and changes in expected future stock returns $(\eta_{h,t})$:

$$x_t - E_{t-1}(x_t) = \eta_{d,t} - \eta_{h,t}. \tag{A1}$$

He defines

$$\eta_{h,t} = (E_t - E_{t-1}) \sum_{k=1}^{\infty} \rho^k x_{t+k}, \tag{A2}$$

where ρ is a number slightly smaller than one, assumed equal to one here for simplicity, and $(E_t - E_{t-1})$ denotes the change in expectation of the quantity it precedes. Consider a transition regime TR_j such that the equity premium at time $t-1$ is $E_{t-1}(x_{t+k}) = \mu_j$, and at time t, $E_t(x_{t+k}) = \mu_{j+1}$, for each period $t+k$ that belongs to the next SR.[16] That is, for $k \geq 1$,

$$(E_t - E_{t-1})x_{t+k} = \begin{cases} \Delta_j & \text{if } t+k \in SR_{j+1} \\ 0 & \text{otherwise.} \end{cases} \tag{A3}$$

---

[16] Formally, our framework requires the TRs to be at least one period long. Letting the premium jump in a single step essentially eliminates the TR. Nevertheless, such a specification simplifies the exposition of our informal argument leading to a useful approximate prior relation between $\eta_{h,t}$ and $\Delta_{(t)}$.

The basic idea is to relate $\eta_{h,t}$ to Δ_j by controlling for the expected durations of the stable and transition regimes. Let $Pr_k = \text{Prob}(t + k \in SR_{j+1}) = p_{2j+1,2j+1}^{k-1}$. As explained in Section I.B.1, $p_{2j+1,2j+1}$ denotes the transition probability of staying in the $(j + 1)$st SR (which is regime $(2j + 1)$ overall). The value of $\eta_{h,t}$ that reflects the expected duration of the stable regime is

$$\eta_{h,t} = \Delta_j \sum_{k=1}^{\infty} Pr_k = \Delta_j \frac{1}{1 - p_{2j+1,2j+1}}$$

$$= \Delta_j E(d_{SR}|p_{2j+1,2j+1}), \quad t \in TR_j, \quad j = 1,\ldots,K. \tag{A4}$$

The last equality is taken from Section I.B.1. In order to obtain a useful approximate prior relation between $\eta_{h,t}$ and Δ_j, we replace the prior conditional expectation of the SR duration with an unconditional expectation. The intuition of the argument leading to equation (A4) holds also for a multiperiod TR, with an adjustment for the length of the TR. Treating all $\eta_{h,t}$s in the same TR equally, we divide equation (A4) by the expected duration of the TR:

$$\eta_{h,t} = \Delta_j \frac{E(d_{SR})}{E(d_{TR})}, \quad t \in TR_j, \quad j = 1,\ldots,K. \tag{A5}$$

Denote $\bar{d} = E(d_{SR})/E(d_{TR})$. Also denote $\Delta_{(t)} = \Delta_j$, for all $t \in TR_j$ and $j = 1,\ldots,K$. Then

$$\eta_{h,t} = \Delta_{(t)}\bar{d}, \quad t \in TR, \tag{A6}$$

where TR denotes the union of all TR_j, $j = 1,\ldots,K$.

A priori, all the b_js are viewed alike, $b_j = b$, $j = 1,\ldots,K$.[17] Based on the definition of b_j in equation (4) and given Δ, the value of b can be viewed as the slope coefficient in a regression of TR equity returns x_t on $\Delta_{(t)}$:

$$b = \frac{\text{Cov}(x_t,\Delta_{(t)})}{\text{Var}(\Delta_{(t)})}, \quad t \in TR. \tag{A7}$$

Combining equations (A6) and (A7),

$$b = \frac{\text{Cov}(x_t,\Delta_{(t)})}{\text{Var}(\Delta_{(t)})} = \frac{\text{Cov}(x_t,\eta_{h,t}/\bar{d})}{\text{Var}(\eta_{h,t}/\bar{d})} = \bar{d}\frac{\text{Cov}(x_t,\eta_{h,t})}{\text{Var}(\eta_{h,t})}, \quad t \in TR. \tag{A8}$$

Campbell's empirical decomposition of the variance of the U.S. equity returns from 1927 to 1988 yields $\text{Var}(\eta_d) = 0.369\sigma_R^2$, $\text{Var}(\eta_h) = 0.285\sigma_R^2$, and $-2\,\text{Cov}(\eta_d,\eta_h) = 0.346\sigma_R^2$, where σ_R^2 denotes the total variance of the un-

---

[17] Of course, in the posterior, the b_js are likely to be different from each other.

expected stock returns. Using these results, taken from the first line of Table II in Campbell (1991), it follows from (A1) that

$$\frac{\text{Cov}(x_t, \eta_h)}{\text{Var}(\eta_h)} = \frac{\text{Cov}(\eta_d, \eta_h)}{\text{Var}(\eta_h)} - 1 = -1.607. \tag{A9}$$

As described in Section I.B.1, the prior expected durations of the SR and TR are 113 and 12 months, respectively, so $\bar{d} = 113/12$. Plugging these values into equation (A8) gives a prior mean for b of

$$\bar{b} = -1.607\bar{d} = -15.13. \tag{A10}$$

Next, we determine a prior mean of the TR variance. The fitted returns from the regression of x_t on $\Delta_{(t)}$ considered in equation (A7) can be viewed as the mean returns in each TR. Therefore, the TR variance can be thought of as the residual variance from that regression. Because $\Delta_{(t)}$ is approximately proportional to $\eta_{h,t}$ (see equation (A6)), this variance is approximately the same as the residual variance σ_ϵ^2 from the regression of x_t on $\eta_{h,t}$ across the TRs. Hence,

$$\sigma_\epsilon^2 = \text{Var}(x_t) - \left(\frac{\text{Cov}(x_t, \eta_{h,t})}{\text{Var}(\eta_{h,t})} \right)^2 \text{Var}(\eta_{h,t}) = 0.264\sigma_R^2. \tag{A11}$$

We take $\sigma_R = 16.94\%$ per year, the unconditional volatility of x_t in the over-all sample. This choice yields $\sigma_\epsilon^2 = 0.000634$, which we use for α^2, the prior mean of the TR variance.

Drawing from the Posterior Distribution

In order to obtain draws of the parameter vector θ from its joint posterior distribution, we use a block-at-a-time version of the Metropolis–Hastings (MH) algorithm.[18] Repeated draws of model parameters from their full con-ditional distributions form a Markov chain. Beyond a burn-in stage, the ele-ments in the chain are draws from the joint posterior distribution.

The changepoints q are drawn using a technique developed by Chib (1998). The parameter vector θ is augmented with a latent state variable indicating the regime from which each observation has been drawn. Conditional on the current draw of θ, Chib's algorithm draws the state variable, which uniquely determines q. One substantial advantage of this technique is that each draw of q requires only two passes through the data, regardless of the number of changepoints. This fact enables us to obtain the posterior distribution of as many as 30 changepoints ($K = 15$), as well as to integrate out the uncer-tainty about the changepoint locations in estimating the equity premium. For more details, see Chib (1998).

---

[18] The algorithm is introduced by Metropolis et al. (1953) and generalized by Hastings (1970). See Chib and Greenberg (1995) for a detailed description of the algorithm.

To implement the MH algorithm, we first perform the change of variables $\lambda \equiv 1/\gamma$ and $\phi_k^2 \equiv 1/\psi_k$. Full conditional posteriors of the model parameters are then obtained by incorporating the change of variables and rearranging the joint posterior, a product of the prior in equation (18) and the likelihood function in equation (19):

$$\bar{\mu}|\cdot \sim N\left(\frac{\iota' V_\mu^{-1} \mu}{\iota' V_\mu^{-1} \iota}, \frac{1}{\iota' V_\mu^{-1} \iota}\right), \quad \bar{\mu} > 0 \tag{A12}$$

$$\phi_i^2|\cdot \sim \frac{\nu + \displaystyle\sum_{t=q_{2i-2}+1}^{q_{2i-1}} \frac{(x_t - \mu_i)^2}{\lambda \mu_i}}{\chi_{\nu+l_{2i-1}}^2}, \quad i = 1,\ldots,K+1 \tag{A13}$$

$$\lambda|\cdot \sim \frac{\dfrac{2}{b_\gamma} + \displaystyle\sum_{i=1}^{K+1} \sum_{t=q_{2i-2}+1}^{q_{2i-1}} \frac{(x_t - \mu_i)^2}{\phi_i^2 \mu_i}}{\chi_{2a_\gamma + \sum_{i=1}^{K+1} l_{2i-1}}^2}, \tag{A14}$$

$$\sigma_{j,j+1}^2|\cdot \sim \frac{(\eta - 2)\alpha^2 + \displaystyle\sum_{t=q_{2j-1}+1}^{q_{2j}} \left(x_t - \left(\frac{\mu_j + \mu_{j+1}}{2} + b_j \Delta_j\right)\right)^2}{\chi_{\eta+l_{2j}}^2}, \tag{A15}$$

$$j = 1,\ldots,K$$

$$b_j|\cdot \sim N(m_{b_j} v_{b_j}, v_{b_j}), \quad j = 1,\ldots,K, \tag{A16}$$

where

$$m_{b_j} = \frac{\bar{b}}{\sigma_b^2} + \frac{\Delta_j l_{2j}}{\sigma_{j,j+1}^2} \left(\frac{\displaystyle\sum_{t=q_{2j-1}+1}^{q_{2j}} x_t}{l_{2j}} - \frac{\mu_j + \mu_{j+1}}{2}\right)$$

$$v_{b_j} = \frac{1}{\dfrac{1}{\sigma_b^2} + \dfrac{\Delta_j^2 l_{2j}}{\sigma_{j,j+1}^2}}.$$

The full conditional posterior of each element of the transition matrix P is

$$p_{k,k}|\cdot \sim \text{Beta}(a_k + n_{kk}, c_k + 1), \tag{A17}$$

where n_{kk} is the number of one-step transitions of the latent state variable from regime k to regime k. The full conditional posterior distribution of μ can be shown to be equal to

$$p(\mu|\cdot) \propto \left(\prod_{i=1}^{K+1} \mu_i^{-(l_{2i-1}/2)} \right)$$

$$\times \exp \left\{ -\frac{1}{2} \left[(\mu - \bar{\mu}\iota)' V_\mu^{-1} (\mu - \bar{\mu}\iota) + \sum_{i=1}^{K+1} \sum_{t=q_{2i-2}+1}^{q_{2i-1}} \frac{(x_t - \mu_i)^2}{\lambda \phi_i^2 \mu_i} \right. \right.$$

$$\left. \left. + \sum_{j=1}^{K} \sum_{t=q_{2j-1}+1}^{q_{2j}} \frac{\left(x_t - \left(\frac{\mu_j + \mu_{j+1}}{2} + b_j \Delta_j \right) \right)^2}{\sigma_{j,j+1}^2} \right] \right\}, \quad \mu > 0$$

$$\propto \left(\prod_{i=1}^{K+1} \mu_i^{-(l_{2i-1}/2)} \right) \exp \left\{ -\frac{1}{2} \left[\mu'(V_\mu^{-1} + D_1 + D_2) \mu \right. \right.$$

$$\left. \left. - 2\mu'(V_\mu^{-1} \bar{\mu}\iota + w + g) + d'\overline{x^2} \right] \right\}, \quad \mu > 0,$$

$$(A18)$$

where d is a $(K + 1)$-vector whose ith element is

$$d_i = \frac{l_{2i-1}}{\lambda \mu_i \phi_i^2}, \tag{A19}$$

w is a $(K + 1)$-vector whose ith element is

$$w_i = -\frac{l_{2i-1}}{2\lambda \phi_i^2}, \tag{A20}$$

$\overline{x^2}$ is a $(K + 1)$-vector whose ith element is

$$\overline{x_i^2} = \frac{1}{l_{2i-1}} \sum_{t=q_{2i-2}+1}^{q_{2i-1}} x_t^2, \tag{A21}$$

$$D_1 = \begin{pmatrix} z_{1,1} & 0 & 0 & \cdots & 0 & 0 \\ 0 & z_{2,1} + z_{1,2} & 0 & \cdots & 0 & 0 \\ 0 & 0 & z_{2,2} + z_{1,3} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & z_{2,K-1} + z_{1,K} & 0 \\ 0 & 0 & 0 & \cdots & 0 & z_{2,K} \end{pmatrix} \tag{A22}$$

$$D_2 = \begin{pmatrix} 0 & -z_{3,1} & 0 & \cdots & 0 & 0 \\ -z_{3,1} & 0 & -z_{3,2} & \cdots & 0 & 0 \\ 0 & -z_{3,2} & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & -z_{3,K} \\ 0 & 0 & 0 & \cdots & -z_{3,K} & 0 \end{pmatrix} \tag{A23}$$

$$g = \begin{pmatrix} z_{4,1} \\ z_{5,1} + z_{4,2} \\ z_{5,2} + z_{4,3} \\ \vdots \\ z_{5,K-1} + z_{4,K} \\ z_{5,K} \end{pmatrix}, \tag{A24}$$

and, for $j = 1,\ldots,K$,

$$z_{1,j} = \frac{l_{2j}\left(b_j - \dfrac{1}{2}\right)^2}{\sigma_{j,j+1}^2} \tag{A25}$$

$$z_{2,j} = \frac{l_{2j}\left(b_j + \dfrac{1}{2}\right)^2}{\sigma_{j,j+1}^2} \tag{A26}$$

$$z_{3,j} = \frac{l_{2j}\left(b_j - \dfrac{1}{2}\right)\left(b_j + \dfrac{1}{2}\right)}{\sigma_{j,j+1}^2} \tag{A27}$$

$$z_{4,j} = -\frac{l_{2j}\left(b_j - \dfrac{1}{2}\right)}{\sigma_{j,j+1}^2}\,\overline{x}_{2j} \tag{A28}$$

$$z_{5,j} = \frac{l_{2j}\left(b_j + \dfrac{1}{2}\right)}{\sigma_{j,j+1}^2}\,\overline{x}_{2j} \tag{A29}$$

$$\overline{x}_{2j} = \frac{1}{l_{2j}} \sum_{t=q_{2j-1}+1}^{q_{2j}} x_t. \tag{A30}$$

Each of the full conditional distributions, except for that of μ in equation (A18), corresponds to a well-known density, so draws from those densities are easily generated. In order to draw μ, a candidate value is drawn from a proposal density and accepted with a probability that ensures convergence of the chain to the target distribution. If a candidate draw is not accepted, the previous draw is retained. The proposal density for μ is a piecewise linear approximation to the target density. We use a two-pass grid method to draw from the proposal density. In the first pass, we use an equally spaced grid (with $f_1 = 40$ gridpoints) to fit a piecewise linear cdf $P_1(\mu|\cdot)$ corresponding to $p(\mu|\cdot)$. The second-pass grid is created by taking $P_1^{-1}(f/f_2|\cdot)$, for $f = 1, \ldots, f_2$, where f_2 determines the fineness of the second-pass grid (we use $f_2 = 25$ gridpoints). In both grids, we also add a few gridpoints in the tails of the support, to prevent undersampling. The second grid increases the accuracy of the procedure by putting more grid points in the regions of greater mass. Through the second-pass grid, we then refit the piecewise linear cdf, denoted as $P_2(\mu|\cdot)$, and use the inverse cdf method to draw μ. That is, we generate a standard uniform variate u and take μ as $P_2^{-1}(u|\cdot)$. The second-pass grid is a very good approximation to the target density, since the rate at which the draws from the proposal density are accepted is over 95 percent. Note that our two-pass grid procedure can, in principle, be used to make draws from virtually any distribution whose integrating constant is unknown. The procedure is marginally feasible in terms of computation time even with a highly dimensional parameter space (here, μ has $K + 1 = 16$ elements).

For each posterior draw of μ, a draw of $\mu_{(j)}$ is constructed as $(\mu_j + \mu_{j+1})/2$ for $j = 1, \ldots, K$. The posterior moments of μ and $\mu_{(j)}$, as well as the posterior probabilities that a particular month belongs in a given regime, are estimated using 40,000 posterior draws, obtained by retaining every 15th draw from a chain of 600,000 draws beyond the first 6,000 draws. For each set of prior parameters, we run two independent Markov chains, with two different seeds in the random number generator. Such an exercise is used to assess the precision of our results. In all cases, the premium estimates produced by the two independent chains are very close, almost always within less than 10 bp (annualized). Minor exceptions with somewhat larger deviations occasionally occur in some cases that are less interesting a priori (e.g., no mean-variance link and $\sigma_\Delta = \infty$), but the overall precision of our results is satisfactory. The reported premium estimate in each month is obtained by averaging the premiums obtained from the two independent chains.

REFERENCES

Albert, James H., and Siddhartha Chib, 1993, Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts, *Journal of Business and Economic Statistics* 11, 1–15.

Andrews, Donald W. K., 1993, Tests for parameter instability and structural change with unknown change point, *Econometrica* 61, 821–856.

Andrews, Donald W. K., and Werner Ploberger, 1994, Optimal tests when a nuisance parameter is present only under the alternative, *Econometrica* 62, 1383–1414.

Ang, Andrew, and Geert Bekaert, 1998, Regime switches in interest rates, Working paper, Stanford University.

Backus, David, and Allan W. Gregory, 1993, Theoretical relations between risk premiums and conditional variances, *Journal of Business and Economic Statistics* 11, 177–185.

Bai, Jushan, 1995, Least absolute deviation estimation of a shift, *Econometric Theory* 11, 403–436.

Bai, Jushan, 1997, Estimating multiple breaks one at a time, *Econometric Theory* 13, 315–352.

Bai, Jushan, Robin L. Lumsdaine, and James H. Stock, 1998, Testing for and dating common breaks in multivariate time series, *Review of Economic Studies* 65, 395–432.

Bai, Jushan, and Pierre Perron, 1998, Estimating and testing linear models with multiple structural changes, *Econometrica* 66, 47–78.

Baillie, Richard T., and Ramon P. DeGennaro, 1990, Stock returns and volatility, *Journal of Financial and Quantitative Analysis* 25, 203–214.

Barry, Daniel, and J. A. Hartigan, 1993, A Bayesian analysis for change point problems, *Journal of the American Statistical Association* 88, 309–319.

Berger, James O., 1985, *Statistical Decision Theory and Bayesian Analysis* (Springer-Verlag, New York).

Bhattacharya, P. K., 1994, Some aspects of change-point analysis, in E. Carlstein, H.-G. Mueller, and D. Siegmund, eds.: *Change Point Problems, IMS Lecture Notes—Monograph Series*, Vol. 23 (Institute of Mathematical Statistics, Hayward, CA).

Bigelow, Erastus B., 1862, *The Tariff Question…* (Little & Brown, Boston).

Brealey, Richard A., and Stewart C. Myers, 1996, *Principles of Corporate Finance* (McGraw-Hill, New York).

Broemeling, Lyle D., and Hiroki Tsurumi, 1987, *Econometrics and Structural Change* (Marcel Dekker, New York).

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

Campbell, John Y., 1991, A variance decomposition for stock returns, *The Economic Journal* 101, 157–179.

Carlin, Bradley P., Alan E. Gelfand, and Adrian F. M. Smith, 1992, Hierarchical Bayesian analysis of changepoint problems, *Applied Statistician* 41, 389–405.

Chan, K. C., Andrew Karolyi, and René M. Stulz, 1992, Global financial markets and the risk premium on U.S. equity, *Journal of Financial Economics* 32, 137–167.

Chen, Jie, and A. K. Gupta, 1997, Testing and locating variance changepoints with application to stock prices, *Journal of the American Statistical Association* 92, 739–747.

Chernoff, H., and S. Zacks, 1964, Estimating the current mean of a normal distribution which is subjected to changes in time, *Annals of Mathematical Statistics* 35, 999–1018.

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

Chib, Siddhartha, and Edward Greenberg, 1995, Understanding the Metropolis–Hastings algorithm, *American Statistician* 49, 327–335.

Cowles, Alfred, III, and Associates, 1939, *Common Stock Indexes*, 2nd ed., *Cowles Commission Monograph* no. 3 (Principia Press, Bloomington, IN).

Diebold, Francis X., and Celia Chen, 1996, Testing structural stability with endogenous break point: A size comparison of analytic and bootstrap procedures, *Journal of Econometrics* 70, 221–241.

Farrell, Maurice L., ed., 1972, *The Dow Jones Averages, 1885–1970* (Dow Jones and Company, New York).

French, Kenneth R., G. William Schwert, and Robert F. Stambaugh, 1987, Expected stock returns and variance, *Journal of Financial Economics* 19, 3–29.

Glosten, Lawrence R., Ravi Jagannathan, and David E. Runkle, 1993, On the relation between the expected value and the variance of the nominal excess return on stocks, *Journal of Finance* 48, 1779–1801.

Goetzmann, William N., and Roger G. Ibbotson, 1994, An emerging market: The NYSE from 1815 to 1871, Working paper, Yale University.

Hamilton, James D., 1989, A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica* 57, 357–384.

Harvey, Campbell R., 1989, Time-varying conditional covariances in tests of asset pricing models, *Journal of Financial Economics* 24, 289–317.

Hastings, W. K., 1970, Monte Carlo sampling methods using Markov chains and their applications, *Biometrika* 57, 97–109.

Homer, Sidney, 1963, *A History of Interest Rates* (Rutgers University Press, New Brunswick, NJ).

Hsu, D. A., 1982, Robust inferences for structural shift in regression models, *Journal of Econometrics* 19, 89–107.

Inclán, Carla, 1993, Detection of multiple changes of variance using posterior odds, *Journal of Business and Economic Statistics* 11, 289–300.

Kandel, Shmuel, and Robert F. Stambaugh, 1990, Expectations and volatility of consumption and asset returns, *Review of Financial Studies* 3, 207–232.

Krishnaiah, P. R., and B. Q. Miao, 1988, Review about estimation of change points, in P. R. Krishnaiah and C. R. Rao, eds.: *Handbook of Statistics*, Vol. 7 (Elsevier, New York).

Liu, Jian, Shiying Wu, and James V. Zidek, 1997, On segmented multivariate regression, *Statistica Sinica* 7, 497–525.

Macaulay, F. R., 1938, *The Movements of Interest Rates, Bond Yields and Stock Prices in the United States since 1856* (NBER, New York).

McCulloch, Robert E., and Ruey S. Tsay, 1994, Bayesian inference of trend- and difference-stationarity, *Econometric Theory* 10, 596–608.

Merton, Robert C., 1980, On estimating the expected return on the market: An exploratory investigation, *Journal of Financial Economics* 8, 323–361.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, 1953, Equations of state calculations by fast computing machines, *Journal of Chemical Physics* 21, 1087–1092.

Pástor, Ľuboš and Robert F. Stambaugh, 1999, Costs of equity capital and model mispricing, *Journal of Finance* 54, 67–121.

Schwert, G. William, 1989, Business cycles, financial crises and stock volatility, *Carnegie-Rochester Conference Series on Public Policy* 31, 83–125.

Schwert, G. William, 1990, Indexes of U.S. stock prices from 1802 to 1987, *Journal of Business* 63, 399–426.

Scruggs, John T., 1998, Resolving the puzzling intertemporal relation between the market risk premium and conditional market variance: A two-factor approach, *Journal of Finance* 53, 575–603.

Siegel, Jeremy J., 1992, The real rate of interest from 1800–1990: A study of the U.S. and the U.K., *Journal of Monetary Economics* 29, 227–252.

Smith, W. B., and A. H. Cole, 1935, *Fluctuations in American Business, 1790–1860* (Harvard University Press, Cambridge, MA).

Sowell, Fallaw, 1996, Optimal tests for parameter instability in the generalized method of moments framework, *Econometrica* 64, 1085–1107.

Stephens, D. A., 1994, Bayesian retrospective multiple-changepoint identification, *Applied Statistician* 43, 159–178.

Tauchen, George, and Robert Hussey, 1991, Quadrature-based methods for obtaining approximate solutions to nonlinear asset pricing models, *Econometrica* 59, 371–396.

Turner, Christopher M., Richard Startz, and Charles R. Nelson, 1989, A Markov model of heteroskedasticity, risk, and learning in the stock market, *Journal of Financial Economics* 25, 3–22.

Viceira, Luis M., 1997, Testing for structural change in the predictability of asset returns, Working paper, Harvard University.

Werner, Walter, and Steven T. Smith, 1991, *Wall Street* (Columbia University Press, New York).

Whitelaw, Robert F., 1994, Time variations and covariations in the expectation and volatility of stock market returns, *Journal of Finance* 49, 515–541.

Zacks, S., 1983, Survey of classical and Bayesian approaches to the change-point problem: Fixed sample and sequential procedures of testing and estimation, in M. H. Rivzi, J. S. Rustagi, and D. Siegmund eds.: *Recent Advances in Statistics* (Academic Press, New York).

Zellner, Arnold, 1971, *An Introduction to Bayesian Inference in Econometrics* (John Wiley and Sons, New York).