The Equity Premium and Structural Breaks

by*

Ľuboš Pástor

and

Robert F. Stambaugh

August 1998

Abstract

Evidence of structural breaks in the historical return distribution raises concerns about averaging a long series to estimate the current equity premium. Data before a break are relevant if one believes that large shifts in the premium are unlikely or that the premium is associated, to some degree, with volatility. The equity excess-return series over two centuries exhibits multiple structural breaks, the latest of which occurs early in the current decade. The average excess return since that break is nearly 10%, but incorporating prior beliefs as described above produces substantially lower estimates of the equity premium.

JEL Classifications: G10, C11

Keywords: equity premium, structural breaks, Bayesian analysis

---

*The Wharton School, University of Pennsylvania (Pástor) and the Wharton School, University of Pennsylvania and the National Bureau of Economic Research (Stambaugh). Stambaugh acknowledges support provided by his appointment during the 1997–98 academic year as a Marvin Bower Fellow at Harvard Business School, where portions of this research were conducted.

1. Introduction

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 (1998) 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.[1] 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 nineteenth 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."

Discarding data before a suspected break reduces the risk of contaminating an estimate of the equity premium with data drawn from a different distribution, possibly having a different mean. That practice seems prudent, but it also encounters the reality that, other things equal, shorter histories yield less precision. The decision to discard data no doubt depends on one's confidence that a break occurred, but it seems that additional judgment must play a role. Suppose, to take an extreme example, that one is confident 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 data drawn from a pre-break distribution, but it hardly seems sensible in estimating the current equity premium. Even if those data are thought to be drawn from a different distribution, it still seems likely that many would judge them to be at least somewhat informative about the current premium. After all, the historical data are still equity returns (as opposed to, say, rainfall data), and judging them to be completely uninformative about the current equity premium would surely be an extreme view.

As time passes since the most recent break, discarding all of the pre-break data seems less radical, but we suggest that similar reasoning still applies. The data before the break offer a

[1] "Mispricing" uncertainty about whether a factor-based model correctly prices the firm's equity can account for a larger fraction, but generally only in cases where the mispricing uncertainty is quite large in economic terms.

tradeoff between imprecision and contamination. Completely discarding the pre-break data is appropriate only when one believes it likely that the break shifted the distribution to such a degree that the pre-break data on equity returns are no more useful than any arbitrarily chosen data set. For those with less extreme beliefs, the pre-break data provide at least some information about the current equity premium, and this study explores methods for incorporating that information.

A shift in equity volatility can occur at a structural break. For example, maximum-likelihood estimation identifies a break near the end of 1991. For the 1992–96 subperiod, the standard deviation of monthly excess market returns is less than 9% (annualized), as compared to about 17% over the entire 1834–1996 period analyzed in this study. Does this recently low equity volatility have implications for the current equity premium? Suppose one believes that, across subperiods separated by structural breaks, there is at least some association between the equity premium and volatility. Then an observation that recent volatility is low by historical standards would tend to lower one's estimate of the current equity premium, relative to an estimate constructed in the absence of such a belief. Of course, the observation that recent volatility is low by historical standards requires the use of data from earlier subperiods. Given a belief in some degree of a premium-volatility link, each earlier subperiod's ratio of equity premium to variance, its "price of risk," provides some information about the current price of risk and, thereby, about the current equity premium. Thus, a belief that the equity premium has a positive association with volatility provides another reason for using data preceding structural breaks, even if one believes that large shifts in the mean and variance could have accompanied those breaks. This study also explores methods for estimating the equity premium when one believes it is positively associated, to some degree, with equity volatility.

We develop and apply a Bayesian framework for estimating the equity premium in the presence of structural breaks. Our methodology can incorporate prior beliefs that the equity premium is associated with volatility or is unlikely to experience extremely large shifts. A few examples illustrate the potential importance of such beliefs. The average excess stock-market return since the most recent structural break, identified at the end of 1991, is nearly 10% (per annum). Suppose one believes that a structural break is unlikely to be accompanied by a shift in the equity premium much greater than 4%, but that the premiums before and after the break are otherwise unrelated. Then, as opposed to the recent 10% average, the estimate of the current equity premium is less than 7% when based on an overall sample beginning in 1834, even though multiple structural breaks are identified within that sample. Suppose one believes instead that, in any earlier subperiod separated from the most recent by one or more structural breaks, there is an 80% probability that the price of risk is equal to the current price of risk multiplied by a number between 0.2 and 2. Then the estimate of the current equity premium is less than 5%, as compared to the recent 10% average, even though the premium's link with volatility specified in the prior seems fairly weak. Combinations of the two types of prior beliefs are also entertained. In general, our investigation reveals the potential importance of using longer return histories, even when such histories contain structural breaks.

The remainder of the paper is organized as follows. The methodology is developed in Section 2, wherein we describe the stochastic setting and discuss the priors used in our Bayesian approach. Section 3 presents the empirical results. Most of that section reports results for a three-break model, with breaks in 1928, 1940, and 1991. We also consider a model with a single break at the end of 1925, which is used as a starting date in many analyses. The locations of the structural breaks, the "breakpoints," are estimated separately and viewed as exogenous in implementing the Bayesian approach in Sections 2 and 3. Section 4 discusses potential extensions in which the Bayesian setting incorporates uncertainty about the number and locations of the breakpoints. As an initial exploration, we present an example with uncertainty about the location of a single break. Section 5 reviews the conclusions.

2. Methodology

We describe here our Bayesian framework for making inferences in the presence of structural change in the distribution of excess market returns. Much of this framework is newly developed, with the objective of incorporating prior beliefs about the equity premium's association with volatility and the potential magnitudes of its shifts. At the same time, our analysis shares features with previous studies dealing with structural change. 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), Bai, Lumsdaine, and Stock (1997), Bai and Perron (1998), Diebold and Chen (1996), Liu, Wu, and Zidek (1997), and Sowell (1996). 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 (1997).[2] Recent studies that investigate structural breaks in some financial time series include Inclán (1993), Chen and Gupta (1997), Viceira

[2]Markov switching models, proposed by Hamilton (1989), are studied in a Bayesian context by Albert and Chib (1993) and McCulloch and Tsay (1994).

(1997), and Ang and Bekaert (1998).

This section first introduces the simple stochastic setting and discusses maximum likelihood estimation. The prior distributions for the model's parameters are then presented and analyzed. The general approach for obtaining posterior distributions is discussed briefly at the conclusion of this section, but the details of the computations are given in the Appendix.

2.1. Stochastic Framework

The data consist of T observations of excess market returns. Let x_t denote the excess return for time t, and define $x = (x_1, \ldots, x_T)$. The overall sample period includes K structural breaks in the probability distribution of excess market returns, and the times at which the breakpoints occur are denoted by $s_1, \ldots, s_K$. The breakpoints divide the sample into $K+1$ subperiods, and within each subperiod the excess market returns are assumed to be normally distributed:

$$\begin{aligned} x_t &\sim N(\mu_1, \sigma_1^2) & t &= 1, \ldots, s_1 \\ x_t &\sim N(\mu_2, \sigma_2^2) & t &= s_1+1, \ldots, s_2 \\ &\vdots & &\vdots \\ x_t &\sim N(\mu_{K+1}, \sigma_{K+1}^2) & t &= s_K+1, \ldots, T. \end{aligned} \tag{1}$$

The number and locations of the breakpoints are viewed as exogenous with respect to the Bayesian setting initially developed here. In the empirical results, reported in the next section, the breakpoints are obtained using maximum-likelihood estimation before conducting the Bayesian analysis. This treatment of the breakpoints is motivated by computational tractability. Later in the study, Section 4 explains how the Bayesian setting can be extended to include uncertainty about the number and locations of the breaks.

Let $\mu = (\mu_1, \ldots, \mu_{K+1})$ denote the $(K+1) \times 1$ vector of equity premiums, and let $\sigma = (\sigma_1, \ldots, \sigma_{K+1})$ denote the $(K+1) \times 1$ vector of standard deviations ("volatilities"). Also define $s = (s_1, \ldots, s_K)$. The likelihood function can be written as a product of $(K+1)$ normal densities:

$$p(x|\mu, \sigma, s) \propto \left(\prod_{k=1}^{K+1} \frac{1}{\sigma_k^{s_k - s_{k-1}}} \right) \exp \left\{ -\frac{1}{2} \sum_{k=1}^{K+1} \sum_{t=s_{k-1}+1}^{s_k} \frac{(x_t - \mu_k)^2}{\sigma_k^2} \right\}, \tag{2}$$

where $s_0 = 0$, $s_{K+1} = T$, and "$\propto$" denotes "proportional to" (up to a factor not involving μ,

σ, or s). Conditional on s, the likelihood function is maximized at

$$\hat{\mu}_k^{(s)} = (\sum_{t=s_{k-1}+1}^{s_k} x_t)/(s_k - s_{k-1}) \tag{3}$$

$$\hat{\sigma}_k^{2(s)} = (\sum_{t=s_{k-1}+1}^{s_k} x_t^2)/(s_k - s_{k-1}) - (\sum_{t=s_{k-1}+1}^{s_k} x_t)^2/(s_k - s_{k-1})^2, \tag{4}$$

for $k = 1, \dots, K+1$. Let $\hat{\mu}^{(s)} = (\hat{\mu}_1^{(s)}, \dots, \hat{\mu}_{K+1}^{(s)})$ and $\hat{\sigma}^{(s)} = (\hat{\sigma}_1^{(s)}, \dots, \hat{\sigma}_{K+1}^{(s)})$. The MLE of s, denoted by $\hat{s}$, is the combination of breakpoints that maximizes $p(x|\hat{\mu}^{(s)}, \hat{\sigma}^{(s)}, s)$, and $\hat{s}$ is computed by searching over all possible breakpoint combinations. The MLE's of μ_k and σ_k^2 are then equal to the estimators in (3) and (4) conditional on $s = \hat{s}$.

2.2. Prior Beliefs

The MLE's discussed above are based only on the likelihood function in (2). In contrast, Bayesian estimators of μ and σ combine the sample information 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 primarily by economic arguments. First, we impose a prior belief that the equity premium is positive. This prior reflects a simple economic 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 non-negativity restriction on the expected excess market return should be imposed in estimating the equity premium. As explained below, our framework also allows informative prior beliefs about the potential magnitudes of changes in the equity premium and about the premium's association with volatility.

2.2.1. Beliefs About Changes in the Premium

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

$$p(\mu|\bar{\mu}) \propto \exp\left\{-\frac{1}{2}(\mu - \bar{\mu}\,\iota)' V_\mu^{-1} (\mu - \bar{\mu}\,\iota)\right\}, \qquad \mu > 0, \tag{5}$$

$$p(\bar{\mu}) \propto 1, \qquad 0 < \bar{\mu} < \kappa, \tag{6}$$

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.[3] The prior for

[3]In the absence of truncation at zero in (5), $\bar{\mu}$ would equal the mean of $p(\mu|\bar{\mu})$.

μ conditional on $\bar{\mu}$ is a truncated normal distribution whose location depends on $\bar{\mu}$. Since κ is set to a very large value, the uniform 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 (5) is informative about differences between the elements of μ. Define $\Delta_k = (\mu_{k+1} - \mu_k),\ k = 1, \ldots, K+1$, and let $\Delta = (\Delta_1, \ldots, \Delta_K)$. The elements of Δ represent the magnitudes by which the market premium changes at the breakpoints. Note that (5) implies that the prior on each Δ_k 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% that the annual equity premium has shifted by more than 10% at any structural break. This type of prior belief can be expressed by specifying a value for the standard deviation of the prior distribution of each Δ_k, denoted by σ_Δ.[4] In the preceding example, $\sigma_\Delta = 5\%$. At one extreme, setting $\sigma_\Delta = \infty$ assigns equal prior probabilities to fixed-width neighborhoods around all values of Δ_k, however large. One consequence of such a noninformative belief about Δ is that, in estimating the current equity premium, all data from subperiods before the last structural break are discarded (in the absence of a volatility link, discussed below). 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 $\Delta_k = 0$ and there has never been a change in the equity premium, in which case data from the entire sample are simply "pooled," roughly speaking, to estimate the current premium (additional discussion is provided in Section 3). In intermediate cases, the smaller the value of σ_Δ, the more attention is paid to data from subperiods before the last break. 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 elements of the covariance matrix V_μ in (5). The

[4]In general, the literature treats the means before and after a structural break as independent of each other (see Carlin, Gelfand, and Smith (1992) and Barry and Hartigan (1993) for Bayesian examples and Liu, Wu, and Zidek (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.

following structure is imposed on V_μ:

$$V_\mu = \sigma_\mu^2 \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^K \\ \rho & 1 & \rho & \dots & \rho^{K-1} \\ \rho^2 & \rho & 1 & \dots & \rho^{K-2} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho^K & \rho^{K-1} & \rho^{K-2} & \dots & 1 \end{pmatrix}, \quad -1 < \rho < 1. \tag{7}$$

Conditional on $\bar{\mu}$, and in the absence of truncation, the prior variance of each μ_k equals σ_μ^2, and the prior correlation between μ_j and μ_k equals $\rho^{|k-j|}$. The unconditional prior variance of Δ_k for any k is equal to $\sigma_\Delta^2 = \text{Var}(\mu_{k+1} - \mu_k)$. The values of σ_μ^2 and ρ that produce a desired value of σ_Δ are computed by simulation, using the uniform and truncated normal distributions. For any positive value of σ_Δ, there is an infinite number of such combinations of σ_μ^2 and ρ. In the empirical results presented in the next section, we select two alternative specifications, distinguished by the choices of ρ. In the first, we set $\rho = 0$ and then solve (numerically) for σ_μ^2 to produce the desired σ_Δ. In the resulting prior, the equity premium is believed to fluctuate independently across subperiods and thereby exhibit "immediate" mean reversion to a grand mean. This setting is denoted as the "$\rho = 0$" specification, in which the correlation between μ_{k+1} and μ_k, conditional on $\bar{\mu}$, is approximately equal to zero. The correlation is not precisely zero, due to the truncation of the conditional normal density in (5), but the difference is very small. In the alternative specification, we fix σ_μ^2 at a large value and then solve for ρ to produce the desired σ_Δ. The resulting values of ρ are very close to 1 for the informative (finite) values of σ_Δ we consider, so this setting is denoted as the "$\rho \approx 1$" specification. With $\rho \approx 1$, the equity premium nearly follows a random walk across subperiods and reverts only negligibly toward a grand mean. In the remainder of the paper, we simply refer to ρ as the "correlation" between μ_{k+1} and μ_k, suppressing the qualifications that the correlation is conditioned on $\bar{\mu}$ and actually deviates slightly from ρ, due to the truncation. Note that the conditional correlation is quite distinct from the unconditional correlation: for a finite σ_Δ, the latter is close to 1 for all values of ρ, due to the high variance of $\bar{\mu}$.

2.2.2. Beliefs About the Premium's Association with Volatility

In a study on 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 a structural-break setting.[5] One might believe that occasional shifts in volatility at structural breaks, separated by a number of years, are associated to some degree with shifts in the equity premium. 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 a subperiod.[6] Moreover, given the relatively small number of subperiods in feasible implementations of structural-break models, the data are likely to be more informative about the presence of a high-frequency relation, if any. The prior link between the equity premium and volatility that we introduce below can take the form of a weak positive association, 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

$$\mu_k = \gamma \psi_k \sigma_k^2, \qquad k = 1, \ldots, K+1, \tag{8}$$

and let $\psi = (\psi_1, \ldots, \psi_{K+1})$. We set $\psi_{K+1} \equiv 1$ to achieve identification, which then implies that γ is the market "price of risk" in the last subperiod, defined as the ratio of the equity premium to the equity variance. The prior on γ is specified as a gamma distribution with parameters $\epsilon/2$ and $2/\epsilon$,

$$p(\gamma) \propto \gamma^{\frac{\epsilon}{2}-1} \exp\left\{-\frac{\gamma\epsilon}{2}\right\}, \qquad \gamma > 0, \tag{9}$$

where ϵ is a positive constant close to zero. Note that, for $\epsilon = 0$, the above prior is a standard diffuse or noninformative prior, $p(\gamma) \propto 1/\gamma$, which is improper.[7] When ϵ is positive, the prior

[5]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.

[6]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.

[7]An "improper" prior density does not have a finite integral.

on γ is proper, which is desirable so that Bayes factors in a later discussion can be defined. With ϵ close to zero, however, the above prior on γ is still noninformative. In other words, this study assumes that there is no prior information about the market price of risk. If such prior information is available, it can easily be incorporated by adjusting the location of $p(\gamma)$ and choosing a larger value of ϵ.

In each earlier subperiod $k = 1, \ldots, K$, the price of risk is equal to $\psi_k \gamma$, with the ψ_k's assumed to be independent across subperiods. The prior on each ψ_k is a gamma distribution, with parameters $\nu/2$ and $2/\nu$,

$$p(\psi_k) \propto \psi_k^{\frac{\nu}{2}-1} \exp\left\{-\frac{\psi_k \nu}{2}\right\}, \qquad \psi_k > 0, \qquad k = 1, \ldots, K. \tag{10}$$

The prior on ψ_k implies that[8]

$$\mathrm{E}(\psi_k) = 1 \tag{11}$$

$$\mathrm{Var}(\psi_k) = \frac{2}{\nu}\,. \tag{12}$$

The desired degree of association between μ_k and σ_k^2 is achieved by specifying the parameter ν. At one extreme, as $\nu \to 0$, the prior on ψ_k approaches a standard diffuse or noninformative prior, $p(\psi_k) \propto 1/\psi_k$. With a noninformative prior on ψ_k, no association between the elements of μ and σ is imposed a priori. At the other extreme, as $\nu \to \infty$, it follows from (12) that $\mathrm{Var}(\psi_k) \to 0$, so $\psi_k = 1$ for all k, which imposes a perfect link between μ_k and σ_k^2 of the form $\mu_k = \gamma\sigma_k^2$. 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. Figure 1 plots prior densities of ψ_k for different values of ν. Observe that, for $\nu = 5$, the prior density is quite disperse, so the prices of risk in earlier periods can differ substantially from γ. As will be demonstrated, however, such a prior can still exert a substantial influence on the posterior distribution of the current equity premium.

Let $\theta = (\mu, \psi, \gamma, \bar{\mu})$. 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

$$p(\theta) = p(\mu|\bar{\mu})\ p(\bar{\mu})\ p(\gamma) \left(\prod_{k=1}^{K+1} p(\psi_k)\right). \tag{13}$$

[8] The two moment equations follow from standard results for the gamma density, such as in Zellner (1971, p.370). The moments exist for all $\nu > 0$, but the density has no mode for $\nu < 2$.

Recall that the densities multiplied on the right-hand side are given in equations (5), (6), (9), and (10).

2.3. Posterior Distributions

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 from (8), the reparameterized likelihood from (2) can be written as

$$p(x|\theta, s) \propto \gamma^{T/2} \left(\prod_{k=1}^{K+1} \psi_k^{(s_k - s_{k-1})/2} \mu_k^{-(s_k - s_{k-1})/2} \right) \exp \left\{ -\frac{1}{2} \sum_{k=1}^{K+1} \sum_{t=s_{k-1}+1}^{s_k} \frac{\gamma \psi_k (x_t - \mu_k)^2}{\mu_k} \right\}. \tag{14}$$

Muliplying the prior in (13) by the likelihood in (14) gives the joint posterior distribution, $p(\theta|x, s)$.

Marginal posterior distributions for parameters of interest, such as the current equity premium μ_{K+1}, are computed numerically using a Metropolis-Hastings algorithm. (The Appendix explains the implementation of that procedure.) The first two moments of the posterior of μ_{K+1}, for example, are estimated as the sample moments of a long series of draws from the posterior distribution of that parameter. The posterior mean provides a point estimate of μ_{K+1}, and the imprecision in that estimate is essentially the posterior standard deviation of μ_{K+1}.[9]

3. Empirical Analysis

3.1. 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 December 1996. The equity series from January 1926 to December 1996 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

[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).

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 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 (1939). Finally, the 1885–1925 data consist of returns on the Dow Jones index of industrial and railroad stocks, taken from Dow Jones (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, essentially because the earlier data do not appear to capture aggregate equity returns. Prior to 1834, the Smith and Cole index is 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 "Late 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%. 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 nineteenth century. Noting different 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 1996, the returns on a one-month Treasury security are taken from CRSP's SBBI file. For 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 non-existent.

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

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

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

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.

3.2. Specifying the Number and Locations of Breaks

This section reports results obtained in a Bayesian setting in which the number and the locations of the breaks are specified exogenously. Section 4 discusses potential extensions that incorporate uncertainty about the breaks. For the primary empirical analysis presented in this section, we specify breakpoint locations that correspond to maximum likelihood estimates. Maximum likelihood estimation of the breakpoints involves maximizing the value of the likelihood function in (2) for each permissible set of breakpoints. With a time series of T returns, there are $\binom{T}{K}$ permissible combinations of K breakpoints. If a break is allowed to occur in any of the $T = 1952$ months used in this study, the number of permissible breakpoint combinations becomes unmanageable even for modest values of K. With almost 2 million combinations for $K = 2$ and over 1 billion combinations for $K = 3$, unrestricted breakpoint estimation is computationally prohibitive. In order to work with models in which $K > 1$, we restrict the permissible break locations to a smaller subset. For $K \leq 3$, maximum likelihood estimation is manageable when breaks are permitted to occur at the end of any year, except that no breaks are allowed in the first three and last three years of the sample, and breaks must be at least three years apart.

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

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

In a model with three breaks, the estimated breakpoints occur at the year ends of 1928, 1940, and 1991. Figure 2 plots the time series of excess market returns for the entire sample period and indicates the locations of the three estimated breakpoints. Note that the first two breakpoints isolate the volatile years during and after the Great Depression, while the third breakpoint precedes the recent period of low volatility. Specifying at least three breaks in a sample beginning in 1834 seems reasonable, and the model with three breaks does have a higher Schwarz criterion (SC) than do models with fewer breaks.[15] With more than three breaks, maximum-likelihood estimation becomes computationally prohibitive when breaks are permitted in any year. To explore the effects of allowing additional breaks, we compute maximum likelihood estimates of breakpoints in a model with six breaks, except that breaks are allowed to occur only at the end of every fifth year, with no breaks in the first five and last five years of the sample period and with breaks at least ten years apart. In this six-break estimation, only three breaks are identified in the twentieth century, and the estimated breakpoints (1930, 1940, and 1990) correspond fairly closely to those in the three-break model. The three estimated breakpoints in the nineteenth century occur in 1855, 1865, and 1875. The subperiod defined by the first two breakpoints includes the Civil War, and the higher volatility during that subperiod is evident in Figure 2. Volatility appears to be lower following that subperiod than following the estimated breakpoint in 1875. Although the data from the nineteenth century can be important in estimating the current equity premium, as will be shown below, the results we report for the three-break model are unlikely to be influenced greatly by specifying additional breaks in that century. Thus, a three-break model appears to provide a tractable setting for obtaining results that are reasonably robust to allowing additional breaks.

3.3. Results When Large Shifts Are Believed To Be Unlikely

Table 1 reports posterior means and standard deviations of μ_4, the equity premium following the most recent (third) break. Results are shown for a range of values for σ_Δ, the prior standard deviation of the shift in the equity premium associated with a break. No prior link between the equity premium and volatility is imposed (i.e., $\nu = 0$). Part A of the table reports two limiting cases. When $\sigma_\Delta = \infty$, the posterior mean of μ_4 equals 9.76% per annum,

[15] This model-selection criterion is due to Schwarz (1978). The computed values of $\log \mathrm{SC}$ (divided by 1000) for a given number of breaks (K) are as follows:

K	0	1	2	3
log SC	3.112	3.132	3.247	3.249

the average excess return over the five-year period from 1992 through 1996.[16] In other words, when one believes that a structural break is likely to cause extremely large shifts in the equity premium, then the data before the most recent break play no role in estimating the current premium. The other limiting case in Part A is $\sigma_\Delta = 0$, which corresponds to a prior belief that structural breaks can cause shifts in volatility but not the equity premium. Inferences about the equity premium in that case, in broad terms, are based on a pooling of the data for the entire 163-year sample period. The posterior mean of 6.06% deviates somewhat from the simple arithmetic average of the excess returns over the entire sample, 5.51%, in essence because the sample averages from the four 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). Mean returns are estimated with less precision in the more volatile subperiods, and those subperiods are given less weight (much as in weighted least squares). When $\sigma_\Delta = \infty$, the posterior standard deviation of μ_4 is 3.62%, as compared to 1.16% when $\sigma_\Delta = 0$. Naturally, posterior uncertainty about the current equity premium is much lower when inferences are based on data from the entire 163-year sample as opposed to the most recent 5-year subperiod.

Results for intermediate values of σ_Δ are reported in the remainder of Table 1. The prior correlation between premiums in adjacent subperiods, ρ, is equal to 0 in Part B and approximately 1 in Part C. When $\sigma_\Delta = 2\%$, i.e., when shifts in the equity premium larger than 4% per annum are judged a priori to be unlikely, the posterior distribution of the current equity premium is affected considerably by the data from earlier subperiods. When $\rho = 0$, for example (Part B), the posterior mean of μ_4 is 6.81%, about 3% lower than the sample average for the last subperiod. Moreover, the posterior standard deviation of μ_4 in that case, 1.84%, is only about half of the value obtained when $\sigma_\Delta = \infty$, 3.62%. As shown later in this section, when the posterior mean of the current premium is based on the average return since 1926, a common practice, the posterior standard deviation of the premium is 2.26%. Thus, even though the most recent subperiod is only five years long, when shifts in the premium greater than 4% are believed unlikely, the current equity premium has a lower standard deviation than a seventy-year average. Even for $\sigma_\Delta = 4\%$, the posterior mean of μ_4 is still nearly 2% lower than the subperiod average, and the posterior standard deviation of the premium is only 2.56%.

When ρ is close to 1, the shift in the equity premium associated with a structural break has a prior mean near zero, conditional on the premium before the break. That is, in contrast to the prior in which $\rho = 0$, specifying ρ close to 1 corresponds to a prior belief that there

[16] The prior restriction that the premium is positive has a negligible effect in this period.

is only the weakest tendency for the equity premium to revert to a "long-run" value when a shift occurs. Results with such a prior are reported in Part C of Table 1. For a given σ_Δ, the posterior mean of μ_4 is higher than in Part B. For example, when $\sigma_\Delta = 2\%$, the posterior mean of μ_4 is 7.96%, about 1.2% higher than the corresponding value in Part B. This difference reflects the fact that the average equity premium is higher in the twentieth century than in the nineteenth century. With $\rho \approx 1$, the data from the more recent subperiods are given more weight, so the posterior mean of μ_4 is then farther from the mean for the overall sample period than when $\rho = 0$. Nevertheless, when $\sigma_\Delta = 2\%$, the posterior mean of μ_4 is still 1.8% below the average for the last subperiod.

3.4. Results When Volatility Is Believed To Play a Role

As the results in Table 1 demonstrate, 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 break can also be relevant if one believes that, across subperiods, the equity premium has at least some degree of positive association with stock-market volatility. The sample volatility of the excess market return in the recent subperiod (1992–96) is low by historical standards: 8.2% versus 17% for the overall sample period. With a prior belief in a link between the equity premium and volatility, an inference that recent volatility is lower than usual would tend to accompany an inference that the equity premium is also lower than usual. Of course, such inferences rely on data before the most recent break. If the data before the most recent break are discarded, then a prior belief in a link between the equity premium and volatility cannot by itself produce a posterior mean different from the sample average excess return, assuming prior beliefs about the price of risk remain noninformative, as in this study.

Recall that the price of risk, defined as the ratio of the equity premium to equity variance is equal to γ in the most recent subperiod and $\psi_k\gamma$ in each of the earlier subperiods, $k = 1, 2, 3$. The prior on γ is noninformative, but the prior on ψ_k is specified as a gamma density with parameters $\nu/2$ and $2/\nu$, and recall that this density becomes more tightly concentrated around 1 as ν increases (cf. Figure 1). Table 2 reports posterior means and standard deviations of μ_4 obtained with priors ranging from $\nu = 0.5$, which is close to the diffuse-prior value of $\nu = 0$ used in Table 1, up to $\nu = \infty$, which produces a prior spiked at $\psi_k = 1$ and corresponds to a perfect link between the equity premium and volatility. In all of the cases in Table 2, we specify $\sigma_\Delta = \infty$ in order to isolate the effect of a prior belief that volatility plays a role.

A striking feature of the results in Table 2 is that a prior belief in even a modest association between the equity premium and volatility produces a posterior mean for the current equity premium that is substantially below the average excess return in the last subperiod. With $\nu = 3$, for example, there is a 10% prior probability that that the price of risk in any of the other subperiods ($\psi_k \gamma$) is less than one fifth of the price of risk in the last subperiod (γ), and there is also a 10% prior probability that $\psi_k \gamma$ is more than twice the value of γ. Nevertheless, the posterior mean of μ_4 under that prior is only 4.71%, which is less than half of the average excess return in the last subperiod. If one believes in a perfect link between the equity premium and volatility ($\nu = \infty$), the estimate of the current equity premium is only 1.76%, due to the low equity volatility in the last subperiod.

The results in Table 2 clearly indicate that volatility can exert a strong effect on the posterior mean of the equity premium. As a result, the posterior uncertainty about the equity premium can also be substantially less than when inferences about the premium are based solely on average returns. Note, for example, that when ν exceeds 10, the posterior standard deviation of μ_4 can be considerably less than the value of 1.16% obtained when $\sigma_\Delta = 0$ and $\nu = 0$ (Table 1). That is, if the equity premium prevailing during the latest five-year subperiod is estimated by essentially averaging returns pooled over the entire 163-year sample period, inferences about the current premium can still be less precise than when that premium is allowed to differ from those in earlier subperiods but believed to obey a volatility link, albeit an imperfect one. Important to realize is that the increased information about the current premium provided by sample volatility comes only in the presence of structural breaks accompanied by differences in volatility across subperiods. Recall that the prior about the current price of risk, γ, is noninformative. Thus, with no structural breaks, or with constant volatility across subperiods, inferences about the equity premium could reflect only the information in sample average returns. We return to this point in the next subsection.

3.5. Additional Results: Combining Both Types of Prior Beliefs

Table 3 reports results with informative prior beliefs about both the equity premium's association with volatility as well as the potential magnitude of its shifts. The parameter σ_Δ, which determines prior beliefs about shifts in the premium (the Δ_k's), is assigned values 2%, 4%, and 6%, while the parameter ν, which determines the strength of the prior beliefs about the premium's association with volatility is assigned values 1, 5, and 30 (recall that a higher ν implies a stronger association). Part A of the table reports results with $\rho = 0$, and Part B reports results with ρ near 1. The effects of both types of prior beliefs are evident in Table

3. In all cases considered, the posterior mean of μ_4 is decreasing in ν, as in Table 2. For $\nu = 1$ and $\nu = 30$, the posterior mean of μ_4 is increasing in σ_Δ, as in Table 1 (where $\nu = 0$), but the mean of μ_4 is decreasing in σ_Δ when $\nu = 5$.

Table 4 displays results across all four subperiods for several combinations of σ_Δ and ν. In the interest of space, we report results only for $\rho = 0$. Part A displays the posterior means and standard deviations of the equity premium (μ_k), while Part B displays the posterior means and standard deviations of equity volatility (σ_k). Observe in Part A that, with the noninformative prior ($\sigma_\Delta = \infty$ and $\nu = 0$), the most recent subperiod has the highest mean equity premium. Hence, lowering σ_Δ, which assigns less prior probability to large shifts in the premium, tends to lower the posterior mean of the premium in that subperiod. Observe in Part B that the most recent subperiod also has the lowest volatility, so raising ν, which tightens the link between the premium and volatility, also tends to lower the mean of the premium in the last subperiod. Also evident in Part B is that, within each subperiod, equity volatility has low posterior uncertainty, and its posterior mean is affected only slightly by the prior parameters.

Recall from the earlier discussion that, as ν increases, the posterior standard deviation of the premium in the recent five-year subperiod can be less than the standard deviation of the premium obtained by averaging returns pooled over the entire 163-year sample period. Observe in Table 4, for example, that the standard deviation of the post-1991 premium is 1.16% with $\sigma_\Delta = 0$, the case in which the data are essentially pooled, but the post-1991 premium has a standard deviation of only 0.60% when a volatility link is imposed with $\nu = 30$. Note, however, that such a result does not obtain in the earlier three subperiods, wherein the posterior standard deviations all exceed 1.16%. The low standard deviation in the last subperiod stems from the low equity volatility in that subperiod. To see the basic point clearly, assume

$$\mu_k = \gamma\sigma_k^2, \quad k = 1, \ldots, K, \tag{15}$$

which is the perfect volatility link corresponding to $\nu = \infty$. Assume also that the posterior variance of σ_k^2 is zero, so that uncertainty about volatility is ignored. (The latter assumption is motivated by the observation that the available monthly data contain more precise information about volatility than about expected returns.) Then,

$$\text{Std}\{\mu_k | x, s\} = \sigma_k^2 \text{Std}\{\gamma | x, s\}, \tag{16}$$

where "Std" denotes standard deviation. In this simplified setting, the posterior standard deviation of the equity premium in a given subperiod is proportional to that subperiod's equity variance. To a rough approximation, the same reasoning applies to an imperfect volatility link, as when $\nu = 30$. In general, if one has noninformative beliefs about the price of risk (γ), then, in an overall sense, the precision of sample information about the equity premium cannot exceed that associated with a sample average return, but the presence of a volatility link can produce rather sharp differences in that precision across subperiods. If, instead, one has an informative prior belief about the price of risk, then the precision of the equity premium in even a high-volatility subperiod could be greater than the precision of the sample average for the overall pooled sample.

A comparison of models in a Bayesian framework is often based on the posterior odds ratio, which is the product of the prior odds ratio and the Bayes factor. For two alternative models, A and B, the Bayes factor is $\mathrm{BF}_{AB} = p(x|A)/p(x|B)$, where x denotes the data.[17] In a Bayesian setting, two models can differ solely in their prior densities for a common parameter vector, θ. That is, the models can share the same likelihood function, $p(x|\theta)$, in which case

$$\mathrm{BF}_{AB} = \frac{\int p(x|\theta)p_A(\theta)d\theta}{\int p(x|\theta)p_B(\theta)d\theta}, \tag{17}$$

where $p_A(\theta)$ and $p_B(\theta)$ denote the priors under each model. Table 5 presents Bayes factors for the three-break models considered above, where the models are distinguished solely by the values of the prior parameters σ_Δ and ν. Reported for each case are two values, arising from different methods of calculating the marginal likelihood, $\int p(x|\theta)p(\theta)d\theta$. (Details of the calculations are presented in the Appendix.) In all cases, "model B" is defined as the noninformative specification, $\sigma_\Delta = \infty$ and $\nu = 0$. With equal prior odds for that model versus an informative specification, a Bayes factor greater than one indicates that the data favor the informative specification, whereas a value less than one indicates the data favor the noninformative prior. Kass and Raftery (1995) suggest that Bayes factors between one and three are "not worth more than a bare mention," and most of the values in Table 5, or their reciprocals, lie within that range.

The Bayes factors for $\nu = 30$ are all 0.06 or less, so the data do not support that degree of positive association between the equity premium and volatility. Such an outcome is probably not surprising, given the patterns of average returns and volatilities across subperiods, as observed in Table 4. Recall, for example, that the 1992–96 subperiod has the highest average return but the lowest volatility. Thus, if one believes a priori that the equity premium and volatility are positively related, that belief is not strengthened by the data. On the other hand, a prior belief in a positive association seems reasonable (e.g., Merton (1980)), and moderate beliefs, as with $\nu = 5$, are not much at odds with the data: the (reciprocals of

[17]See, for example, Poirier (1995), p.380.

the) Bayes factors for $\nu = 5$ and $\rho = 0$ are between two and three for the finite values of σ_Δ. Recall that such beliefs can have very substantial effects on the posterior mean of the equity premium.

3.6. Results with a Single Break in 1925

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 and Ibbotson Associates. In essence, given the availability of the earlier data, using just the post-1925 data is equivalent to specifying a break in December 1925 and having noninformative priors about Δ_k and ψ_k. We investigate here the extent to which the equity premium estimated in such a manner is robust to informative prior beliefs about the potential magnitude of shifts in the premium or about the premium's association with volatility.

Table 6 reports posterior moments obtained using the same 1834–1996 excess-return series as before but in a model with only a single break, specified at December 1925. For this model, the results with $\rho \approx 1$ are nearly identical to those with $\rho = 0$, so only the latter results are reported. With non-informative priors ($\sigma_\Delta = \infty$ and $\nu = 0$), the equity premium for the post-1925 period has a posterior mean of 7.97%, similar to standard textbook values.[18] Because the posterior mean of the premium in the 1834–1925 subperiod is lower, 3.65% with noninformative priors, specifying an informative prior for Δ lowers the mean of the post-1925 equity premium compared to that produced with the noninformative prior. The mean equity premium is lower by nearly 1% with $\sigma_\Delta = 4\%$ and by nearly 2% with $\sigma_\Delta = 2\%$. Note that such results occur even though the post-break period is seventy years in length. Thus, the analogous results reported previously for the three-break model (Table 1) are not driven solely by the short length of the final subperiod in that case.

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 6 for $\sigma_\Delta = \infty$ and $\sigma_\Delta = 0$. On one hand, discarding the pre-1926 data produces an estimate closer to the posterior mean than does pooling the data if one's σ_Δ is 4% or more. On the other hand, pooling the data is the "lesser of the two

[18] For example, at several places in their popular text, Brealey and Meyers (1996) use an equity premium of 8.4%, which they report is an estimate based on the 1926–1994 period (p. 145).

evils" if σ_Δ is 2% or less. (The "breakeven" is in the neighborhood of $\sigma_\Delta = 3\%$.)

Prior beliefs about the volatility link play a less important role in this single-break example, chiefly because the estimated post-1925 volatility of about 19% is not much different from the pre-1926 volatility estimate of about 15%. Raising ν from 1 to 30 changes the mean of the post-1925 premium by less than 40 basis points.

4. Uncertainty About the Number and Locations of Breaks

In frequentist analyses of structural breaks, the typical approach involves estimating the break locations first and then estimating the remaining parameters of interest conditional on those break locations. Recent examples include Bai (1995, 1997), Bai and Perron (1998), and Liu, Wu, and Zidek (1997). A typical 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 a Bayesian approach, one can, in principle, integrate with respect to the uncertainty about the locations of the breakpoints, as discussed below.[19]

In order to extend the Bayesian framework in the previous sections to include uncertainty about break locations, we assume that, in the prior, the locations of the breakpoints are independent of μ and σ, the vectors containing the subperiod means and standard deviations. The prior probability that a random K-tuple of structural breaks, S, occurred at a particular combination of break locations, s, is denoted by

$$p(S = s) = p_s, \qquad \text{for all } s \text{ with strictly increasing elements.} \tag{18}$$

In other words, a prior probability p_s is assigned to every combination of K breakpoints, and the sum of p_s across all s must be equal to one. For example, if one believes that, with $K = 2$, the two breaks definitely occurred in 1855 and 1925, then $p_s = 1$ for that pair of breakpoints and $p_s = 0$ for all the other combinations. A noninformative prior on S can be imposed by setting each p_s to the reciprocal of the number of permissible breakpoint combinations. The joint posterior of θ and S is then given by

$$p(\theta, S = s|x) \propto p(x|\theta, S = s)p(\theta)p_s, \tag{19}$$

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

where the prior of θ and the likelihood function $p(x|\theta, S = s)$ are as given earlier. The marginal posterior probability $p(S = s|x)$ is obtained by integrating (19) with respect to θ, and the marginal posterior of θ, which accounts for the uncertainty about break locations, can be obtained by summing (19) over all permissible breakpoint combinations:

$$p(\theta|x) = \sum_s p(\theta|S = s, x)p(S = s|x). \tag{20}$$

Such computations can, in principle, be performed using essentially the same Metropolis-Hastings algorithm used to obtain the previously reported results, except that, in addition to drawing μ and σ, values of S must be drawn as well. Generating the latter draws is straightforward, since the full conditional distribution $p(S = s|\theta, x)$ is discrete.

Recall that, in computing maximum likelihood estimates of the breakpoints, we restrict breakpoints to occur only at year ends. For $K \leq 3$, the number of permissible breakpoints then becomes manageable, but our initial efforts using a noninformative prior for S reveal slow mixing of the Metropolis-Hastings Markov chain of draws from the conditional posterior distributions. We judge the computer time required for accurate computation of the posterior distribution to be unreasonable as of this writing, although one can imagine that advances in computing capabilities could soon make such an exercise feasible.

In order to illustrate the nature of the results that such an approach can provide, we present results for a model with a single break ($K = 1$), where the location of the break has a noninformative (flat) prior across all year ends in the overall 1834-1996 sample period. (Breaks in the first and last three years of the sample are prohibited.) Figure 3 displays marginal posterior distributions of the breakpoint. The distribution in the upper graph is obtained using the same form of prior on μ and σ used in the previous analysis, with the priors of Δ and ψ specified as noninformative ($\sigma_\Delta = \infty$ and $\nu = 0$). The posterior probability for the break location is rather concentrated in a few years beginning around 1940. As a result, uncertainty about the break location has only a modest effect on the estimate of the current equity premium. The estimate that accounts for break uncertainty, 8.46%, is close to 8.37%, the estimate obtained when the maximum likelihood estimate of the breakpoint is treated as the true value. The distribution in the lower graph of Figure 3 is obtained under the same specification, with the key exception that volatilities before and after the break are required to be the same ($\sigma_1 = \sigma_2$). Observe that, with the latter restriction, the posterior distribution of the breakpoint is quite disperse, covering essentially the entire sample period. When the single structural break is restricted to be purely a shift in the equity premium, with no change in volatility, the return history does not provide much information about the location of the breakpoint. The information in the data regarding a structural break evidently comes largely through a shift in volatility.

The methods discussed so far have conditioned on a specified value of K, as is often done. Even though the number of structural breaks is rarely known with certainty, a common frequentist approach is to select a model supported by one or more selection criteria and then proceed assuming that model is the true one.[20] Such an approach ignores the potentially substantial uncertainty about whether the selected model is indeed true. For example, if two competing models yield similar values of model-selection criteria, choosing just one of the models could discard useful information about the parameter of interest. Bayesian analysis can account for model uncertainty without forcing such a choice. In our setting, the unconditional posterior of θ can be obtained by weighting the posterior distributions of θ in models with different numbers of breaks by the posterior model probabilities:

$$p(\theta|x) = \sum_k p(\theta|K = k, x)p(K = k|x). \tag{21}$$

For example, the estimate of the current equity premium that accounts for model uncertainty is a weighted average of estimates from different models. For each k, the model probability is given by

$$p(K = k|x) = \frac{p(K = k)p(x|K = k)}{\sum_k p(K = k)p(x|K = k)}, \tag{22}$$

where $p(x|K = k)$ is the marginal likelihood, used earlier in obtaining Bayes factors for a given K, and $p(K = k)$ is the prior probability of a model with K breaks.

5. Conclusions

Even when a long return history contains structural breaks, the entire history can still contain information that is useful in estimating the current equity premium. If one believes that the shift in the equity premium accompanying a structural break is unlikely to be extremely large, then average returns in earlier subperiods contain information about the current premium. The earlier returns also provide information about the current premium if one believes there exists, to at least some degree, a positive association between the equity premium and volatility.

[20] Among the most common model-selection criteria are AIC, the Akaike information criterion (Akaike (1973)), and the previously discussed Schwarz criterion, SC. Yao (1988) establishes weak consistency of an estimator of the number of breaks based on SC. Christiano (1992) uses AIC and SC to estimate the number of breaks and obtains conflicting signals from the two criteria. Liu, Wu, and Zidek (1997) estimate the number of breaks using a modified SC.

We find that prior beliefs of either type can produce estimates of the current equity premium that are substantially less than the sample average for the subperiod following the most recent break, estimated to occur near the end of 1991. The average return for that recent subperiod is nearly 10%. If one believes, for example, that the premium is unlikely to shift by more than 4%, the estimate of the current premium lies between 6.8% and 8%, depending on one's belief about the persistence of the premium across breaks. Greater reductions are obtained if one believes there exists even a modest association between the equity premium and volatility. Suppose, for example, that, with an 80% prior probability, the price of risk before one or more structural breaks is greater than 0.2 times the current price of risk but less than 2 times the current value. The estimate of the current equity premium is then less than 5%. If, with the same 80% prior probability, the prices of risk in the earlier subperiods are between 0.5 and 1.6 times the current value, then the current equity premium is estimated to be only 2.4%. Moreover, in that case, the precision of the estimated premium is comparable to that of the sample average computed by pooling all of the data since 1834.

The widespread practice of averaging excess returns beginning in 1926 overstates the equity premium if one believes that large shifts in the premium are unlikely. On the other hand, simply averaging the data beginning in 1834 is probably too low an estimate, unless one believes that no shift in the premium occurred. If one believes that the pre-1926 equity premium could be lower than the more recent value, but probably not by more than, say, 4%, then the estimate of the current premium is about 2% less than the post-1925 average.

One area for potential extensions to the Bayesian framework in this study is incorporating uncertainty about the number and locations of breaks, as discussed in Section 4, but other issues also offer potentially interesting extensions. Perhaps chief among these is permitting conditional moments to depend on observable state variables. This study deals with what are presumed to be relatively infrequent breaks in the return-generating process, but the i.i.d. setting within each subperiod abstracts from any temporal variation in moments between the structural breaks. The conditional distribution of returns within each subperiod could instead depend on a set of state variables, which could be functions of lagged returns or other time series. Included in such extensions are models in which the conditional volatility fluctuates, producing fat-tailed marginal distributions of excess returns within the subperiods.

Extending the model to a multivariate setting offers another direction for future research. If, for example, structural breaks in the distribution of U.S. returns are associated with economic factors and events of international importance, then returns from a number of countries could exhibit contemporaneous structural breaks, and examining those return series in a multivatiate setting could yield more precise inferences about the parameters of interest and the number and locations of breaks. A multivariate approach to the structural-break problem is introduced by Bai, Lumsdaine, and Stock (1997), who develop frequentist confidence intervals for the location of a single common break in multiple time series. They find that a multivariate approach can yield substantial gains in inferences, and a multivariate approach could also be pursued in our Bayesian setting.

Appendix

Posteriors

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.[21] Repeated draws of model parameters from their full conditional distributions form a Markov chain of parameter draws. Beyond a burn-in stage, the elements in the chain are draws from the joint posterior distribution.

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 (13) and the likelihood function in (14):

$$\bar{\mu}|\cdot \sim N\left(\frac{\iota' V_\mu^{-1}\mu}{\iota' V_\mu^{-1}\iota}, \frac{1}{\iota' V_\mu^{-1}\iota}\right), \qquad 0 < \bar{\mu} < \kappa \tag{A.1}$$

$$\phi_k^2|\cdot \sim \frac{\nu + \sum_{t=s_{k-1}+1}^{s_k}(x_t - \mu_k)^2/(\lambda\mu_k)}{\chi^2_{s_k - s_{k-1}+\nu}}, \qquad k = 1, \ldots, K \tag{A.2}$$

$$\lambda|\cdot \sim \frac{\sum_{k=1}^{K+1}\sum_{t=s_{k-1}+1}^{s_k}(x_t - \mu_k)^2/(\phi_k^2\mu_k) + \epsilon}{\chi^2_{T+\epsilon}}, \tag{A.3}$$

$$\begin{aligned} p(\mu|\cdot) &\propto \left(\prod_{k=1}^{K+1}\mu_k^{-(s_k - s_{k-1})/2}\right) \\ &\quad \times \exp\left\{-\frac{1}{2}\left[(\mu - \bar{\mu}\,\iota)' V_\mu^{-1}(\mu - \bar{\mu}\,\iota) + \sum_{k=1}^{K+1}\sum_{t=s_{k-1}+1}^{s_k}(x_t - \mu_k)^2/(\lambda\phi_k^2\mu_k)\right]\right\}, \quad \mu > 0 \\ &\propto \left(\prod_{k=1}^{K+1}\mu_k^{-(s_k - s_{k-1})/2}\right) \\ &\quad \times \exp\left\{-\frac{1}{2}\left[\mu'(V_\mu^{-1} + D)\mu - 2\mu'(V_\mu^{-1}\bar{\mu}\iota + D\bar{x}) + \iota' D\overline{x^2}\right]\right\}, \quad \mu > 0, \end{aligned} \tag{A.4}$$

where D is a $(K+1) \times (K+1)$ diagonal matrix whose kth diagonal element is equal to

$$d_{kk} = \frac{s_k - s_{k-1}}{\lambda\mu_k\phi_k^2}, \tag{A.5}$$

[21] 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 as well as for a justification of its block-at-a-time version.

$\bar{x}$ is a $(K+1)\times 1$ vector whose kth element is

$$\bar{x}_k = \left(\sum_{t=s_{k-1}+1}^{s_k} x_t\right)/(s_k - s_{k-1}), \tag{A.6}$$

and $\overline{x^2}$ is a $(K+1)\times 1$ vector whose kth element is

$$\overline{x^2}_k = \left(\sum_{t=s_{k-1}+1}^{s_k} x_t^2\right)/(s_k - s_{k-1}). \tag{A.7}$$

Each of the full conditionals above, except for that of μ, 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 product of independent inverted gamma densities:

$$\left(\prod_{k=1}^{K+1} \mu_k^{-(s_k-s_{k-1})/2}\right) \exp\left\{-\frac{1}{2}\left[\iota' D(\overline{x^2}-a)\right]\right\}, \tag{A.8}$$

where a is a $(K+1)\times 1$ vector whose k-th element a_k is a fixed point of the contraction mapping,

$$a_k^{(i+1)} = \left[\left(\overline{x^2}_k - a_k^{(i)}\right)/(\lambda\phi_k^2)\right]^2, \tag{A.9}$$

where the sequence is initialized at $a_k^{(1)} = 0$. This adjustment is found to increase the acceptance rate in the chain, as opposed to setting $a = 0$, which appears more natural from (A.4). The acceptance rates range from around 25% to around 97%, depending on the values of the prior parameters.

The posterior moments of μ and σ are estimated across 10,000 posterior draws, obtained by retaining every 250th draw from a chain of 2,500,000 draws beyond the first 5,000 draws. In each case reported, the average of the 10,000 draws of the current equity premium, reported as the posterior mean, has a standard error less than 5 basis points (per annum). Since the Markov chain in this framework mixes fairly slowly, as many as 250 draws have to be skipped for each draw retained in order to obtain that degree of accuracy.

Marginal Likelihood Estimates

Recall that, in order to compute the Bayes factor for one model versus another, it is necessary to obtain estimates of the marginal likelihoods of the data under both models.[22] Let $p(x)$

[22] The material in this section draws heavily on the corresponding discussions in Newton and Raftery (1994) and Kass and Raftery (1995). See these two references for more details on the calculation of marginal

denote the marginal likelihood of the data under a particular model. Since the subsequent discussion pertains to that model, the dependence of the marginal likelihood on the model is suppressed for notational convenience. In general,

$$p(x) = \int p(x|\theta)p(\theta)d\theta, \tag{A.10}$$

where $p(x|\theta)$ is the likelihood function and $p(\theta)$ is the prior distribution over the set of model parameters θ. The marginal likelihood can therefore be viewed as the expected likelihood, where the expectation is taken with respect to the prior density. The simplest Monte Carlo estimate of $p(x)$ is then

$$\hat{p}_1(x) = \frac{1}{m}\sum_{i=1}^{m} p(x|\theta^{(i)}), \tag{A.11}$$

where $\left\{\theta^{(i)} :\ i = 1, \dots, m\right\}$ is a sample from the prior distribution. A major shortcoming of $\hat{p}_1(x)$ is that, if the data are informative relative to the prior, most of the draws of θ from the prior produce small likelihood values. Since $\hat{p}_1(x)$ is then dominated by a small number of large likelihood values, the variance of the marginal likelihood estimate is large, and the simulation can be quite inefficient.

One way to improve the precision of the simplest Monte Carlo estimate is to draw $\theta^{(i)}$ from a density $p^*(\theta)$ that is likely to produce higher likelihood values and then weight each likelihood value $p(x|\theta^{(i)})$ in (A.11) by $p(\theta^{(i)})/p^*(\theta^{(i)})$. This approach is known as importance sampling, and the density $p^*(\theta)$ from which the $\theta^{(i)}$'s are drawn is generally known as the importance sampling density.[23] If the data are informative relative to the prior, a good candidate for the importance sampling density seems to be the posterior density of θ. Using $p^*(\theta) = p(\theta|x)$ and simplifying the resulting estimate yields

$$\hat{p}_2(x) = \left\{\frac{1}{m}\sum_{i=1}^{m} p(x|\theta^{(i)})^{-1}\right\}^{-1}, \tag{A.12}$$

the harmonic mean of the likelihood values. As $m \to \infty$, $\hat{p}_2(x)$ converges almost surely to the correct value $p(x)$. The estimator $\hat{p}_2(x)$ does not, however, generally satisfy the Gaussian central limit theorem. Intuitively, drawing $\theta^{(i)}$ from the posterior can occasionally produce a value of $p(x|\theta^{(i)})$ close to zero, which has a large impact on $\hat{p}_2(x)$. Despite its instability, $\hat{p}_2(x)$ has been shown to produce accurate results in many cases.

In order to avoid the instability of the harmonic mean estimator in (A.12), Newton and Raftery (1994) recommend using a mixture of the prior and posterior densities as the

likelihood estimates.

[23] See, for example, Geweke (1989) for a discussion of importance sampling.

importance sampling density: $p^*(\theta) = \delta p(\theta) + (1-\delta)p(\theta|x)$, where $0 < \delta < 1$. The resulting estimator, $\hat{p}_3(x)$, is more stable than $\hat{p}_2(x)$ and satisfies the Gaussian central limit theorem. Newton and Raftery ultimately propose a modification of $\hat{p}_3(x)$ that avoids the need for generating draws from the prior. Instead, all m values of θ are drawn from the posterior, and it is imagined that a further $\delta m/(1-\delta)$ values of θ are drawn from the prior, all with likelihoods $p(x|\theta^{(i)})$ equal to their expected value $p(x)$. The resulting marginal likelihood estimate can be evaluated by an iterative scheme based on

$$\hat{p}_4(x) = \frac{\delta m/(1-\delta) + \sum_{i=1}^{m} p(x|\theta^{(i)})/\left\{\delta \hat{p}_4(x) + (1-\delta)p(x|\theta^{(i)})\right\}}{\delta m/(1-\delta)\hat{p}_4(x) + \sum_{i=1}^{m}\left\{\delta \hat{p}_4(x) + (1-\delta)p(x|\theta^{(i)})\right\}^{-1}}. \qquad \text{(A.13)}$$

The implementation of the scheme requires evaluating the likelihood at a large number of posterior draws. By increasing m, the marginal likelihood can be estimated to any degree of precision.[24] This study uses $m = 10,000$.

There seems to be no consensus in the statistics literature as to what value of δ in (A.13) produces the most efficient marginal likelihood estimates. On one hand, Newton and Raftery recommend small values of δ and report that $\hat{p}_4(x)$ is stable and performs very well even for $\delta = 0.01$. On the other hand, Rosenkrantz (1992) finds that values of δ close to 1.0 result in the best performance. This study reports results (Table 5) for both $\delta = 0.1$ and $\delta = 0.9$.

[24] Alternative ways of estimating marginal likelihoods are proposed by Chib (1995), Lewis and Raftery (1997), DiCiccio, Kass, Raftery, and Wasserman (1997), and others. The method proposed by Chib (1995) is feasible only if all integrating constants of the full conditional posterior distributions are known, which is not the case in our problem. Both the Laplace-Metropolis estimator proposed by Lewis and Raftery and the volume-corrected Laplace-Metropolis estimator proposed by DiCiccio et al. essentially approximate the posterior distribution of the parameters of interest by a normal distribution centered at the posterior mode. Another approximation to the marginal likelihood is based on the Schwarz criterion discussed earlier, although such an approximation does not take prior beliefs into account.

Table 1

Estimates of the Current Equity Premium with Priors About the Magnitude of Changes in the Premium Across Breaks

The table reports posterior means and standard deviations of the current equity premium (μ_4) in a model with $K = 3$ structural breaks, the latest of which occurs in December 1991. The overall period extends from January 1834 through December 1996, and the break locations are determined by maximum-likelihood estimation. 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_{k+1} denotes the equity premium following the k-th structural break. The k-th break is associated with a shift in the equity premium given by $\Delta_k = \mu_{k+1} - \mu_k$. For $k = 1 \ldots, K$, the prior standard deviation of Δ_k is σ_Δ (annualized in the table), and the prior correlation between μ_k and μ_{k+1} is equal to ρ. No prior link between the premium and volatility is imposed.

	Posterior moments of μ_4 (% per annum)	
σ_Δ (%)	Mean	Standard deviation
	A. Limiting cases	
0	6.06	1.16
∞	9.76	3.62
	B. $\rho = 0$	
2	6.81	1.84
4	7.87	2.56
6	8.62	2.98
10	9.31	3.28
	C. $\rho \approx 1$	
2	7.96	2.14
4	8.97	2.83
6	9.29	3.14
10	9.62	3.40

Table 2

Estimates of the Current Equity Premium with Priors About the Premium's Association with Volatility

The table reports posterior means and standard deviations of the current equity premium (μ_4) in a model with $K = 3$ structural breaks, the latest of which occurs in December 1991. The overall period extends from January 1834 through December 1996, and the break locations are determined by maximum-likelihood estimation. 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_{k+1} denotes the equity premium following the k-th structural break. The variance of the excess stock return following the k-th break is denoted by σ^2_{k+1}. In the period following the most recent break, the price of risk is defined as $\gamma = \mu_{K+1}/\sigma^2_{K+1}$. In each earlier period ($k = 1, \ldots, K$), the relation between the equity premium and variance is given by $\mu_k = \psi_k \gamma \sigma^2_k$, where the prior for each ψ_k is a gamma distribution with parameters $\nu/2$ and $2/\nu$. The k-th structural break is associated with a shift in the equity premium given by $\Delta_k = \mu_{k+1} - \mu_k$, and the prior standard deviation of Δ_k is set to infinity for $k = 1 \ldots, K$.

Prior for ψ_k			Posterior moments of μ_4 (% per annum)	
	Percentiles			Standard
ν	10%	90%	Mean	Deviation
0.5	0.00	3.00	8.88	3.70
1	0.02	2.71	7.94	3.69
2	0.10	2.30	6.03	3.38
3	0.20	2.08	4.71	2.87
5	0.32	1.85	3.26	1.81
10	0.49	1.60	2.40	1.00
30	0.69	1.34	2.00	0.65
∞	1.00	1.00	1.76	0.48

Table 3

Estimates of the Current Equity Premium with Priors About the Premium's Association with Volatility and About the Magnitude of Changes in the Premium Across Breaks

The table reports posterior means and standard deviations of the current equity premium (μ_4) in a model with $K = 3$ structural breaks, the latest of which occurs in December 1991. The overall period extends from January 1834 through December 1996, and the break locations are determined by maximum-likelihood estimation. 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_{k+1} denotes the equity premium following the k-th structural break. The k-th break is associated with a shift in the equity premium given by $\Delta_k = \mu_{k+1} - \mu_k$. For $k = 1 \ldots, K$, the prior standard deviation of Δ_k is σ_Δ (annualized in the table), and the prior correlation between μ_k and μ_{k+1} is equal to ρ. The variance of the excess stock return following the k-th break is denoted by σ^2_{k+1}. In the period following the most recent break, the price of risk is defined as $\gamma = \mu_{K+1}/\sigma^2_{K+1}$. In each earlier period ($k = 1, \ldots, K$), the relation between the equity premium and variance is given by $\mu_k = \psi_k \gamma \sigma^2_k$, where the prior for each ψ_k is a gamma distribution with parameters $\nu/2$ and $2/\nu$.

Parameters in the priors for Δ_k and ψ_k		Posterior moments of μ_4 (% per annum)	
σ_Δ (%)	ν	Mean	Standard Deviation
	A. $\rho = 0$		
2	1	6.48	1.86
2	5	4.49	2.00
2	30	1.32	0.63
4	1	6.95	2.69
4	5	3.58	1.96
4	30	1.60	0.60
6	1	7.39	3.04
6	5	3.43	1.93
6	30	1.72	0.60
	B. $\rho \approx 1$		
2	1	7.33	2.26
2	5	4.26	2.23
2	30	1.42	0.60
4	1	7.83	2.92
4	5	3.49	2.01
4	30	1.61	0.58
6	1	7.87	3.23
6	5	3.31	1.92
6	30	1.72	0.59

Table 4

Estimates of the Equity Premium and Volatility Across Subperiods

The table reports posterior moments of the equity premium and equity volatility in a model with $K = 3$ structural breaks. The overall period extends from January 1834 through December 1996, and the break locations are determined by maximum-likelihood estimation. 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_{k+1} denotes the equity premium following the k-th structural break. The k-th break is associated with a shift in the equity premium given by $\Delta_k = \mu_{k+1} - \mu_k$. For $k = 1 \ldots, K$, the prior standard deviation of Δ_k is σ_Δ (annualized in the table), and the prior correlation between μ_k and μ_{k+1} is equal to zero. The standard deviation (volatility) of the excess stock return following the k-th break is denoted by σ_{k+1}. In the period following the most recent break, the price of risk is defined as $\gamma = \mu_{K+1}/\sigma^2_{K+1}$. In each earlier period ($k = 1, \ldots, K$), the relation between the equity premium and variance is given by $\mu_k = \psi_k \gamma \sigma^2_k$, where the prior for each ψ_k is a gamma distribution with parameters $\nu/2$ and $2/\nu$.

Parameters in the priors for Δ_k and ψ_k		Period between breaks			
σ_Δ (%)	ν	1/1834–12/1928	1/1929–12/1940	1/1941–12/1991	1/1992–12/1996
A. Mean (standard deviation) of the equity premium (% per annum)					
0	0	6.06	6.06	6.06	6.06
		(1.16)	(1.16)	(1.16)	(1.16)
∞	0	4.17	8.83	8.25	9.76
		(1.55)	(6.54)	(2.01)	(3.62)
∞	∞	5.10	25.48	4.77	1.76
		(1.01)	(5.45)	(0.96)	(0.48)
4	1	4.89	7.10	7.78	6.95
		(1.39)	(2.96)	(1.78)	(2.69)
4	30	4.45	10.53	5.22	1.60
		(1.24)	(2.35)	(1.62)	(0.60)
B. Mean (standard deviation) of equity volatility (% per annum)					
∞	0	15.13	34.36	14.46	8.23
		(0.32)	(2.04)	(0.41)	(0.77)
4	30	15.14	32.85	14.60	8.98
		(0.32)	(1.82)	(0.41)	(0.87)

Table 5

Bayes Factors for Various Priors in the Three-Break Model

For various informative priors in a model with $K = 3$ structural breaks, the table reports Bayes factors relative to a model with noninformative priors. The overall period extends from January 1834 through December 1996, and the break locations are determined by maximum-likelihood estimation. 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_{k+1} denotes the equity premium following the k-th structural break. The k-th break is associated with a shift in the equity premium given by $\Delta_k = \mu_{k+1} - \mu_k$. For $k = 1 \ldots, K$, the prior standard deviation of Δ_k is σ_Δ (annualized in the table), and the prior correlation between μ_k and μ_{k+1} is equal to ρ. The standard deviation (volatility) of the excess stock return following the k-th break is denoted by σ_{k+1}. In the period following the most recent break, the price of risk is defined as $\gamma = \mu_{K+1}/\sigma^2_{K+1}$. In each earlier period $(k = 1, \ldots, K)$, the relation between the equity premium and variance is given by $\mu_k = \psi_k \gamma \sigma^2_k$, where the prior for each ψ_k is a gamma distribution with parameters $\nu/2$ and $2/\nu$. The first value reported is based on a 10-90 weighting of the prior and the posterior in forming the importance-sampling density, and the second value is based on a 90-10 weighting.

	ν			
σ_Δ	0	1	5	30
	A. Limiting cases for σ_Δ			
0	0.841	—	—	—
	0.430	—	—	—
∞	1	0.677	0.133	0.024
	1	0.771	0.161	0.026
	B. $\rho = 0$			
2	1.509	1.348	0.376	0.008
	1.064	0.954	0.400	0.023
4	1.963	1.494	0.411	0.039
	1.333	1.092	0.344	0.050
6	1.921	1.438	0.361	0.054
	1.301	1.081	0.293	0.052
	C. $\rho \approx 1$			
2	2.307	1.693	0.245	0.017
	1.543	1.336	0.343	0.030
4	2.180	1.790	0.325	0.048
	1.483	1.289	0.321	0.053
6	1.813	1.572	0.330	0.056
	1.359	1.197	0.284	0.055

Table 6

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

The table reports posterior moments of the equity premium and equity volatility 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), and the prior correlation between μ_1 and μ_2 is equal to zero. The standard deviations (volatilities) of the excess stock return before and after the break are denoted by σ_1 and σ_2. In the period following the break, the price of risk is defined as $\gamma = \mu_2/\sigma_2^2$. In the earlier period, the relation between the equity premium and variance is given by $\mu_1 = \psi\gamma\sigma_1^2$, where the prior for ψ is a gamma distribution with parameters $\nu/2$ and $2/\nu$.

Parameters in the priors for Δ and ψ		Posterior Mean		Posterior Standard Deviation	
σ_Δ (%)	ν	1834–1925	1926–1996	1834–1925	1926–1996
		A. Equity premium (% per annum)			
0	0	5.06	5.06	1.31	1.31
2	0	4.57	6.09	1.38	1.67
4	0	4.09	7.08	1.47	1.98
6	0	3.91	7.52	1.50	2.10
10	0	3.81	7.93	1.50	2.20
∞	0	3.65	7.97	1.51	2.26
2	1	4.55	6.09	1.37	1.65
2	5	4.52	6.25	1.32	1.59
2	30	4.42	6.47	1.22	1.54
4	1	4.13	7.07	1.42	1.95
4	5	4.24	7.13	1.32	1.85
4	30	4.42	7.16	1.15	1.73
∞	1	3.78	7.90	1.46	2.20
∞	5	4.01	7.82	1.31	2.09
∞	30	4.40	7.56	1.13	1.87
		B. Equity volatility (% per annum)			
∞	0	15.16	19.01	0.33	0.46
4	30	15.16	19.03	0.32	0.45

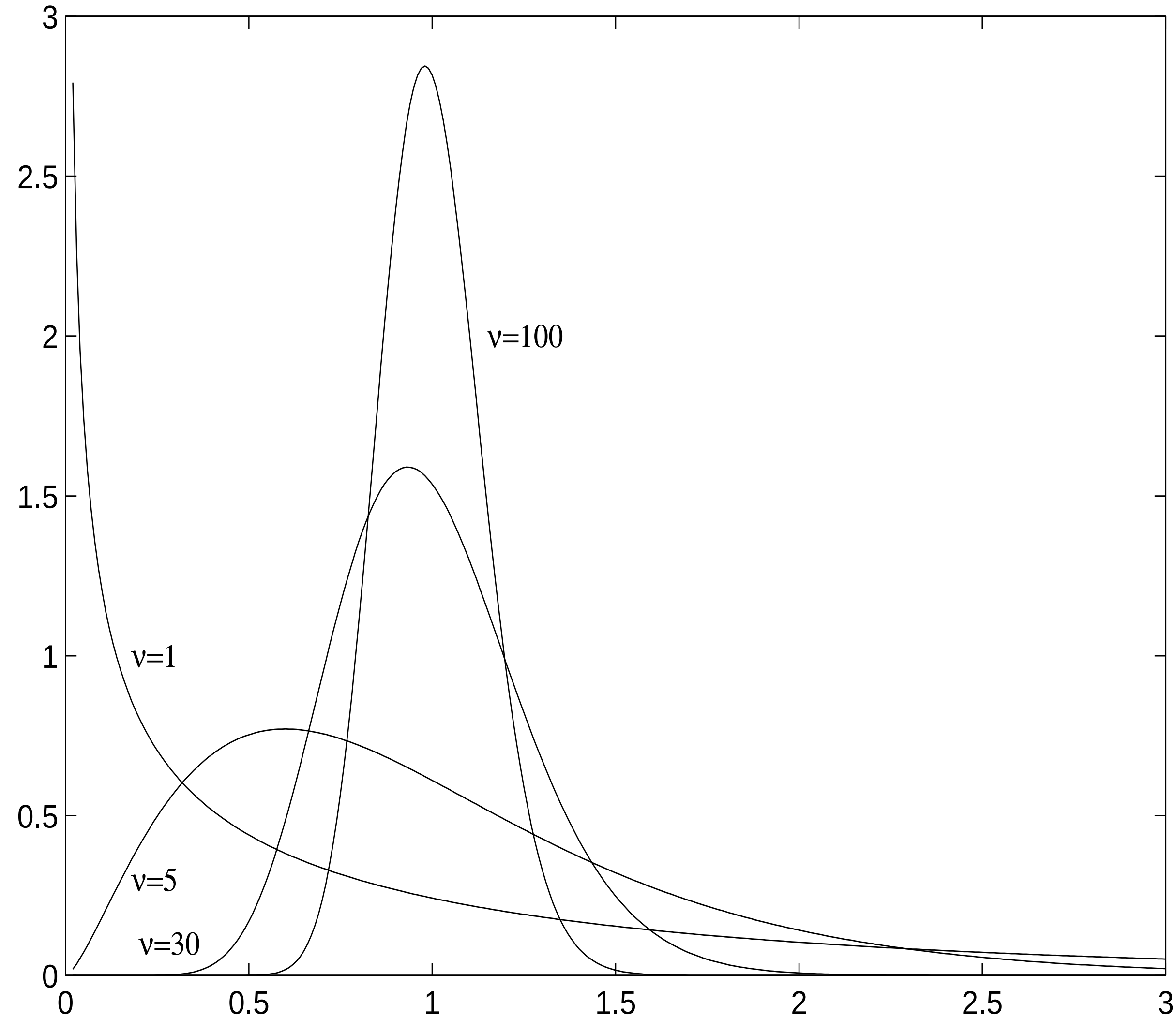

Figure 1. Prior densities of ψ_k for different values of ν. The prior density of ψ_k determines the strength of the link between the equity premium and variance. In the period following the most recent of K breaks, the price of risk is defined as $\gamma = \mu_{K+1}/\sigma^2_{K+1}$. In each earlier period, the relation between the equity premium and variance is given by $\mu_k = \psi_k \gamma \sigma^2_k, k = 1, \ldots, K$, where the prior for each ψ_k is a gamma distribution with parameters $\nu/2$ and $2/\nu$.

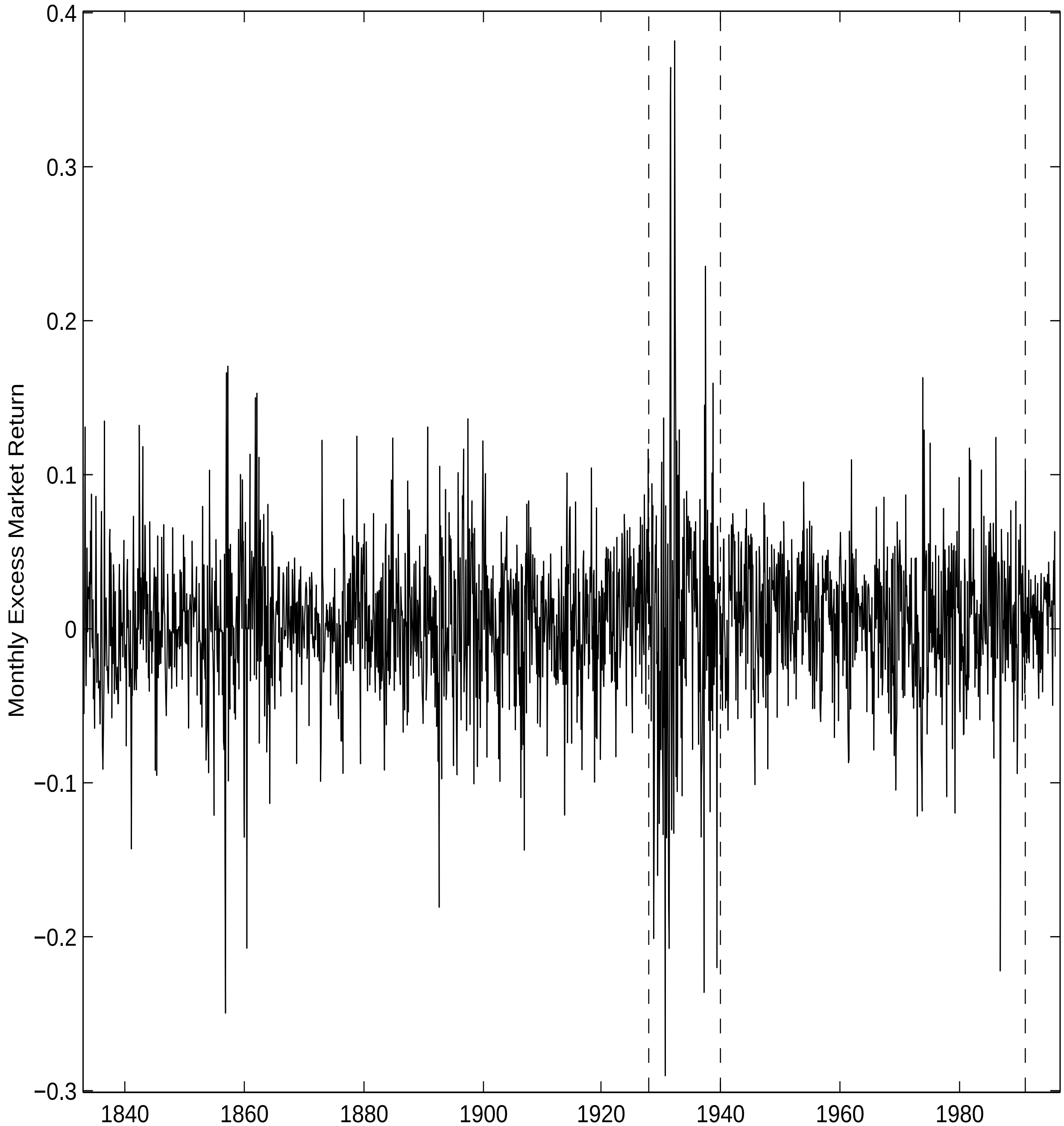

Figure 2. Excess market returns and structural breaks. The plot displays the time series of monthly excess returns on the U.S. stock market (solid line) from January 1834 through December 1996. The vertical dashed lines indicate the maximum-likelihood estimates of the breakpoints in a three-break model.

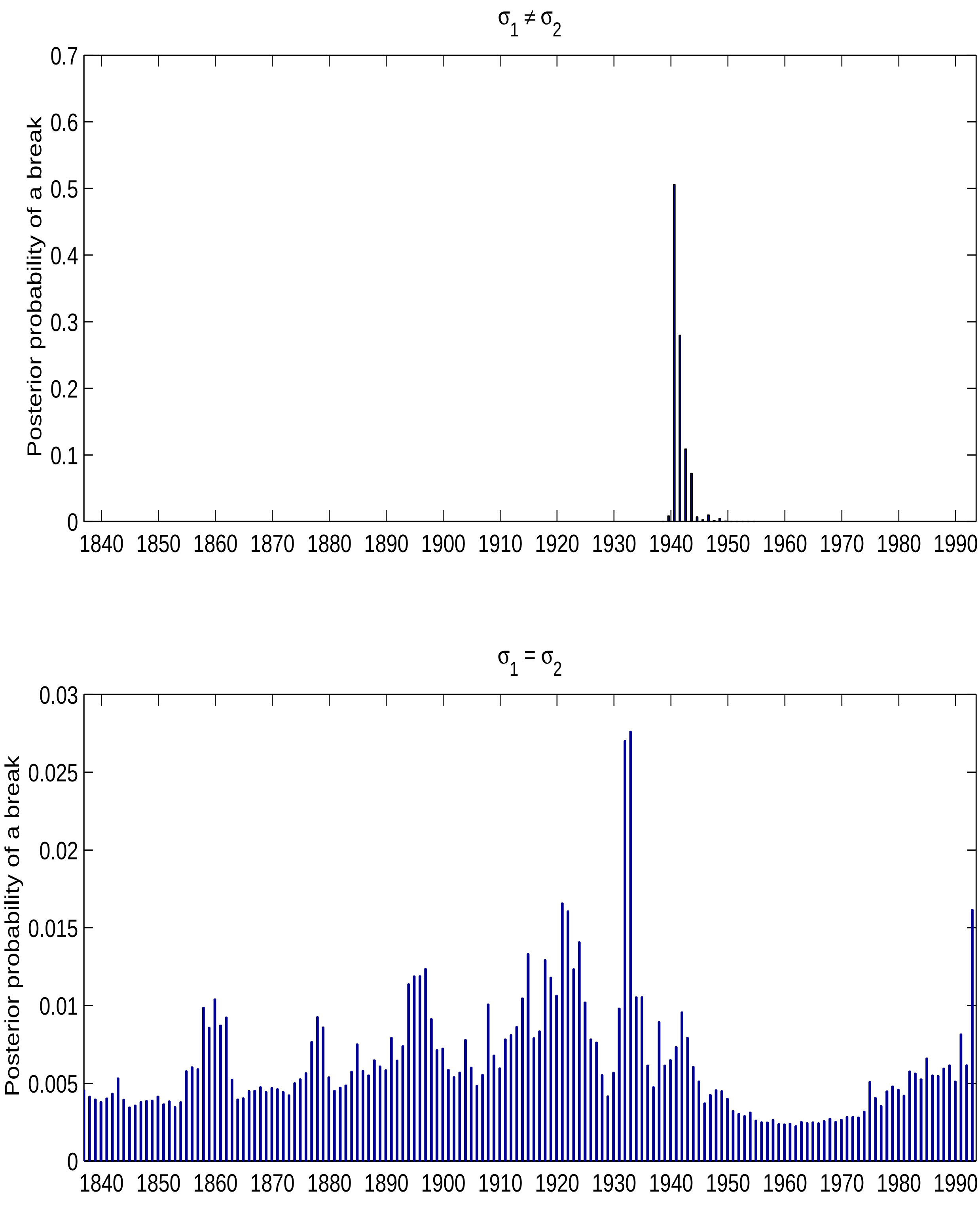

Figure 3. Posterior distributions for the location of a single break. The upper graph displays the posterior when volatility is permitted to shift at the break ($\sigma_1 \neq \sigma_2$), and the lower graph displays the posterior when volatility is restricted to be the same before and after the break ($\sigma_1 = \sigma_2$).

References

Akaike, H., 1973, Information theory and the extension of the maximum likelihood principle, in *Proceedings of the 2nd International Symposium on Information Theory*, ed. by B. N. Petrov and F. Csaki, Budapest: Akademi Kiado, 267–281.

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, D. W. K. and W. 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, J., 1995, Least absolute deviation estimation of a shift, *Econometric Theory* 11, 403–436.

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

Bai, Jushan, Robin L. Lumsdaine, and James H. Stock, 1997, Testing for and dating common breaks in multivariate time series, *Review of Economic Studies*, forthcoming.

Bai, J. and P. 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, NY).

Bhattacharya, P. K., 1994, Some aspects of change-point analysis, in *Change Point Problems, IMS Lecture Notes – Monograph Series*, Vol. 23, ed. by E. Carlstein, H.-G. Mueller, and D. Siegmund. Hayward, CA: Institute of Mathematical Statistics, 28–56.

Bigelow, Erastus B., 1862, *The tariff question...* Boston, MA.

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, Inc., New York.

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

Carlin, B. 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, *Journal of Financial Economics* 32, 137–167.

Chen, J. 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, 1995, Marginal likelihood from the Gibbs output, *Journal of the American Statistical Association* 90, 1313–1321.

Chib, Siddhartha, 1997, Estimation and comparison of multiple change point models, working paper, Washington University

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

Christiano, Lawrence J., 1992, Searching for a break in GNP, *Journal of Business and Economic Statistics* 10, 237–250.

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

DiCiccio, Thomas J., Robert E. Kass, Adrian E. Raftery, and Larry Wasserman, 1997, Computing Bayes factors by combining simulation and asymptotic approximations, *Journal of the American Statistical Association* 92, 903–915.

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

Dow Jones and Co., 1972, *The Dow Jones Averages, 1885–1970*, ed. by Maurice L. Farrell, New York: Dow Jones.

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

Geweke, John, 1989, Bayesian inference in econometric models using Monte Carlo integration, *Econometrica* 57, 1317–1339.

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, J. 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, S. and R.F. Stambaugh, 1990, Expectations and volatility of consumption and asset returns, *Review of Financial Studies* 3, 207–232.

Kass, Robert E., and Adrian E. Raftery, 1995, Bayes factors, *Journal of the American Statistical Association* 90, 773–795.

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

Lewis, Steven M. and Adrian E. Raftery, 1997, Estimating Bayes factors via posterior simulation with the Laplace-Metropolis estimator, *Journal of the American Statistical Association* 92, 648–655.

Liu, J., S. Wu, and J. 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*, New York: NBER.

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.

Newton, Michael A. and Adrian E. Raftery, 1994, Approximate Bayesian inference with the weighted likelihood bootstrap, *Journal of the Royal Statistical Society* B, 3–48.

Pástor, Ľuboš and Robert F. Stambaugh, 1998, Costs of equity capital and model mispricing, *Journal of Finance*, forthcoming.

Poirier, Dale J., 1995, *Intermediate Statistics and Econometrics* (MIT Press, Cambridge, MA).

Rosenkrantz, S., 1992, *The Bayes factor for model evaluation in a hierarchical Poisson model for area counts*, PhD dissertation, Department of Biostatistics, University of Washington, Seattle, WA.

Schwarz, Gideon, 1978, Estimating the dimension of a model, *Annals of Statistics* 6, 461–464.

Schwert, G. William, 1989, Business cycles, financial crises and stock volatility, *Carnegie-Rochester Conference Series on Public Policy* 31 (Autumn): 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*, Cambridge, Mass.: Harvard University Press.

Sowell, F., 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.

Yao, Yi-Ching, 1988, Estimating the number of change-points via Schwarz' criterion, *Statistics and Probability Letters* 6, 181–189.

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

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