Title
Structural Breaks, Incomplete Information and Stock Prices

Author
Timmermann, Allan

Publication Date
2001

2001-02

UNIVERSITY OF CALIFORNIA, SAN DIEGO

DEPARTMENT OF ECONOMICS

STRUCTURAL BREAKS, INCOMPLETE INFORMATION
AND STOCK PRICES

BY

ALLAN TIMMERMANN

DISCUSSION PAPER 2001-02
JANUARY 2001

Structural Breaks, Incomplete Information and Stock Prices*

Allan Timmermann
University of California San Diego[†]

August 4, 2000

Abstract

This paper presents empirical evidence on the existence of structural breaks in the fundamentals process underlying US stock prices. We develop an asset pricing model that represents breaks in the context of a Markov switching process with an expanding set of non-recurring states. Different hypotheses on how investors form expectations about future dividends after a break are proposed and analyzed. A model in which investors do not have full information about the parameters of the dividend process but gradually update their beliefs as new information arrives is shown to induce skewness, kurtosis, volatility clustering and serial correlation in stock returns after a break.

*I am grateful to Frank Diebold, Graham Elliott, Rob Engle, Eric Ghysels, Massimo Guidolin, Jim Hamilton, Simon Potter, Garey Ramey, Doug Steigerwald and seminar participants at UCSD, UCLA, UC Riverside, Penn State, University of Pennsylvania, Caltech, the Econometric Society Meetings in Pasadena (June 1997) and Canberra (July 1998) for discussions and comments on the paper. The comments of two anonymous referees were also very helpful in revising the paper.

[†]Department of Economics UCSD, 9500 Gilman Drive, La Jolla CA 92093-0508. E-mail: atimmerm@ucsd.edu.

1. Introduction

Is the fundamentals process underlying US stock prices stable over several decades? This stability assumption is implicitly made in the vast majority of papers in the empirical asset pricing literature that tests present value models. Recent studies have questioned this assumption, however. Discussing the mean return on US stocks since 1926, Brennan (1997) argues that "... there are good reasons to doubt that this parameter has remained constant for almost three quarters of a century which has witnessed the most dramatic economic, technological and social change of any comparable period in history" (Brennan, page 5). Observations like these suggest that a full understanding of asset prices requires careful consideration of the stability of the underlying fundamentals process.

This paper proposes a new approach to modeling stock market prices which links structural breaks in the underlying dividend process with the assumption that investors have imperfect information about the new dividend growth rate after a break. Our approach is based on new empirical tests which suggest that there are multiple breaks in the fundamentals process underlying US stock prices and the paper considers their importance in the context of an asset pricing model.

Structural breaks in the dividend process, if present, can affect stock prices in two important ways. First, like any shock to the endowment process, breaks will affect future dividends. The main difference between breaks and ordinary shocks to dividends is that the former are low frequency events that lead to rare level shifts in dividends which remain in effect for a long time. This is the 'persistence' effect of breaks.

Breaks also give rise to an information effect which concerns how much information investors have and how they revise their expectations about future dividends following a structural break. Under full information, investors instantaneously observe the new parameters of the dividend process after each break. While this is an important benchmark, it seems far from empirically plausible. Episodes linked to breaks in the dividend process, such as the Great Depression or the two world wars, were associated with substantial uncertainty over future prospects of the economy. Such uncertainties, we argue, can better be modeled by assuming that investors have incomplete knowledge of the new dividend growth rate and undertake a recursive updating process which gradually provides them with more precise growth estimates as new data emerges.

This imperfect information hypothesis has important empirical implications. In the period following a structural break investors cannot rely on historical data to produce an estimate of the new mean dividend growth rate. Large revisions in investors' parameter estimates are more likely to occur immediately after such breaks since the 'learning clock' runs fast and this produces a clustering in the volatility of asset prices through their dependence on investors' beliefs. Under full information a break in the dividend process will only show up as a single outlier in the return distribution in the period where the break occurs. The combination of multiple breaks in the fundamentals process and endogenous learning effects is important. Genotte (1986) calls for a model where the market's learning affects the underlying return process, while Lewis (1989) conjectures that multiple break points are needed to better explain movements in asset prices. Our analysis incorporates both of these elements.

Earlier studies such as Cecchetti, Lam & Mark (1990) and Veronesi (1999) have modeled instabilities in the fundamentals process in the context of switches between two recurring states in the drift of US dividends. By assuming that states repeat and that switches do not represent clean breaks with the past, investors in these models can use historical information to update their beliefs, although they face the filtering problem associated with identifying the underlying, but unknown, state. Our approach models the dividend process as a Markov switching model with an expanding set of non-recurring states. Each time a break occurs, the new state is characterized by a different set of parameter values. Consequently, revisions in investors' parameter estimates and the volatility of asset prices will be greater immediately after a break.

The two approaches are not mutually exclusive. The assumption of repeated states is particularly appealing at the business cycle frequency, c.f. Perez-Quiros & Timmermann (2000) and Veronesi (1999). But this does not rule out the presence of breaks in the underlying dividends that occur at a much lower frequency. Indeed while there are eight official post-war recessions, our empirical analysis only identifies one or two post-war breaks. However, although structural breaks are rare, we find that they can have very important effects on the moments of stock returns.

The plan of the paper is as follows. Section 2 presents empirical evidence on breaks in the dividend process underlying US stocks while Section 3 develops an asset pricing model under such breaks. Incomplete information and recursive learning effects are introduced in Section 4. Section 5 reports results from simulations of the model under complete and incomplete information and compares the simulations to actual data on US stock returns. Section 6 concludes.

2. Structural Breaks in Fundamentals: Empirical Evidence

To formally test for breaks in the endowment process we present the outcome of two different econometric approaches. The first approach considers breaks as deterministic events and hence does not make any assumptions about the distribution from which the breaks were drawn. The advantage of not imposing a probability model for the breaks is the resulting robustness with respect to possible misspecification of such a "meta model".

We also consider a Markov switching approach to break point estimation. In contrast to the nonstationary model which regards breaks as deterministic events, this approach views breaks as draws from a stationary meta distribution. This distinction is very important from a theoretical perspective. Stock prices depend on expectations of all future dividend growth rates. Therefore they can only be modeled under some assumption about the distribution from which the new growth rate after a break is obtained.

We apply both breakpoint methods to investigate the presence of breaks in the US dividend series provided by Shiller (2000). This data consists of monthly dividends from 1871-1999 giving a total of 1548 observations. Dividends are scaled by the consumer price index to get real dividends, D_t, and we model the change in the logarithm of these to get real dividend growth rates, $d_t = \Delta \log(D_t)$.

2.1. *Deterministic Breaks*

First consider the procedure for consistent estimation of multiple breakpoints in linear regression models developed by Bai & Perron (1998). Let $\mathbf{x}_t$ be a vector of factors whose coefficients in the linear regression of d_t on $\mathbf{x}_t$ change at m discrete (break) points in time:

$$
\begin{array}{ll}
d_t = \mathbf{x}_t^{'}\boldsymbol{\delta}_1 + u_t & t = 1, 2, ...T_1 \\
d_t = \mathbf{x}_t^{'}\boldsymbol{\delta}_2 + u_t & t = T_1 + 1, ...T_2 \\
\ldots\ldots\ldots\ldots\ldots\ldots & \ldots\ldots\ldots\ldots \\
d_t = \mathbf{x}_t^{'}\boldsymbol{\delta}_{m+1} + u_t & t = T_m + 1, ...T.
\end{array} \tag{1}
$$

Here T is the sample size, $T_1 < T_2 < ... < T_m < T$ and u_t is a disturbance term. Bai and Perron develop tests for the consistent estimation of the number and location of breakpoints (T_1,...., T_m) and the parameters (δ_1'**,....,** δ_{m+1}').

Currently available econometric techniques do not facilitate consistent estimation of multiple unknown breaks in the variance of a process. To capture a possible break in the volatility of the dividend process we also consider the absolute value of the dividend growth rate, $|d_t|$.

Table 1 reports the number of break points identified by three separate criteria. A Gauss program provided by Bai and Perron was used in the estimations. The maximum number of breakpoints was set to eight and we allowed for heteroskedasticity in the residuals. The sequential break point test uses a significance level of five percent, while the two information criteria are based on the penalized likelihood function. In the absence of a well established structural model for the dividend process, we test for breaks in models with an intercept and a first-order autoregressive representation. These capture the essential dynamics of the dividend growth process. The number of breaks identified by the tests varies considerably depending on which test is used. While the AIC always chooses the maximum number of breaks permitted (eight), the sequential approach chooses between zero and six breaks. However, even for the specifications without a lag where the sequential approach chooses zero or one break, the battery of additional tests described by Bai and Perron (1998) strongly indicate the presence of breaks. For instance, in the model for d_t with only an intercept term, the SupF test rejects the null of zero versus seven breaks at the 1% level. When adopted sequentially it also rejects the null of 2 versus 3 breaks. Consistent with this, both the $UD\max$ and $WD\max$ tests are significant at the 2.5% level.

The BIC results are somewhat more consistent across model specifications and lead to four breaks for three of the four models. The break dates for the models selected by BIC are identified around 1920, 1931, 1938 and 1950. Breaks are thus associated with the period after WWI, the Great Depression, and the beginning and end of WWII.

To illustrate the extent of the parameter variation, and to make our results comparable to those from the Markov switching specification, we report the parameter estimates and their standard errors from the simplest model with an intercept, μ, in the dividend regression:

$$d_t = \begin{array}{lll} \widehat{\mu} & (s.e.) & \text{interval} \\ 0.19 & (0.12) & 1871{:}1 - 1911{:}5 \\ -0.71 & (0.29) & 1911{:}6 - 1920{:}5 \\ 0.71 & (0.10) & 1920{:}6 - 1930{:}11 \\ -0.53 & (1.32) & 1930{:}12 - 1938{:}11 \\ 0.14 & (0.07) & 1938{:}12 - 1999{:}12 \end{array}. \tag{2}$$

For the volatility proxy we have

$$|d_t| = \begin{array}{lll} \mu & (s.e.) & \text{interval} \\ 1.42 & (0.07) & 1871{:}1 - 1921{:}1 \\ 0.85 & (0.06) & 1921{:}2 - 1931{:}12 \\ 2.12 & (0.43) & 1932{:}1 - 1939{:}11 \\ 1.11 & (0.15) & 1939{:}12 - 1951{:}11 \\ 0.39 & (0.03) & 1951{:}12 - 1999{:}12 \end{array} \tag{3}$$

Volatility is exceptionally high during the Great Depression and becomes far smaller than at any other time during the post-war sample.

2.2. *A Non-recurring State Model*

An obvious alternative to the structural break interpretation of the dividend growth series is to consider the data as the outcome of either a finite-state or an expanding state Markov switching process. Recurring state models have been successfully used to capture repeated patterns in asset returns at the cyclical frequency, c.f. Veronesi (1999) and Perez-Quiros & Timmermann (2000), or at the somewhat higher frequency of high and low volatility episodes.

In this paper we are instead concerned with modelling low frequency breaks in the endowment process that are unlikely to be repeated and hence cannot be well represented by a recurring state model. To represent breaks in the context of a Markov switching model, we build on the approach of Chib (1998) towards change points. The mean and variance of the growth in the dividend process are driven by a latent state variable, s_t:

$$d_t = \mu_{s_t} + \sigma_{s_t}\varepsilon_t, \qquad s_t = 1, ..., n_s \tag{4}$$

where n_s is the number of states. We do not include any lags in this specification since we are interested in testing for breaks in the mean growth rate, while allowing also its volatility to differ across regimes. This allows us to keep low the number of parameters that have to be estimated, an important consideration for large values of n_s. Movements across non-recurring states are controlled by the state transition probability matrix

$$\Pi = \begin{pmatrix} p_{11} & 0 & 0 & ... & 0 \\ 1-p_{11} & p_{22} & 0 & ... & 0 \\ \vdots & 1-p_{22} & \vdots & \vdots & \vdots \\ ... & \vdots & 0 & p_{n_s-1,n_s-1} & 0 \\ 0 & 0 & ... & 1-p_{n_s-1,n_s-1} & 1 \end{pmatrix}, \tag{5}$$

where $p_{ji} = \Pr(s_{t+1} = j|s_t = i)$. Notice that s_t can either remain at its current value (i) or move to its subsequent value $(i+1)$. Since we are conditioning on the existence of n_s states, the stayer probability of the terminal state is set to one. However, unconditionally, as the sample size, T, goes to infinity, the number of states will also increase. An advantage of this specification is that the number of parameters that have to be estimated is $3n_s - 1$. This only grows linearly with n_s. With a recurring state specification Π is given as

$$\Pi = \begin{pmatrix} p_{11} & p_{12} & p_{13} & \cdots & p_{1n_s} \\ p_{21} & p_{22} & p_{23} & \cdots & p_{2n_s} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ ... & \vdots & 0 & p_{n_s-1,n_s-1} & p_{n_s,n_s-1} \\ p_{n_s1} & p_{n_s2} & \cdots & 0 & p_{n_s,n_s} \end{pmatrix}, \tag{6}$$

so the number of parameters that have to be estimated, $n_s(n_s+1)$, grows quadratically in n_s.

Conditional on being in a given state the density of the dividend growth process is assumed to be Gaussian with state-specific mean (μ_j) and volatility (σ_j):

$$f(d_t|s_t = j) = \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp(\frac{-(d_t - \mu_j)^2}{2\sigma_j^2}), \qquad j = 1, ..., n_s.$$

We estimate this model by maximum likelihood conditional on a given value of n_s, the number of states, and then apply BIC to select the best model across different values of n_s. Panel A in Table 2 presents the outcome of this exercise. A choice of $n_s = 7$ maximizes the BIC. For this choice of n_s the values of the parameters estimates shown in Panel B of Table 2 are sufficiently different to justify the use of a multi-state model. To facilitate interpretation of the states and to identify the switches between them, Figure 1 plots the filtered probabilities of being in states one to seven conditional on the full-sample parameter estimates and period t information, $\Pr(s_t|\widehat{\theta}_T, \Omega_t)$. An even clearer separation of states occurs from Figure 2 which plots the full-sample smoothed state probabilities, $\Pr(s_t|\widehat{\theta}_T, \Omega_T)$, using the algorithm of Kim (1994). The identified states span the following periods: 1871-1919, 1919-1920, 1920-1931, 1931-1951, 1952-1966, 1966-1975, 1975-2000.

Since the two-state Markov switching model with recurring states has become very popular in the empirical literature, it is also of interest to compare this model with the results in Table 2 and Figures 1 and 2. In early work, Kandel & Stambaugh (1991) represent the mean and volatility of the logarithm of the consumption growth rate as a recurring state model. The model analyzed by Cecchetti et al. (1990) and Veronesi (1999) assumes two drift parameters and a single volatility parameter. This restricted model generates values of the log-likelihood function and the BIC of 4497.2 and -8957.8, respectively. These values are far below that of the seven state model with non-recurring states. Allowing σ to vary with the state does not change this conclusion. Another indication of the value of adopting a break point approach to identify a low frequency component in the dividend growth process comes from the plot of the two-state model's state probabilities which are shown in Figure 3. The first state is predominantly a post-1960 state and vice versa with state 2, indicating a fundamental change in the post-war data.

3. Stock Prices Under Breaks in the Dividend Process: A Theoretical Model

The empirical evidence in Section 2 suggests that a model of US stock prices must account for multiple breaks in fundamentals. Common to the Great Depression and the world wars is that these events were rapidly recognized once they had occurred. Furthermore, these events appear to be sufficiently unique to make it unlikely that they are repeated draws from the same (two-state) switching process.

In this section we propose a simple asset pricing model based on the Markov switching model with an expanding set of non-recurring states. To acknowledge the uniqueness of the breaks we instead assume that, after each break, the parameters of the dividend process are drawn from a continuous distribution. This also guarantees that parameter uncertainty will not be eliminated even asymptotically. If the data were generated by a finite state Markov process, investors would eventually learn the parameter values arbitrarily well, although of course they need not know the true state of the economy which gives rise to a filtering problem.

Stock prices, P_t, are assumed to be determined by a present value relation based on a representative investor with power utility

$$u(C_t) = \begin{cases} \frac{C_t^{1-\alpha}-1}{1-\alpha} & \alpha \geq 0 \\ \ln C_t & \alpha = 1 \end{cases}. \tag{7}$$

Here C_t is real consumption at time t, and α is the coefficient of relative risk aversion. Standard equilibrium asset pricing models (e.g., Lucas (1978)) assume that non-storable dividends from a single endowment source are the economy's only source of income, i.e., $C_{t+i} = D_{t+i}$. Subject to a budget constraint, the representative agent chooses stock holdings to maximize the discounted value of expected future utilities from consumption, $E_t\left[\sum_{k=0}^{\infty} \beta^k u(D_{t+k})\right]$, where $\beta = \frac{1}{1+\rho}$ and ρ is the rate of impatience. This yields the following Euler equation

$$P_t = \beta E_t[(P_{t+1} + D_{t+1}) \left(\frac{D_{t+1}}{D_t}\right)^{-\alpha}], \tag{8}$$

where E_t is the expectation operator conditional on the information set at time t. Fluctuations in stock prices are driven by shocks to dividends and revisions in investors' beliefs.

Real dividends (D_t) are assumed to follow a geometric random walk process:

$$\ln(D_{t+1}) = \ln(D_t) + \mu_{t+1} + \sigma_{t+1}\varepsilon_{t+1}, \tag{9}$$

where μ_{t+1} is a drift term, σ_{t+1} is the volatility parameter and $\varepsilon_{t+1} \sim N(0,1)$ is a normally distributed innovation term. Consistent with the change point model proposed in Section 2, let s_{t+1} be a 'break indicator' such that $s_{t+1} = s_t$ implies that there is no switch between states in the dividend process, while if $s_{t+1} = s_t + 1$, a break has occurred in period $t+1$. Also let $\Pr(s_{t+1} = s_t|\varepsilon_{t+1}) = \pi_t$ and $\Pr(s_{t+1} = s_t + 1|\varepsilon_{t+1}) = 1 - \pi_t$, be the probabilities of no switch and a switch from the state prevailing at time t, respectively, for all possible realisations of ε_{t+1}. π_t is then the diagonal element of the transition probability matrix, Π in (5). Finally assume that the process for s_{t+1} is independently and identically distributed and also independent of the ε's. The process for $(\mu_{t+1}, \sigma^2_{t+1})$ is given by

$$\begin{array}{lll} \Pr(\mu_{t+1} = \mu_t, \sigma^2_{t+1} = \sigma^2_t | s_{t+1} = s_t) & = 1 & \text{(no break)} \\ \Pr(\mu_{t+1} \leq \overline{\mu}, \sigma^2_{t+1} \leq \overline{\sigma}^2 | s_{t+1} = s_t + 1) & = H(\overline{\mu}, \overline{\sigma}^2) & \text{(break)} \end{array} \tag{10}$$

where $H(.,.)$ is the bivariate cumulative density function for the new values of μ_{t+1} and σ^2_{t+1}. Under these assumptions we have $E_t[D_{t+1}/D_t|s_{t+1} = s_t] = \exp(\mu_t + \sigma^2_t/2) \equiv (1 + g_t)$, one plus the mean growth rate conditional on no break in the dividend process.

To simplify, we assume that each time a break occurs, the new dividend growth rate, $g_{t+1} \equiv \exp(\mu_{t+1} + \sigma^2_{t+1}) - 1$, is drawn from a univariate density, $G(g_{t+1})$ defined on the support $[\underline{g}, \overline{g}]$. For example, if the density is uniform, $U(g_{t+1})$, equation (10) becomes

$$\begin{array}{lll} \Pr(\exp(\mu_{t+1} + \frac{\sigma^2_{t+1}}{2}) = \exp(\mu_t + \frac{\sigma^2_t}{2}) | s_{t+1} = s_t) & = 1 & \text{(no break)} \\ \Pr(\exp(\mu_{t+1} + \frac{\sigma^2_{t+1}}{2}) \leq 1 + g | s_{t+1} = s_t + 1) & = \frac{g - \underline{g}}{\overline{g} - \underline{g}} & \text{(break)} \end{array} \tag{11}$$

for all $g \in [\underline{g}, \overline{g}]$. Likewise, the persistence of the new regime, π_{t+1}, is drawn from a density $F(\pi_{t+1})$ defined on $[\underline{\pi}; \overline{\pi}]$. For simplicity we assume that $F(.)$ and $G(.)$ are independent. The possibility of breaks in the mean and persistence parameters is the only non-standard part of the specification of the dividend process and the innovation term is homoskedastic and serially uncorrelated. The changes in the mean dividend growth rate that we have in mind with this dividend specification are rare structural breaks like the ones identified in the empirical analysis. Breaks in the discount rate will have a symmetric effect on asset prices and can be analyzed accordingly. Alternatively, breaks can be thought of as occurring in the differential between the discount rate and the growth rate, $\rho - g$.

Investors are assumed to observe if a break has occurred in a given period, so their information set is $\Omega_t = \{D_t, D_{t-1}, ..., P_t, P_{t-1}, ..., s_t, s_{t-1}, ..\}$. In practice investors may either have superior information that allows them to anticipate a break or, conversely, only gradually realize that a break has occurred. The advantage of our informational assumption is that it allows us to study the clean effect of a break on stock prices.

In the appendix we prove that, under full information and with breaks in the dividend process, the stock price is given by the following proposition:

Proposition

Suppose that with probability $1 - \pi_t$ the mean growth rate of the dividend process breaks from the state prevailing at time t. After a break, the new mean growth rate, g_{t+1}, is drawn from a density $G(.)$ with support $[\underline{g}, \overline{g}]$ while the persistence parameter, π_{t+1}, is drawn from a density $F(.)$ defined on $[\underline{\pi}, \overline{\pi}]$. Then the full information stock price (8) is given by

$$P_t = \left(\frac{D_t}{(1+\rho) - \pi_t(1+g_t)^{1-\alpha}}\right)\Bigg\{\pi_t(1+g_t)^{1-\alpha} + (1-\pi_t)\int_{\underline{g}}^{\overline{g}}(1+g_{t+1})^{1-\alpha}dG(g_{t+1}) \qquad (12)$$
$$+ \frac{(1-\pi_t)\int_{\underline{\pi}}^{\overline{\pi}}\int_{\underline{g}}^{\overline{g}}\left(\frac{\pi_t(1+g_t)^{2-2\alpha} + (1-\pi_t)\int_{\underline{g}}^{\overline{g}}(1+g_{t+1})^{1-\alpha}dG(g_{t+1})(1+g_t)^{1-\alpha}}{1+\rho-\pi_t(1+g_t)^{1-\alpha}}\right)dG(g_t)dF(\pi_t)}{1 - \int_{\underline{\pi}}^{\overline{\pi}}\int_{\underline{g}}^{\overline{g}}\frac{(1-\pi_t)(1+g_t)^{1-\alpha}}{1+\rho-\pi_t(1+g_t)^{1-\alpha}}dG(g_t)dF(\pi_t)}\Bigg\}.$$

In the special case where investors are risk neutral ($\alpha = 0$), the density of the dividend growth rate after a break is uniform and $\pi_t = \pi$ is constant, the stock price simplifies to

$$P_t = \left(\frac{a + \pi(1+g_t)}{1+\rho-\pi(1+g_t)}\right)D_t \qquad (13)$$

where a is a constant defined by

$$a = \frac{(1-\pi)\left((1+\frac{\underline{g}+\overline{g}}{2})(\overline{g}-\underline{g}) + \pi A\right)}{\overline{g}-\underline{g}-(1-\pi)B},$$

and

$$
\begin{aligned}
A &= \frac{-(\overline{g}^2 - \underline{g}^2 + 2(\overline{g} - \underline{g}))}{2\pi} - \frac{(1+\rho)(\overline{g} - \underline{g})}{\pi^2} + \frac{(1+\rho)^2}{\pi^3} \ln\left(\frac{1+\rho-\pi(1+\underline{g})}{1+\rho-\pi(1+\overline{g})}\right) \\
B &= \frac{1}{\pi}\left((\frac{1+\rho}{\pi}) \ln\left(\frac{1+\rho-\pi(1+\underline{g})}{1+\rho-\pi(1+\overline{g})}\right) + \overline{g} - \underline{g}\right).
\end{aligned}
$$

Notice the tradeoff involved in the choice of π, the parameter determining the breakpoint frequency. If π is low, breaks occur frequently but their effect tends to be smaller since they are expected to influence dividends over a shorter future horizon. If π is close to one, breaks will be rare but they also have a much larger effect when they do occur.

4. Stock Prices Under Incomplete Information and Recursive Learning

The solution to the stock price in Section 3 was derived under the assumption that, at each point in time, investors know the true mean and persistence parameters of the dividend growth rate (g_t, π_t). This assumption becomes less plausible in the presence of breaks in the dividend process. After a break investors no longer have access to a large sample of historical data points that can provide them with a precise estimate of the new parameters of the dividend process. If investors do not know the true parameter values it is plausible to assume that they attempt to learn them through efficient use of information after the break. To make investors' estimation problem tractable we assume that only the drift of the dividend process (μ_t) is unknown and subject to breaks, i.e. $\sigma_t^2 = \sigma^2$ and $\pi_t = \pi$ are constants known by investors. As argued by Merton (1980) and Brennan (1997), the mean parameter of the fundamentals process is typically imprecisely estimated in small samples, while the volatility can be precisely estimated by more frequent data sampling. Likewise, the variation in π_t documented in Section 2 is sufficiently small that this is unlikely to be an important source of variation in stock prices.

Investors use a Bayesian updating procedure and are interested in calculating the stock price as a function of μ_t, $\lambda(\mu_t|\sigma, \alpha, \rho, \pi)$. The form of λ follows from the proposition and the mapping from μ_t to the growth rate, g_t. Let $\boldsymbol{\xi}_t = (\xi_t, \xi_{t-1}, ...,$ $\xi_{t-n+1})$, where $\xi_t = \Delta \log(D_t)$ and n is the number of observations of the dividend process since the most recent break. Using the assumption that $\boldsymbol{\xi}_t$ is normally distributed, the likelihood function for μ_t conditional on $\boldsymbol{\xi}_t$ is given by $L(\mu_t;\boldsymbol{\xi}_t) = (1/\sqrt{2\pi\sigma^2/n})\exp(\frac{-(\overline{\xi}_t-\mu_t)^2}{2\sigma^2/n})$, where $\overline{\xi}_t = (1/n)\sum_{i=0}^{n-1}\xi_{t-i}$. Let $p(\mu_t)$ be the prior distribution for μ_{t+1}. Then the stock price follows from Bayes' rule:

$$E[\lambda(\mu_t)|\boldsymbol{\xi}_t] = \frac{\int \lambda(\mu_t)L(\mu_t;\boldsymbol{\xi}_t)p(\mu_t)d\mu_t}{\int L(\mu_t;\boldsymbol{\xi}_t)p(\mu_t)d\mu_t} \tag{14}$$

This expression accounts for investors' estimation uncertainty. We set the prior in our model equal to the unconditional density for the mean dividend growth rate, i.e. the indicator function $I_{[\underline{g};\overline{g}]}$ scaled by $1/(\overline{g}-\underline{g})$. The mean growth rate is bounded between $\underline{g}$ and $\overline{g}$; $1+\underline{g} \leq \exp(\mu_t+\sigma^2/2) \leq 1+\overline{g}$, so the true value of the unknown drift, μ_t, lies between the following bounds: $\underline{l} \equiv \ln(1+\underline{g})-\sigma^2/2 \leq \mu_t \leq \ln(1+\overline{g})-\sigma^2/2 \equiv \overline{l}$. It follows that the expression in the denominator of (14) reduces to

$$\begin{aligned} &\frac{1/(\overline{g}-\underline{g})}{\sqrt{2\pi\sigma^2/n}}\int_{\underline{l}}^{\overline{l}} \exp(\frac{-(\overline{\xi}_t-\mu_t)^2}{2\sigma^2/n})d\mu_t \\ = \; &\frac{1}{\overline{g}-\underline{g}}\left(\Phi(\frac{\overline{l}-\overline{\xi}_t}{\sigma/\sqrt{n}})-\Phi(\frac{\underline{l}-\overline{\xi}_t}{\sigma/\sqrt{n}})\right) \end{aligned} \tag{15}$$

The simplicity of the updating problem confronting investors is helpful in simulations based on a large number of computations of the stock price.

Our model is closely related to the analysis of Lewis (1989) which considers the market's forecast error process arising from a one-off break in the drift parameter of the first-differenced fundamentals process. Investors learn gradually about the shift through a Bayesian updating rule and, as in our model, also know the time where the fundamentals process may have changed. Lewis analyses separate scenarios depending on whether the new drift parameter is known or unknown to investors. Compared to the case where the market knows the drift after the switch, she finds that learning evolves much more slowly when investors have to estimate this parameter. This observation will be important to our simulation results.

It is possible to extend the setup to allow agents to use pre-break data when estimating the new growth rate after a break or even when testing for a break in the first instance. For example, agents could use a reverse ordered Cusum test to detect the most recent break and then use a stopping rule to determine the optimal estimation window. The cost of introducing these layers of complexity is, however, that the learning problem becomes progressively more complicated and the simulation results more difficult to interpret.

There is an alternative approach that does not condition on knowing the underlying state. Authors such as Cecchetti, Lam & Mark (1993) and Veronesi (1999) have developed rational expectation equilibrium models to explain asset prices when these are driven by dividends whose drift switches between two unobservable states. These models give rise to a filtering problem that is closely related to the recursive parameter estimation problem considered here, since the state probabilities are also updated through Bayes rule. The vector of filtered state probabilities under Markov switching follows from the updating equations

$$\begin{aligned} \mathbf{p}_{t|t} &= \frac{\mathbf{p}_{t|t-1} \odot \boldsymbol{\eta}_t}{\iota'(\mathbf{p}_{t|t-1} \odot \boldsymbol{\eta}_t)}, \\ \mathbf{p}_{t+1|t} &= \Pi \mathbf{p}_{t|t}, \end{aligned} \tag{16}$$

where $\mathbf{p}_{t|t} = \Pr(s_t | s_{t-1}, d_t, d_{t-1}, ...)$, $\odot$ represents element by element multiplication and $\boldsymbol{\eta}_t$ is the vector of state densities evaluated at the dividend growth realization:

$$\boldsymbol{\eta}_t = \begin{pmatrix} \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp(-(d_t - \mu_1)^2 / 2\sigma_1^2) \\ .... \\ \frac{1}{\sqrt{2\pi\sigma_{n_s}^2}} \exp(-(d_t - \mu_{n_s})^2 / 2\sigma_{n_s}^2) \end{pmatrix}. \tag{17}$$

The stock price can be derived by summing the discounted value of future expected dividends across the states weighted by the filtered state probabilities. This probability weighting has the effect of smoothing the price series relative to the full information case where the full impact of a switch in the state is immediately incorporated into the stock price.

5. Simulations

Structural breaks introduce non-linearities in dividends and stock prices and recursive learning effects introduce non-stationarities in returns. This rules out standard econometric tests of our model, c.f. Bossaerts (1995). Instead we evaluate the model by simulating dividends and forming stock prices according to the formulae in sections 3 and 4. The purpose of this analysis is not to calibrate the moments of stock returns but rather to study some of the qualitative features associated with breaks and different models for investors' expectation formation. The following set of monthly dividend parameter values are assumed in the experiments

$$\rho = 0.0062, \overline{g} = 0.0050, \underline{g} = -0.0034, \sigma = 0.015, \pi = 0.997. \quad (18)$$

The annualized real discount rate is 7.5%, and the minimum and maximum values of the dividend growth rate are -4% and 6%, respectively, yielding an average growth rate of one percent per annum. These parameter values as well as the monthly volatility of 1.5% match the real dividend growth data over the period 1871 - 1999. The choice of interval for the dividend growth rate is based on our assumption that the dividend growth rate is drawn from a uniform distribution with support $[\underline{g}; \overline{g}]$, such that $r < \overline{g}$. The value of π means that the drift of the dividend process on average changes about once every thirty years. The historical returns used as a benchmark for the simulations comprise 129 years of monthly data or 1548 observations. We did not experiment with the parameters in order to obtain the best fit to the actual return series.

5.1. *Diagnostic Tests and Moments*

Figure 4 plots the outcome of a particular simulation. The upper windows present excess returns under full information, filtering and Bayesian learning. Since dividends are identical for the return series, the difference between the plots reflects differences in investors' growth estimates, plots of which are provided in the second row. For this particular simulation there were seven changes in the dividend growth rate. The volatility in the growth estimate under Bayesian learning following a break and the subsequent gradual adjustment towards the true value is clear from the middle picture. Under the filtering model, the dividend growth rate adjusts more gradually. It increases initially in response to the high growth rate prevailing between observations 200 and 300, only to decline gradually and increase towards the end as a result of the higher terminal growth rate. Volatility clustering in returns around the breaks is visibly present in the simulations under learning. In contrast, under full information there is no tendency for outliers to carry over as higher volatility in subsequent periods. The smoother adjustment under filtering gives rise to a return series that more closely resembles white noise.

Figure 5 confirms that learning can give rise to ARCH effects and serial correlation. For the simulation used to construct Figure 4, the upper row plots the time series of a twelfth-order ARCH test. To track evicence of local volatility clustering in returns we use rolling regressions with a window length of 120 months, or ten years of data. There is only weak evidence of ARCH effects in the simulated returns from the full information model shown in the first window of Figure 5. A very different picture emerges from the second window which plots the estimated conditional volatility under Bayesian learning against the one percent critical value of the test statistic. The large variations in the growth estimate between observations 400 and 700 give rise to high values of the ARCH test. Under filtering there is no evidence of volatility clustering.

To measure local serial correlation in returns we calculate Ljung-Box statistics for twelfth order serial correlation, again using rolling regressions with a window length of 120 months. In this particular simulation local serial correlation never occurs under full information while it shows up both under Bayesian learning and under filtering.

Table 3 provides a more systematic set of results for the monthly S&P500 returns and the simulated data based on sample sizes of 1548 observations. As documented in many previous studies, monthly stock returns are characterized by skewness, fat tails, first order serial correlation and strong volatility clustering, c.f. the significant ARCH effects. The strong evidence of first order serial correlation in the actual returns data is likely to reflect non-synchronous trading effects as opposed to genuine predictable patterns in returns. For this reason we do not attempt to replicate this feature of the data.

Consider next the simulated data. Under full information and no breaks ($\pi = 1$) this model is unable to match the high volatility, skewness and fat tails observed in the data. This simply reflects the common finding in the asset pricing literature that - in the context of a stationary dividend growth model - dividend variations alone do not fully explain movements in observed stock prices. These results are identical for different values of the coefficient of risk aversion, α.

Introducing breaks, but maintaining the full information assumption and setting $\alpha = 0$, the volatility of stock returns increases from 1.5 to 2.9 percent and the skewness and kurtosis also go up dramatically, exceeding the estimates for US returns by an order of magnitude. This happens because of the outliers in stock returns observed after a break in the dividend growth rate, c.f. Figure 4. Since isolated outliers is the opposite of volatility clustering, this model does not give rise to the ARCH effects observed in the data. However, the comparison of the full information model with and without breaks clearly demonstrates the importance to the distribution of stock prices of allowing for breaks in fundamentals irrespective of informational assumptions.

Next consider stock returns under Bayesian learning. This model generates average volatility of 3.3 percent, close to the sample estimate of 4.1 percent. Compared with the full information case, Bayesian learning decreases the skewness and kurtosis to a level more closely in line with the data. It is easy to understand why: Under full information a jump in the dividend growth rate is instantly recognized by investors and shows up as a major revision in the stock price. In contrast, under Bayesian learning, new dividend information after a break will only gradually be incorporated into the price and is weighted against investors' prior beliefs. This gives rise to a more gradual price adjustment and hence decreases the skewness and kurtosis of returns. Despite this gradual adjustment, the Bayesian learning model does not seem to generate much full-sample serial correlation in the *level* of returns.

The Bayesian learning model also produces volatility clustering. Close to 40 percent of the simulations generate significant ARCH effects when investors are risk neutral. To measure the persistence in the conditional volatility of excess returns we sum the coefficients of the squared residuals in an ARCH(12) regression of squared residuals on a constant and twelve lags. Under either full information or filtering the mean value of the persistence estimate is -0.01, while under Bayesian learning this figure increases to 0.27. This compares with an estimated persistence of the conditional volatility of the returns data of 0.59.

Increasing α to 0.5, the Bayesian learning model is still able to generate values of the diagnostic tests that are not too far from those of the actual returns data. However, once α increases to 2, the model becomes very similar to the full information model with breaks taking the form of isolated outliers. Consequently, when investors' degree of risk aversion is too high and the mapping from the growth estimate to the asset price is no longer convex, learning effects appear less able to fit historical returns. As pointed out by Cecchetti et al. (1990), when $\alpha > 1$, the equilibrium price-dividend multiple is lower when the dividend growth rate is high and agents' intertemporal smoothing incentive leads them to sell out stocks and suppresses their price. The positive correlation between dividend shocks and the price dividend ratio thus no longer prevails and the two effects tend to cancel out, thereby removing the strong effects from learning. At any rate, such high values of α do not appear to characterize the US stock returns data very well. Hansen & Singleton (1982) report estimates of α around 0.8.

Under filtering, the return distribution is close to Gaussian, the main difference being that there is significant evidence of serial correlation−though not of volatility clustering−in the returns generated by this model. Naturally if we were to introduce state dependence in the volatility parameter, σ, then the filtering model would give rise to ARCH effects. However, again such effects are concerned with volatility clustering at the cyclical frequency as opposed to clustering around unique breaks such as the Great Depression.

5.2. *Tests of Euler Equation*

Following Hansen & Singleton (1982), it has become standard to test intertemporal restrictions implied by asset pricing models estimated by GMM. The present value model (8) implies an Euler equation

$$E[\left(\beta(\frac{P_{t+1}+D_{t+1}}{P_t}\left(\frac{D_{t+1}}{D_t}\right)^{-\alpha}-1\right)\mathbf{Z}_t]=0, \tag{19}$$

where $\mathbf{Z}_t$ is a vector of instruments known at time t. This methodology is based on stationarity of the underlying return series. Asymptotically breaks therefore cannot explain rejections of the moment condition. However, in samples with relatively few breaks it is possible that these can distort the size of such tests. Similarly, learning about the parameters or about the underlying state can also introduce serial correlation in returns which could lead to more rejections of the moment condition.

To explore the impact of breaks and incomplete information on tests of the Euler equation, we undertook 500 simulations of the three models and estimated the parameters (α, β) by GMM. As instruments we use a constant, the dividend yield and the lagged return. These are standard instruments and have been established empirically to have reasonable power in testing moment conditions. With two parameters to be estimated, there is one overidentifying restriction to be tested. Table 4 reports the percentage of rejections at one, five and ten percent critical levels as well as the mean value of Hansen's J test.

There is a systematic tendency for breaks to lead to overrejections irrespective of the underlying model for agents' expectation formation. For example, the rejection frequency at the 5% critical level is close to 10 percent in many cases. The rejection rates are highest under the filtering model which introduces most serial correlation in returns. These results lend some credibility to the notion that breaks and learning effects may in part be responsible for rejections of moment tests of the Euler equation.

5.3. *Timing of the Breaks*

Two conclusions can be drawn from the simulations in the previous section. First, independent of how much information investors' hold, breaks in fundamentals affect the distribution of stock returns in important ways and may be helpful in understanding the kurtosis and fat tails in the observed data. Furthermore, imperfect information and gradual updating of investors' beliefs after a break seem plausible candidates to an explanation of the clustering of volatility observed in US stock returns.

Authors such as Veronesi (1999) have previously suggested that recursive updating of the probability of the underlying state may give rise to volatility clustering in returns. Our results add to this analysis by actually predicting the timing of the volatility clustering in stock returns. It provides an *ex ante* identification of the point in time, namely after a break in the dividend process, where ARCH effects can be expected to occur in stock returns. The model also predicts serially correlated returns after a break.

It is instructive to compare our model to the dividend growth rate process and the learning problem analyzed by Barsky & DeLong (1993). These authors argue that long-run movements in the price-dividend ratio of US stocks can be explained by investors' projections of future dividends modeled as a long moving average of their own past with geometrically declining weights. The magnitude of learning effects does not change over time in this setup. In contrast, in our model learning effects are far stronger in the periods right after a break.

To test these predictions on US data, Figure 6 plots the twelfth-order LM and Ljung-Box statistics for ARCH and serial correlation based on a 120 month rolling window adopted to the returns data. Again we use a high order (12) of the diagnostic test since in practice investors' knowledge of breaks are likely to be less precise than what was assumed in the theoretical model in Sections 3 and 4. The conditional volatility of excess returns is very high around 1918-1920, 1929-1933 and in 1952. There is evidence of particularly strong serial correlation in excess returns over the period 1922-52 and again in the early seventies. We would not expect these points to coincide exactly with the break dates identified in Section 2 since investors could either have anticipated a breakpoint (if they have superior information) or failed to immediately identify a break in real time since historically they did not have access to the full sample information. Although the coincidence of the two series is not perfect, it is nevertheless suggestive of the importance of learning effects and structural breaks to an explanation of some of the most important episodes for US stock prices in the twentieth century.

6. Conclusion

This paper has presented new empirical evidence on low-frequency breaks in the fundamentals process underlying US stock prices. Through simulations of an asset pricing model that accounts for such rare breaks, we showed that how much information investors possess about such breaks can strongly affect the dynamics of asset prices. Our findings also suggest that although breaks may be drawn from a stationary meta model, their rare occurrence and potentially large effect can lead to small sample distortions of standard econometric tests of asset pricing models.

Dividend growth and discount rates are typically modeled as simple stationary processes without breaks. However, mounting empirical evidence suggests that this is too simple a representation. Donaldson & Kamstra (1996) provide an effective demonstration of the importance of correctly modelling the dividend process. They proxy investors' beliefs about the dividend growth process by means of a neural network and argue that the resulting nonlinearities in the dividend process could have led to the 1929 stock market crash. Interestingly, our results justify their use of 1920 as a starting point in the estimations since a dividend break is identified in 1919. Donaldson and Kamstra condition on the parameter estimates of their model and hence do not consider investors' uncertainty about their very complicated nonlinear model, so it is hard to say how learning would affect their results. The two approaches are closely related, nevertheless, since our approach concludes that the dynamic specifications that ignore breaks are also likely to be misspecified.

Many alternative explanations have been proposed for the the seemingly high first and second moments of stock returns. Wang (1993) suggests asymmetry of information between noise traders and rational investors which leads uninformed traders to rationally behave like price chasers. This introduces serial correlation in stock returns and increases volatility and risk premiums. Campbell & Cochrane (1999) propose an asset pricing model in which consumption growth follows a lognormal process but with habit formation effects. Their model fits both the unconditional equity premium and the risk free interest rate as well as a range of other moments. Cecchetti, Lam & Mark (2000) introduce belief distortions that vary over expansions and contractions and leads to predictability in returns. These models form an important part of a complete story for the variations in US stock prices. For example, we have not attempted to address the equity premium puzzle or explain the short-run dynamics of stock prices.

What this paper has suggested, is that episodes such as the Great Depression and the world wars have fundamental effects on the statistical properties of fundamentals underlying US stock prices. This is consistent with Kim, Nelson & Starz (1991) who find evidence of a fundamental change in the stock return process after WWII and go on to "conjecture that it may be due to the resolution of the uncertainties of the 1930's and 1940s" (page 515). Attempts to model stock prices over the last century should therefore pay careful attention to structural breaks and their impact on investors' expectations.

Appendix

Proof of the Proposition

Suppose that the solution for P_t takes the form $P_t = \gamma(g_t, \pi_t)D_t$ for some function $\gamma(.)$. Taking expectations conditional on information at time t it follows from (9) that

$$
\begin{aligned}
& (1+\rho)\gamma(g_t,\pi_t)D_t = \sum_{i=0}^{1} E_t[(P_{t+1}+D_{t+1})\left(\frac{D_{t+1}}{D_t}\right)^{-\alpha}|s_{t+1}=s_t+i]\Pr(s_{t+1}=s_t+i) \\
& = \pi_t D_t \int_{-\infty}^{\infty} (1+\gamma(g_t,\pi_t))(1+g_t)^{1-\alpha}\exp((1-\alpha)(\sigma_t\varepsilon_{t+1}-\frac{\sigma_t^2}{2}))\phi(\varepsilon_{t+1}|\sigma_t^2)d\varepsilon_{t+1} \\
& +(1-\pi_t)D_t \int_{\underline{\pi}}^{\overline{\pi}}\int_{\underline{g}}^{\overline{g}}\int_{-\infty}^{\infty} (1+\gamma(g_{t+1},\pi_{t+1}))(1+g_{t+1})^{1-\alpha} \times \\
& \exp((1-\alpha)(\sigma_{t+1}\varepsilon_{t+1}-\frac{\sigma_{t+1}^2}{2}))\phi(\varepsilon_{t+1}|\sigma_{t+1}^2)d\varepsilon_{t+1}dG(g_{t+1})dF(\pi_{t+1}) \\
= \; & \pi_t D_t(1+g_t)^{1-\alpha}(1+\gamma(g_t,\pi_t)) + (1-\pi_t)D_t\int_{\underline{g}}^{\overline{g}}(1+g_{t+1})^{1-\alpha}dG(g_{t+1}) \\
& +(1-\pi_t)D_t\int_{\underline{\pi}}^{\overline{\pi}}\int_{\underline{g}}^{\overline{g}}(1+g_{t+1})^{1-\alpha}\gamma(g_{t+1},\pi_{t+1})dG(g_{t+1})dF(\pi_{t+1}), \qquad \text{(A1)}
\end{aligned}
$$

where $G(g_{t+1})$ is the cdf of g_{t+1} which has support $[\underline{g};\overline{g}]$, $F(\pi_{t+1})$ is the cdf of π_{t+1} with support $[\underline{\pi};\overline{\pi}]$, $\phi(.|\sigma_t^2)$ is the normal density function with mean zero and variance σ_t^2, and the last equality follows by using the independence of ε_{t+1} and g_{t+1} and integrating out ε_{t+1}. Dividing through by (D_t) in (A1) and simplifying we get

$$
\begin{aligned}
& \left(1+\rho-\pi_t(1+g_t)^{1-\alpha}\right)\gamma(g_t,\pi_t) = \pi_t(1+g_t)^{1-\alpha} + (1-\pi_t)\int_{\underline{g}}^{\overline{g}}(1+g_{t+1})^{1-\alpha}dG(g_{t+1}) \\
& +(1-\pi_t)\int_{\underline{\pi}}^{\overline{\pi}}\int_{\underline{g}}^{\overline{g}}(1+g_{t+1})^{1-\alpha}\gamma(g_{t+1},\pi_{t+1})dG(g_{t+1})dF(\pi_{t+1}). \qquad \text{(A2)}
\end{aligned}
$$

Next multiply by $(1+g_t)^{1-\alpha}dG(g_t)dF(\pi_t)/\left(1+\rho-\pi_t(1+g_t)^{1-\alpha}\right)$ and integrate over $[\underline{g};\overline{g}]$ and $[\underline{\pi};\overline{\pi}]$

$$\int_{\underline{\pi}}^{\overline{\pi}}\int_{\underline{g}}^{\overline{g}} (1+g_t)^{1-\alpha}\gamma(g_t,\pi_t)dG(g_t)dF(\pi_t) \tag{A3}$$

$$= \int_{\underline{\pi}}^{\overline{\pi}}\int_{\underline{g}}^{\overline{g}} \frac{\pi_t(1+g_t)^{1-\alpha} + (1-\pi_t)\int_{\underline{g}}^{\overline{g}} (1+g_{t+1})^{1-\alpha}dG(g_{t+1})}{1+\rho-\pi_t(1+g_t)^{1-\alpha}}(1+g_t)^{1-\alpha}dG(g_t)dF(\pi_t) +$$

$$\int_{\underline{\pi}}^{\overline{\pi}}\int_{\underline{g}}^{\overline{g}} \frac{(1-\pi_t)(1+g_t)^{1-\alpha}}{1+\rho-\pi_t(1+g_t)^{1-\alpha}}dG(g_t)dF(\pi_t) \int_{\underline{\pi}}^{\overline{\pi}}\int_{\underline{g}}^{\overline{g}} (1+g_{t+1})^{1-\alpha}\gamma(g_{t+1},\pi_{t+1})dG(g_{t+1})dF(\pi_{t+1}).$$

Under the assumption that the underlying densities $F(.), G(.)$ do not vary through time, we must have

$$\int_{\underline{\pi}}^{\overline{\pi}}\int_{\underline{g}}^{\overline{g}} (1+g_t)^{1-\alpha}\gamma(g_t,\pi_t)dG(g_t)dF(\pi_t) = \int_{\underline{\pi}}^{\overline{\pi}}\int_{\underline{g}}^{\overline{g}} (1+g_{t+1})^{1-\alpha}\gamma(g_{t+1},\pi_{t+1})dG(g_{t+1})dF(\pi_{t+1}). \tag{A4}$$

This gives an equation which can be used to assess the integral in (A3):

$$\int_{\underline{\pi}}^{\overline{\pi}}\int_{\underline{g}}^{\overline{g}} (1+g_t)^{1-\alpha}\gamma(g_t,\pi_t)dG(g_t)dF(\pi_t) \tag{A5}$$

$$= \frac{\int_{\underline{\pi}}^{\overline{\pi}}\int_{\underline{g}}^{\overline{g}} \left(\frac{\pi_t(1+g_t)^{2-2\alpha}+(1-\pi_t)\int_{\underline{g}}^{\overline{g}}(1+g_{t+1})^{1-\alpha}dG(g_{t+1})}{1+\rho-\pi_t(1+g_t)^{1-\alpha}}\right) dG(g_t)dF(\pi_t)}{1-\int_{\underline{\pi}}^{\overline{\pi}}\int_{\underline{g}}^{\overline{g}} \frac{(1-\pi_t)(1+g_t)^{1-\alpha}}{1+\rho-\pi_t(1+g_t)^{1-\alpha}}dG(g_t)dF(\pi_t)}.$$

Inserting this expression in (A2), we have

$$P_t = \left(\frac{D_t}{(1+\rho)-\pi_t(1+g_t)^{1-\alpha}}\right)\{\pi_t(1+g_t)^{1-\alpha} + (1-\pi_t)\int_{\underline{g}}^{\overline{g}} (1+g_{t+1})^{1-\alpha}dG(g_{t+1}) \tag{A6}$$

$$+\frac{(1-\pi_t)\int_{\underline{\pi}}^{\overline{\pi}}\int_{\underline{g}}^{\overline{g}} \left(\frac{\pi_t(1+g_t)^{2-2\alpha}+(1-\pi_t)(1+g_t)^{1-\alpha}\int_{g_{t+1}}(1+g_{t+1})^{1-\alpha}dG(g_{t+1})}{1+\rho-\pi_t(1+g_t)^{1-\alpha}}\right)dG(g_t)dF(\pi_t)}{1-\int_{\underline{\pi}}^{\overline{\pi}}\int_{\underline{g}}^{\overline{g}} \frac{(1-\pi_t)(1+g_t)^{1-\alpha}}{1+\rho-\pi_t(1+g_t)^{1-\alpha}}dG(g_t)dF(\pi_t)}\}.$$

In the special case with risk-neutral agents, $\alpha = 0$, constant break probability, $1-\pi$, and a uniform cdf, $F = U$, we can use that $\int_{\underline{g}}^{z} dU(g_t) = \frac{z}{\overline{g}-\underline{g}}$ to simplify (A5)

$$\int_{\underline{g}}^{\overline{g}} (1+g_t)\gamma(g_t)dU(g_t) = \frac{\int_{\underline{g}}^{\overline{g}} \left(\frac{\pi(1+g_t)^2+(1-\pi)(1+(\overline{g}+\underline{g})/2)(1+g_t)}{1+\rho-\pi(1+g_t)}\right) dg_t}{\overline{g}-\underline{g}-(1-\pi)\int_{\underline{g}}^{\overline{g}} \frac{(1+g_t)}{1+\rho-\pi(1+g_t)}dg_t}. \tag{A7}$$

To evaluate $\int_{\underline{g}}^{\overline{g}} \frac{(1+g_t)^2}{1+\rho-\pi(1+g_t)}dg_t$ and $\int_{\underline{g}}^{\overline{g}} \frac{(1+g_t)}{1+\rho-\pi(1+g_t)}dg_t$, change variables and notice that

$$\begin{aligned} A &\equiv \int_{\underline{g}}^{\overline{g}} \frac{(1+g_t)^2}{1+\rho-\pi(1+g_t)}dg_t = \frac{-y_t^2}{2\pi} - \frac{(1+\rho)y_t}{\pi^2} - \frac{(1+\rho)^2\ln(1+\rho-\pi y_t)}{\pi^3}\Bigg|_{1+\underline{g}}^{1+\overline{g}} \\ B &\equiv \int_{\underline{g}}^{\overline{g}} \frac{1+g_t}{1+\rho-\pi(1+g_t)}dg_t = \frac{-y_t}{\pi} - \frac{(1+\rho)\ln(1+\rho-\pi y_t)}{\pi^2}\Bigg|_{1+\underline{g}}^{1+\overline{g}}. \end{aligned} \tag{A8}$$

c.f. Gradshteyn & Ryzhik (1994). After some algebra we see that

$$\begin{aligned} A &= \frac{-(\overline{g}^2-\underline{g}^2+2(\overline{g}-\underline{g}))}{2\pi} - \frac{(1+\rho)(\overline{g}-\underline{g})}{\pi^2} + \frac{(1+\rho)^2}{\pi^3}\ln\left(\frac{1+\rho-\pi(1+\underline{g})}{1+\rho-\pi(1+\overline{g})}\right) \\ B &= \frac{1}{\pi}\left(\frac{1+\rho}{\pi}\ln\left(\frac{1+\rho-\pi(1+\underline{g})}{1+\rho-\pi(1+\overline{g})}\right)+\underline{g}-\overline{g}\right) \end{aligned} \tag{A9}$$

and hence from (A7)

$$\int_{\underline{g}}^{\overline{g}} (1+g_t)\gamma(g_t)dU(g_t) = \frac{\pi A + (1-\pi)(1+\frac{\overline{g}+\underline{g}}{2})B}{(\overline{g}-\underline{g})-(1-\pi)B}. \tag{A10}$$

Therefore in this special case the price-dividend ratio $\gamma(g_t)$ is given by

$$\gamma(g_t) = \frac{\pi(1+g_t)+(1-\pi)(1+\frac{\overline{g}+\underline{g}}{2})}{1+\rho-\pi(1+g_t)} + \frac{(1-\pi)(\pi A+(1-\pi)(1+\frac{\overline{g}+\underline{g}}{2})B)}{(\overline{g}-\underline{g}-(1-\pi)B)(1+\rho-\pi(1+g_t))}. \tag{A11}$$

References

Bai, J. & Perron, P. (1998), 'Estimating and testing linear models with multiple structural changes', *Econometrica* **66**, 47–78.

Barsky, R. & DeLong, J. (1993), 'Why does the stock market fluctuate?', *Quarterly Journal of Economics* **108**, 291–312.

Bossaerts, P. (1995), 'The econometrics of learning in financial markets', *Econometric Theory* **11**, 151–189.

Brennan, M. (1997), 'The role of learning in dynamic portfolio decisions', *UCLA, AGSM working paper* **3-97**.

Campbell, J. & Cochrane, J. (1999), 'By force of habit: A consumption-based explanation of aggregate stock behavior', *Journal of Political Economy* **107**, 205–251.

Cecchetti, S., Lam, P. & Mark, N. (1990), 'Mean reversion in equilibrium asset prices', *American Economic Review* **80**, 394–418.

Cecchetti, S., Lam, P. & Mark, N. (1993), 'The equity premium and the risk-free rate: Matching the moments', *Journal of Monetary Economics* **31**, 21–45.

Cecchetti, S., Lam, P. & Mark, N. C. (2000), 'Asset pricing with distorted beliefs: Are equity returns too good to be true?', *Mimeo, Ohio State University* .

Chib, S. (1998), 'Estimation and comparison of multiple change-point models', *Journal of Econometrics* **86**, 221–242.

Donaldson, G. & Kamstra, M. (1996), 'A new dividend forecasting procedure that rejects bubbles in asset prices: The case of 1929's stock crash', *Review of Financial Studies* **9**, 333–383.

Genotte, G. (1986), 'Optimal portfolio choice under incomplete information', *Journal of Finance* **41**, 733–746.

Gradshteyn, I. & Ryzhik, I. (1994), 'Table of integrals, series, and products', *Academic Press* .

Hansen, L. & Singleton, K. (1982), 'Generalized instrumental variables estimation of nonlinear rational expectations models', *Econometrica* **50**, 1269–1286.

Kandel, S. & Stambaugh, R. (1991), 'Asset returns and intertemporal preferences', *Journal of Monetary Economics* **27**, 39–71.

Kim, C.-J. (1994), 'Dynamic linear models with markov-switching', *Journal of Econometrics* **60**, 1–22.

Kim, M., Nelson, C. & Starz, R. (1991), 'Mean reversion in stock prices? a reappraisal of the empirical evidence', *Review of Economic Studies* **58**, 515–578.

Lewis, K. (1989), 'Changing beliefs and systematic rational forecast errors with evidence from foreign exchange', *American Economic Review* **79**, 621–636.

Lucas, R. (1978), 'Asset prices in an exchange economy', *Econometrica* **46**, 1429–1445.

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

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

Shiller, R. (2000), 'Irrational exuberance', *Princeton University Press* .

Veronesi, P. (1999), 'Stock market overreaction to bad news in good times: A rational expectations equilibrium model', *Review of Financial Studies* **5**, 975–1007.

Wang, J. (1993), 'A model of intertemporal asset prices under asymmetric information', *Review of Economic Studies* **60**, 249–282.

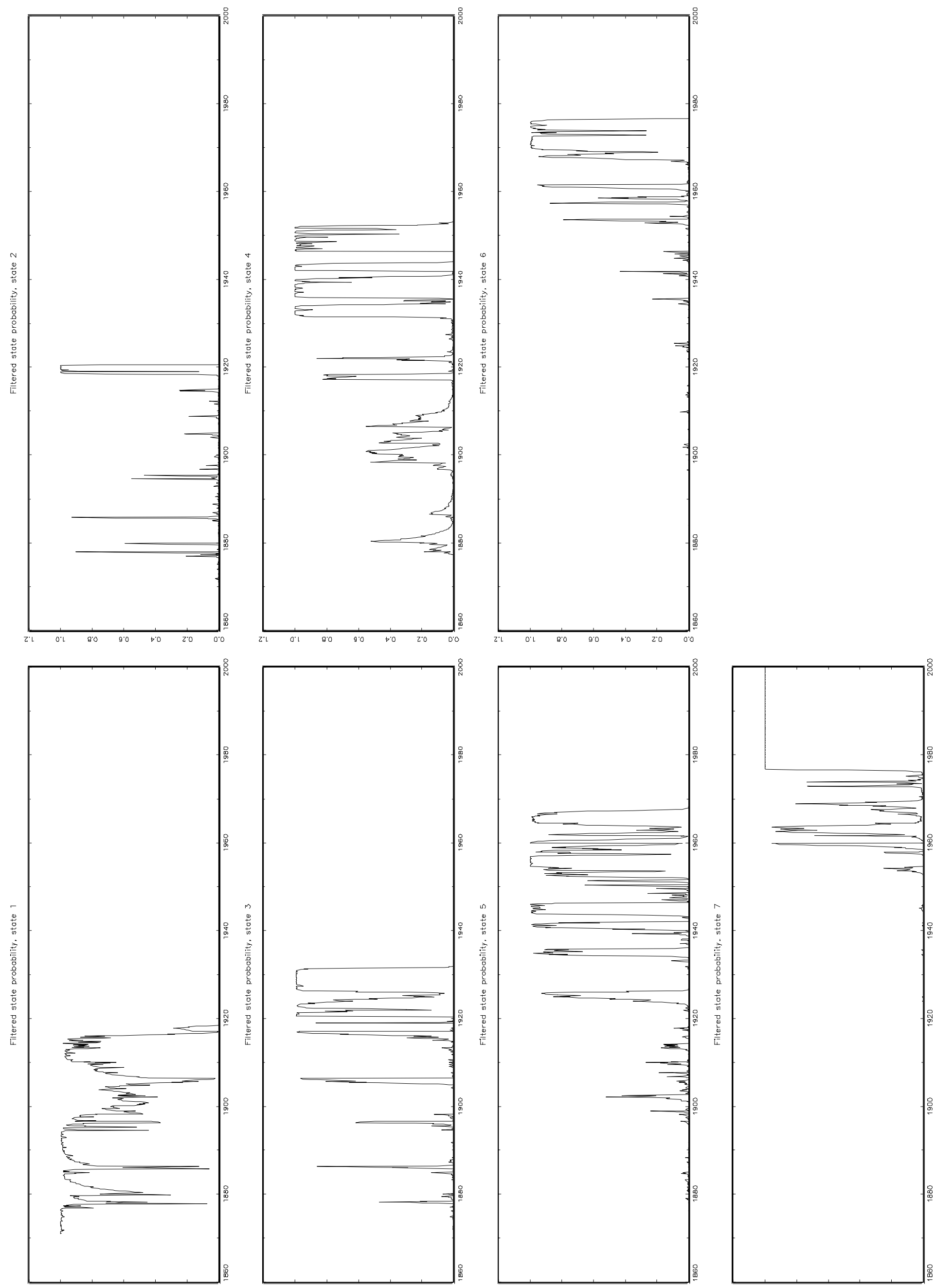

Figure 1: Filtered state probabilities from Markov switching model with seven non-recurring states

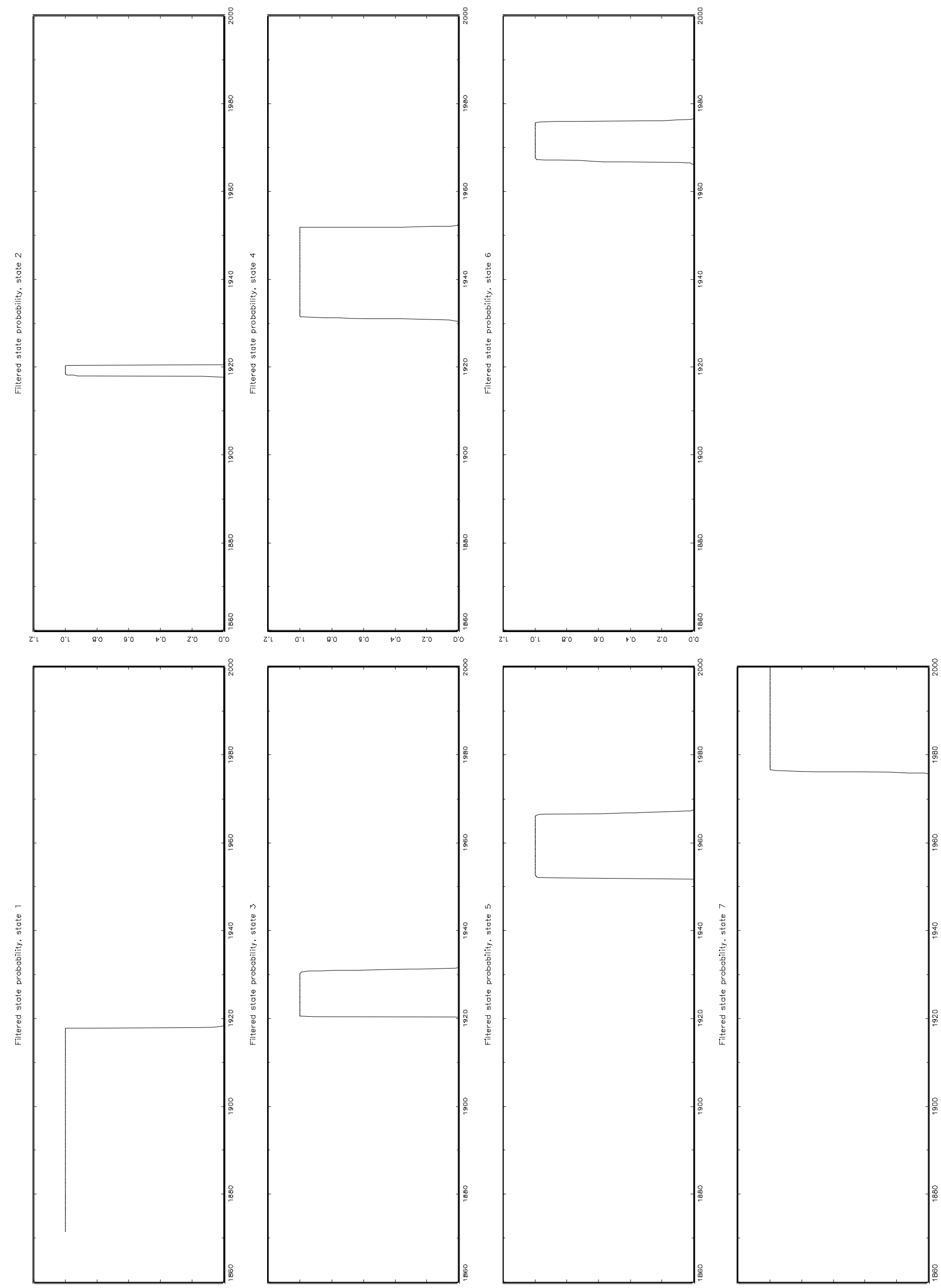

Figure 2: Smoothed state probabilities from Markov switching model with seven non-recurring states

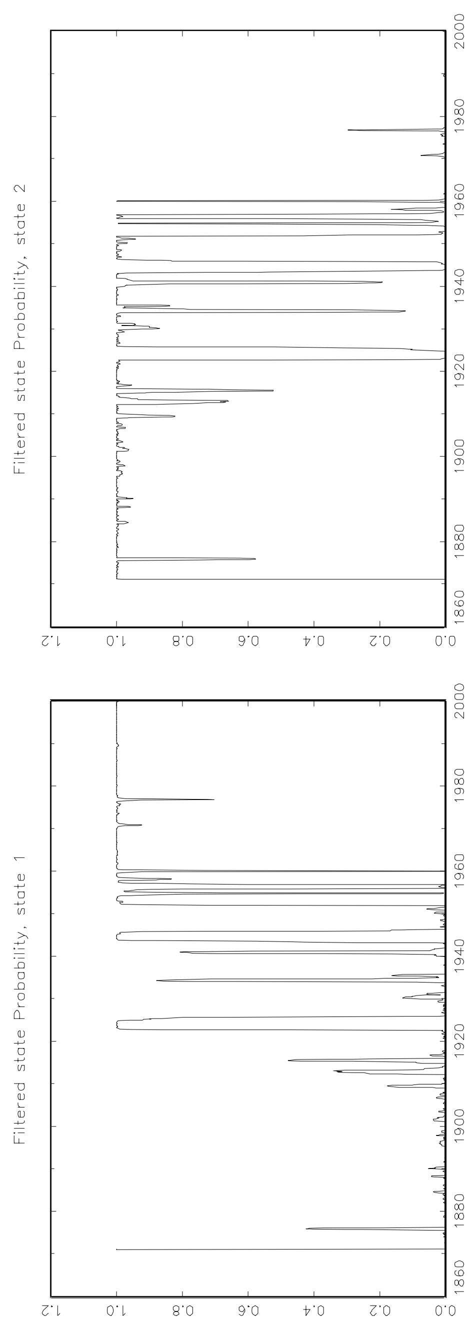

Figure 3: Smoothed state probabilities from Markov switching model with two repeated states

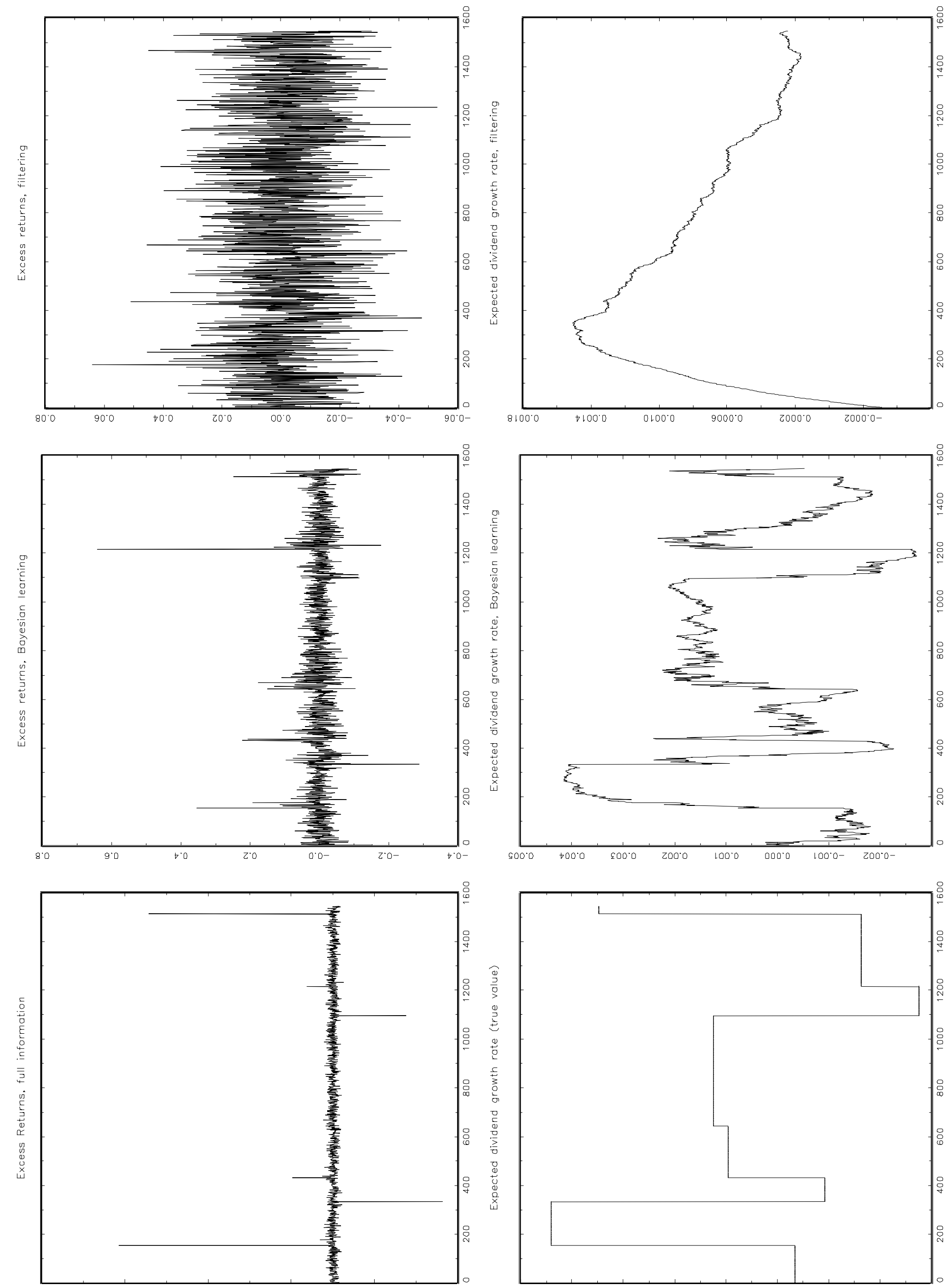

Figure 4: Returns and expected dividend growth under three assumptions on investors' expectations

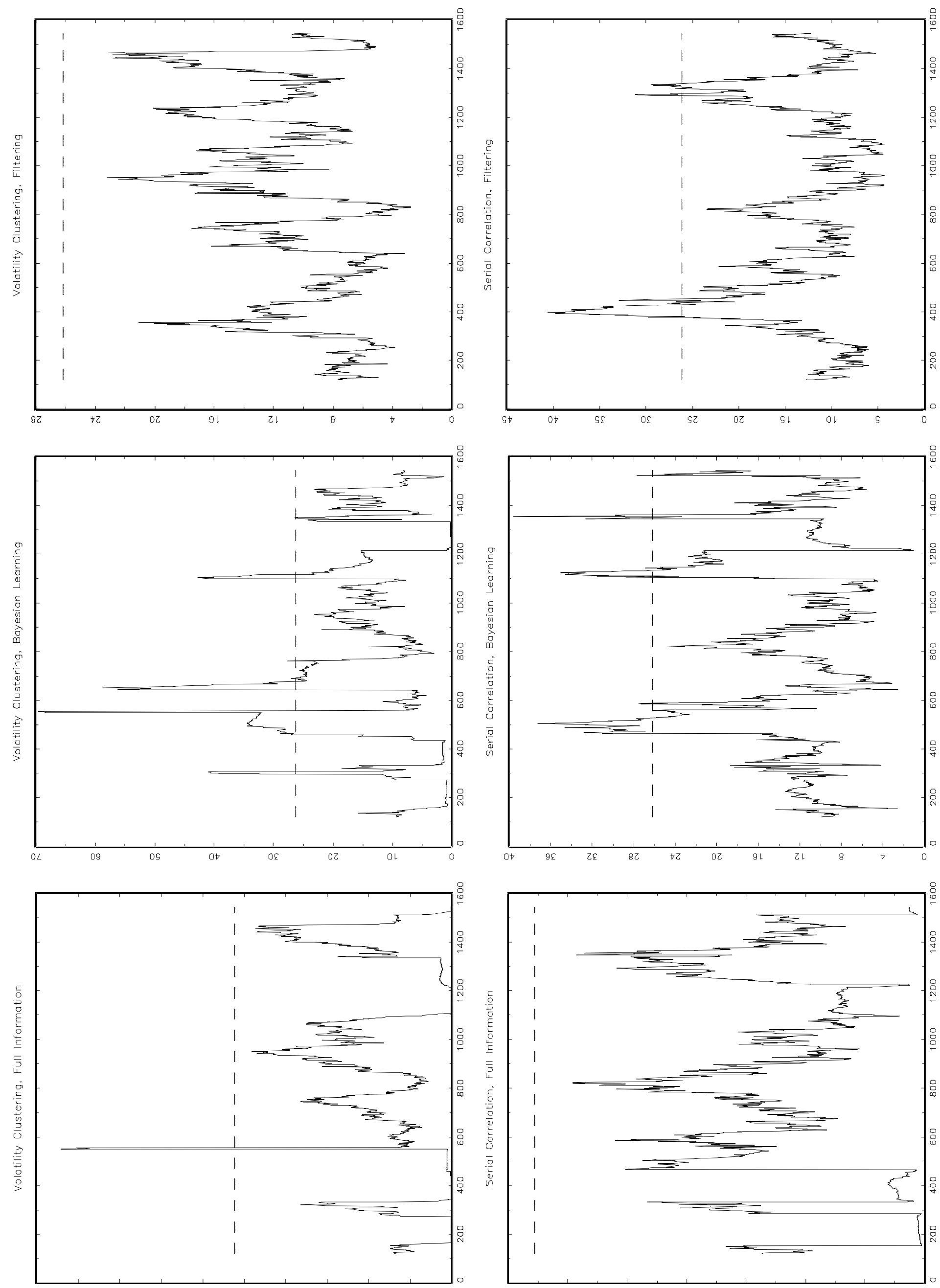

Figure 5: Volatility clustering and serial correlation under three assumptions on investors' expectations

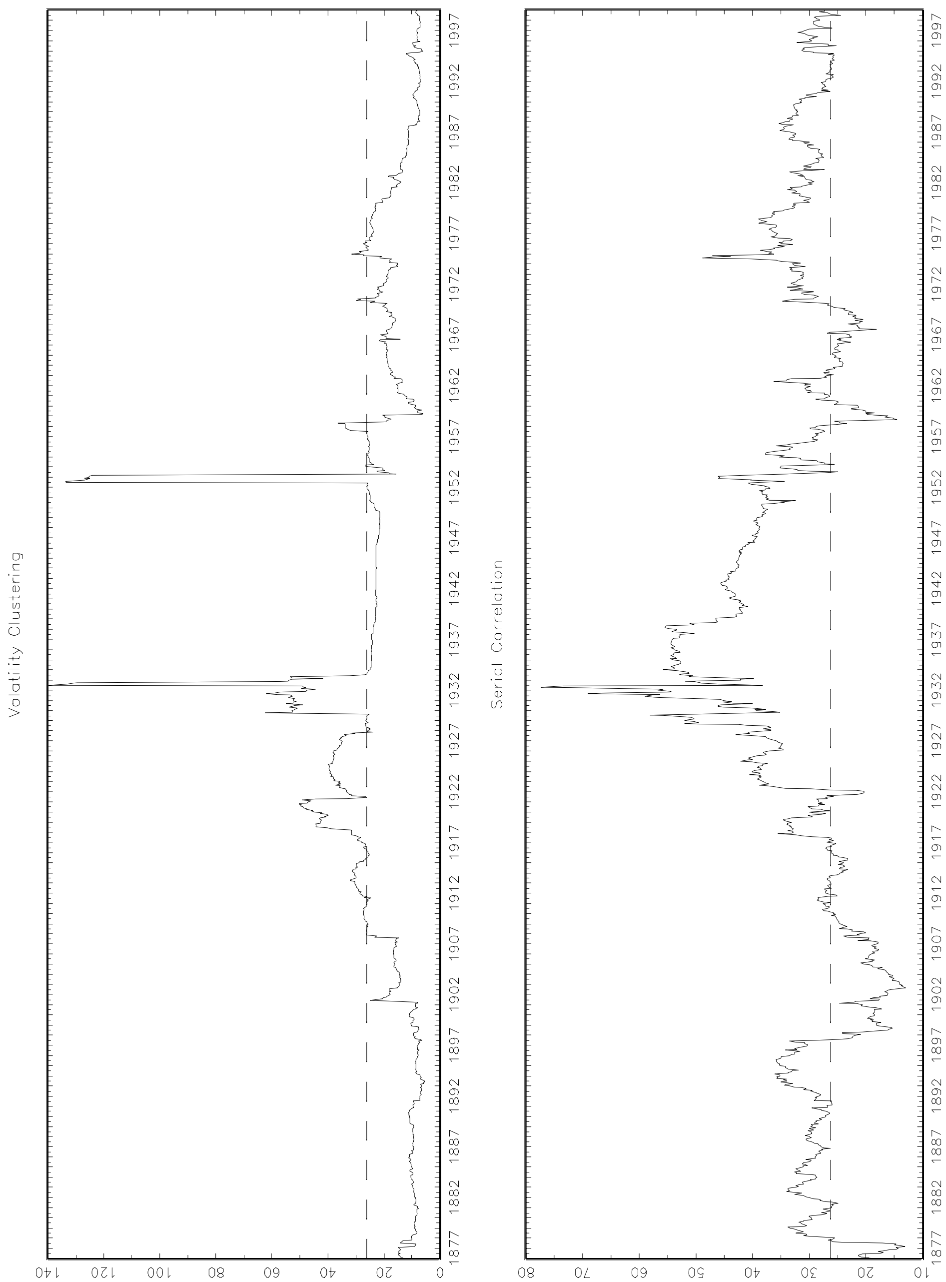

Figure 6: Volatility clustering and serial correlation in US stock returns: Rolling window estimates

Table 1. Estimated number of break points, Bai-Perron Method.

Process	Break point criterion Sequential	AIC	BIC	Break Dates (BIC)
Dividend growth	0	8	4	1911, 1920, 1930, 1938
Absolute dividend growth	1	8	4	1921, 1931, 1939, 1951
Dividend growth w. lag	3	8	1	1917
Absolute dividend growth w. lag	6	8	4	1920, 1931, 1938, 1950

Note: This table reports the number of break points detected by the Bai-Perron (1998) method adopted to the logarithm of the monthly growth in real dividends. The results are based on univariate specifications with an intercept term or an intercept term and a single lag of the dependent variable as regressors.

The sequential approach to determining the number of breaks using a significance level of 5 percent. AIC and BIC are the Akaike and Bayesian penalized likelihood model selection criteria.

Table 2. Estimation Results for the Markov switching models with non-recurring states.

A. Likelihood function

Number of states	Log-likelihood	BIC
1	3534.5	-7054.3
2	4720.7	-9404.7
3	4751.6	-9444.4
4	4788.6	-9496.4
5	4798.0	-9493.2
6	4813.5	-9502.2
7	4870.4	-9593.9
8	4878.2	-9587.5

B. Fitted parameters for the seven state model

	$s = 1$	$s = 2$	$s = 3$	$s = 4$	$s = 5$	$s = 6$	$s = 7$
μ	0.161	-2.327	0.688	-0.048	0.280	-0.254	0.121
(s.e.)	(0.077)	(0.226)	(0.074)	(0.131)	(0.050)	(0.035)	(0.022)
σ	1.825	1.214	0.816	2.071	0.661	0.364	0.374
(s.e.)	(0.054)	(0.16)	(0.052)	(0.093)	(0.035)	(0.025)	(0.016)
p_{ii}	99.82	96.66	99.22	99.60	99.44	99.10	1.00
(s.e.)	(0.18)	(3.31)	(0.79)	(0.40)	(0.56)	(0.93)	NA

Note: Panel A reports the outcome of fitting to the monthly dividend growth rate data a Markov switching model with an expanding set of non-recurring states. Panel B reports the parameter estimates and their standard errors for the selected seven state model.

Table 3. Statistical Properties of Monthly Stock Returns. US data (1871-1999) and simulations.

	Data	Simulations									
		No Break	Break								
Moments of Excess Returns	S&P 500	Full Information	Full Info.	Filtering	Bayesian Learning	Full Info.	Filtering	Bayesian Learning	Full Info.	Filtering	Bayesian Learning
			$\alpha = 0$			$\alpha = 0.5$			$\alpha = 2$		
Standard deviation	0.041	0.015	0.029	0.015	0.033	0.019	0.015	0.022	0.025	0.015	0.017
Skewness	0.82	0.05	8.70	0.05	1.42	1.22	0.05	0.24	5.66	0.05	4.66
Kurtosis	22.42	2.98	304.96	2.98	31.01	62.42	2.98	8.54	215.20	2.99	178.10
Serial Correlation	115.23	0.48	0.22	2.64	0.94	0.43	1.18	0.66	0.51	0.64	1.72
R^2 in Yield Regression	0.005	0.000	0.001	0.003	0.002	0.003	0.004	0.002	0.026	0.017	0.044
ARCH(1)	12.11	1.00	0.28	1.05	7.96	0.55	1.03	6.79	0.43	1.03	0.88
ARCH(4)	68.13	3.92	0.61	4.04	21.84	1.50	4.03	20.82	1.69	4.09	2.99
ARCH(12)	173.32	12.00	3.40	11.91	38.27	5.65	11.90	41.00	4.62	12.03	5.82
Percentage of Simulations with Significant Value of the Diagnostic Statistic											
Serial Correlation		4.9	2.0	38.4	15.6	3.9	21.7	8.8	7.5	9.6	31.2
ARCH(1)		4.3	0.6	5.4	31.8	0.8	4.9	42.5	0.6	5.1	4.2
ARCH(4)		5.0	0.9	6.3	41.5	1.6	6.5	61.6	1.8	7.4	5.5
ARCH(12)		4.9	2.7	3.8	41.2	3.2	4.1	67.1	3.8	4.9	6.1

Note: The first three rows (standard deviation, skewness and kurtosis of excess returns) give the estimates of the first three (centered) moments of the actual and simulated data. Serial correlation is the estimate of the first order Ljung-Box test statistic. R^2 in yield regression is the estimated R^2 from a regression of excess returns on a constant and the lagged dividend yield. The ARCH statistics give the values of the LM test for ARCH suggested by Engle (1982). These are chi-squared distributed with degrees of freedom equal to the order of the ARCH test.

The last four rows provide the percentage of simulations that generated values of a given diagnostic test that were significant at the 5 percent critical level.

The following annualized parameter values were used in the (1000) simulations:

$\pi = .997$, $g = -.04$, $g = .06$, $\rho = .075$, $\sigma = .015$.

Table 4. Method of Moments test of Euler equations

	Full Information	Bayesian learning	Filtering	Full information	Bayesian learning	Filtering	Full information	Bayesian learning	Filtering
		$\alpha = 0$			$\alpha = 0.5$			$\alpha = 2.0$	
Rejection rates									
1%	2.8	3.8	7.6	0.6	2.8	3.6	2.8	3.2	9.6
5%	10.4	11.6	13.2	6.0	7.0	9.8	9.2	8.4	17.4
10%	19.0	16.0	19.0	13.2	13.4	17.0	16.2	14.6	24.8
Mean value of *J*-test	1.6	1.5	3.0	1.2	1.4	1.6	1.5	1.5	4.6

Note: This table reports the outcome of 500 Monte Carlo simulations of the three asset pricing models under full information, Bayesian learning and filtering. Each set of simulations tests the Euler equation using a constant, lagged return and dividend yield as instruments.