We examine a variety of models in which the variance of a portfolio's excess return depends on a state variable generated by a first-order Markov process. A model in which the state is known to economic agents is estimated. It suggests that the mean excess return moves inversely with the level of risk. We then estimate a model in which agents are uncertain of the state. The estimates indicate that agents are consistently surprised by high-variance periods, so there is a negative correlation between movements in volatility and in excess returns.

1. Introduction

Risk-averse agents require compensation for holding risky assets. In a simple two-asset world, where one asset is risky with normally distributed returns and the other is riskless, the nondiversifiable risk is simply the anticipated variance of the excess return above the riskless rate. If the excess return has a constant variance, the risk premium is constant.

The normal-return/constant-variance model of asset prices does not adequately explain the behavior of asset markets such as the stock market. The returns from many assets, including an asset consisting of a portfolio of stocks from any of the common indices, appear to be drawn from nonnormal unconditional distributions [Fama (1963), Mandlebrot (1963)]. In particular,
the empirical distributions of returns from these assets tend to have pronounced peaks and heavy tails [Gallant and Hsieh (1988), Schwert (1987, 1988)].

This shape is typical of unconditional densities of normal observations subject to heteroskedasticity. The sample variance of the density is a weighted average of the variances of the individual observations. It is larger than the smallest variance and smaller than the largest variance. As a result, some observations are drawn from densities with smaller variances than the sample variance; these will be more peaked than a normal density with the sample variance. Likewise, some observations will be drawn from densities with larger variances than the sample variance; these will have more mass in their tails than does a normal with the sample variance. Since the unconditional density of the data is a linear combination of these normal densities, it will have more mass in its peak and tails than the simple normal.

A large literature suggests that the variance of asset prices is not only heterogeneous but also predictable [cf. Bollerslev et al. (1987), Engle et al. (1987), Schwert (1987, 1988)]. Engle and Bollerslev demonstrate the predictability of these variances with a generalized autoregressive conditional heteroskedasticity (GARCH) model. Schwert explores this aspect with an autoregression on squared errors and a Markov model on nominal returns. Pagan and Schwert (1989) compare these and other models. Together, these papers imply that a properly specified model of the risk premium must allow a time-dependent variance with a predictable element. This in turn implies that the risk premium is time-dependent, since future risk moves in a predictable fashion.

We introduce a model of the stock market in which the excess return is drawn from a mixture of two normal densities. In our model the market is assumed to switch between two states. The state in each period determines which of two normal distributions is used to generate the excess return for that period. The states are characterized by the variances of their densities, as a high-variance state and a low-variance state. The state itself is assumed to be generated by a first-order Markov process. This approach was proposed by Hamilton (1989) in a different context. Like Bollerslev et al. (1987), this model leads to a variance that is a function of the variance of prior periods. Our model, however, allows the conditional variance to be a stochastic function of the prior period's variance.1

We use the model to explore the relation between the time-dependent variance and the risk premium in the stock market. We develop two models based on the heteroskedastic structure discussed above. Each incorporates a

1Cechetti, Mark, and Lam (1988, 1989) also use Hamilton's method in modeling stock returns. Although their papers also use a two-state Markov process, they focus on changes in mean real returns, whereas our paper focuses on changes in price volatility and therefore risk.
different assumption about agents' information sets. We use post-World War II data on excess returns from the portfolio of stocks in Standard and Poor's index.

In the first model we assume that economic agents know the realization of the Markov process underlying the state, even though the econometrician does not observe it. There are two risk premiums in this specification. The first is the difference between the mean of the distribution of the low-variance state and the riskless return. Agents require an increase in return over the riskless rate to hold an asset with a random return. The second premium is the added return necessary to compensate for increased risk in the high-variance state. This is the standard model in which agents know the variance extended to the case in which the return on the risky asset has a heterogeneous variance. The parameter estimates from this model suggest that whereas the first risk premium is positive, the second is negative. This result may be due to the imputation of too much knowledge to economic agents.

We assume that neither economic agents nor the econometrician observe the states directly in the second model. In each period, agents form probabilities of each possible state in the following period conditional on current and past excess returns, and use these probabilities in making their portfolio choices. The parameter of interest is the increase in return necessary to compensate the agents for a given percentage increase in the prior probability of the high-variance state. This model yields parameter estimates that are largely consistent with asset pricing theory.

In section 2 we explore the two simple models of the risk premium discussed above. In section 3 we develop the statistical specification of the model. Section 4 discusses maximum-likelihood estimation of the specification. In section 5 we report estimates of the parameters and interpret them. Here we report the full-sample posterior distribution of the state in each period. We also discuss the implications of the parameter estimates for the risk premium.

2. An economic model of excess returns in a two-state world

Consider a two-asset economy. The first asset is riskless, yielding a sure return $r$. The second asset yields a normally distributed return per dollar invested, $q_t$, with time-dependent expectation $\theta_t$ and variance

\[ \sigma_t^2 = \begin{cases} \sigma_0^2 & \text{if } S_t = 0, \\ \sigma_1^2 & \text{if } S_t = 1, \end{cases} \]  

where $S_t$ is an index of the state and $\sigma_1 \geq \sigma_0$. The excess return of the risky asset at time $t$ is then simply $y_t = q_t - r$. The expected value of excess returns is $\mu_t = \theta_t - r$, and its variance is $\sigma_t^2 = \sigma_t^2$. The states, $S_t$, are generated by a
realization of a first-order Markov process with transition probabilities

\[
P(S_t = 1 | S_{t-1} = 1) = p,
\]
\[
P(S_t = 0 | S_{t-1} = 1) = 1 - p,
\]
\[
P(S_t = 0 | S_{t-1} = 0) = q,
\]
\[
P(S_t = 1 | S_{t-1} = 0) = 1 - q.
\]

(2)

The expected value of excess returns, \( \mu_t \), is the premium agents require at time \( t \) for accepting the variance in returns associated with the risky asset. In general, \( \mu_t \) is thought to be positive and to be positively related to the variance \( \sigma_t^2 \). The nature of this relationship, however, depends on the information agents acquire.

2.1. Agents know the states

Assume agents know the realization of the Markov process generating the state, so they know the extent of risk in each period. In this case the excess return is given by

\[
y_t = \mu_t + \xi_t, \quad \xi_t \sim N(0, \sigma_t^2),
\]

(3)

where \( \mu_t \) is the risk premium in time \( t \). It is expected to be positively related to \( \sigma_t^2 \). Because \( \sigma_t^2 \) is a deterministic function of the state, the risk premium \( \sigma_t \) is, too. Thus, the risk premium in each period is simply the mean of the normal distribution determined by that period's state. That is, \( \mu_t = E(y_t | S_t = i) \), \( i = 0, 1 \). Let

\[
\mu_t = \begin{cases} 
\mu_0 & \text{if } S_t = 0, \\
\mu_1 & \text{if } S_t = 1.
\end{cases}
\]

(4)

If agents are risk-averse, we expect that \( \mu_1 \geq \mu_0 \geq 0 \) as \( S_t = 1 \) is the high-variance state.

2.2. Agents are unsure of the states

If agents are unsure of the state, \( S_t \), then the process by which they form their expectations must be specified. Here we assume that agents are unsure of the prevailing state in the past, present, and future. We assume they know the structure generating the states, that is, they know (2) and the parameters of the normal densities from which the excess returns are drawn. Agents base their
buying and selling decisions in period \( t \) on a prior distribution of the state in that period. Each period they update their beliefs about that period's state with current information using Bayes' theorem. Agents' prior distribution of the state in period \( t \) will be based on information through \( t - 1 \).

Let \( \Psi_t \) be the information set through period \( t \). The agents' prior distribution of the state is \( P(S_t = i \mid \Psi_{t-1}) \), \( i = 0,1 \). In period \( t \) they observe \( \Psi_t \) and update their prior distribution using Bayes' theorem

\[
P(S_t = i \mid \Psi_t) = \frac{P(S_t = i \mid \Psi_{t-1}) \times f(\Psi_t \mid S_t = i, \Psi_{t-1})}{f(\Psi_t \mid \Psi_{t-1})},
\]

for \( i = 0,1 \). Here \( f(\Psi_t \mid S_t = i, \Psi_{t-1}) \) is the distribution of the information set conditional on the state of the system, \( f(\Psi_t \mid \Psi_{t-1}) \) is the unconditional distribution of the information set, and \( P(S_t = i \mid \Psi_t) \), \( i = 0,1 \), is the posterior distribution of the state conditional on all the information through period \( t \).

The Markov structure underlying the state ensures that the prior distribution for the state in the following period is simply a linear transformation of the posterior,

\[
P(S_{t+1} = i \mid S_t = j) = \sum_{j=0}^{1} P(S_{t+1} = i \mid S_t = j)P(S_t = j \mid \Psi_t),
\]

for \( i = 0,1 \). \( P(S_{t+1} = i \mid S_t = j) \) is given by the appropriate transition probabilities in (2).

The prior distribution may be summarized by the probability of the high-variance state, \( P(S_t = 1 \mid \Psi_{t-1}) \), without loss of information, since the model has only two states. Agents' portfolio choice may be specified as a simple function of this probability. That is, agents require an increase in the excess return in period \( t \) when faced with an increase in their prior probability that the high-variance state will prevail in that period. We model the risk premium when agents are unsure of the state simply as

\[
\mu_t = \alpha + \gamma P(S_t = 1 \mid \Psi_{t-1}),
\]

where \( \gamma \) is positive. The constant, \( \alpha \), represents agents' required excess return for holding an asset in the low-variance state.

3. Specification

We estimate three specifications based on the models discussed above. The models are estimated on post-World War II monthly returns from a portfolio consisting of the stocks in Standard and Poor's index. The first two specifica-
tions are direct translations from the economic models discussed previously. The third takes into account agents' behavior during the period.

In the model in which the states are known with certainty, no change is necessary for estimation. Eqs. (3) and (4) may be rewritten as

\begin{align}
y_t &= (1 - S_t) \mu_0 + S_t \mu_1 + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2_t), \\
\sigma^2_t &= (1 - S_t) \sigma_0^2 + S_t \sigma_1^2
\end{align}

where \( \mu_0 \) and \( \mu_1 \) are the risk premiums in the low- and high-variance states. \( S_t \) is given by the first-order Markov process with (2) as the transition probabilities. Again, since agents are risk-averse, we expect both \( \mu_0 \) and \( \mu_1 \) to be nonnegative, and since we have normalized the state so that \( \sigma_1^2 \geq \sigma_0^2 \), we expect \( \mu_1 \geq \mu_0 \).

The model in which agents are unsure of the state, eq. (7), may be specified as

\begin{align}
y_t &= \alpha + \gamma P(S_t = 1 \mid \Psi_{t-1}) + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2_t), \\
\sigma^2_t &= (1 - S_t) \sigma_0^2 + S_t \sigma_1^2
\end{align}

The risk premium in period \( t \) is agents' expectation of the excess return conditional on information through period \( t - 1 \). As before, it is \( \alpha + \gamma P(S_t = 1 \mid \Psi_{t-1}) \). It should always be positive and increasing in the anticipated variance, so that we expect both \( \alpha \) and \( \gamma \) to be positive.

The preceding specification assumes that agents are able to trade assets only once each period. With monthly data this assumption should be questioned, as agents may make many trades within each period. At the beginning of period \( t \), agents value their assets on the basis of their prior distribution of the state in that period, \( P(S_t = 1 \mid \Psi_{t-1}) \). During the period agents continue to observe trades. Agents' posterior distribution of the state based on this data will affect the price and return of the asset. Since all we observe is the distribution at the end of the period \( t \), and this is a function of \( y_t \), we cannot include the posterior in our specification of \( y_t \) [Pagan and Ullah (1988)]. Since agents know the structure of the system, we can model their behavior using the true value of the state as a proxy for their posterior distribution. This leads to the specification

\begin{align}
y_t &= (1 - S_t) \alpha_0 + S_t \alpha_1 + \gamma P(S_t = 1 \mid \Psi_{t-1}) + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2_t), \\
\sigma^2_t &= (1 - S_t) \sigma_0^2 + S_t \sigma_1^2
\end{align}
where $S_i$ is generated by the first-order Markov process with (2) as the transition probabilities.

We can sign all the parameters in (10). Our basic premise is that the stock price at time $t$ should reflect all available information. This requires that the price at $t$ should fall below its value at $t-1$ if some new unfavorable information about fundamentals, such as an increase in variance, arrives between the periods. This fall is necessary to assure that the return from time $t$ to $t+1$ is expected to be higher than usual so as to compensate stockholders for added risk. According to this scenario, the return between $t-1$ and $t$ will be negative on average for those periods in which adverse information is newly acquired, and positive on average when favorable information is acquired. This is essentially the same argument as in French, Schwert, and Stambaugh (1987) and Schwert (1989).

In terms of eq. (10), the parameter $\gamma$ on $P(S_i = 1 | \Psi_{t-1})$ represents the effect when agents anticipate as of time $t-1$ that the return of time $t$ will be drawn from the high-variance distribution. According to standard mean-variance theory, foreknowledge of a high-variance should be compensated by a higher expected return. The expected variance in this model is simply

$$
E(\sigma^2_t | \Psi_{t-1}) = P(S_t = 0 | \Psi_{t-1}) \sigma^2_0 + P(S_t = 1 | \Psi_{t-1}) \sigma^2_1.
$$

Thus when $P(S_t = 1 | \Psi_{t-1})$ is high, the expected excess return should be positive so that the parameter $\gamma$ is positive. On the other hand, it could be that today's high-variance state, $S_t = 1$, was not anticipated in the previous period. In this case $P(S_t = 1 | \Psi_{t-1})$ is small so that the average return between $t-1$ and $t$ is dominated by $\alpha_1$ in eq. (10). During a period in which agents are surprised by the event $S_t = 1$, the stock price must fall below what would have been seen had $S_t = 0$ occurred instead. This will make the return between $t-1$ and $t$ lower and will show up as a negative value for $\alpha_t$. Similar reasoning suggests that if the variance unexpectedly decreases, the return between $t-1$ and $t$ will turn out to be higher than usual, suggesting that $\alpha_0$ should be positive.

We may also sign a linear combination of the parameters. The risk premium in $t$, $\mu_r$, is given by the expected value of $\gamma_t$ conditional on the current information set $\Psi_{t-1}$. Thus, the risk premium is

$$
E(\gamma_t | \Psi_{t-1}) = \alpha_0 P(S_t = 0 | \Psi_{t-1}) + (\alpha_1 + \gamma) P(S_t = 1 | \Psi_{t-1}).
$$

If agents are risk-averse, this equation should always be positive and increase with $P(S_t = 1 | \Psi_{t-1})$. The expectation will always be positive as long as $\alpha_0 \geq 0$ and $\gamma + \alpha_1 \geq 0$. Finally, if both of these conditions hold with inequality and
\[ \gamma + \alpha_t > \alpha_0, \text{ then} \]
\[
\frac{dE(\gamma | \Psi_{t-1})}{dP(S_t = 1 | \Psi_{t-1})} = \gamma + \alpha_t - \alpha_0 > 0. \tag{13}
\]

That is, the risk premium will increase with agents' prior probability of the high-variance state.

To complete the model, agents' information set must be specified. In this case, \( \Psi_t = (y_1, y_2, \ldots, y_t) \) for \( t = 1, 2, \ldots, T \). We assume agents observe only past realizations of the market's excess return when forming their prior distribution of the state. This assumption is made simply for convenience. An extension of the model in which agents use other variables in forming their prior is in preparation.

4. Estimation

Models in which observations are chosen from a small set of distributions are not new. In statistics they are called finite-mixture distributions and their estimation is one of the oldest applied problems. Pearson derived the first solution: an application of the method of moments that involved finding the roots of a nonic polynomial. In Pearson's problem and in the statistics literature in general, the distribution governing the state is generally binomial [Everitt and Hand (1981)].

In econometrics the use of finite-mixture distributions is discussed in Goldfeld and Quandt (1973), who call them switching regressions. They suggest that a Markov process can be used to generate the states. More recently, Hamilton (1988, 1989) models the growth rate of nonstationary time series, such as gross national product, subject to occasional discrete shifts in rate of growth or in variance using a Markov process. Specifically he considers models of the same form as (8), though with autoregressive terms common to both states. Schwert (1988) uses Hamilton's model to study the instability of nominal stock-market returns.

Cosslett and Lee (1985) derive the likelihood function for this model. They use the rule of elimination to derive the joint density of the data from the density of the data conditional on the state vector and unconditional distribution of the state vector. Alternatively the likelihood \( f(y_1, y_2, \ldots, y_T) \) may be defined by the forecast error decomposition approach as
\[
f(y_1, y_2, \ldots, y_T) = f(y_1) f(y_2 | y_1) \cdots f(y_T | y_1, y_2, \ldots, y_{T-1}), \tag{14}
\]
where the conditional densities on the right-hand side are defined by
\[
f(y_t | y_1, \ldots, y_{t-1}) = \sum_{i=0}^{1} P(S_t = i | y_1, \ldots, y_{t-1}) f(y_t | S_t = i, x_t). \tag{15}
\]
since \( y_i \) is uncorrelated except for the state and \( x_i \). In the model where agents are assumed to know the state, \( x_i \) is the null vector for all \( t \). In the model where agents are uncertain of the state it is their prior probability of the high-variance state, that is, \( x_i = P(S_i = 1 | y_1, \ldots, y_{t-1}) \). The forecast of the state, that is, \( P(S_i = i | y_1, \ldots, y_{t-1}) \), \( i = 0, 1 \), is easy to obtain in a Markov setting. It is simply

\[
P(S_i = i | y_1, \ldots, y_{t-1}) = \sum_{j=0}^{1} P(S_i = i | S_{i-1} = j) P(S_{i-1} = i | y_1, \ldots, y_{t-1}),
\]

for \( i = 0, 1 \). The latter probability is obtained from Bayes' theorem,

\[
P(S_i = i | y_1, \ldots, y_i) = \frac{P(S_i = i | y_1, \ldots, y_{i-1}) f(y_i | S_i = i, x_i)}{f(y_i | y_1, \ldots, y_{t-1})},
\]

for \( i = 0, 1 \).

Eq. (14) could be maximized directly, but for ease and flexibility we adopt the EM-algorithm. This approach has also been suggested by Hamilton (1989). It consists of three steps: [1] the nomination of starting values, [2] the evaluation of the expectation of the likelihood function, conditional on the current parameter estimates, and [3] the maximization of the log-likelihood's expectation. Steps [2] and [3] are iterated until convergence. For this model the algorithm is essentially iteratively reweighted least squares. The weights for each iteration are simply the probability of the appropriate state conditional on all the data, that is \( P(S_i = i | y_1, y_2, \ldots, y_T) \), \( i = 0, 1 \). Baum et al. (1970) show that each iteration increases the log-likelihood unless a critical point has been reached. The algorithm is discussed in detail, as it relates to this problem, in Turner (1989).

5. Results

We analyze monthly data from Standard and Poor's composite index. This is a value-weighted index of 500 stock prices. The series analyzed is the percentage nominal total return adjusted for dividends less the three-month T-bill rate of return, that is, the monthly excess return of the portfolio times 100. The method of collection is described in Ibbotson and Sinquefield (1976). The estimation period is from January 1946 to December 1987. The results for the model in which agents know the state are presented in tables 1 and 2. Estimates from the models in which agents are uncertain of the state are
presented in table 3. In section 5.1, we assess the implications of the estimated model for heteroskedasticity in excess returns, and in the following section we examine the implications for the risk premium.

5.1. Basic characteristics of the two-state variance model

Our basic hypothesis is that there are two states in the volatility of stock-market returns, that is, the density of excess returns is a mixture of two normals with different variances. Further, the distribution of the state has a time-dependent element. In this section we test the hypotheses of two states and time-dependence.

Unfortunately, under the null hypothesis of only one state, $p$ and $q$ are unidentified parameters. In this circumstance the likelihood-ratio test is not asymptotically distributed $\chi^2$. However, Wolfe (1971) suggests a modified likelihood-ratio statistic for testing the hypothesis of a mixed multivariate normal distribution against the null of simple multivariate normality. In our situation the normals are not multivariate, so the statistic simplifies to

$$\lambda_* = -\frac{2}{T}(T-3)(I_\omega - I_\Omega).$$

where $I_\omega$ is the log-likelihood of the one-state model, the constant-distribution model, in table 1, and $I_\Omega$ is the log-likelihood of the full two-state model, the Markov mean and variance model. This statistic is approximately distributed $\chi^2$ with two degrees of freedom. It's value is 30.57, which is significant at any reasonable level of significance.\(^2\)

Fig. 1 plots the probability of the high-variance states conditional on all the data, $P(S_1 = 1 \mid y_1, y_2, \ldots, y_T)$. These posterior probabilities provide a visual test of the mixture hypothesis. In general if the null hypothesis of simple normality is true, then the plots of these probabilities should indicate uncertainty of the state in most periods. They should be relatively flat and centered at 0.5. There are few periods in the samples in which the probability hovers around 0.5 and the full sample posterior is between 0.20 and 0.80 in only 7% of the sample.

Statistically, the Markov mean and variance model requires that states be characterized by different means and/or different variances. The variances in the two states are very well defined. The variance in state $S_1 = 1$, the high-variance state, is more than four times that in the low-variance state. As discussed below, the standard error of this parameter and all the high-state parameters is

\(^2\)Simulations by Everitt [Everitt and Hand (1981)] show that this test has low power for models in which $\sigma_1^2 = \sigma_2^2 = \sigma^2$ unless $(|\mu_1 - \mu_2|/\sigma^2) > 2$. However, since we can easily reject, low power does not appear to be a problem for our data.
Excess Return from Standard & Poor's 500 Stock Portfolio
Posterior Probability of the High-Variance State

Fig. 1. The top figure is a time series plot of the monthly excess return from Standard and Poor's stock index from January 1946 through December 1987. The bottom figure is drawn assuming each period's excess return is taken from one of two normal densities with different means and variances according to an unobserved dichotomous state variable $S_t$. If $S_t = 1$, then the return is taken from the high-variance density, otherwise it is drawn from the low-variance density. The bottom figure plots the econometrician's probability that the excess return in that month is taken from the high-variance density, $P(S_t = 1 | y_1, y_2, \ldots, y_T)$. 
Table 1

Estimation results for the state-dependent model of the excess return in the stock market.

Estimation results for the excess return of Standard and Poor's index for eqs. (2) and (8). The variable $S_t$ is an unobserved dichotomous state variable generated by a first-order Markov process with transition probabilities given by (2). This is a model of the risk premium for the case in which agents know the state in each period, current and future. Asymptotic standard errors are in parentheses. The statistic $Q_{12}$ is the Box–Pierce portmanteau statistic on the forecast error of the variance of $y_t$. Under the null hypothesis these errors are white noise.

Sample period: January 1946 to December 1987

Observations: 504

$y_t = \mu_0(1 - S_t) + \mu_1 S_t + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2_t)$  \hspace{1cm} (8)

$\sigma^2_t = \sigma_0^2 (1 - S_t) + \sigma_1^2 S_t$

$P(S_t = 1 | S_{t-1} = 1) = \rho, \quad P(S_t = 0 | S_{t-1} = 0) = q$ \hspace{1cm} (2)

<table>
<thead>
<tr>
<th>Model</th>
<th>Low-state variance $\sigma_0^2$</th>
<th>High-state variance $\sigma_1^2$</th>
<th>Low-state mean $\mu_0$</th>
<th>High-state mean $\mu_1$</th>
<th>High-state transition probability $p$</th>
<th>Low-state transition probability $q$</th>
<th>Log-likelihood $l$</th>
<th>$Q_{12}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant distribution</td>
<td>17.6965</td>
<td>0.5983</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-1438.73</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.1148)</td>
<td>(0.1874)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Markov mean</td>
<td>16.5874</td>
<td>0.6445</td>
<td>-21.6888</td>
<td>0.0000</td>
<td>0.9979</td>
<td></td>
<td>-1430.83</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.0439)</td>
<td>(0.1809)</td>
<td>(4.2933)</td>
<td>(0.2631)</td>
<td>(0.0016)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Markov variance</td>
<td>14.7658</td>
<td>0.6723</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-1424.64</td>
<td>9.0075</td>
</tr>
<tr>
<td></td>
<td>(1.0631)</td>
<td>(0.1756)</td>
<td>(4.2933)</td>
<td>(0.1951)</td>
<td>(0.0083)</td>
<td></td>
<td></td>
<td>(0.70)</td>
</tr>
<tr>
<td>Markov mean and variance</td>
<td>14.2480</td>
<td>0.7798</td>
<td>-1.9380</td>
<td>0.7892</td>
<td>0.9839</td>
<td></td>
<td>-1423.31</td>
<td>9.4372</td>
</tr>
<tr>
<td></td>
<td>(1.3612)</td>
<td>(0.2600)</td>
<td>(0.6745)</td>
<td>(0.2085)</td>
<td>(0.0143)</td>
<td></td>
<td></td>
<td>(0.72)</td>
</tr>
</tbody>
</table>

Constant distribution: $\mu_1 = \mu_0$, $\sigma_1^2 = \sigma_0^2$

Markov mean: $\sigma_1^2 = \sigma_0^2$

Markov mean and variance: $\mu_1 = \mu_0$
quite large. We can test the hypothesis $\sigma_1^2 = \sigma_0^2$, while letting $\mu_1 \neq \mu_0$. This is a likelihood ratio test of the Markov mean and variance model against the Markov mean model. We reject the null at a reasonable level of significance: the statistic is 15.04. Tests of the difference between the means of the two distributions yield conflicting tests. This is a test of the hypothesis that $\mu_1 = \mu_0$ against the alternative, while letting $\sigma_1^2 \neq \sigma_0^2$. This yields a likelihood ratio statistic of 2.66, not significant at the 5% level. The $t$-statistic for this hypothesis, 5.29, is significant, however.

The estimates of the transition probabilities suggest the low-variance state will dominate. The estimates of the transition probabilities, $p$ and $q$, of the Markov process suggest that the stationary, or unconditional, probabilities of the states will be 0.9290 and 0.0710 for the low- and high-variance states. Thus, for any given sample only about 7% of the observations will be expected to fall into the high-variance state.

The improbability of the high-variance state makes inference on high-variance state parameters difficult. The sample size in estimating these parameters depends, of course, on the number of observations that fall into the high-variance state. Fig. 1 shows that there are relatively few of these periods. More formally, the sample size in estimating these parameters is effectively the sum of the weights, $\sum P(S_t = 1 | y_1, \ldots, y_T)$ or 33 [Turner (1989)]. Thus, because the high-variance state is likely in only a relatively few periods, we will not be able to estimate any of the parameters associated with it precisely.

The point estimates of the probabilities $p$ and $q$ suggest a strong time-dependence in the Markov process generating the states. The large standard error of $\hat{p}$ suggests, however, that $p$ can lie almost anywhere between 0 and 1. This possibility makes a formal test of time-dependence in the model particularly interesting. Recall that a binomial process is simply a Markov process with $p = 1 - q$. A binomial process removes the dependence of the current state's probability distribution on past states. We may reject this hypothesis: the $t$-statistic is 3.51.

The point estimate of $p$ suggests that once in the high-variance state, the state is expected to persist. Since $\hat{p}$ is greater than 0.5, the high-variance state is expected to persist for at least two periods. More specifically, we wish to find the smallest value of $j$ for state $i$ that satisfies

$$P(S_{t+j} = i, S_{t+j-1} = i, \ldots, S_{t+1} = i | S_t = i) < 0.5. \quad (19)$$

For any $j$ satisfying (19) the probability of remaining in state $i$ for $j$ consecutive periods is less than one-half. This is the median duration in state $i$ of the process. Once in the high-variance state the median number of periods the system will remain there is approximately three months. The low-variance

---

3The parameter estimates of the latter model depend largely on the October 1987 crash.
High-variance periods in the excess return of the stock market.

The excess return of Standard and Poor's index is assumed to be drawn from one of two normal densities with different means and variances. The density is chosen each period by an unobserved dichotomous state variable $S_t$. If $S_t = 1$ the return is drawn from the high-variance density, otherwise it is drawn from the low-variance density. This table describes the posterior distribution of the state conditional on all the data in the full sample. The first column lists the dates of the periods in which the probability of the high-variance state exceeded one half. The second column lists the length of these periods. The last column lists the lengths of intervening periods in which the probability of the low-variance state exceeded one half.

<table>
<thead>
<tr>
<th>High-variance episodes</th>
<th>Length of high-variance episodes (months)</th>
<th>Length of low variance episodes (months)</th>
</tr>
</thead>
<tbody>
<tr>
<td>May to June 1962</td>
<td>2</td>
<td>196</td>
</tr>
<tr>
<td>April 1970</td>
<td>1</td>
<td>93</td>
</tr>
<tr>
<td>June 1974 to January 1975</td>
<td>8</td>
<td>151</td>
</tr>
<tr>
<td>September to December 1987</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>3.8</td>
<td>97.7</td>
</tr>
<tr>
<td>Median</td>
<td>3.0</td>
<td>93.0</td>
</tr>
</tbody>
</table>

state is much more persistent, with an estimated median duration of approximately 43 months. Since $q$ is near unity, small changes in it imply large changes in the minimum value of $j$ satisfying (19).

The persistence of the states and general behavior of the stock market are described in a nonparametric way in fig. 1 and table 2. They summarize the posterior distribution of the state in each period, conditional on the entire sample used in estimation. Generally, this plot and table indicate that both states are persistent, and that the low-variance state is very persistent, having existed, in expectation, without a break from January 1946 until May 1962. Further, the probability of the high-variance state does not exceed 0.2 during the 1950's. This period heavily influences the estimates of $p$ and $q$.

For each model estimated in which the variance is assumed to change between states the statistic $Q_{12}$ is reported (with p-values below) in tables 1 and 3. This is the Box–Pierce portmanteau statistic for the residuals from the regression

$$y_t^2 = \beta_0 + \beta_1 P(S_t = 1 | y_1, \ldots, y_{t-1}) + u_t.$$  \hspace{1cm} (20)

If the Markov model is correctly specified, then $P(S_t = 1 | y_1, \ldots, y_{t-1})$ should be the optimal forecast of the state, and therefore the variance of $y_t$. Since the best estimate we have of the variance is $y_t^2$, under the null the residuals $u_t$ should be white noise [Pagan and Schwert 1989]. Otherwise we could construct a better forecast of $y_t^2$. This statistic is insignificant in each case which
indicates that our model satisfactorily represents the correlation in the volatility of the data.4

5.2. Implications for the risk premium

Estimates of the Markov mean and variance model, in which agents are assumed to know the state, do not support an increasing risk premium. The parameter estimates indicate that agents require an increase in annual return over T-bills of approximately 10% to hold the risky asset in low-variance periods. The estimates also suggest, however, that the premium declines as the level of risk increases, that is, \( \hat{\mu}_1 < \hat{\mu}_0 \). Further, not only is \( \hat{\mu}_1 \) significantly less then \( \hat{\mu}_0 \), it is also significantly negative. We can reject the hypothesis of a risk premium increasing in the variance. These parameter estimates are in general agreement with those found by Schwert (1988) in his analysis of nominal returns from stocks using Hamilton's (1989) autoregressive model.5 The results are also consistent with French, Schwert, and Stambaugh's (1987) finding of a negative correlation between innovations in the volatility of the stock market and the return, though they employ a two-step procedure.

Misspecification is a likely explanation for this result. If agents are uncertain of the state, so that they are basing their decisions on forecasts of the state in the following period, as in the uncertainty model given by (9), estimates assuming they know the state with certainty will be inconsistent. Agents' expectation, or forecast, of the state is \( P(S_t = 1 | \Psi_{t-1}) \). If agents are uncertain of the state, then the Markov mean and variance model, which assumes certainty, suffers from the usual errors-in-variables problem. The forecast error, \( S_t - P(S_t = 1 | \Psi_{t-1}) \), is included on the right side in (8). Alternatively, French, Schwert, and Stambaugh (1987) and others argue that a negative relation between the level of excess returns and their volatility may be explained by changing financial leverage.

The parameter estimates for the uncertainty model, in which agents are uncertain of the state in the variance, but do not act on information obtained during the period, are described in table 3. They provide support for a risk premium rising as the anticipated level of risk rises. In this model, the level of risk is measured by the probability of the high-variance state. This model is similar in spirit to a GARCH-M model in that it does not allow current information to affect the excess return. Our results are consistent with these models [French, Schwert, and Stambaugh (1987), Engle, Lilien, and Robins (1987)]. If the economic agents are certain next period's return will be drawn

---

4 The Box-Pierce statistic has been adjusted for heteroskedasticity. We use the expression \( \sum_{t=1}^{12} (\hat{\sigma}_t^2 / \hat{\sigma}_k^2) \) for \( Q_{12} \), where \( \hat{\sigma}_k^2 \) is White's heteroskedasticity-consistent estimate of the variance of \( \hat{\mu}_k \) [White (1980), Pagan and Schwert (1989)].

5 Schwert's analysis is based on a different data set. He uses stock prices beginning in the mid-nineteenth century. His sample includes the Great Depression. In the same paper he notes that this period was characterized by a much higher variance than the postwar period; hence his estimate for \( p \) is much higher than our estimate.
Table 3

Estimation results for the state-dependent model of the risk premium in the stock market when the state is unknown to the agents.

Estimation results for the excess return of Standard and Poor's index for eqs. (2) and (10). The variable $S_t$ is an unobserved dichotomous state variable generated by a first-order Markov process with transition probabilities given by (2). In these models, agents are unsure of the state, $S_t$. In the uncertain model, agents make trades based on a prior distribution of the state using last period's excess return. In the learning model, they make trades based on this prior and on a posterior distribution using trades during the period. In both models, the parameter $\gamma$ indicates the effect of the prior probability that this period's return will be drawn from the high-variance density on the excess return. Asymptotic standard errors are in parentheses. The statistic $Q_{12}$ is the Box–Pierce portmanteau statistic for the forecast error of the variance of $\eta_t$. Under the null hypothesis these errors are white noise.

Sample period: January 1946 to December 1987
Observations: 504

\[
y_t = \alpha_0(1 - S_t) + \alpha_1 S_t + \gamma P(S_t = 1 | \eta_1, \eta_2, \ldots, \eta_{t-1}) + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2 \epsilon)
\]

\[
\sigma^2 = \alpha_0^2(1 - S_t) + \alpha_1^2 S_t
\]

\[
P(S_t = 1 | S_{t-1} = 1) = p, \quad P(S_t = 0 | S_{t-1} = 0) = q
\]

<table>
<thead>
<tr>
<th>Model</th>
<th>Low-state variance $\sigma^2_0$</th>
<th>High-state variance $\sigma^2_1$</th>
<th>Low-state intercept $\alpha_0$</th>
<th>High-state intercept $\alpha_1$</th>
<th>$\gamma$</th>
<th>$p$</th>
<th>$q$</th>
<th>Log-likelihood</th>
<th>$Q_{12}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uncertainty</td>
<td>14.6372 (1.0255)</td>
<td>100.9216 (28.6918)</td>
<td>0.4391 (0.7754)</td>
<td>8.5079 (2.4131)</td>
<td>0.7008 (0.1770)</td>
<td>0.9891 (0.0080)</td>
<td>-1422.46 (0.16)</td>
<td>8.1225</td>
<td></td>
</tr>
<tr>
<td>Learning</td>
<td>13.6507 (1.2090)</td>
<td>46.6369 (11.6701)</td>
<td>0.7028 (0.7205)</td>
<td>-3.3663 (1.3412)</td>
<td>2.8834 (1.3327)</td>
<td>0.7818 (0.1994)</td>
<td>0.9800 (0.0155)</td>
<td>-1420.52 (0.24)</td>
<td>8.7925</td>
</tr>
</tbody>
</table>

Uncertainty: $\alpha_1 = \alpha_0$
C.M. Turner et al., Heteroskedasticity, risk, and learning

from the low-variance density, then \( P(S_t = 1 | y_1, \ldots, y_{t-1}) = 0 \). Hence, agents anticipate a monthly return of \( \alpha \) percent. Given the parameter estimates in table 3, this translates into an annual return of 5\%. Likewise, if agents are certain next period's return will be drawn from the high-variance density, then \( P(S_t = 1 | y_1, \ldots, y_{t-1}) = 1 \). According to eq. (7), agents will require a monthly return of \( \alpha + \gamma \) percent, or 180\% annually. These estimates suggest that agents perceive stocks to be a very risky asset during high-variance periods. The unconditional probability of the high-variance state is only 0.0352, however. Following (7), this suggests the risk premium will average approximately 9\% on an annual basis. This number is close to the average excess return, 7.5\%.

The learning model given by (10) generalizes the case in which agents are unsure of the state to allow learning during the period. The model also suggests the direction of bias of the estimate of the risk premium when agents are assumed to know the state. Estimating the Markov mean and variance model under this regime would yield parameter estimates that smear the risk premium and the news effect. Since an increasing risk premium and the news effect have opposite effects on excess returns, we expect \( \hat{\mu}_0 \) to be an upwardly biased estimate of the risk premium and \( \hat{\mu}_1 \) to be a downwardly biased estimate.

Further evidence consistent with a news effect may be found in the conditional probabilities derived from the parameter estimates of the Markov mean and variance model. If agents are risk-averse and learn about the state during the period, we would expect high-variance episodes to be marked first by a sharp drop in the stock market and later by a strong rise. During the early part of the episode, investors do not recognize the change in regime until after it has taken place. Agents revalue their stock holdings downwards on the basis of this news, so the excess return for these periods should be negative. Recall that \( S_t \) is our proxy variable for agents' posterior distribution based on intramonth observations and \( P(S_t = 1 | y_1, \ldots, y_T) \) summarizes all our knowledge of \( S_t \). Then the average return for those months in which agents learn the market is in a high-variance state should be negative. That is

\[
\sum_t P(S_t = 1 | y_1, y_2, \ldots, y_T) y_t / \sum_t P(S_t = 1 | y_1, y_2, \ldots, y_T) < 0. \quad (21)
\]

For the probabilities derived from the Markov mean and variance model this expression is \(-1.9380\). It is more than twice its approximate standard error.

\^6The left side of inequality (21) is identical to the expression used in the EM-algorithm for \( \hat{\mu}_1 \). We approximate the variance using the expression

\[
\sum_t P(S_t = 1 | y_1, y_2, \ldots, y_T) \bar{s}^2_t / \left( \sum_t P(S_t = 1 | y_1, y_2, \ldots, y_T) \right)^2.
\]

where \( \bar{s}^2_t \) is given in eq. (11). The variance of the left side of (22) was found analogously. This formulation allows us to approximate the covariance between the left side of (21) and that of (22) and the unweighted mean.
0.7550. The unweighted mean of the excess return over the sample period is 0.5983, with a standard error of 0.0343. Expression (21) is significantly less than the unweighted mean at any reasonable significance level. The t-statistic is $-3.69$. Likewise, for those periods in which agents have forecast returns drawn from the high-variance density, we expect a positive return. That is, we expect the weighted mean, where the weights are agents' probabilities of the high-variance state prior to the period, to be positive. We expect

$$\sum_t P(S_t = 1 | y_1, y_2, \ldots, y_{t-1}) y_t / \sum_t P(S_t = 1 | y_1, y_2, \ldots, y_{t-1}) > 0.$$  

(22)

The value of (22) is 0.9836. Its approximate standard error is 0.5025. It is significantly different from zero at the 5% level, and significantly greater than (21) at a reasonable level: the t-statistic is 5.87. It is not, however, significantly greater than the unweighted mean. These estimates suggest the stock market does in fact fall when it unexpectedly moves into a high-variance episode. When it is expected to be in the high-variance state, however, it does not fall.

The parameter estimates from the learning model indicate that the model has sorted out the risk premium and the news effect. In general, the signs on the parameters are as predicted in section 2.3 and suggested above. The parameters $\gamma$ and $\alpha_0$ are significantly greater than zero, while $\alpha_1$ is significantly negative. The latter two parameters are also significantly different from each other: the t-statistic is 3.01. The news parameters, $\alpha_0$ and $\alpha_1$, capture the effect of the state on the stock's return during the period - the high-variance state is bad news and the low-variance state is good news. The estimates predict that the excess return will fall by more than 33% on an annual basis if the market makes an unforeseen jump into the high-variance state, that is, if $S_t = 1$, while $P(S_t = 1 | y_1, \ldots, y_{t-1}) = 0$. This result is qualitatively consistent with that in French, Schwert, and Stambaugh (1987).

Recall from (12) that the risk premium is given by the expectation of the excess return. Section 2.3 showed that the risk premium is increasing with the anticipated variance if the derivative, (13), is positive. This is true if $\gamma + \alpha_1 > \alpha_0$. The point estimates indicate that the risk premium does not increase with the anticipated variance. The above inequality is reversed for the estimates from the learning model. The variance of the linear combination $\hat{\gamma} + \hat{\alpha}_1 - \hat{\alpha}_0$ is large in relation to the point estimate, the t-statistic is $-0.21$, so that the model provides no evidence for a risk premium changing with or against the variance. As before, this result is consistent with that of French, Schwert, and Stambaugh (1987): they also find little direct evidence of a relation between the risk premium and volatility. As with the uncertainty model, the unconditional probability of the high-variance state may be used to derive the average
risk premium. In this case this probability is 0.0838. Using this value and (12), the risk premium is predicted to average approximately 7.5% per year.

Examining fig. 1 closely, one can see that whenever the stock market enters the high-variance state, it falls. In the next period it generally recovers what it has lost. The parameter estimates summarize this tendency. Our model provides a basis for understanding this behavior: the probability that a high-variance period will follow a low-variance period is quite small, so that agents are always surprised. Our parameter estimates also indicate that in the following periods, the probability of remaining in the high-variance state is quite large, so that agents generally anticipate it, and collect their risk premium.

6. Conclusion

We show that an adequate model of the excess return from the stock market may be constructed with a mixture of normal densities with different means and variances. The heteroskedasticity this mixture implies has a strong time-dependence, suggesting that the conditional variance of the market can be forecasted.

This result suggests that the risk premium will move over time in response to agents' perception of the market's riskiness. Agents do not always forecast the market's variance accurately, so information about the state gained during periods is important to explaining the overall return.

References


Cecchetti, Stephen G., Pok-sang Lam, and Nelson C. Mark. 1988. Mean reversion in equilibrium asset prices. Unpublished manuscript (Ohio State University, Columbus, OH).

Cecchetti, Stephen G., Pok-sang Lam, and Nelson C. Mark. 1989. The equity premium and the risk free rate: Matching the moments. Unpublished manuscript (Ohio State University, Columbus, OH).


