NBER WORKING PAPER SERIES

A MARKOV MODEL OF HETEROSKEDASTICITY, RISK, AND LEARNING IN THE STOCK MARKET

Christopher M. Turner

Richard Startz

Charles R. Nelson

Working Paper No. 2818

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 1989

Nelson's participation was supported in part by the Center for the Study of Banking and Security Markets University of Washington. This research is part of NBER's research program in Financial Markets and Monetary Economics. Any opinions expressed are those of the authors not those of the National Bureau of Economic Research.

NBER Working Paper #2818
January 1989

A MARKOV MODEL OF HETEROSKEDASTICITY, RISK, AND LEARNING IN THE STOCK MARKET

ABSTRACT

Risk premia in the stock market are assumed to move with time varying risk. We present a model in which the variance of the excess return of a portfolio depends on a state variable generated by a first-order Markov process. A model in which the realization of the state is known to economic agents, but unknown to the econometrician, is estimated. The parameter estimates are found to imply that the risk premium declines as the variance of returns rises. We then extend the model to allow agents to be uncertain about the state. Agents make their decisions in period t using a prior distribution of the state based only on past realizations of the excess return through period $t-1$ plus knowledge of the structure of the model. The parameter estimates from this model are consistent with asset pricing theory.

Christopher M. Turner
Department of Economics
DK-30
University of Washington
Seattle, WA 98195

Richard Startz
Department of Economics
DK-30
University of Washington
Seattle, WA 98195

Charles R. Nelson
Department of Economics
DK-30
University of Washington
Seattle, WA 98195

1.0 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 while 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 then the risk premium is constant.

The normal return/constant variance model of asset prices does not provide an adequate explanation of 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 stock market indices, appear to be drawn from non-normal unconditional distributions (Fama, 1961). In particular the empirical distributions of returns from these assets tend to have a pronounced peaks and heavy tails. (Gallant 1988; Schwert, 1987, 1988). This is demonstrated for returns from a portfolio consisting of stocks from Standard and Poor's index by the histogram shown in Figure (1). In this distribution 39% of the probability mass of the empirical density lies within one–half of a standard deviation from the mean, 30% more than in the normal density.

This shape is typical of unconditional densities of normal observations subject to heteroskedasticity. The sample variance of the density will be a weighted average of the variances of the individual observations. It will be 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 then does a normal with the sample variance. As 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 heterogenous but also is predictable, c.f. Bollerslev, *et. al.* (1987), Mandlebroit (1963), Engle, *et. al.* (1987), Schwert (1987, 1988). Engle and Bollerslev demonstrate the predictability of these variances with an autoregressive conditional heteroskedasticity model and a generalization. Schwert explores this aspect with an autoregression on squared errors and a Markov model on nominal returns. Their conclusions 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 will be 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 stock market is assumed to switch between two states. The state of the market 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. There is 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 first proposed by Hamilton (1987A, 1987B) in a different context. Like Bollerslev (1987), this model leads to a variance which is a function of the variance of prior periods. However, our model will allow the conditional variance to be a stochastic function of the prior period's variance. This model will allow us to examine both the heteroskedasticity of excess returns, and their time dependence.

We use the model to explore the relationship between the time dependent variance and the risk premium in the stock market. We will develop two models based on the heteroskedastic structure discussed above. Each model will be based on a different assumption on agents' information sets. We will estimate each model using postwar data from excess returns based on a portfolio of stocks in Standard and Poor's index.

In the first model we will simply assume that economic agents know the realization of the Markov process underlying the generation of states, even though the econometrician does not observe the state. There are two risk premia in this specification. The first is simply the difference between the mean of the distribution of the low variance state and the riskless return. Agents will require an increase in return over the riskless asset to hold an asset with a random return. The second premium is given by the difference between the mean of the distribution of the high variance state and that of the low variance state. This is the added return necessary to compensate for increased risk in the high variance state. Note that this is the standard model where agents know the variance extended to the case when the return on the risky asset has a heterogenous variance.

We will assume that neither economic agents nor the econometrician observe the states directly in the second model. Each period they form probabilities of each possible state in the following period conditional on current and past excess returns. They use these probabilities in making their portfolio choices in those periods. 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.

In *Section 2.2* we explore the two simple models of the risk premium discussed above. In *Section 2.3* we develop the statistical specification of the model. *Section 2.4* discusses maximum likelihood estimation of the specification. This is further developed in *Appendix A*. In *Section 2.5* we report estimates of the parameters and interprete them. Here we report the full sample posterior distribution of the state in each period.

2.0 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_t. The second asset yields a normally distributed return per dollar invested q_t with time dependent expectation θ_t and variance

$$v_t = \begin{cases} v_0, & \text{if } S_t = 0 \\ v_1, & \text{if } S_t = 1 \end{cases} \tag{1}$$

where S_t is an index of the state and where $v_1 \geq v_0$. The excess return of the risky asset at time t is then simply $y_t = q_t - r_t$. The expected value of excess returns is then $\mu_t = \theta_t - r_t$, while its variance is $\sigma_t^2 = v_t$. The states, S_t, are generated by a realization of a first order Markov process with transition probabilities

$$\begin{aligned} &\mathrm{P}(S_t = 1|S_{t-1} = 1) = p \\ &\mathrm{P}(S_t = 0|S_{t-1} = 1) = 1 - p \\ &\mathrm{P}(S_t = 0|S_{t-1} = 0) = q \\ &\mathrm{P}(S_t = 1|S_{t-1} = 0) = 1 - q. \end{aligned} \tag{2}$$

The expected value of excess returns, μ_t, is the premium agents require at time t for accepting the variance in returns associated with the risky asset. In general, μ_t is thought to be positive and to be positively related to the variance σ_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 states, thus they know the extent of risk in each period. In this case the excess return will be given by

$$y_t = \mu_t + \zeta_t, \qquad \zeta_t \sim N(0, \sigma_t^2) \tag{3}$$

where μ_t is the risk premium in time t. It is expected to be positively related to σ_t^2. Note that σ_t^2, is a deterministic function of the state hence the risk premium μ_t will also be a deterministic function of the state. Thus, the risk premium in each period is simply the mean of the normal distribution determined by that period's state. That is, $\mu_t = \mathrm{E}(y_t|S_t = i)$, $i = 0, 1$. Letting,

$$\mu_t = \begin{cases} \mu_0, & \text{if } S_t = 0 \\ \mu_1, & \text{if } S_t = 1. \end{cases} \tag{4}$$

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 agents form their expectations must be specified. Here we will assume that agents are unsure of the prevailing state in the past, present, and future. We assume agents know the structure generating the states, i.e. they know equations (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' rule. Agents' prior distribution of the state in period t will be based on information through $t - 1$.

Let Φ_t be the information set through period t, then agents' prior distribution of the state is $\mathrm{P}(S_t = i|\Phi_{t-1})$, $i = 0, 1$. In period t they observe Φ_t and update their prior distribution using Bayes theorem

$$\mathrm{P}(S_t = i|\Phi_t) = \frac{\mathrm{P}(S_t = i|\Phi_{t-1}) \times f(\Phi_t|S_t = i, \Phi_{t-1})}{f(\Phi_t|\Phi_{t-1})} \tag{5}$$

for $i = 0, 1$. Here $f(\Phi_t|S_t = i, \Phi_{t-1})$ is the distribution of the information set conditional on the state of the system, $f(\Phi_t\ \Phi_{t-1})$ is the unconditional distribution of the information set, and $\mathrm{P}(S_t = i|\Phi_t)$, $i = 0, 1$ 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

$$\mathrm{P}(S_{t+1} = i|\Phi_t) = \sum_{j=0}^{1} \mathrm{P}(S_{t+1} = i|S_t = j)\mathrm{P}(S_t = j|\Phi_t) \tag{6}$$

for $i = 0, 1$. $\mathrm{P}(S_{t-1} = i|S_t = j)$ is given by the appropriate transition probabilities in equations (2).

The prior distribution may be summarized by the probability of the high variance state, $P(S_t = 1|\Phi_{t-1})$ 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 state in that period will be the high variance state. We model the risk premium, when agents are unsure of the state, as simply

$$\mu_t = \alpha + \beta P(S_t = 1|\Phi_{t-1}) \tag{7}$$

where β is positive. The constant, α, represents agents' required excess return for holding an asset in the low variance state.

3.0 Specification

We will estimate three specifications based on the models discussed above. The models will be estimated on postwar monthly returns from a portfolio consisting of the stocks in Standard and Poor's index. The first two specifications will be direct translations from the economic models discussed previously. The third will take into account agents behavior during the period.

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

$$\begin{aligned} y_t &= (1 - S_t)\mu_0 + S_t\mu_1 + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma_t^2) \\ \sigma_t^2 &= (1 - S_t)\sigma_0^2 + S_t\sigma_1^2 \end{aligned} \tag{8}$$

where μ_0 and μ_1, are the risk premia in the low and high variance states, respectively. S_t is given by the first–order Markov process with equations (2) as the transition probabilities. Again, since agents are risk averse we expect both μ_0 and μ_1 to be non–negative and $\mu_0 \geq \mu_1$.

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

$$\begin{aligned} y_t &= \alpha + \beta P(S_t = 1|\Phi_{t-1}) + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma_t^2) \\ \sigma_t^2 &= (1 - S_t)\sigma_0^2 + S_t\sigma_1^2. \end{aligned} \tag{9}$$

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

The above specification of the model assume that agents are only able to trade assets 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 based on their prior distribution of the state in that period, $P(S_t = 1|\Phi_{t-1})$. During the period agents continue to observe trades. Agents' posterior distribution of the state based on this data will effect the price and return of the asset. Since all we observe is the posterior 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 agents' posterior distribution. This leads to the specification

$$\begin{aligned} y_t &= (1 - S_t)\alpha_0 + S_t\alpha_1 + \gamma P(S_t = 1|\Phi_{t-1}) + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma_t^2) \\ \sigma_t^2 &= (1 - S_t)\sigma_0^2 + S_t\sigma_1^2 \end{aligned} \tag{10}$$

where S_t is generated by the first-order Markov process with equations (2) as the transition probabilities.

We can sign all the parameters in equations (10). Agents react to an increase in the anticipated variance in time t by decreasing the asset's value at the beginning of the period. Since σ_1^2 is the high variance, and the anticipated variance is given by

$$E(\sigma_t^2|\Phi_{t-1}) = P(S_t = 1|\Phi_{t-1})\sigma_1^2 + P(S_t = 0|\Phi_{t-1})\sigma_0^2 \tag{11}$$

we expect γ to be positive. Equivalently we may note that γ expresses the increase in excess return risk averse agents require to hold the risky asset for a given increase in their prior probability that this period's state will be the high variance state. If in period t the true state is the low variance state, the return of the asset will rise as agents realize this is the case and alter their portfolios, in favor of the risky asset. This behavior drives its price up at the end of t relative to the asset's price at the beginning of the period. We expect α_0 to be positive. Likewise, if the true state is the high variance state, the return of the asset will fall as agents become convinced this is the case and revalue it downwards. Thus, α_1 should be be negative.

We may also sign a linear combination of the parameters. Note that the risk premium in t, μ_t, is given by the expected value of y_t conditional on the current information set Φ_{t-1}. Thus, the risk premium is

$$E(y_t|\Phi_{t-1}) = \alpha_0 P(S_t = 0|\Phi_{t-1}) + (\alpha_1 + \gamma)P(S_t = 1|\Phi_{t-1}) \tag{12}$$

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

$$\frac{d\mathrm{E}(y_t|\Phi_{t-1})}{d\mathrm{P}(S_t = 1|\Phi_{t-1})} = \gamma \cdot \alpha_1 - \alpha_0 > 0, \tag{13}$$

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, $\Phi_t = (y_1, y_2, ..., y_t)$, for $t = 1, 2, ..., T$. We assume agents only observe past realizations of the excess return of the stock market when forming their prior distribution of the state. This assumption is simply made for convenience. However, it is tenable—the stock market has often been modeled as a crap game, independent of the real economy. An extension of the model in which agents use other variables in forming their prior is in preperation.

4.0 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 which 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 was discussed in Goldfeld and Quandt (1973), who called them switching regressions. They suggested that a Markov process could be used to generate the states. More recently Hamilton (1987A, 1987B) modeled 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 considered models of the same form as equations (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) derived 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. In our case, the likelihood is given by $f(y_1, y_2, ..., y_T|x_1, x_2, ..., x_T)$. In the model where agents are certain of the state, x_t is the null vector for $t = 1, 2, ..., T$. In the model where agents are uncertain of the state x_t is their prior probability of the high variance state, i.e.

$$x_t = \mathrm{P}(S_t = 1|y_1, y_2, ..., y_{t-1}). \tag{14}$$

for $t = 2,3..,T$. With this notation, the likelihood is given by an enumeration of all possible states weighted by their probabilities.

$$f(y_1, y_2, ..., y_T | x_1, x_2, ..., x_T) = \sum_{i_1=0}^{1} \sum_{i_2=0}^{1} \cdots \sum_{i_T=0}^{1} \Big\{ f(y_1, ..., y_T | S_1 = i_1, ..., S_T = i_T, x_1, ..., x_T) \times \mathrm{P}(S_1 = i_1, S_1 = i_2, ..., S_T = i_T) \Big\} \tag{15}$$

The terms in this equation are easy to describe. Since y_t conditional on x_t is serially uncorrelated except for the state, the density of the data vector, $y_1, y_2, ..., y_T$, conditional on the state vector is given by

$$f(y_1, ..., y_T | S_1 = i_1, ..., S_T = i_T, x_1, ..., x_T) = \prod_{j=1}^{T} f(y_j | S_j = i_j, x_t) \tag{16}$$

for $i_k = 0, 1$, $k = 1, 2, ..., T$. Given the Markov structure underlying the probability model, the unconditional distribution of the state vector is given by

$$\mathrm{P}(S_1 = i_1, S_2 = i_2, ..., S_T = i_T) = \prod_{j=2}^{T} \mathrm{P}(S_j = i_j | S_{j-1} = i_{j-1}) \mathrm{P}(S_1 = i_1) \tag{17}$$

for $i_k = 0, 1$, $k = 1, ..., T$.

Direct maximization of the log of the likelihood function requires the evaluation of 2^T terms in every iteration of maximization routine. It is computationally intractable for any reasonable sample size. We adopt the EM-algorithm to maximize equation (15). For this problem, 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. The algorithm is discussed in detail, as it relates to this problem, in *Appendix A*.

5.0 Results

The analysis was carried out for monthly data from Standard and Poor's index of 500 stock prices. The series analyzed was the percentage nominal total return less the three month T-bill rate of return, i.e. the monthly excess return of the portfolio times 100. The period of estimation was from January 1946 through December 1987. The results of the estimation 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 states

Model	σ_0^2	σ_1^2	μ_0	μ_1	p	q	ℓ	R^2
I	17.6965		0.5983				-1438.73	
	(1.1148)		(0.1874)					
II	13.3101	43.9681	0.8451	-1.0762	0.8641	0.9771	-1423.69	0.0377
	(1.4545)	(9.8076)	(0.2075)	(0.3987)	(0.3359)	(0.0552)		

Model I: Constant Mean, Constant Variance
Model II: Markov Mean, Markov Variance

Sample Period: January 1946—December 1987
Observations: 504

Table 1

Estimation results for model in which agents know the state in each period. Asymptotic standard errors in parantheses.

are presented in Table (3). In *Section 2.5.1* we will assess the implications the estimated model has for heteroskedasticity in excess returns. In the following section we will examine the models' implications for the risk premium.

5.1 Basic characteristics of the two state variance model

The basic hypothesis upon which this paper is founded is that there are two states in the volatility of stock market returns, i.e. 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 will test the hypotheses of two states and time dependence.

Unfortunately, the test of the hypothesis of only one state forces p and q to the edge of their parameter space: under the null one must be zero and the other unity. Under these conditions, the likelihood ratio test is not asymptotically distributed χ^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

$$\lambda^* = -\frac{2}{T}(T-3)(\ell_\omega - \ell_\Omega) \tag{18}$$

where ℓ_ω is the log likelihood of the one state model, Model I, in Table (1), and ℓ_Ω is the log likelihood of the full two state model, Model II. This statistic is asymptotically distributed χ^2 with two degrees of freedom. The value of this statistic is 29.9010. This value is significant at any reasonable level of signficance. Unfortunately, simulations by Everitt (1981) show that this test has low power unless $|\mu_1 - \mu_0| > 2$. Further, the power of the test has also been questioned for heteroskedastic models such as ours.

Figure (2) plots the probability of the high variance states conditional on all the data, $P(S_t = 1|y_1, y_2, ..., y_T)$, for the full and sub-samples respectively. 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 hover around 0.5. The full sample posterior is between 0.20 and 0.80 in only 18% of the sample.

Statistically, the mixture model, Model II, requires that the two states be characterized by different means and/or different variances. The variances in the two states are very well defined. The high variance state is more than three times that in the low variance state. As will be discussed below, the standard error of this parameter and all the high state parameters are quite large. We may test the hupothesis $\sigma_1^2 = \sigma_0^2$, while letting $\mu_1 \neq \mu_0$.♠ Despite the relatively large standard error of $\widehat{\sigma}_1^2$ we may reject the null at a reasonable level of confidence, the t-statistic is 3.2024. Though not as widely seperated as the variances, the means of the distributions are distinct. A test of the hypothesis that $\mu_1 = \mu_0$ against the alternative, while letting $\sigma_1^2 \neq \sigma_0^2$, yields a t-statistic of -4.6375.

The estimates of the transition probabilities suggest the low variance state will predominate. The estimates of the transition probabilities, p and q, of the Markov process suggest that the stationary, or unconditional, probabilities of the states for the full sample will be 0.8557 and 0.1443 for for the low and high variance states, respectively. Thus, for any given sample only about 14% of the observations will be expected to fall into the high variance state.

♠ We may test the hypotheses $\sigma_1^2 \neq \sigma_0^2$, and $\mu_1 \neq \mu_0$ however, we cannot test them jointly as this is equivalent to a test of the hypothesis that $1 - p = q = 0$.

High Variance Epsiodes $P(S_t = 1) > 0.5$	*Length of High Variance Episodes*	*Length of Low Variance Episodes*
		7 months
August--September 1946	2 months	186
April--July 1962	4	88
December 1969—June 1970	7	40
November 1973—February 1975	16	137
August 1986--January 1987	6	7
September— December 1987	4	
Mean	6.5 months	77.5 months
Median	5.0	64.0

Table 2

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 exeeded 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 state exceeded one half.

The improbability of the high variance state makes inference on high variance state parameters difficult. The sample size in estimating the high variance state parameters is, of course, dependent on the number of observations that fall into the high variance state. Figure (2) show that there are relatively few of these periods. More formally, *Appendix A*, shows that the sample size in estimating these parameters is effectively $\sum_t P(S_t = 1|y_1, ..., y_T)$, 64.7385 in this case. Thus, due to the relatively few periods in which the high variance state is likely, 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. However, the large standard error of $\widehat{p}$ suggests 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 probability distribution of the current state on past states. Fortunately, the null hypothesis of a binomial is nested within the Markov and does not require p and q to lie near the boundaries of their possible values. We may reject this hypothesis, the t-statistic is -2.5603.

The point estimate of p, suggests that once in the high variance state, the state is expected to persist. Since $\hat{p}$ is greater then 0.5, in both Table (1) and (2), 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 which

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

i.e. the probability of remaining in the state i for j consecutive periods is less then one half. For our simple first-order Markov process the number of periods, following Hamilton (1987A), j is given by $1/(1-\mathrm{P}(S_{t+1} = i|S_t = i))$. For the high variance state, $i = 1$ so that $j = 1/(1-p)$, or 7.3586 months That is, once in the high variance state, the stock market is expected to stay in that state for about seven months. The low variance state is much more persistent, setting $i = 0$, we then calculate $j = 1/(1-q)$, or 43.6205 months. Note that as $\hat{q}$ 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 is described in a non-parametric way in Figure (2) and Table (2). They summarize the posterior distribution of the state in each period, conditional on the entire sample used in estimation. Generally, these plots and tables indicate that both states are persistent, and that the low variance state is very persistent. It persisted, in expectation, without break from October 1946 until November 1962, fifteen and half years. 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.

5.2 Implications for the risk premium

Estimates of Model II, where 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 11% to hold the risky asset in low variance periods. However, the estimates also suggest the premium *declines* as the level of risk increases, i.e. $\hat{\mu}_1 < \hat{\mu}_0$. 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 agreement with those found by Schwert (1988) in his analysis

Model	σ_0^2	σ_1^2	α	β	p	q	ℓ	R^2
III	13.0458	52.9963	0.3364	3.0321	0.8072	0.9728	-1423.26	0.0056
	(1.3023)	(13.8229)	(0.0097)	(0.0261)	(0.3048)	(0.0370)		

Model	σ_0^2	σ_1^2	α_0	α_1	γ	p	q	ℓ	R^2
IV	12.7085	49.9850	0.5218	-1.1939	2.3802	0.8248	0.9729	-1421.41	0.0454
	(1.5247)	(16.2129)	(0.2356)	(0.5340)	(1.0119)	(0.3142)	(0.0618)		

Model III: Agents are unsure of the state
Model IV: Agents learn about the state during the period

Sample Period: January 1946—December 1987
Observations: 504

Table 1

Estimation results for the models in which agents are unsure of the states. In Model III agents make trades based on a prior distribution of the state using last period's excess return. In Model IV they make trades based on this prior and on a posterior distribution using trades during the period.

of nominal returns from stocks using Hamilton's (1987B) autoregressive model.♣

Mis-specification is a likely explanation for this result. If agents are uncertain of the state, so that Model III is the correct model, then estimates based on Model II will be inconsistent. Agents' expectation, or forecast, of the state is $P(S_t = 1|\Phi_{t-1})$. If agents are uncertain of the state, then this model suffers from the usual error in variables problem since the forecast error, $S_t - P(S_t = 1|\Phi_{t-1})$, is included on the right hand side in equations (8).

The parameter estimates for Model III, equations (9), 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 predicts agents will require an annual return of approximately 4%, if certain next period's return will be drawn from the low variance density. For a one percent increase in the probability of the high variance state, agents require an increase in monthly return

♣ Schwert's analysis was based on a different dataset. He used stock prices beginning in the mid-nineteenth century.

of 0.03%. Agents perceive the stock market in the high variance state to be very risky. If certain of the high variance state, they require an annual return of about 49%. However, the unconditional probability of the high variance state is only 0.1236. This suggests the risk premium will average approximately 9% on an annual basis.

Though the estimates of Model III are consistent with theory, the estimated model explains very little of the variance in excess returns. The model always predicts a positive return, thus its R^2 is less then 0.6%. The reason why it cannot predict a negative return is that the specification ignores the news effect. Agents acquire information by observing trades during the period. If agents don't know period t is drawn from the high variance density, then this piece of information is bad news. As agents observe trades within period t they will adjust their prior distribution of the state and revalue stocks downwards. Likewise, if agents don't know period t is drawn from the low variance density it is good news. They will adjust their prior distribution of the state and revalue their stocks upwards. Model IV generalizes the case where agents are unsure of the state to allow learning during the period.

Note that Model IV also suggests the direction of bias of the estimate of the risk premium when agents are assumed to know the state. Estimating Model II under this regime would yield parameter estimates which smear the risk premium and this news effect together. Since an increasing risk premium and the news effect have opposite effects on excess returns, we would expect $\widehat{\mu}_0$ to be an upwardly biased estimate of the risk premium and $\widehat{\mu}_1$ to be a downwardly biased estimate.

The estimated results indicate that we have 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 β and α_0 are significantly greater than zero, while α_1 is signficantly negative. The latter two parameters are also significantly different from each other, the t-statistic is -2.9737. They capture the effect of the state on the return of the stock during the period—the high variance state is bad news and the low variance state is good news. The high variance state is very bad news. All else constant, the estimates predict that the returns from the stock market will drop by more then 15% on annual basis relative to the market for T-bills. Allowing for a news effect, $\alpha_0 \neq \alpha_1$, greatly improves the fit. The R^2 rises to 4.5%.

In the model agents are assumed to know the parameters, thus we should expect that no matter how large the fall in the market, a perfectly forseen high variance period should lead to a positive expectation of the excess return. That is, we should expect $\gamma + \alpha_1 > 0$. The estimates indicate that such a period leads to an expected annual excess return of 15%. Note however, due to the large standard errors of the high variance parameters this increase is not statistically significant.

Recall from equation (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 increase with the anticipated variance. Again due to the large standard errors of the high variance parameters this increase is not statistically significant from zero. As with Model III, the unconditional probability of the high variance state may be used to derive the average risk premium. In this case this probability is 0.1338. Employing this value and equation (12) the risk premium is predicted to average approximately 7.5% per year for the full and sub-samples, respectively.

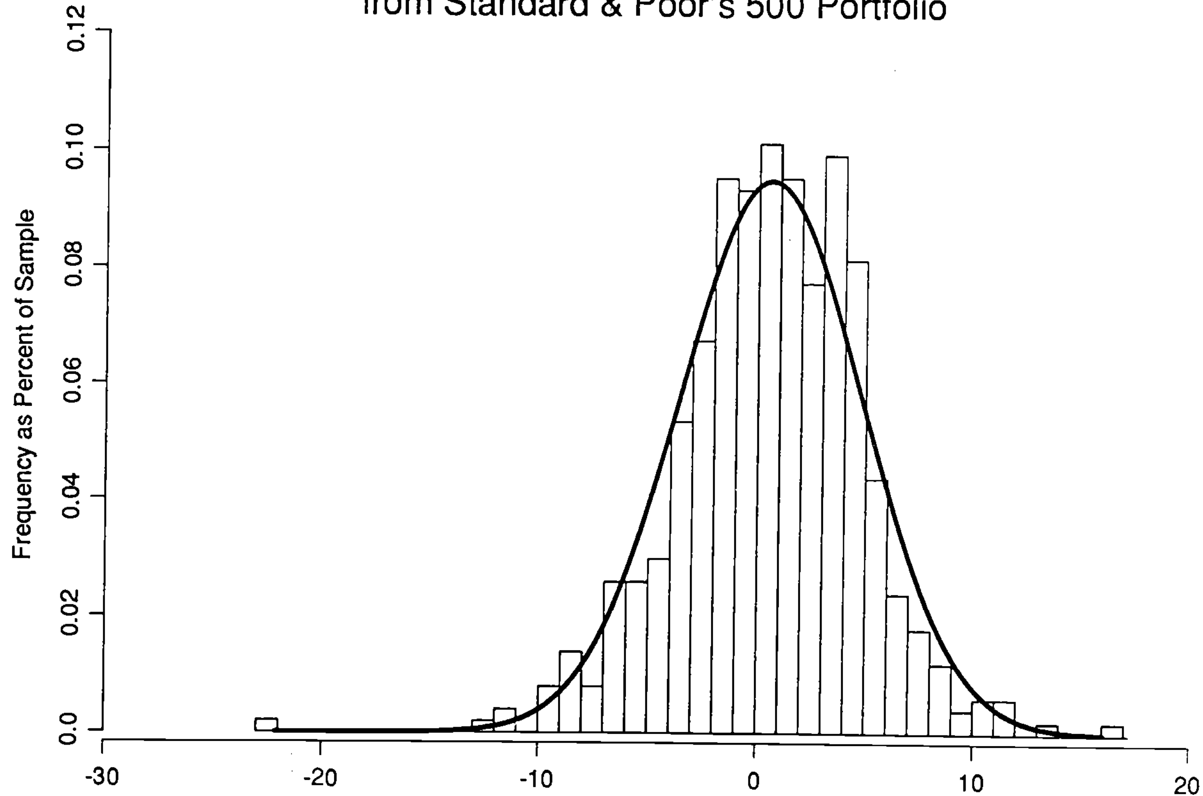
Examining Figure (2) closely, it becomes apparent that whenever the stock market enters the high variance state, it falls. In the next period it generally recovers more then it has lost. The parameter estimates summarize this tendancy. Our model provides a basis for understanding this behavior: The probability of the high variance state following a low variance state is quite small, so that agents are always surprised. Since $\widehat{\alpha}_1$ is negative this leads to a big drop in the market. In the following period, the probability of the high variance state is quite large, so that agents anticipate it, and collect their risk premium. The estimates in Table (3) indicate the risk premium will nearly double in the period that is, *ex post*, high variance, rising to 13.6%. This fact is made clear by Figure (3). This figure compares agents' prior distribution with the posterior conditional on all the data. This econometrician's posterior leads the agent's prior. That is, most periods in which the agents' prior gives significant weight to the high variance state, follow periods in which the posterior gives weight to that state. In short, agents are often surprised by the move from low variance to high variance. They are not surprised if the system remains in the high variance state.

6.0 Conclusion

We have shown 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 that 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 forecast of the market's variance is not always successful, so information about the state gained during periods is important to explaining the overall return.

Postwar Excess Return
from Standard & Poor's 500 Portfolio



with Corresponding Normal Density

Figure 1

Excess Return of Standard & Poor's Index
Posterior Distribution of the High Variance State

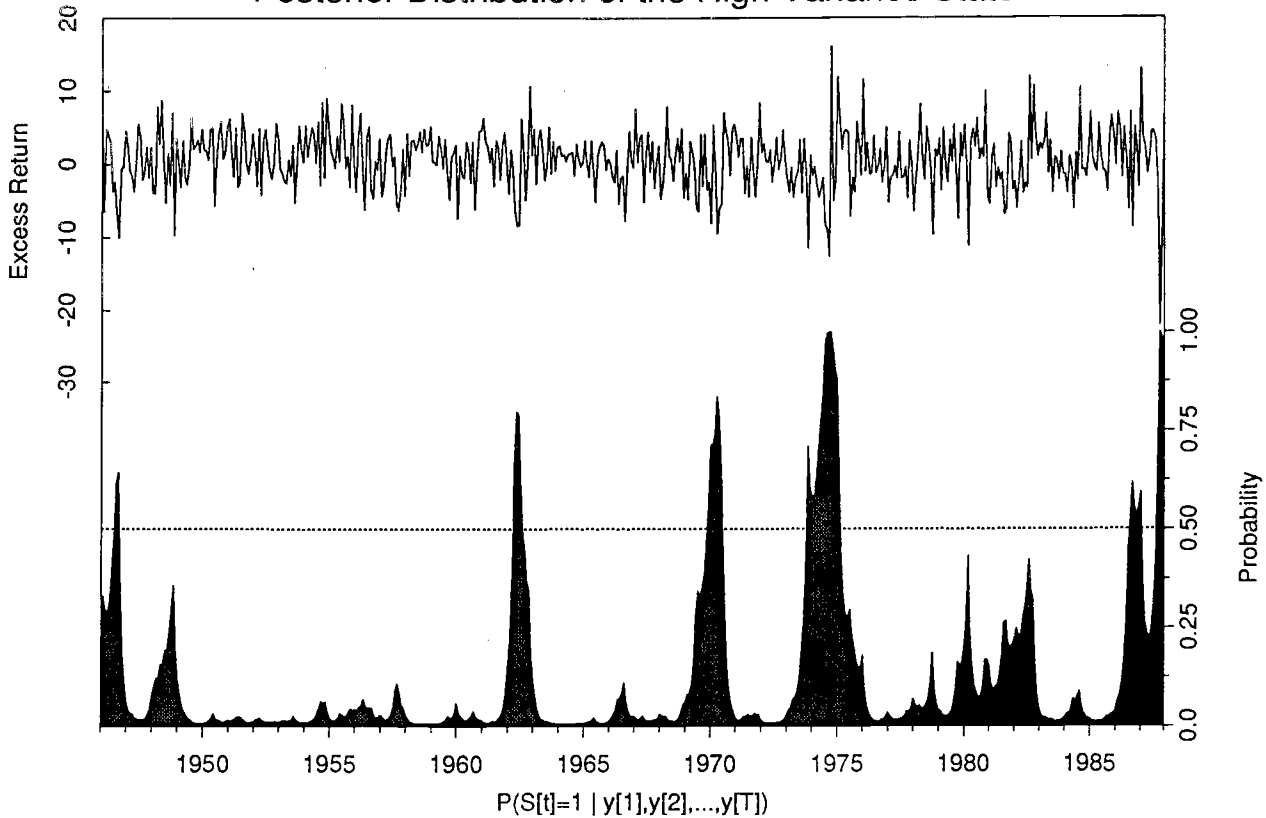


Figure 2

Agents' Prior Probability of the High Variance State
with the Econometrician's Posterior Probability

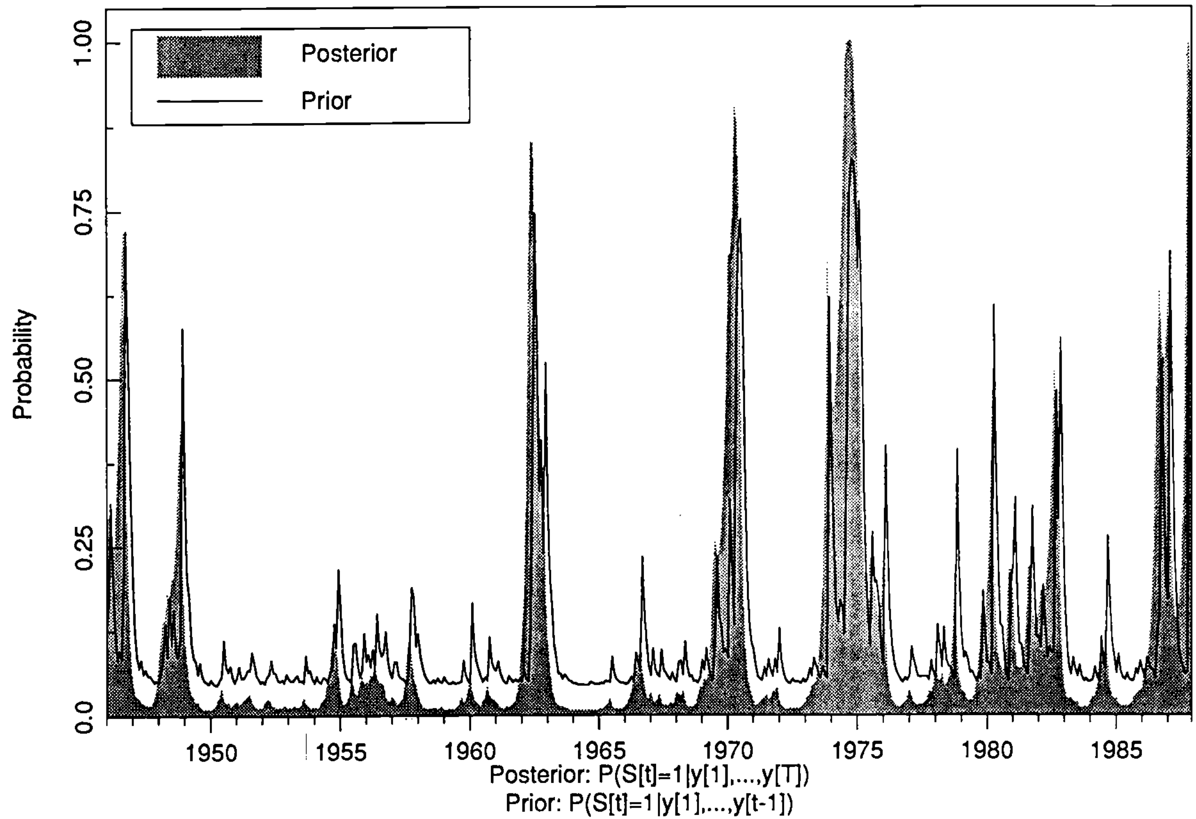


Figure 3

BIBLIOGRAPHY

Baum, Leonard E., Ted Petrie, George Soules, and Norman Weiss. (1970). "A Maximization Technique Occuring in the Statistical Analysis of Probabilistic Functions of Markov Chains," *The Annals of Mathematical Statistics*, **41**, No. 1, pp. 164-171.

Bollerslev, Tim., Robert F. Engle, and Jeffrey M. Wooldridge. (1987). "A Capital Asset Model with Time-varying Covariances," *Journal of Political Economy* **96**, No. 1, pp. 116–131.

Cosslett, Stephen R. and Lung-Fei Lee. (1985). "Serial Correlation in Latent Discrete Variable Models," *Journal of Econometrics*, **27**, pp. 79–97.

Engle, Robert F., David M. Lilien, and Russel P. Robins. (1987). "Estimating Time Varying Risk Premia in the Term Structure: The ARCH-M Model," *Econometrica*, **55**, No. 2, pp. 391–407.

Everitt, B. S. (1981). "A Monte Carlo Investigation of the Likelihood Ratio Test for the Number of Components in a Mixture of Normal Distributions," *Multiv. Behav. Res.*

Hamilton, James D. (1987A). "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, forthcoming.

——. (1988). "Analysis of Time Series Subject to Changes in Regime," *Unpublished Manuscript*,

——. (1987B). "Rational-Expectations Econometric Analysis of Changes in Regime: An Investigation of the Term Structure of Interest Rates," *Unpublished Manuscript*,

Lindley, D. V. (1984). *Bayesian Statistics, A Review*. Society for Industrial and Applied Mathematics.

Little, Roderick J. A. and Donald B. Rubin. (1987). *Statistical Analysis with Missing Data*. John Wiley & Sons.

Pagan. Adrian and Aman Ullah (1988). "The Econometric Analysis of Models with Risk Terms," *Journal of Applied Econometrics*, Vol. 3, pp. 87-105.

Schwert, G. William, (1988). "Business Cycles, Financial Crises, and Stock Volatility," *Unpublished Manuscript*

——. (1987). "Why Does Stock Market Volatility Change Over Time?" *Unpublished Manuscript*.

Romano, Joseph P. and Andrew F. Siegal. (1985). *Counterexamples in Probability and Statistics.* Wadsworth & Brooks/Cole.

Sundberg, Rolf. (1976). "An Iterative Method for Solution of the Likelihood Equations for Incomplete Data from Exponential Families," *Communications in Statistics*, **B5** No. 1, pp. 55–64

Wolfe, J. H. (1971). "A Monte Carlo Study of the Sampling Distribution of the Likelihood Ratio for Mixtures of Multinormal Distributions," Naval Personnel and Training Research Laboratory, *Technical Bulletin*, **STB-72-2**.

Appendix A

MAXIMUM LIKELIHOOD ESTIMATION OF MARKOV MODELS WITH THE EM-ALGORITHM

A.1 *Introduction*

As noted in the body of the paper, direct maximization of the likelihood function as defined in equation (2-15) requires the evaluation of 2^T terms in every iteration. It is computationally intractable for any reasonable sample size. We employ the EM-algorithm to maximize the likelihood function. The algorithm was developed from an old ad hoc idea for handling missing data. (1) replace the missing values by estimated values; (2) estimate the parameters; (3) re-estimate the missing values assuming the estimated values are correct; (4) iterate over (2) and (3) until convergence. Missing data methods are relevant for our purposes because the states, S_t, may be interpreted as missing data.

The algorithm differs from this technique in that the missing values are not filled in, rather they are replaced by sufficient statistics or, as in our case, the likelihood function is approximated in step (3). This is called the *E*-step, or expectation step, while step (2) is the *M*-step, or maximization step. A good introduction to the algorithm is provided by Little and Rubin (1987). In general, if the underlying distribution is from an exponential distribution, each iteration of the algorithm will yield a higher value of the likelihood, unless it is at a maximum. Dempster *et. al.* (1977) show this for the general missing data problem. Baum,*et. al.* (1970), shows the EM-algorithm maximizes the likelihood function if such a maximum exists, when the data is a mixture of exponential distributions and the underlying state is generated by a Markov process. A basic problem with the algorithm is that its rate of convergence is proportional to the missing information. As the missing state variable contains much information, the algorithm's convergence will be slow in our case. However, as will be demonstrated below, the ease in interpretation and coding make-up for the lack of speed in computation.

Section A-2 presents the *M*-step for estimation of the parameters of the Markov model. We show that the expected value of the likelihood function may be maximized by the simultaneous solution of normal equations developed from the first-order conditions. The following section presents the *E*-step. We derive the distribution of the state conditional on the parameters. The final section combines these steps and presents the formal EM-algorithm.

A.2 The M step: Maximum likelihood estimates of the parameters when the distribution of the state is known

If the state were known in each period, then the likelihood function for each observation y_t would simply be given by the expression

$$(1 - S_t)\phi(\epsilon_{0,t}/\sigma_0) + S_t\phi(\epsilon_{1,t}/\sigma_1). \tag{A-1}$$

The error, $\epsilon_{i,t}$, is given by

$$\epsilon_{i,t} = \begin{cases} y_t - \mu_i & \text{Model II} \\ y_t - \alpha - \beta x_t & \text{Model III} \\ y_t - \alpha_i - \gamma x_t & \text{Model IV} \end{cases} \tag{A-2}$$

For $i = 0, 1$, and where x_t is a regressor common to both states. Maximization of the full likelihood with respect to the parameters is trivial.

However, we don't know the state. If we knew the probability distribution of the state of the system prior to observing the realization y_t, then the expected value of the likelihood for an observation is

$$g_t = \sum_{i=0}^{1} \mathrm{P}(S_t = i)\phi(\epsilon_{i,t}/\sigma_i) \tag{A-2}$$

where $\mathrm{P}(S_t = i)$, $i = 0, 1$, is the prior probability of the state in period t. The log of the full expected likelihood is $\ell = \sum_t \ln g_t$.

The first order conditions for maximizing the likelihood for the model in which agents learn about the state are given by

$$\begin{aligned} \frac{\partial \ell}{\partial \alpha_i} &= \sum_{t=1}^{T} \frac{1}{g_t \sigma_i^2} \mathrm{P}(S_t = i)\phi(\epsilon_{i,t}/\sigma_i)(y_t - \alpha_i - \gamma)(-1) = 0, \qquad i = 0, 1 \\ \frac{\partial \ell}{\partial \gamma} &= \sum_{t=1}^{T} \frac{1}{g_t} \sum_{i=0}^{1} \frac{1}{\sigma_i^2} \mathrm{P}(S_t = i)\phi(\epsilon_{i,t}/\sigma_i)(y_t - \alpha_i - \gamma x)(-x) = 0. \end{aligned} \tag{A-3}$$

Note that the posterior distribution of the state in period t upon observing y_t is simply

$$\mathrm{P}(S_t = i|y_t) = \frac{1}{g_t}\mathrm{P}(S_t = i)\phi(\epsilon_{i,t}/\sigma_i), \qquad i = 0, 1. \tag{A-4}$$

This suggests that the solutions to equations (A-3) may be obtained by weighted least squares. Defining the weights

$$\begin{aligned} C_{i,t} &= \mathrm{P}(S_t = i|y_t)\frac{1}{\sigma_i^2}, \qquad i = 0, 1 \\ D_t &= C_{0,t} + C_{1,t} \end{aligned} \tag{A-5}$$

the first order conditions suggest the normal equations

$$\begin{pmatrix} \sum_t C_{0,t} & 0 & \sum_t C_{0,t}x_t \\ 0 & \sum_t C_{1,t} & \sum_t C_{1,t}x_t \\ \sum_t C_{0,t}x_t & \sum_t C_{1,t}x_t & \sum_t D_t x_t^2 \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \gamma \end{pmatrix} = \begin{pmatrix} \sum_t C_{0,t}y_t \\ \sum_t C_{1,t}y_t \\ \sum_t D_t y_t x_t \end{pmatrix}. \tag{A-6}$$

Solving for the parameter estimates, $\widehat{\alpha}_0$, $\widehat{\alpha}_1$, and $\widehat{\gamma}$, is of course, trivial.

Note that in the case when agents know the state, the appropriate normal equations are given by a subset the equations (A-6). These may be solved to yield the estimates

$$\begin{aligned} \widehat{\mu}_i &= \frac{\sum_{t=1}^{T} \mathrm{P}(S_t = i|y_t)y_t}{\sum_{t=1}^{T} \mathrm{P}(S_t = i|y_t)}, \qquad i = 0, 1 \\ \widehat{\sigma}_i^2 &= \frac{\sum_{t=1}^{T} \mathrm{P}(S_t = i|y_t)(y_t - \widehat{\mu}_i)^2}{\sum_{t=1}^{T} \mathrm{P}(S_t = i|y_t)}, \qquad i = 0, 1 \end{aligned} \tag{A-7}$$

Note that the effective sample size in equations (A-7) for state i is simply the sum of the weights, $\sum_t \mathrm{P}(S_t = i|y_t)$. Note also that if the posterior distribution of the state variable is degenerate—if S_t is known with certainty—then the estimates take on the intuitively pleasing forms

$$\begin{aligned} \widehat{\mu}_i &= \frac{\sum_{\forall S_t = i} y_t}{\sum_{\forall S_t = i} 1}, \qquad i = 0, 1 \\ \widehat{\sigma}_i^2 &= \frac{\sum_{\forall S_t = i}(y_t - \widehat{\mu}_i)^2}{\sum_{\forall S_t = i} 1}, \qquad i = 0, 1. \end{aligned} \tag{A-8}$$

The parameters where agents don't know the state may be estimated in the same way.

For all three models the estimates of the parameters of the density in each state are not dependent on the transition probabilities, p and q. This implies that the maximum likelihood estimates of the probabilities conditional on the maximum likelihood estimates of the density parameters will be the same as the unconditional estimates. Thus, a two-step technique may be employed to maximize the expectation of the likelihood function with respect to all of the parameters of the model. First, the appropriate normal equations are solved to estimate the parameters of the density of y_t. Second, the expected value of the likelihood function is maximized with respect to p and q conditional on the estimates of density parameters. Hamilton (1988) extends this method by deriving normal equations for the transition probabilities.

A.3 *The E step: The distribution of the state when the parameters are known*

With real data, of course, the distribution of the state of the system, S_t, will never be known. However, Bayes theorem may be employed to derive the distribution conditional on *all* the parameters and the data. Recall that Bayes theorem is given by the law of conditional expectations. That is, if we are interested in some parameter ψ, then the density of that parameter given an observation y_t is

$$p(\psi|y_t) = \frac{p(\psi)p(y_t|\psi)}{p(y_t)}, \tag{A-9}$$

where the unconditional density of y_t is given by

$$p(y_t) = \int_{\Psi} p(y_t|\psi)p(\psi)d\psi. \tag{A-10}$$

Where Ψ is the parameter space of ψ. It is customary to refer to $p(\psi)$ as the *prior* density, as it is held prior to the datum y_t. Likewise, $p(\phi|y_t)$ is the density *posterior* to y_t.

In time series, the posterior density in period t becomes the prior density for period $t+1$. Bayes' theorem for the distribution of ψ using both y_t and y_{t+1} is given by

$$p(\psi|y_{t+1}, y_t) = \frac{p(\psi) \times p(y_t, y_{t+1}|\psi)}{p(y_{t+1}, \psi, y_t)} \tag{A-11}$$

but this is just

$$\begin{aligned} p(\psi|y_{t+1}, y_t) &= \frac{p(\psi) \times p(y_t|\psi) \times p(y_{t+1}|\psi, y_t)}{p(y_{t+1}|y_t) \times p(y_t)} \\ &= \left(\frac{p(\psi) \times p(y_t|\psi)}{p(y_t)}\right) \frac{p(y_{t+1}|\psi, y_t)}{p(y_{t+1}|y_t)}. \end{aligned} \tag{A-12}$$

The parantheses is just Bayes' theorem, so that

$$p(\psi|y_{t+1}, y_t) = \frac{p(\psi|y_t) \times p(y_{t+1}|\psi, y_t)}{p(y_{t+1}|y_t)}. \tag{A-13}$$

This equation is the basis for Bayesian sequential updating. When we have a posterior distribution of ψ, based on observations $y_1, y_2, ..., y_{t-1}$ and we observe y_t, we may update it simply by allowing the posterior distribution at time t to become the distribution prior to observing y_{t+1}.

In Markov models, when the parameter of interest is the state, Bayes theorem takes on an especially tractable form. These models are characterized by a finite number of states, in the case at hand two. This results in a discrete prior distribution. Furthermore, the distribution of the state is dependent only on the realized state in the previous period. The previous state is unknown, however, we have its posterior distribution, $P(S_{t-1} = i|\mathbf{y}_{t-1})$, $i = 0, 1$, $\mathbf{y}_{t-1} = (y_1, y_2, ..., y_{t-1})$, from the previous application of Bayes theorem. The prior distribution will simply be last period's posterior updated with the appropriate transition probabilities

$$P(S_t = i|\mathbf{y}_{t-1}) = \sum_{j=0}^{i} P(S_t = i|S_{t-1} = j)P(S_{t-1} = j|\mathbf{y}_{t-1}) \qquad \text{(A-14)}$$

Note that in the initial period, there is no posterior from the previous period. This observation is most easily handled by assuming that the Markov process began infinitely far into the past. Thus, the prior distribution of the first observation is simply the steady state probability distribution of the state. That is, the prior for the first observation, y_1, is defined to be

$$\begin{aligned} P(S_0 = 1) &= \frac{1-q}{2-p-q} \\ P(S_0 = 0) &= \frac{1-p}{2-p-q}. \end{aligned} \qquad \text{(A-15)}$$

Conditional on the state, y_t is distributed *iid* Normal. Thus the likelihood of y_t is given by the set of equations defined by the normal densities

$$p_i(y_t|S_t = i, \mathbf{y}_{t-1}) = \phi(e_{i,t}/\sigma_i). \qquad i = 0, 1. \qquad \text{(A-16)}$$

where as before $e_{i,t}$ is defined by equation (A-2).

The Markov structure also simplifies the structure of the the unconditional density of y_t, $p(y_t)$. Due to the discrete prior, the integral of equation (A-10) is replaced by the summation

$$p(y_t|\mathbf{y}_{t-1}) = P(S_t = 0|\mathbf{y}_{t-1}) \times p_0(y_t|S_t = 0) + P(S_t = 1|\mathbf{y}_{t-1}) \times p_1(y_t|S_t = 1). \qquad \text{(A-17)}$$

Note that weights on the densities sum to unity by definition, since $P(S_t = i|y_{t-1})$, $i = 0,1$ is a well defined distribution function. Thus, $p(y_t|\mathbf{y}_{t-1})$ is simply a mixed density: it is a proper density function that integrates to one.

Simple application of Bayes theorem gives the posterior distribution of the state conditional on information through period t,

$$P(S_t = i|\mathbf{y}_t) = \frac{P(S_t = i|\mathbf{y}_{t-1}) \times p_i(y_t\ S_t = i)}{p(y_t|\mathbf{y}_{t-1})} \qquad i = 0, 1 \qquad \text{(A-18)}$$

We now want to update this distribution to find the distribution of the state conditional on the data through period T. That is, we wish to evaluate the probability $P(S_t = 1|\mathbf{y}_T)$. We now let expression (A–18) be the prior distribution of the state, and update this distribution for $t+1, t+2, ..., T$ using Bayes theorem. Suppose we have performed the update through $t+j-1$. We wish to add observation y_{t+j} to our posterior. Then the components of Bayes theorem are given by

$$\text{Prior:} \quad P(S_t = 1|\mathbf{y}_{t+j-1}) \tag{A–19}$$

$$\text{Likelihood:} \quad f(y_{t+j}|S_t = 1, \mathbf{y}_{t+j-1}) \tag{A–20}$$

$$\text{Unconditional:} \quad f(y_{t+j}|\mathbf{y}_{t+j-1}) \tag{A–21}$$

Note that the unconditional distribution, expression (A–21) may be derived by integrating the state out of the likelihood. In general, the likelihood itself is difficult to evaluate. Recall that the data, y_t, is uncorrelated except for the state and that the state is generated by a first order Markov proces. This implies that

$$\begin{aligned} f(y_{t+j}|S_{t+j-1} = i_{j-1}, ..., S_{t+1} = i_1, S_t = 1, \mathbf{y}_{t+j-1}) \\ = f(y_{t+j}|S_{t+j-1} = i_{j-1}), \end{aligned} \tag{A–22}$$

for $i = 0, 1$. We can then obtain the likelihood using the rule of elimination,

$$\begin{aligned} &f(y_{t+j}|S_t = 1, \mathbf{y}_{t+j-1}) = \\ &\sum_{i_{j-1}=0}^{1} f(y_{t+j}|S_{t+j-1} = i_{j-1}) P(S_{t+j-1} = i_{j-1}|S_t = 1, \mathbf{y}_{t+j-1}). \end{aligned} \tag{A–23}$$

The expression $P(S_{t+j-1} = i_{j-1}|S_t = 1, \mathbf{y}_{t+j-1})$, for $i = 0, 1$ is readily evaluated using Bayes theorem. All we need do is follow the algorithm for updating the probability $P(S_{t+k} = 1|\mathbf{y}_{t+k})$, for $k = 1, 2, ..., j-1$, conditional on $S_t = 1$. That is,

$$\begin{aligned} &P(S_{t+k} = 1|S_t = 1, \mathbf{y}_{t+k}) = \\ &\frac{P(S_{t+k} = 1|S_t = 1, \mathbf{y}_{t+k}) \times f(y_{t+k}|S_t = 1, S_{t+k} = 1, \mathbf{y}_{t+k})}{f(y_{t+k}|S_t = 1, \mathbf{y}_{t+k})} \end{aligned} \tag{A–24}$$

Each component on the right hand side of equation (A–24) may be easily evaluated. The likelihood is

$$f(y_{t+k}|S_t = 1, S_{t+k} = 1, \mathbf{y}_{t+j-2}) = \phi(e_{1,t+k}/\sigma_1). \tag{A–25}$$

The "unconditional" density of the new datum, y_{t+k} is

$$f(y_{t+k}|S_t = 1, \mathbf{y}_{t+k-1}) = \sum_{i=0}^{1} P(S_{t+k}|S_t = 1, \mathbf{y}_{t+k-1})\phi(e_{i,t+k}/\sigma_i). \tag{A–26}$$

Finally, the prior probability of the state, $S_{t+k} = 1$, conditional on S_t and the data, i.e. the expression $P(S_{t+k} = 1|S_t = 1, \mathbf{y}_{t+k-1})$, is simply

$$pP(S_{t+k-1} = 1|S_t = 1, \mathbf{y}_{t+k-1}) + (1-q)P(S_{t+k-1} = 0|S_t = 1, \mathbf{y}_{t+k-1}) \qquad (A\text{-}27)$$

Applying equation (A-24), Bayes theorem for making inference on the state repeatedly allows us to evaluate equation (A-22), and thus the likelihood. Once we have the likelihood it is easy to evaluate the unconditional distribution of the state. Evaluating Bayes theorem is then just a matter of substitution.

A.4 *The EM algorithm: Maximum likelihood estimates of the parameters when the distribution of the state is unknown*

So far we have derived a method of obtaining maximum likelihood estimates of the parameters of the density function of each state and transition probabilities *given* the probability distribution of the state. We have also found a method of obtaining the probability of each state *given* the parameters of density functions and the transition probabilities. Combining these two techniques and iterating give us the EM-algorithm.

The combined algorithm is as follows: (1) Nominate estimates of the parameters. Denote the nominated estimates, $\theta^{[0]}$. In Model II this is

$$\theta^{[0]} = \left(\mu_0^{[0]}, \mu_1^{[0]}, \sigma_0^{2[0]}, \sigma_1^{2[0]}, p^{[0]}, q^{[0]}\right); \qquad (A\text{-}28)$$

(2) Use Bayes theorem to derive the probability distribution of S_t, $t = 1, 2, ..., T$ conditional on the parameter estimates $\theta^{[0]}$, $P(S_t = i|\theta^{[0]}, \mathbf{y}_T)$, $i = 0, 1$; (3) Set the weights employed in the weighted least squares estimation, equations (A-6) equal to the probabilities associated with the distribution derived by Bayes theorem. Thus, we are asserting that the known prior distribution of the state in equation (A-2) is $P(S_t = i|y_1, ..., y_{t-1}, y_{t+1}, ..., y_T)$. Thus, the posterior distribution of equation (A-4) is simply $P(S_t = i|\mathbf{y}_T)$. This is presented formally in Hamilton (1988). (4) Use the two-step estimation technique discussed in *Section* 3.4.1 to obtain new estimates of the parameters. Set $\theta^{[1]}$ equal to the resulting estimates of the parameters,

$$\theta^{[1]} = \left(\widehat{\mu}_0^{[0]}, \widehat{\mu}_1^{[0]}, \widehat{\sigma}_0^{2[0]}, \widehat{\sigma}_1^{2[0]}, \widehat{p}^{[0]}, \widehat{q}^{[0]}\right); \qquad (A\text{-}29)$$

Where $\widehat{\mu}_0^{[i]}$, etc... represent the maximum likelihood estimates of the parameters of Model II conditional on the posterior distribution $P(S_t = j|\theta^{[i]})$, $j = 0, 1$; (5) Iterate steps (2) through (4) until an appropriate convergence criteria is met. In our implementation of the algorithm the stopping condition was met when

$$||\theta^{[i]} - \theta^{[i-1]}|| < 0.001. \qquad \text{(A-30)}$$

This technique not only yields maximum likelihood estimates of the parameters but application of Bayes theorem gives us the posterior distribution $P(S_t = i|\mathbf{y}_T)$, $i = 0, 1$. This allows us to make inferences concerning the state of the system, and to evaluate agents' prior distribution of the state.