Is the Ex ante Premium Always Positive? Evidence and Analysis from Australia

Kathleen D Walsh

School of Banking and Finance
University of New South Wales

This Draft: Oct 2004

Abstract:
An implicit assumption of the conditional CAPM is that the ex ante market risk premium is positive in all states of the world. Recent studies on US portfolios by Boudoukh, Richardson and Smith (1993) and on World portfolios by Ostdiek (1998) find violations of this assumption. This paper seeks to test the sign of the market risk premium in the Australian Market using two parallel tests. Firstly, the series is examined for the presence of two regimes using a test developed in Bayesian inference. Truncated normal priors are applied to the means in this test to specifically detect means of opposite sign. Secondly, we applied the test developed by Boudoukh, Richardson and Smith (1993) which allows the moments implied by the model to be conditioned on observable information, using an instrumental variables approach. This test was conducted using a contemporaneous information set to simulate perfect foresight. We then lagged the instrumental variables to test whether forming dynamic portfolios would result in positive economic returns. The parallel testing from economic and econometric perspectives allowed us to gain a deeper understanding of the sign of the risk premium. We were able to reject the null of a single regime in favour of the two regime model using the regime switching test. In addition the inequality tests rejected the restriction of a positive risk premium. However the dynamic portfolios formed using this information did not yield a significant economic return. The combination of a dual regime and a negative ex ante risk premium is a rejection of the assumption of positivity required for the conditional CAPM.

* The author would like to thank Professor Tom Smith and Dr Ron Guido for helpful comments and suggestions.
Is the Ex ante Premium Always Positive? Evidence and Analysis from Australia

1 Introduction

The ex ante equity risk premium is an important element in asset pricing. It is a primary component in the estimation of a firm’s cost of equity capital and as such is relevant to all capital budgeting analyses. It is also of interest to investment managers in the context of strategic and tactical asset allocation. Therefore considerable research effort has been spent on the estimation of the risk premium. Recent studies have attempted to estimate the ex ante risk premium. Amongst these Claus and Thomas (2001) estimate the premium by equating a discount rate with the present value of prevailing forecasts of future cash flows, Fama and French (2002) apply an estimation procedure relying on fundamentals stemming from the dividend growth model and Arnott and Bernstein (2002) use economic fundamentals to estimate the premium. The above researchers make the important distinction between ex ante and ex post returns. In particular Fama and French (2002) note that ex post average return on stocks are up to 60% higher than their expectations model and therefore realisations do not equate to expectations.

Further, several recent studies have highlighted the possibility of a negative risk premium (see Arnott and Ryan (2001)). In particular Arnott and Bernstein (2002, pp64) note that

“...the forward looking risk premium is nowhere near the level of the past; today, it may well be near zero, perhaps even negative.”
This is of particular interest to academic researchers because it is a direct contradiction of one of the assumptions of the conditional Capital Asset Pricing Model (CAPM). Specifically, the conditional CAPM has the restriction that the risk premium is always positive but has no restriction on its magnitude. Therefore the predictions by Arnott and Bernstein (2002) and Arnott and Ryan (2001) of a negative risk premium imply a rejection of the CAPM. However it is difficult to separate the testing of the CAPM from the specification of the expectations model. Does the negative risk premia imply that CAPM fails or does it imply model specification error? The impact of model specification error is highlighted by the range of estimates of the risk premium, from 6% (Ibbotson and Chen (2001) to -1.1% (Arnott and Bernstein (2002)).

A recent study on a US portfolio by Boudoukh, Richardson and Smith (1993) (henceforth BRS), uses an instrumental variables approach which does not require specification of an expectations model. In this way they are able to directly test for the negativity of the ex ante risk premium. This testing procedure was also applied on an international portfolio by Ostdiek (1998). Both studies identified states of the world where the ex ante market risk premium was negative, providing sufficient evidence to reject the conditional CAPM in their samples.

In this paper we also restrict our analysis to the sign of the premium and we seek to provide a robust test of the sign of the risk premium in Australia. We consider the problem by addressing two key issues. Firstly, if it is possible for the risk premium to be negative then one interpretation is that regime shifts are present in the series. Secondly, if two regimes are present, can we identify them using a nominated information set? The following develops each argument separately.
As discussed earlier, two recent studies identified states of the world where the *ex ante* market risk was negative. This can be interpreted as regime shifts in the data series. Specifically, there was one state of the world where the theory held and the equity premium was positive and another state of the world where the theory failed and the premium was negative. This would be evident if the data exhibited regime switches where the mean of each regime was either positive or negative. Given that the CAPM relies on the assumption of a positive risk premium we would only expect the premium to shift into the negative regime a small proportion of the time. This was also noted by Arnott and Bernstein (2002), who suggested that a negative risk premium would be abnormal in the extreme.

Previous approaches to estimating the parameters of a regime shifting model have utilised Maximum Likelihood methodology as in Gray (1996a), Hamilton (1988) and Hamilton (1998). However, this approach requires specification of the likelihood function which is often analytically intractable. A computationally attractive technique would be to adopt Markov Chain Monte Carlo (MCMC) estimation. This procedure allows estimation without the explicit specification of the Likelihood function. The procedure generates posterior distributions of all unknown parameters by simulating from standard distributions. It provides more information than the Maximum Likelihood estimation as it generates full conditional distributions rather than a point estimate. In addition it allows the econometrician to impose informative priors on the estimates thus increasing the power of the test.

A primary innovation of this paper is the imposition of truncated normal priors on the means of each regime. These priors allow the test to reflect the information held by the econometrician. Specifically, the estimation can complement the prior beliefs by restricting the means of each regime to be either positive or negative. The first stage of our analysis, therefore, is the possible rejection of a single regime model in favour of a dual regime model.
The second set of tests relates the regime shifts to economic variables. In other words, given that we find two regimes in the data series can we identify an information set that will detect the second regime? As highlighted earlier, tests of the positivity of the risk premium have been hindered because conditional expectations are unobservable. This usually requires the econometrician to specify a conditional model, which complicates most testing procedures. A recent test developed by BRS overcomes the need to specify a conditional expectations model. They develop a test of inequality constraints allowing moments to be conditioned on observable information using an instrumental variables approach. The test is consistent with the procedures of Hansen and Singleton (1982) and Gibbons and Ferson (1985).

Using this approach we can test whether fundamental economic variables, such as the shape of the yield curve or the level of the risk free rate, can be used to identify the regime shifts in the data. We apply the tests in two stages to answer separate questions. Firstly, we test for the negativity of the risk premium using contemporaneous data. This is equivalent to assuming perfect foresight in the information set and applying one step ahead forecasts. If we are able to reject the restriction of a positive risk premium then, given perfect foresight, we consider the risk premium to be negative.

Although this procedure gives us a deeper understanding of the sign of the risk premium a more important question would be whether we can earn an economic return using the information set. Therefore the second stage of the inequality analysis uses the lagged information set to form dynamic portfolios. We can then test whether or not we can earn a positive economic return by using the information set. This concept stems from work by Kirby (1998) and Whitelaw (2000).
One drawback of this approach is that the econometrician can never be sure that the selection of conditioning agents is complete. In BRS’s case the selection of agents was based on clear economic theory and they were successful in identifying states of the world where the CAPM theory failed. Similarly Ostdiek (1998) (using a comparable set of agents) was able to identify states where the world *ex ante* premium was negative. However, it is possible that, for a different sample, these agents do not hold and that a wider set of agents is needed to identify states of the world where the MRP is negative. Therefore the combination of the regime switching analysis and the inequality tests provides a more robust approach to testing the negativity of the risk premium. In essence the regime switching test considers the statistical properties of the series whilst the inequality tests consider the problem from an economic theory perspective.

This paper is organised as follows. Section 2 describes the testing methodology, Section 3 describes the data, Section 4 presents the results and I draw conclusions in Section 5.

2 **Methodology**

As this is a two-pronged analysis I will discuss each procedure separately. Section 2.1 develops the testing procedure for the regime shifts and Section 2.2 outlines the inequality testing developed by BRS and describes its application to this problem.

2.1 **Regime Switching Methodology**

The first prong in our analysis of the sign of the risk premium is an investigation of regime shifts in the series. The rationale for this is straightforward. If it is possible for the risk premium to be negative, as we
have seen in US (BRS (1993)) and World portfolio studies (Oestdiek (1998)),
then one interpretation is that two regimes exist in the data. Specifically, one
regime exhibits a negative mean whilst the other is positive. It would
therefore be prudent to test for the existence of regime shifts in our data.

Markov regime switching models have often been used to capture cyclical
effects or regime shifts in finance data. The most popular method for
estimation of regime shifts has been Maximum Likelihood as in Hamilton
(1988) and Gray (1996a). However Maximum Likelihood estimation requires
closed form representations of the densities which are often analytically
intractable. A computationally simple alternative to Maximum Likelihood is
Bayesian estimation which computes such densities using simulation.
Specifically, the unobserved states and the unobserved parameters are
obtained using the Markov Chain Monte Carlo tool of Gibbs sampling.

Another benefit of Bayesian analysis is that traditional tests of regime
switching do not allow the econometrician to impose prior information.
Specifically, if our contention is that the market risk premium is negative is
some states, then it would be appropriate to apply a restriction on the sign of
the premium. Bayesian analysis allows this restriction to be incorporated into
the analysis through a prior distribution thus restricting the means of each
regime could be restricted to being either purely positive or purely negative.

A full description of the process of MCMC analysis is given in Casella and
George (1992), Chib and Greenberg (1996), Smith and Gelfand (1992) and
Tanner (1996). The following provides an explanation of the process as it
applies to this problem.

The series in question is modelled as a simple linear process with two regimes
in which the parameters (mean and variance) differ. We have kept the model
simple in this case as we wanted to be careful not to bias the results with
model specification error. However this procedure can be extended to represent a more complex model such as that developed by Gray (1996a).

The following explains the Bayesian framework in which both the unobserved states and the unobserved parameters are obtained using the Markov Chain Monte Carlo tool of Gibbs sampling. The Gibbs sampler is a technique for generating random variables from a (marginal) distribution indirectly, without having to calculate the density. Specifically, given a pair of random variables \((X,Y)\), the Gibbs sampler generates a sample from \(f(x)\) by sampling instead from the known conditional distributions \(f(x | y)\) and \(f(y | x)\). This method is expedient in our case because the conditional posterior distribution of the parameters given the states and the conditional posterior of the states given the parameters all have a form amenable to the Gibbs sampling technique.

### 2.1.1 Implementation of the Markov Chain Monte Carlo technique

As noted earlier, we have characterised the process governing the market risk premium as a model with a regime shift and transition probabilities that are state dependent. Specifically, we have assumed that the data follows the following distribution:

\[
y_i = \mu_k(t) + \sigma_k(t) e_i
\]

\textbf{Equation 1}

where

\[e_i \sim N(0,1)\] and

\[
k(t)
\] can take the values of 1 or 2

To detect the presence of 2 regimes it is necessary to identify the following parameters:
- The probability of switching from regime 1 to regime 2: \( P_{12} \) where
\[
P_{12} = P(k(t+1) = 2|k(t) = 1) \quad \text{and} \quad P_{11} = 1 - P_{12}.
\]
The probability of switching from regime 2 to regime 1: \( P_{21} \) where
\[
P_{21} = P(k(t+1) = 1|k(t) = 2) \quad \text{and} \quad P_{22} = 1 - P_{21}.
\]
These four probabilities form the transition matrix, \( \Pi \).
- The means of each regime \( \mu_1 \) & \( \mu_2 \)
- The standard deviations of each regime \( \sigma_1 \) & \( \sigma_2 \)
- The regime indicator variable, \( K \), identifying whether the observation belongs to regime 1 or 2.

Consistent with previous literature beta priors are placed on the transition probabilities and inverse gammas on the conditional variances. However, an interesting innovation of this paper is the use of truncated normal priors on the means. Recall that our initial contention was that in general the mean risk premium would be positive however there may be states of the world where the theory failed and the mean risk premium would be negative. Given this, we would expect a positive mean for regime 1 and a negative mean for regime 2 and we need to select appropriate priors to reflect this. One candidate prior for a positive parameter would be a lognormal distribution however this is not tractable in a Bayesian framework. We decided that a more appropriate choice would be a truncated normal. This formulation allows us to easily sample from the density as well as incorporate the prior information about the properties of the parameter.

We were able to construct the conditional distributions that form the basis of the simulation. Given the parameter set \( \theta = (\mu_1, \mu_2, \sigma_1, \sigma_2, p_{12}, p_{21}) \) the conditional distributions are:
\[ P(\mu_i | Y, \theta, K) \sim N\left( \mu_i, \frac{\sigma_i}{A} \right) I(\mu_i > 0) \tag{Equation 2} \]

Where \( i = 1 \) or 2

\[ P(\sigma_i^2 | Y, \theta, K) \sim IG\left( a_i + \frac{T}{2}, S + n(\bar{y} - \mu)^2 + b_i \right) \tag{Equation 3} \]

Where \( i = 1 \) or 2

\[ P(p_{1i} | Y, \theta, z_i, K) \sim Beta(a + n_{12}, b + n_{11}) \tag{Equation 4} \]

\[ P(p_{2i} | Y, \theta, z_i, K) \sim Beta(a + n_{21}, b + n_{22}) \tag{Equation 5} \]

A full derivation of the conditional distributions (along with details of the notation) is contained in Appendix A.

Each of these complete conditionals can be simulated via the Gibbs sampler leading to a posterior sample from the joint distribution of the parameters and the states. Estimation of the parameters is done for both the two regime and one regime specifications and the means of the distributions are reported as the parameter estimates.

The method used to for the simulation of the unobservable switching vector is discussed in the next section.

### 2.1.2  Simulation of the regime switching vector

This section attempts to simulate the unobservable variable \( K^n \). In order to make the sampling as efficient as possible we adopt a block sampling approach that simulates \( K^n \) from the joint distribution \( p[K^n | Y^n, \theta] \) developed by Carter and Kohn (1994).
Carter and Kohn (1994) show how to generate a switching vector (K) conditional on the data set Yn. The vector k1….km contains the possible values assumed by K(t) which in our case is 1 and 2 as we restrict this analysis to 2 regimes.

As y(t) is observed then

\[ p\{K(t)|Y^t\} \propto p\{y(t)|K(t)\} p\{K(t)|Y^{t-1}\} \]  \hspace{1cm} \text{Equation 6}

Using the recursive filtering equations of Anderson and Moore (1979, Ch 8), Carter and Kohn develop an Algorithm to calculate \( p\{K(t)|Y^t, K(t+1)\} \).

\[ p\{K(t)|Y^t, K(t+1)\} = \frac{p\{K(t+1)|K(t)\} p\{K(t)|Y^t\}}{p\{K(t+1)|Y^t\}} \] \hspace{1cm} \text{Equation 7}

Using Equation 7, we recursively generate one iteration of the switching vector K, given the data and obtain desired samples from \( p[K^n| Y^n, \theta] \).

### 2.1.3 The test statistic

Recall from Bayes rule that, assuming \( \theta \) and \( K_t \) are independent:

\[ P(\theta,K_t|Y_t) \propto P(Y_t|\theta,K_t)P(\theta,K_t) \] \hspace{1cm} \text{Equation 8}

where \( P(\theta,K_t|Y_t) \) is the posterior distribution and \( P(\theta,K_t) \) is the prior distribution.

However \( P(Y_t|\theta,K_t) \) is not the likelihood function because if there is more than 1 regime the K vector needs to be “integrated out” to obtain the
likelihood function \( P(y_i|\theta) \). This is achieved by summing across both regimes:

\[
P(y_i|\theta) = \sum_{i=1}^{k=2} P(y_i|k_i, \theta) \cdot P(k_i = i|y_i) \tag{Equation 9}
\]

However, it is not possible to apply the standard maximum likelihood ratio test to assess the significance of the evidence against a null hypothesis as it is not a nested model. Specifically, the parameters associated with the second regime are not identified under the null of a single regime Gray (1996a).

Model selection can also be determined by using an information criteria rather than formal maximum likelihood ratios. Information criteria penalises the maximum likelihood, based on the number of fitted parameters. Two popular tests are the Schwarz Information Criteria (SIC) and the Akaike Information Criteria (AIC). The SIC is an approximation to the logarithm of the marginal, which selects one model over another if its information criteria exceeds the other (or the change is positive). Specifically, the information criteria require the increase in the maximum log-likelihood, \( \Delta \ln L \), to exceed a penalty based on the number of additional parameters, \( k \). Note that the AIC penalty is independent of the sample size, while the SIC penalty increases with sample size.

\[
\Delta AIC = \Delta \ln L - k > 0 \tag{Equation 10}
\]

\[
\Delta SIC = \Delta \ln L - \frac{1}{2} k \ln (N) > 0 \tag{Equation 11}
\]

where \( N \) is the sample size and \( k \) is the number of additional parameters

Therefore a positive AIC or SIC is evidence of a rejection of the null of one regime.
2.2 Inequality Testing

The second prong of our analysis applies the inequality testing procedure developed by BRS (1993). Recall that an explicit implication of the conditional CAPM is that the ex ante Market risk premium is positive in all states of the world. This implies an inequality restriction on the model that the expected return on the market is greater the return on the riskless asset over the same period.

The procedure used by BRS is attractive because it allows the econometrician to identify agents which theory suggests provide some information about the ex ante risk premium but doesn’t require an explicit functional form. Agents used in previous literature have included the level of interest rates, the shape of the yield curve, the dividend yield and the level of volatility. BRS report evidence that the US risk premium is negatively related to T-bill rates and positively related to volatility, dividend yield and the slope of the term structure.

Using this methodology the econometrician is able to condition the ex ante MRP on a suitable economic agents and formally test whether the premium is positive. Hence this methodology provides a test for the conditional ex ante market risk premium. Rejection of the null implies that the ex ante market risk premium is negative in some states of the world (identified by the conditioning agent used). Rejection of the null also implies that the conditional CAPM does not hold.

Specifically we can identify elements of
Where $R_{m_{t+1}}$ is the return on the market portfolio at $t+1$, $R_f$ is the risk free rate from $t$ to $t+1$ and $z_t^+$ is a purely positive set of instrument at time $t$. Equation 12 provides a set of moment conditions that can be used to estimate $\theta_{\mu z_t^+}$. It is important to note that the fact that $\mu_t$ is unobservable is not an issue because the vector of observables $(R_{m_{t+1}}, R_f, z_t^+)$ is enough to identify $\theta_{\mu z_t^+}$. The test implies that the econometrician can observe an positive instrument at $t$ and use this to form expectations of the ex ante market risk premium. Using this approach we can test whether economic variables, such as the shape of the yield curve or the level of the risk free rate, can be used to identify the regime shifts in the data.

In the first stage of our inequality analysis we apply the testing procedure under the assumption of perfect foresight of the instruments. In other words we assume that we can forecast the instrument at $t+1$ with no error. The element of our analysis will therefore be $R_{m_{t+1}}, R_f$ and $z_{t+1}^+$ and using Equation 12 results in a set of moment conditions that can be used to estimate $\theta_{\mu z_{t+1}^+}$. If we are able to reject the restriction of a positive risk premium using a contemporaneous information set then, given perfect foresight, we consider the risk premium to be negative.

Perhaps a more important questions question would be whether we can earn an economic return using the information set. Therefore the second stage of the inequality analysis uses the lagged information set to form dynamic portfolios.

\[ E_t\left[ \left( R_{m_{t+1}} - R_f \right) \otimes z_t^+ - \theta_{\mu z_t^+} \right] = 0 \]  
\[ \theta_{\mu z_t^+} = E \left[ \mu \otimes z_t^+ \right] \geq 0 \]
Kirby (1998) notes that the covariance between \( r_t \) and \( z_{t-1} \) is the expected excess payoff on a dynamic trading strategy:

\[
\text{cov}(r_t, z_{t-1}) = E\left[ r_t (z_{t-1} - \mu_z) \right]
\]  \hspace{1cm} \text{Equation 14}

The trading strategy exploits the information conveyed by the realisation of \( z_{t-1} - \mu_z \). The construct of our inequality testing procedure provides a direct application of this. By employing a lagged information we are able to observe the payoff of a dynamic trading strategy. However, we note that the restriction of a positive information set requires that our holdings are all positive and hence we cannot profit from a short sales strategy. Therefore the economic return earned on the dynamic portfolio is restricted due to positive holding requirements. Whitelaw (2000) applies this strategy using the consumption CAPM.

In summary the inequality testing has a two step approach. Firstly we test whether, given perfect foresight on the instruments, we can identify a negative risk premium and secondly using lagged instruments we test for a positive economic return on a dynamic trading strategy.

Importantly, the inequality tests are restricted by the identification of candidate conditioning agents. As noted earlier a major innovation of this paper is the use of the regime switching procedure, which does not require such identification and therefore provides a robust parallel test to the inequality analysis.

3 Data

Data was sourced from the Centre for Research in Finance (CRIF) at the Australian Graduate School of Management. The Share Price and Price Relative database is an historical record from December 1973 of share prices
and calculated price relatives of all Australian listed and previously listed companies with fully paid shares. Specific data for the return on the market \( R_m \) was sourced from the month end price relatives on the CRIF index and the risk free rate proxy \( R_f \) was the monthly yields on 13 week Treasury Bills. The difference between these two yielded the observed risk premium.

It was also necessary to construct a set of purely positive instruments. To be consistent with previous US and world portfolio research, 4 conditioning agents were considered which reflect periods of implied low risk premiums.

- condition on times when the risk free rate is high (above the long run average)
- the term structure is downward sloping (long rate less short rate < 0),
- volatility is low (below the long run average, calculated as absolute value of \( R_m \))
- the dividend yield is low (below the long run average).

Dividend yields were constructed from the CRIF index as monthly averages across all valid dividend yields\(^1\). Yield curve data and the risk free rate were also sourced from the CRIF database\(^2\).

Informative and uninformative information sets were constructed for each of the conditioning agents.

**Table 0-1: Transformation of information set**

<table>
<thead>
<tr>
<th>Transformation</th>
</tr>
</thead>
</table>

---

\(^1\) Yield is calculated as dividend divided by price. Zero Yields are admitted. Yields based on stale prices are excluded.

\(^2\) Original source was the Reserve Bank Bulletins
| Non-Informative | $Z_t^* = 1$ when $R_{ft}$ is high  
= 0 otherwise  

| $Z_{2t}^* = 1$ when term structure is downward sloping  
= 0 otherwise  

| $Z_{3t}^* = 1$ when $\sigma_{mt}$ is low  
= 0 otherwise  

| $Z_{4t}^* = 1$ when $\frac{D_{mt}}{P_{mt}}$ is low  
= 0 otherwise  

| Transformation | $Z_{it}^* = \max(0, R_{ft} - E[R_{ft}])$  

| Informative | 

| $Z_{2t}^* = \max(0, -\Delta r_{ft})$  

| $Z_{3t}^* = \left( \frac{1}{\sigma_{mt}} \right)$  

| $Z_{4t}^* = \left( \frac{P_{mt}}{D_{mt}} \right)$  

Each instrument was normalised by dividing through by the expected value of $Z^*$. Specifically $Z_{it}^\prime = \frac{Z_{it}^*}{E[Z_{it}^*]}$.

In summary we have constructed an observed equity risk premium $(R_{mt+1} - R_{ft})$ to be used in both the Regime-switching analysis and the inequality analysis. In addition and a set of 4 purely positive normalised instruments have been constructed for use in the inequality analysis.
4 Results

The Bayesian Regime-switching analysis was run for 1000 warm-up and 10,000 sample iterations. The extensive iteration process ensured that the estimates were drawn from reliable simulations of the posterior distributions of the parameters. The distributions of each simulated parameter are presented visually in Figure 2 and as estimates in Table 2.

Firstly we will consider the point estimates. The results in Table 2 show that the annualised mean of regime 1 ($\mu_{12}$) was 9.64% and regime 2 ($\mu_{22}$) was 24.18%. This reflects the imposition of positive and negative priors on the means of each regime. As expected the mean of the single regime model ($\mu_{11}$) fell between these two means at 4.16%.

Interestingly, the variances of each of the means varied considerably. In the two regime model the annualised standard deviation of mean 1 ($\sigma_{12}^2$) was quite low at 0.1296 whilst the standard deviation of mean 2 ($\sigma_{22}^2$) was several times larger at 0.366. This is possibly due to the small number of observations in the second regime. Again, as expected, the standard deviation of the mean of the single regime model ($\sigma_{11}^2$) fell between these estimates at 0.1469.

We also consider the full conditional posterior distributions of the parameters. Two important aspects of the analysis warrant a mention. Firstly the starting values of the parameter estimates were arbitrary so speedy convergence during the warm-up period was integral to the analysis. Visual inspection of the plots of warm-up period iterates show that iterates converged to reasonable estimates well within the warm-up period.

Secondly, the Bayesian construct resulted in a simulation of the posterior distributions of each parameter. A histogram of each is reported in Figure 2. It is important to note that all histograms reflect the shape of the proposed
posterior. For example the posterior distribution of the mean of regime 1 was a product of a normal likelihood and a truncated normal prior. The resulting simulation reflects this.

Of particular interest is the regime switching vector $K$ and the probabilities associated with this vector Equation 7. Both parameters are plotted in Figure 3. The first panel plots the probabilities of a data point belonging to regime 1 using the Algorithm developed by Carter and Kohn (1994). Also presented is the plot of the switching vector $(K)$ constructed using the Carter and Kohn Algorithm. The two plots provide a visual link between the probabilities of being in a regime with the selection of that regime. As per our initial contention the switching vector selects two regimes (one positive and one negative) however the negative regime is only selected a small proportion of the time.

The persistence of each regime is noted in Panel C of Table 2. The number of transitions between each regime is noted as $n_{ij}$ (where $i$ and $j$ are the regimes). For example the number of transitions from regime 1 to regime 2 is $n_{12}$ and is reported as 7. Similarly the transitions from 2 to 1 $n_{21}$ were also low at 8. As expected the persistence of regime 1 was evident with $n_{11}$ showing 292 from 328 possible shifts. These results are also summarised in the transition probability estimates of $P_{12}$ and $P_{21}$ in Panel A of Table 2. The persistence of regime 1 is evident with the probability of switching from regime 1 to 2 ($P_{12}$) estimated at only 3%.

We now turn to the test for the presence of two regimes. We report two Information Criteria in Table 2 both with a critical value of 0. The SIC and the AIC both were positive thus providing evidence of a rejection of the null of one regime. In other words the information criteria of the 2 regime model exceeded that of the single regime model even after imposing penalties for additional parameters and sample size.
This is of particular importance due to the large difference between the estimates of the positive and negative regimes. Recall that the estimate of the risk premium is used to make capital budgeting and asset allocation decisions and therefore large deviations in estimates could lead to incorrect investment decisions. We note that the annualised mean of the positive regime in the two regime model is 9.64% whilst the annualised mean of the mean of the single regime model is 4.16%. This highlights the potential for incorrect investment decisions if an estimate of the risk premium is drawn under the assumption of a single regime model.

Initial interest in the inequality testing lies in the choice of the instrumental variable. As discussed earlier our choice extended to risk free rate, shape of the yield curve, level of volatility and level of dividend yield. The first analysis was a simple calculation of the annualised conditional risk premium and is detailed in Table 3. As expected, given economic theory the unconditional risk premiums were positive whilst conditioning on risk free rate, yield curve and volatility levels yielded a negative result. Interestingly, despite economic theory and previous results to the contrast, dividend yield did not result in a negative conditional risk premium and was therefore discarded from the analysis. We therefore ran inequality tests of the positive nature of the risk premium using 3 instruments – high risk free rate, downward sloping yield curve and low levels of volatility.

Results of the contemporaneous inequality tests are included in Table 4. The first panel reports the results of using uninformative instruments whilst the second reports the informative instruments.

As expected (from Table 3) the mean of the risk premium conditioned on each uninformative instrument yielded a negative result. However, this is only suggestive of negative ex ante risk premiums. We therefore formally test the
joint multiple inequality and find a Wald statistic of 3.1686. This equates to a
p value 0.114296 when tested against the critical value of the weighted $\chi^2$. Therefore, although initial analysis implied a negative risk premium when conditioned on the contemporaneous information set, when tested jointly the result is not statistically significant and we are unable to reject the null of a positive risk premium.

However, when we consider an informative contemporaneous information set we find the results are somewhat different. The results of the empirical tests provide the conditional means of the risk premium, weighted by the magnitudes of each instrumental variable. Although the conditional means are negative for the level of risk free rate and the level of volatility, the means are actually positive for the shape of the yield curve. When the system is jointly tested we find a Wald statistic of 10.777 equating to a p value of .002525. Therefore, although the informative analysis only yielded a negative premium for two instruments the results were statistically significant. In other words the informative instruments lead to a rejection of the \textit{ex ante} premium being positive.

The informative results can be interpreted in the following way. If an econometrician had perfect foresight and was able to make one step ahead forecasts (without error) of the level of the risk free rate and the level of volatility then they would be able to identify states of the world where the \textit{ex ante} risk premium was negative.

We therefore need to discover whether a simple trading strategy can earn a significant economic return. The second stage of the inequality tests rely on a lagged information set of the instruments used in the contemporaneous analysis. The results of the uninformative and informative tests are included in Table 5.
For each of the tests the conditioning information set was lagged one period. This equates to forming a dynamic portfolio based on information known and time \( t \) and realising the return at \( t+1 \). The conditional means from the uninformative tests are reported in the first panel of Table 5. In this analysis the conditional means based on the risk free rate and the shape of the yield curve were negative whilst the level of volatility was positive. However, when jointly tested the results were insignificant with a \( p \) value of 0.203. Therefore as with the contemporaneous results, although there is a suggestion of a negative premium we were unable to reject the null of positivity.

The informative instruments result in conditional means that are weighted by the magnitudes of each instrumental variable. Surprisingly, all means in the informative analysis are positive resulting in a Wald statistic of zero. This means that there is no difference between the restricted and unrestricted models and therefore no \( p \) value can be calculated. In essence whether we use informative or uninformative instruments we are unable to reject the null of a positive \textit{ex ante} risk premium.

The inequality results can be interpreted as follows. We are able to identify a significant negative \textit{ex ante} risk premium if we assume perfect foresight of the instruments. However, if we employ a simple portfolio formation strategy based on a lagged information set then we are unable to earn a significant economic return. It is important to note that the dynamic portfolios are restricted to positive elements and that perhaps allowing for negative holdings would alter the results.

5 Conclusions

A two pronged analysis was performed to explore the nature of the \textit{ex ante} equity risk premium. The first set of tests explored the nature of regime
shifts in the observed equity premium. The motivation behind this analysis stemmed from evidence in the literature of negative expectations of the risk premium. We considered that if the risk premium was negative in some states of the world then perhaps two regimes would be evident in the data. Bayesian analysis was required for the regime switching tests to allow for the imposition of means of opposite sign for each regime. Using an information criterion we found that the two regime model was statistically significant and exceeded the information content of the single regime model. Of interest was the large difference in means between the two models. We noted earlier that estimation of the risk premium was an important element in capital budgeting and investment decisions therefore failure to accurately account for a two regime model could lead to estimation error in the risk premium.

The second set of tests followed the work of BRS (1993) and is a somewhat backdoor approach to testing the conditional CAPM, as a positive ex ante MRP is an implicit assumption of the model. The inequality constraints were tested allowing sample moments to be conditioned on observable information. Using an informative contemporaneous information set we were able to reject the positive restriction on the ex ante equity risk premium. However, this is insufficient evidence to reject the conditional CAPM as it requires perfect foresight in forecasting the information set. An additional inequality test was conducting using a lagged information set. This set of tests was tantamount to forming dynamic portfolios based on information in our information set at time t. However, this simple strategy did not yield a significant economic return.

In summary, by analysing the data from two perspectives we have provided much deeper insight into the nature of the market risk premium in Australia. We support the conclusions found in the international literature that the ex ante risk premium can be negative and attempt to explain it in part by
modelling regime shifts in the data. Further understanding of the nature of regime shifts should lead to improved estimates of the *ex ante* risk premium.
Table 2 Regime switching results

This table reports the parameters estimates from the regime switching model. Panel A reports the annualised means and standard deviations for the 2 regime and single regime case. The means and standard deviations are the average of the 10,000 iterates in the sampling period. Panel B reports the test statistics measuring the difference between the 2 regime and 1 regime model. The Schwarz (SIC) and Aikake (AIC) Information Criteria tests have a critical value of 0. Rejection of the single regime model in favour of the 2 regime model is denoted by *. Panel C reports the number of changes between regimes. For example N12 reports the number of times the series jumped from regime 1 to regime 2.

Panel A

<table>
<thead>
<tr>
<th></th>
<th>Annualised Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Single Regime Specification</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regime 1</td>
<td>4.16%</td>
<td>0.1469694</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Two Regime Specification</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regime 1</td>
<td>9.64%</td>
<td>0.1296148</td>
</tr>
<tr>
<td>Regime 2</td>
<td>-24.18%</td>
<td>0.3666061</td>
</tr>
</tbody>
</table>

Switching Probabilities

| Probability 12 | 0.0322 |
| Probability 21 | 0.1892 |

Panel B

<table>
<thead>
<tr>
<th>Test Statistic</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SIC</td>
<td>77.3447*</td>
</tr>
<tr>
<td>AIC</td>
<td>84.9368*</td>
</tr>
</tbody>
</table>

Panel C

<table>
<thead>
<tr>
<th>changes between regimes</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>292</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>21</td>
</tr>
</tbody>
</table>
Table 3 reports the annualised risk premium conditioned (contemporaneously) on each of the instruments. Specifically, the unconditional risk premium is the average of all monthly risk premia in the sample. The monthly average is then annualised assuming monthly compounding. The conditional risk premium averages are calculated as the average of all risk premia at t conditioned on information at time t. (The risk premia was included in the sample if the contemporaneous conditioning information was equal to 1).

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Annualised Conditional risk premium</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unconditional</td>
<td>4.16%</td>
</tr>
<tr>
<td>High risk free rate</td>
<td>-5.97%</td>
</tr>
<tr>
<td>Downward sloping yield curve</td>
<td>-6.77%</td>
</tr>
<tr>
<td>Low Volatility</td>
<td>-3.02%</td>
</tr>
<tr>
<td>Low Dividend Yield</td>
<td>2.20%</td>
</tr>
</tbody>
</table>
Table 4  Inequality Tests of whether the ex ante risk premium is always positive
(using contemporaneous instruments)

The sample covered in this study includes monthly (end) data on CRIF value weighted portfolio returns and dividend yields and the 90 Bank accepted bill from 1973 to 2001. The statistic W is a joint test of multiple inequality restrictions corresponding to contemporaneous high 90 day BAB rates, downward sloping term structures and low volatility periods. The estimators, $\theta_{\mu Z_t}$, represent the annualised conditional mean of the risk premium in these states. Also given are the probability of these states and the standard errors of the conditional means. All estimates are adjusted for conditional heteroskedasticity and serial correlation using the method of Newey and West (1987).

<table>
<thead>
<tr>
<th></th>
<th>Uninformative Instruments</th>
<th>Informative Instruments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Level of rf rate</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Probability of state</td>
<td>0.42</td>
<td>0.42</td>
</tr>
<tr>
<td>Annualised Mean $\theta_{\mu Z_t}$</td>
<td>-5.97%</td>
<td>-2.23%</td>
</tr>
<tr>
<td>(standard error)</td>
<td>(0.02305)</td>
<td>(0.01872)</td>
</tr>
<tr>
<td><strong>Shape of yield curve</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Probability of state</td>
<td>0.23</td>
<td>0.23</td>
</tr>
<tr>
<td>Annualised Mean $\theta_{\mu Z_t}$</td>
<td>-6.76%</td>
<td>0.18%</td>
</tr>
<tr>
<td>(standard error)</td>
<td>(0.02925)</td>
<td>(0.02643)</td>
</tr>
<tr>
<td><strong>Level of volatility</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Probability of state</td>
<td>0.61</td>
<td>0.61</td>
</tr>
<tr>
<td>Annualised Mean $\theta_{\mu Z_t}$</td>
<td>-3.02%</td>
<td>-6.24%</td>
</tr>
<tr>
<td>(standard error)</td>
<td>(0.00525)</td>
<td>(0.00565)</td>
</tr>
<tr>
<td>Multiple inequality restrictions statistic W</td>
<td>3.16859829</td>
<td>10.77675626</td>
</tr>
<tr>
<td>(p-value)</td>
<td>0.114296</td>
<td>0.002525</td>
</tr>
</tbody>
</table>
Table 5  
**Inequality Tests of whether the ex ante risk premium is always positive**  
((using lagged instruments))

The sample covered in this study includes monthly (end) data on CRIF value weighted portfolio returns and dividend yields and the 90 Bank accepted bill from 1973 to 2001. The statistic $W$ is a joint test of multiple inequality restrictions corresponding to lagged high 90 day BAB rates, downward sloping term structures and low volatility periods. The estimators, $\theta_{t+Z}$, represent the Annualised conditional mean of the risk premium in these states. Also given are the probability of these states and the standard errors of the conditional means. All estimates are adjusted for conditional heteroskedasticity and serial correlation using the method of Newey and West (1987).

<table>
<thead>
<tr>
<th></th>
<th>Uninformative Instruments</th>
<th>Informative Instruments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Level of rf rate</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Probability of state</td>
<td>0.42</td>
<td>0.42</td>
</tr>
<tr>
<td>Annualised Mean $\theta_{t+Z}$</td>
<td>-2.19%</td>
<td>9.75%</td>
</tr>
<tr>
<td>(standard error)</td>
<td>(0.02167)</td>
<td>(0.01924)</td>
</tr>
<tr>
<td><strong>Shape of yield curve</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Probability of state</td>
<td>0.23</td>
<td>0.23</td>
</tr>
<tr>
<td>Annualised Mean $\theta_{t+Z}$</td>
<td>-3.49%</td>
<td>6.92%</td>
</tr>
<tr>
<td>(standard error)</td>
<td>(0.02284)</td>
<td>(0.02496)</td>
</tr>
<tr>
<td><strong>Level of volatility</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Probability of state</td>
<td>0.61</td>
<td>0.61</td>
</tr>
<tr>
<td>Annualised Mean $\theta_{t+Z}$</td>
<td>7.00%</td>
<td>5.37%</td>
</tr>
<tr>
<td>(standard error)</td>
<td>(0.0.1116)</td>
<td>(0.01979)</td>
</tr>
<tr>
<td>Multiple inequality</td>
<td>0.20259250</td>
<td>0.000000</td>
</tr>
</tbody>
</table>
| restrictions statistic W | (p-value)                 | 0.186508                | N/A
Figure 1 Regime-switching parameter estimates of $\mu_i$ and $\sigma^2_i$.

This figure shows the Gibbs Sampling iterates of $\mu_i$ and $\sigma^2_i$. Where $i$ denotes the regime and $j$ denotes the total number of regimes. For example, $\mu_1$ is the mean of regime 1 in the two regime model. The warm-up period consisted of 1000 iterations and the plots in panels A to F show the speed of convergence during this period. The Sampling period consisted of 10,000 iterations and histograms in panels A to F depict the simulated posterior distributions of each parameter.

Panel A

<table>
<thead>
<tr>
<th>Iterates of $\mu_{12}$ during warm-up period</th>
<th>Iterates of $\mu_{12}$ during sample period</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Graph of $\mu_{12}$ during warm-up period" /></td>
<td><img src="image2" alt="Histogram of $\mu_{12}$ during sample period" /></td>
</tr>
</tbody>
</table>

Panel B

<table>
<thead>
<tr>
<th>Iterates of $\mu_{22}$ during warm-up period</th>
<th>Iterates of $\mu_{22}$ during sample period</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image3" alt="Graph of $\mu_{22}$ during warm-up period" /></td>
<td><img src="image4" alt="Histogram of $\mu_{22}$ during sample period" /></td>
</tr>
</tbody>
</table>

Panel C

<table>
<thead>
<tr>
<th>Iterates of $\mu_{11}$ during warm-up period</th>
<th>Iterates of $\mu_{11}$ during sample period</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image5" alt="Graph of $\mu_{11}$ during warm-up period" /></td>
<td><img src="image6" alt="Histogram of $\mu_{11}$ during sample period" /></td>
</tr>
</tbody>
</table>
Figure 1 continued

Panel D
Iterates of $\sigma_{12}^2$ during warm-up period
Iterates of $\sigma_{12}^2$ during sample period

Panel E
Iterates of $\sigma_{22}^2$ during warm-up period
Iterates of $\sigma_{22}^2$ during sample period

Panel F
Iterates of $\sigma_{11}^2$ during warm-up period
Iterates of $\sigma_{11}^2$ during sample period
Figure 2 Regime switching graphs

Figure 2 plots the probabilities of a data point belonging to regime 1  $\{K(t)\mid Y^n\}$ using the Algorithm developed by Carter and Kohn (1994). Also presented is the plot of the switching vector (K) constructed using the Carter and Kohn Algorithm. The two plots provide a visual link between the probabilities of being in a regime with the selection of that regime.

Panel A: $\{K(t)\mid Y^n\}$: plot of the probability of being in regime 1

Panel B: Switching Vector K: Assignment of regime (1 or 2) based on the Carter Kohn Algorithm
The parameter set estimated using Gibbs sampling included the transition matrix of switching probabilities. The plots in panels A and B of Figure 3 show the iterates of $p_{12}$ (the probability of switching from regime 1 to regime 2) and $p_{21}$ (the probability of switching from regime 2 to regime 1) during the warm-up period. These plots highlight the speed of convergence in the warm-up iterates. Panels A and B also depict a histogram of iterates of each transition probability during the sample period. The warm-up period consisted of 1000 iterations while the sampling period consisted of 10,000 iterations.
Appendix A Derivation of the posterior parameter densities

It is necessary to obtain the posterior densities for:

1. the conditional means \( \mu_1 \) and \( \mu_2 \)
2. the conditional variance \( \sigma_1^2 \) and \( \sigma_2^2 \)
3. the transition probabilities \( p_{12} \) and \( p_{21} \)

1. The conditional means

The conditional mean for each regime, \( P(\mu_i | Y, \theta, K) \) for \( i = 1, 2 \), needs to be constructed. By Bayes Theorem:3

\[
P(\mu_i | Y, \theta, K) \propto P(Y | \theta, K) P(\mu_i | K)
\]

Equation A 1

Where \( i = 1 \) or 2 representing the two regimes. We will fully consider the case where \( i = 1 \). Therefore we require a prior on the mean \( P(\mu_i) \) and the likelihood of the data given the regime \( P(Y | \theta, K) \).

Our initial contention was that most of the time the mean risk premium would be positive however there may be states of the world where the theory failed and the mean risk premium would be negative. Given this, we would expect a positive mean for regime 1 and a negative mean for regime 2 and we need to select appropriate priors to reflect this. One candidate prior for a positive parameter would be a lognormal distribution however this is not tractable in a Bayesian framework. A more appropriate choice would be a truncated normal. This formulation allows us to easily sample from the density as well as incorporate the prior information about the properties of the parameter. Details are shown below.

The prior of the mean in regime 1 \( P(\mu_1) \) is a truncated normal to restrict the set to purely positive estimates: \( \mu_i \sim N(0, c \sigma_i^2) I(\mu_i > 0) \). Where \( c \) is a hyperparameter.4

---

3 For notational convenience we will omit K
4 Prior work (see Kohn etc) suggest to impose a hyperparameter c to “stretch out” the normal by the degree of prior information. By imposing a large hyperparameter (eg \( c=100 \)) the normal prior approximates a uniform distribution and thus acts as an uninformative prior.
Where the denominator $\left(1 - \Phi_{0,c} (\mu_i, c \sigma_i^2)\right)$ is a normalising constant (the area under the pdf over the relevant parameter space). In a symmetrical distribution with mean 0, truncating the cdf to purely positive space equals $\frac{1}{2}$. As we are only considering the proportional distribution, this constant can be ignored. Therefore:

$$P(\mu_i) \propto \frac{1}{\sqrt{2\pi c \sigma_i^2}} \cdot \exp \left[-\frac{1}{2c \sigma_i^2}(\mu_i - 0)^2\right] \cdot I(\mu_i > 0) \quad \text{Equation A 2}$$

Under the Bayesian construct we must also define the likelihood of the data in each regime. The likelihood of the data set $Y = \{y_1, y_2, \ldots, y_T\}$ is the product of the likelihood of each individual data points and we therefore consider a candidate point $y_1$. For this data point, conditioning on states (K) is equivalent to conditioning on the state at time t ($K_t$).

$$P\left(y_1 | \theta, K\right) = P\left(y_1 | \theta, K_t\right) \quad \text{Equation A 4}$$

where $K = \{k_1, k_2, \ldots, k_T\}$

Our model states that the data exhibits a mixture of normals and therefore given the state, returns are conditionally normal:

$$P\left(y_i | \theta, K_t = 1\right) \propto \frac{1}{\sqrt{2\pi \sigma_i^2}} \cdot \exp \left[-\frac{1}{2\sigma_i^2}(y_i - \mu_i)^2\right] \quad \text{Equation A 5}$$

Therefore the likelihood of the entire set of data $Y = \{y_1, y_2, \ldots, y_T\}$ is:

---

5 Consequently truncating the cdf to purely negative space is also equal to a half. Therefore we have omitted the derivation of the posterior on the means for the second (negative) regime.<<need to check that we are using the correct portion of the cdf>>
\[ P(Y | \theta, K) \propto \frac{1}{(2\pi \sigma^2)^{T/2}} \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^{T} (y_i - \mu_i)^2 \right] \]  \hspace{1cm} \text{Equation A 6}

Now that we have a function for the likelihood and the prior we can construct the posterior by Bayes rule.

\[ P(\mu | Y, \theta, K) \propto P(Y | \theta, K).P(\mu | K) \]  \hspace{1cm} \text{Equation A 7}

\[ \propto \frac{1}{(2\pi \sigma_i^2)^{T/2}} \exp \left[ -\frac{1}{2\sigma_i^2} \sum_{i=1}^{T} (y_i - \mu_i)^2 \right]. \]  \hspace{1cm} \text{Equation A 8}

\[ \frac{1}{(2\pi \sigma_i^2)^{1/2}} \exp \left[ -\frac{1}{2\sigma_i^2} (\mu_i)^2 \right] I(\mu_i > 0) \]

Which reduces to

\[ \propto \exp \left[ -\frac{(A(\mu_i - \tilde{\mu})^2)}{2\sigma_i^2} \right] I(\mu_i > 0) \]  \hspace{1cm} \text{Equation A 9}

Where

\[ A \equiv T + \frac{1}{c} \] \hspace{1cm} \text{and} \hspace{1cm} \tilde{\mu} \equiv \frac{T \bar{y}}{T + \frac{1}{c}}

Hence the conditional mean for regime 1 is normally distributed\(^6\).

\[ P(\mu | Y, \theta, K) \sim N\left(\tilde{\mu}, \frac{\sigma^2}{A}\right) I(\mu_i > 0) \]  \hspace{1cm} \text{Equation A 10}

\(^6\) quadratic of $\mu_i$ ’s divided by a value ~ normal
2. The conditional variances

By Bayes Theorem:

\[ P\left(\sigma_i^2 \mid Y, \theta, \sigma_i^2, K\right) \propto P(Y \mid \theta, K)P(\sigma_i^2 \mid K) \quad \text{Equation A 11} \]

And once again we will restrict the detail to the variance of regime 1. In the previous section we described the likelihood of the data given the regime \( P(Y \mid \theta, K) \).

\[ P(Y \mid \theta, K) \propto \frac{1}{2\sigma_1^2} \exp \left[ -\frac{1}{2\sigma_1^2} \sum_{i=1}^{r}(y_i - \mu_i)^2 \right] \]

This can also be expressed as

\[ P(Y \mid \theta, K) \propto \frac{1}{(2\pi\sigma_1^2)^{r/2}} \exp \left[ -\frac{1}{2\sigma_1^2} \left( S + n(\bar{y} - \mu)^2 \right) \right] \quad \text{Equation A 12} \]

Therefore we only require a prior on the variance \( P(\sigma_i) \) in order to construct the posterior distribution.

An appropriate prior for a variance is an inverse gamma (IG) as the product of an inverse gamma and a normal is an IG and hence results in a posterior that is amenable to sampling. However, unless we have prior beliefs about the shape of the IG we should set the IG parameters to reflect a state of relative ignorance.

That is the

\[ P(\sigma_i^2 \mid K) \sim IG(a_i, b_i) \quad \text{Equation A 13} \]

Where the hyperparameters \( a_i \) and \( b_i \) are set close to 1. Hence the prior on the variance for regime 1 is:
\[ P\left( \sigma_i^2 \mid K \right) = \left( \sigma_i^2 \right)^{(1-a_i)} \exp \left[ -\frac{b_i}{\sigma_i^2} \right] \quad \text{Equation A 14} \]

Then by Bayes Theorem the posterior is:

\[ P\left( \sigma_i^2 \mid Y, \theta_{-\sigma_i^2}, K \right) \propto \]
\[ \frac{1}{(2\pi\sigma_i^2)^{T/2}} \exp \left[ -\frac{1}{2\sigma_i^2} \left( S + n(\bar{y} - \mu)^2 \right) \right] \]
\[ \cdot \left( \sigma_i^2 \right)^{(1-a_i)} \exp \left[ -\frac{b_i}{\sigma_i^2} \right] \quad \text{Equation A 15} \]

Which reduces to:

\[ \propto \left( \sigma_i^2 \right)^{(1-a_i)-T/2} \exp \left[ -\frac{1}{2\sigma_i^2} \left( S + n(\bar{y} - \mu)^2 \right) + b_i \right] \quad \text{Equation A 16} \]

And hence the posterior distribution of the variance of each regime is distributed as an inverse gamma.

\[ P\left( \sigma_i^2 \mid Y, \theta_{-\sigma_i^2}, K \right) \sim IG\left( a_i + \frac{T}{2}, S + n(\bar{y} - \mu)^2 + b_i \right) \quad \text{Equation A 17} \]

3. The transition probabilities

The probability of switching from regime 1 to regime 2 is defined as \( P_{12} \) where \( P_{12} = P\left(k(t+1) = 2 \mid k(t) = 1\right) \) and \( P_{11} = 1 - P_{12} \). The probability of switching from regime 2 to regime 1 is defined as \( P_{21} \) where \( P_{21} = P\left(k(t+1) = 1 \mid k(t) = 2\right) \) and \( P_{22} = 1 - P_{21} \). These four probabilities form the transition matrix, \( \Pi \).

In this section we will detail the construction of \( p12 \):

\[ P\left( p_{12} \mid Y, \theta_{-p_{12}}, K \right) \quad \text{Equation A 18} \]
As the transition probabilities with respect to Y are fully described by the state, K, we can ignore Y and simplify to:

\[ P\left( p_{12} \mid \theta_{p_{12}}, K \right) \]

Equation A 19

By Bayes rule:

\[ P\left( p_{12} \mid \theta_{p_{12}}, K \right) \propto P\left( K \mid \theta \right) P\left( p_{12} \right) \]

Equation A 20

We therefore need to define the likelihood \( P\left( K \mid \theta \right) \). Recall that we are considering a 2 regime case and therefore \( K_t = 1 \) or \( 2 \). At a given point \( y_t \) where the data is in regime 1 we stay in regime 1 (with probability \( p_{11} \)) or move to regime 2 (with probability \( p_{12} \)). Therefore the distribution of \( K \) is:

\[ \left( p_{11} \right)^{n_{11}} \left( p_{12} \right)^{n_{12}} \]

Equation A 21

where

\( n_{12} \) is the number of times the data switched from regime 1 to regime 2

\( n_{11} \) is the number of times the data remained in regime 1

An appropriate prior for transition probability would be a beta distribution.

\[ p_{12} \sim Beta(a, b) \]

Equation A 22

The beta parameters are also set to reflect low information content.

\[ p\left( p_{12} \right) = p_{12}^a \left( 1 - p_{12} \right)^b \]

Equation A 23

Therefore the posterior distribution:
\begin{align*}
P(p_{12} | Y, \theta_{-p_{12}}, K) &\propto (1 - p_{12})^{n_{11}} \cdot p_{12}^a \cdot p_{12}^{n_{12}} (1 - p_{12})^{b} \tag{Equation A 24} \\
\text{Or} \\
&\propto (1 - p_{12})^{b+n_{11}} \cdot p_{12}^{a+n_{12}} \tag{Equation A 25} \\
\text{And therefore} \\
P(p_{12} | Y, \theta_{-p_{12}}, K) &\sim Beta(a + n_{12}, b + n_{11}) \tag{Equation A 26}
\end{align*}
References


Casella, G and E I George, 1992 “Explaining the Gibbs Sampler”, The American Statistician, Vol 46, 167 - 190


