Bond Term Premium Analysis in the presence of Multiple Regimes

RON GUIDO and KATHLEEN WALSH
School of Banking and Finance, University of New South Wales

ABSTRACT

This papers addresses whether observed violations in the Liquidity Preference Hypothesis (LPH) can be explained by the presence of multiple regimes in the term premia. The investigation proceeds by directly testing the LPH via a series of inequality tests which allow the moments to be conditioned on observable information using an instrumental variables approach. The apparent rejection of the LPH is then investigated by modelling the term premia over time using a simple Bayesian Markov mixture model. The results suggest the presence of time varying term premia and multiple regimes which may explain the violations of the LPH.

The Liquidity Preference Hypothesis (LPH) states that the *ex ante* return on government bonds is a monotonically increasing function of time to maturity. In other words, conditional on all available information, the expected holding period return on a 10 year bond should be greater than that of a 7 year bond which is greater than that of a 5 year bond and so on. The intuition underpinning the LPH is that longer maturity bonds are more risky than shorter maturity and therefore a risk premium is included in the expected holding period return (see Hicks (1946) and Kessel (1965)). Therefore tests of the LPH amount to testing a series of inequality restrictions on the set of risk premiums.

Tests of the LPH have fallen into two broad categories. Firstly, unconditional tests of the LPH have been conducted by Fama (1984), McCulloch (1987) and Richardson,

Richardson and Smith (1992) with mixed results. Moreover, the power of these tests is questionable as the econometrician is discarding information available to the economic agents. The LPH makes inferences about the monotonicity of conditional expected returns; so unconditional tests lack the power to fully test the theory.

Secondly, the theory relates ex ante returns, which are unobservable. Many econometricians have attempted to address this issue by forming expectations models and testing the fitted values of the expected returns (see Fama (1986), Fama and Bliss (1987), Stambaugh (1988), Fama and French (1989) and Klemkosky and Pilotte (1992)). However, when the tests of the LPH are complicated with an expectations model it is difficult to decipher the true implications. That is, it is necessary to consider the joint statistical properties of both the LPH test and the expectations model. This complication makes the test results difficult to interpret.

More recently, statistical methods for testing inequality constraints have been developed. Boudoukh, Richardson, Smith and Whitelaw (1999) (hereafter BRSW) developed a test of inequality constraints allowing moments to be conditioned on observable information using an instrumental variables approach consistent with Hansen and Singleton (1982). This procedure overcomes the problem of unobservable ex ante risk premia and allows the econometrician to condition the returns on available information. This procedure was particularly applicable to tests of the LPH as it accounts for cross correlation amongst the different maturities of bond returns.

In the BRSW tests, bond returns were conditioned on an information set available to the econometrician. In this way they overcame the need to form an expectations model of expected return and at the same time they incorporated information available to economic agents. BRSW chose conditioning information that was drawn from economic theory. Specifically, they conditioned the ex ante returns on the shape of the yield curve due to its relation to the marginal rate of substitution. Investigating the sample period 1972 to 1994, BRSW found only weak evidence of a violation of LPH.

The conditional tests performed by BRSW have an interesting interpretation. When the yield curve is flat or upward sloping, as it is most of the time, then they were unable to reject the LPH. However, when the yield curve was downward sloping there was weak evidence to reject the LPH. It would seem that there were two states of the world evident in the risk premia, one where the LPH holds and the premium is positive and one where it is violated indicated by a negative premium, as identified by the conditioning agent. An interpretation of this result is that there are two regimes and that the premia switches between these positive and negative states over time. Recent analysis by Walsh (2004) demonstrated an approach for testing the existence of multiple regimes using a Bayesian framework, whilst specifying the regimes to be of opposite sign. This test can be directly applied in this context as we wish to identify two states of the world, one where the term premium is positive and supports LPH and one where the term premium in negative in violation of LPH. Identification of regime switching of the term premium would confirm the existence of two states of the world.

Prior research has considered analysis of the regime switching behaviour of interest rate markets. Multiple regimes in US interest rates have been studied by Hamilton (1988) and Gray (1996a). Of particular interest to this study, Gray (1996b) found evidence of regime switching in the Australian 90 day Bank accepted bill rates using weekly observations from 1978 to 1995. However, the regime switching analysis in each of these papers uses a classical framework whereas this study applies a Bayesian analysis in estimating the parameters of the regime switching process.

In addition, several authors have conducted studies of Australian short-term interest rates including Brailsford and Maheswaran (1998) and Gray and Treepongkaruna (2002). Other researchers have studied the nature of the term structure of interest rates (see Bhar (1996), Heaney (1994) and Alles (1995)). However, a separate analysis of either holding period returns or the spread of such returns has not been conducted using Australian data.

In summary we will conduct direct tests of the LPH using Australian data and then apply a complimentary analysis of regime switching using Bayesian techniques. The paper is structured as follows. Section I introduces the Methodology of the inequality and regime switching tests, Section II outlines the data used in the analysis, Section III discusses the results and the conclusions are drawn in Section IV.

I Methodology

Section A describes the multiple inequality testing procedure developed by BRSW (1999) and explains how it is applied in this paper. Section B outlines the application of Bayesian estimation and model selection techniques to regime switching models and explains how they can be applied to investigate the term premium and the slope of the yield curve.

A Inequality testing methodology

The LPH implies that the ex ante return on government bonds is a monotonically increasing function of time to maturity. Specifically, the LPH suggests that the expected holding period return on a j period bond is greater than that of a *j*-1 period bond. Testing this model is difficult due to two issues. Firstly, the LPH implies a set of inequality restrictions on the parameters to be estimated. Secondly, conditional expected returns are unobservable to the econometrician. BRSW developed a testing methodology that overcomes these issues and requires only weak assumptions on the underlying processes and little knowledge of conditional moments.

For consistency we have adopted the same notation used by BRSW. Let us define the term premium for a bond with maturity τ as the conditional expected one-period return in excess of the yield on a one-period bond:

$$E(P_{\tau,t+1}) \equiv E_t\left(r_{t,t+1}(\tau) - r_{t,t+1}(1)\right) \quad \textbf{(1)}$$

The LPH implies:

$$E\left[P_{\tau_k,t+1}\right] \geq E\left[P_{\tau_{k-1},t+1}\right] \geq \cdots \geq E\left[P_{\tau_1,t+1}\right], \tau_i > \tau_{i-1} \quad \textbf{(2)}$$

Or:

$$E_t\left(P_{\tau_k,t+1} - P_{\tau_{k-1},t+1}\right) \geq 0 \quad \textbf{(3)}$$

This methodology is particularly interesting as it allows us to test the inequality constraints in (1) without a model of return expectations.

We define I_t as a purely positive conditioning agent and normalize it to form the information set z_t. The conditioning agents used in this process can be dichotomous (0 or 1) or informative (using a positive measure of magnitude).

$$z_t = I_t \big/ E[I_t] \text{ so that } E[z_t] = 1 \tag{4}$$

As z_t is a non-negative random variable, multiplying both sides of (3) will not change the sign. Therefore we can write:

$$E_t\left(P_{\tau_k,t+1} - P_{\tau_{k-1},t+1}\right) \times z_t \geq 0, \tag{5}$$

Rearranging (5) and applying the law of iterated expectations,

$$E\left[\left(P_{\tau_k,t+1} - P_{\tau_{k-1},t+1}\right) \times z_t - \theta\right] = 0 \tag{6}$$

Under the null hypothesis, the parameter vector is positive, $\theta \geq 0$. To test this hypothesis we first estimate θ as the sample mean of the term premiums, conditional on z_t:

$$\hat{\theta}_t = \frac{1}{T} \sum_{t=1}^{T} \left(P_{\tau_k,t+1} - P_{\tau_{k-1},t+1}\right) \times z_t \tag{7}$$

Equation (7) provides a set of moment conditions that identify the vector θ in terms of observables - the ex post returns on bonds and the shape of the term structure. The next step is to estimate the same mean, but now under the restriction that it must be nonnegative, which we denote by $\hat{\theta}^R$. We then compare the vector of restricted and unrestricted means using a multivariate one-sided Wald statistic:

$$W \equiv T\left(\hat{\theta}^R - \hat{\theta}\right)\hat{\Omega}^{-1}\left(\hat{\theta}^R - \hat{\theta}\right) \quad \textbf{(8)}$$

where $\hat{\Omega}^{-1}$ is the sample covariance matrix of the conditional term premiums. We then evaluate significance using:

$$\sum_{k=0}^{N} \Pr\left[\chi^2 \geq c\right] w\left(N, N-k, \frac{\hat{\Omega}}{T}\right) \quad \textbf{(9)}$$

This multiple inequality test of the LPH incorporates conditioning information that does not require a structural model of the return series. However, as BRSW note, for the tests to be powerful the selection of the conditioning set must be founded on economic theory. BRSW identified two information sets both derived from information contained in the zero coupon yield curve. The first information set contained instances of non-monotonic yield curves and the second was downward sloping yield curves (where downward sloping was a subset of the non-monotonic set). In addition we considered a third information set of downward sloping with a negative change. That is, if the yield curve was downward sloping and had steepened in the previous period it was considered to carry additional information about the holding period returns. Therefore a series of information sets were constructed. We denote $I_{1Ut} = 1$ if the term structure is inverted or humped and 0 if the term structure is monotonically upward sloping, $I_{2Ut} = 1$ if downward sloping and 0 otherwise, and $I_{3Ut} = 1$ if the curve is downward sloping and the change in yield spread from the previous period is negative and 0 otherwise. We then construct the informative data sets. These are defined as the maximum difference between yields when that section of the curve is inverted (I_{1It}), the maximum difference between yields when the curve is downward sloping (I_{2It}) and the maximum difference between yields when the curve is downward sloping and the change is negative (I_{3It}).

As noted above, these information sets were then normalized by the expected value of I_t. A complimentary test to the inequality analysis is to directly consider the statistical properties of the time series. The analysis of the LPH under a regime switching model is discussed next.

B Bayesian Analysis of Multiple Regimes

i A Markov Mixture model of the Holding Period Return.

The implication of the multiple inequality tests is that conditional on an information set it may be possible to reject the LPH. An interpretation of this is that there may be multiple regimes present in the data generating process of the term premium, hence establishing the presence of time varying term premia. We postulate that it may be the case that the majority of the time the LPH holds and the expected risk premium on bond returns is positive. However, it may be possible for the process to switch to an alternative regime where the LPH doesn't hold, in which case the premium would be negative. To determine whether there are multiple regimes in the term premium and their effect on the LPH, we propose a Markov mixture model following the work of Hamilton (1989, 1994) and Gray (1996a) to describe the transition dynamics of the term premium. We briefly consider the motivation for the model and its development, followed by a discussion of its estimation and testing.

Once again, denote the term premium on a τ period bond over the one period bond as

$$E_t[P_{\tau,t+1}] \equiv E_t[r_{t,t+1}(\tau) - r_{t,t+1}(1)] \tag{10}$$

Under the LPH, this value should be strictly positive and time invariant:

$$E_t[P_{\tau,t+1}] = \mu \geq 0 \tag{11}$$

Our aim is to model the term premium, $E_t[P_{\tau,t+1}]$ conditionally by describing its evolution over time. Specifically we model the term premium as a state dependent process, governed by an unobservable discrete random variable, S_t. This state variable characterises the prevailing state of nature, where S_t is equal to one when the expected term premium is positive, and equal to 2 when the expected term premium is negative:

$$E_t[P_{\tau,t+1}] = \mu_t = \begin{cases} \mu_1 \text{ if } S_t = 1 \\ \mu_2 \text{ if } S_t = 2. \end{cases} \tag{12}$$

Following Hamilton (1989, 1994) and Gray (1996a), we assume that S_t evolves according to a first-order Markovian process. As a result, conditional on all previous information, the probability of a certain state of nature occurring is time varying, which we assume is governed by a transition probability matrix given by

$$\Pi = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} \tag{13}$$

where: $p_{ij} = \Pr(S_t = j \mid S_{t-1} = i)$ and $\sum_{j=1}^{2} p_{ij} = 1$ for all i. Given this Markov process, we describe the evolution of $P_{\tau,t}$ as

$$P_{\tau,t+1} = \mu(S_t) + \sigma(S_t)\varepsilon_t \qquad \textbf{(14)}$$

where

ε_t is a Gaussian random variable distributed $N(0,1)$,

$\mu(S_t) = \mu_i$ when $S_t = i$, and $\sigma(S_t) = \sigma_i$ when $S_t = i$.

The expression in (14) allows us to model the term premium and the variance of $P_{\tau,t}$ conditional on the state variable S_t :

$$E_t[P_{\tau,t+1} \mid S_t = i] = \mu_i \qquad \textbf{(15)}$$

$$Var_t[P_{\tau,t+1} \mid S_t = i] = \sigma_i^2 \qquad \textbf{(16)}$$

The central idea behind the modelling and subsequent estimation strategy is to view the term premia, $\mu(S_t)$, as signals which are measured with noise using the observed return of the multiperiod bond in excess of the single period bond, $P_{\tau,t}$. The value of $P_{\tau,t}$ in a given period can be influenced by a range of market forces. The noise in the measurement is governed by the innovation variable ε_t and the state dependent volatility, $\sigma(S_t)$. As such the observed process, $P_{\tau,t}$, is able to deviate from expected values at any point in time. The task is to therefore develop an approach to jointly estimate the state dependent parameters, and the latent state variable driving the data generating process.

ii Estimating the Model using Bayesian MCMC techniques

The simple model described above belongs to a general class where the distribution of observations depends upon a latent Markovian switching process on a discrete state space, which we seek to estimate and test using Bayesian Markov Chain Monte Carlo

(MCMC) methods, namely through the application of the Gibbs Sampler (see Albert and Chib (1993), Carter and Kohn (1994) and Chib (1995)). For the purposes of the current model specification, we define the following quantities. Let $\theta = \left\{ \mu_i, \sigma_i^2, p_{ij} \right\}_{\forall i,j}$ represent the parameter vector for the model in question, $X^t = (X_1,...,X_t)$ represent the observable data vector of bond holding period returns up to time *t* where we define $X_t = P_{\tau,t}$, and $S^t = (S_1,...,S_t)$ represents the vector of the latent state variables up to time *t*. In this context we seek to derive the joint posterior of the parameter vector and the latent state vector given the set of observable data: $f(S, \theta|X)$. Although analytically intractable, the Gibbs sampler can be applied to this problem by obtaining draws of θ and *S* which can be viewed as being drawn form the joint density of interest. By firstly augmenting the parameter vector by the latent vector *S*, the decomposition of the joint posterior density according to Bayes theorem, $f(S, \theta|X) \propto f(X \mid S, \theta) f(S| \theta) f(\theta)$, leads to the following algorithm:

1. Initialise θ
2. Sample S from $f(S|\theta)$.
3. Sample θ from $f(\theta|S, X)$.
4. Repeat steps 2 and 3.

Under mild regularity conditions, the iterates generated from this sampling algorithm will converge to their invariant target distribution. Given a sufficiently large number of draws, the parameters' marginal posterior distributions can be constructed. Furthermore, by averaging subsets of these simulations Bayesian estimators of the parameters can also be formed. For details, refer to Casella and George (1992), Tanner (1996), and Chib and

Greenberg (1996). Obtaining the Bayesian estimators for the model's parameters entails sampling from the set of full conditional posterior distributions. Sampling from these full conditional distributions form the basis for the Gibbs Sampler which leads to the generation of iterates from the joint distribution of the parameters governing the models which are now presented.

We adopt the block-sampling scheme developed by Carter and Kohn (1994) and Chib (1995) to generate the state variables, S^n. Suppressing θ for notational convenience we generate S^n from the distribution Pr($S^n | X^n$). By noting the decomposition Pr($S^n | X^n$) = $\Pr(S_n \mid X_n) \prod_{t=1}^{n-1} \Pr(S_t \mid X_t, S_{t+1})$, the algorithm samples the vector S^n as a block using the joint conditional distribution, Pr($S^n | X^n$), rather than from the set of individual full conditional distributions $\Pr(S_t \mid X_n, S_{-t})$. Since the process is Markov, and therefore correlated, such blocking will lead to faster convergence to the posterior distribution, and is therefore preferred to the single move sampling.

Conditional on having generated the latent state variable, S, it is a relatively straightforward task to sample from the full conditional distributions of the parameters that form the transition probability matrix, Π. Let $\Pi_i \equiv (p_{i1},.., p_{iK})$ represent the ith row of Π; the vector of state transition probabilities given $S_t = i$. By construction, these probabilities must sum to unity. The full conditional distribution for Π_i can then be expressed by Bayes rule as:

$$\Pr(\Pi_i \,|X_n,\ S^n, \theta_{/\Pi_i}) = \Pr(\Pi_i \,|\ S^n) \propto \Pr(S^n \,|\Pi_i)\Pr(\Pi_i). \tag{17}$$

Given that S_t evolves according to a first order Markov process, the joint likelihood for S^n, $\Pr(S^n \,|\Pi_i)$, can be expressed as a Dirichlet process. By adopting conjugate priors for $\Pr(\Pi_i)$, the posterior denisty too will be Dirichlet, and so that parameters for Π_i can be jointly sampled from the following Dirichlet distribution :

$$\Pi_i \,|\ S^n \sim \mathrm{Dir}(d_{i1}, d_{i2}, \ldots, d_{ik}) \tag{18}$$

where $d_{ij} = n_{ij} + u_{ij}$, n_{ij} represents the number of transitions from state i to state j : $n_{ij} = \sum_{t=2}^{n} \mathbf{I}_{it-1}\mathbf{I}_{jt}$, and u_{ij} are the hyperparameters of the Dirichlet prior, where $\mathbf{I}_{it}$ equals one when S_t equals i, and zero otherwise.

Given that the joint sampling conditional density or conditional likelihood for X_t is Gaussian, using uninformative conjugate priors for $\mu_i : N(\xi, \kappa^{-1})$ and $\sigma_i^2 :$ $\mathrm{IG}(\alpha_i, \beta_i)$, and applying Bayes rule, it is straightforward to construct the full conditional densities for μ_i and σ_i^2 :

$$\mu_i \,|\, X_T, S^T, \sigma_i^2 \sim N(BA^{-1}, A^{-1}) \tag{19}$$

where $A \equiv \sigma_i^{-2}\sum_{t=1}^{T}\mathbf{I}_{it} + \kappa$, $B \equiv \sigma_i^{-2}\sum_{t=1}^{T} X_t\mathbf{I}_{it} + \kappa\xi$, N is a Normal Distribution ; and

$$\sigma_i^2 \,|\, S^T, X_T, \mu_i \sim \mathrm{IG}(\frac{1}{2}\sum_{t=1}^{T}\mathbf{I}_{it} + \alpha_i, S), \tag{20}$$

where $S = \beta_i + 0.5\sum_{t=1}^{T} \frac{(X_t - \mu_i)^2}{T}\mathbf{I}_{it}$ and IG is an Inverse Gamma.

iii Informative versus Uninformative Bayesian Priors

The priors adopted in the previous section typically have their hyper parameters set such that they are disperse or flat to reflect the lack of prior information possessed by the experimenter. Given the theoretical structure of the model and the LPH, it may be of interest to consider an alternative to uninformative priors and suggest an informative structure which the model may imply. Specifically, by investigating whether there is positive and negative term premia, we are able to suggest an informative prior structure on the state dependent means: $\Pr(\mu_i \mid S_t)$. Under S_t = 1, a positive term premium state, our prior for μ_i would be such that the variable is restricted to be positive. Similarly, for S_t = 2, a negative term premium state, the prior for $\Pr(\mu_i \mid S_t)$ could reflect that μ_i is strictly negative. We therefore suggest the following proper truncated normal distributions for the state dependent priors:

$$\mu_1 \sim N_{tr}(0, \kappa^{-1})\mathbf{I}_{\mu_1 > 0} \qquad \textbf{(21)}$$

$$\mu_2 \sim N_{tr}(0, \kappa^{-1})\mathbf{I}_{\mu_2 < 0} \qquad \textbf{(22)}$$

where the variance κ^{-1} is chosen to be sufficiently large, to ensure μ_i has sufficient support. Thus the prior for μ_1 is a truncated normal with mean zero, variance κ^{-1} that is bound to be strictly positive, and the prior for μ_2 is truncated normal with mean zero, variance κ^{-1} that is bound to be strictly negative. Adopting these priors then results in the following conditional distributions for μ_1 and μ_2 which are used in the Gibbs sampler:

$$\mu_1 \mid X_T, S^T, \sigma_1^2 \sim N_{tr}(B_1 A_1^{-1}, A_1^{-1})\mathbf{I}_{\mu_1 > 0} \qquad \textbf{(23)}$$

where $A_1 \equiv \sigma_1^{-2}\sum_{t=1}^{T}\mathbf{I}_{1t} + \kappa$, $B \equiv \sigma_1^{-2}\sum_{t=1}^{T} X_t \mathbf{I}_{1t}$, and

$$\mu_2 \mid X_T, S^T, \sigma_1^2 \sim N_{tr}(B_2 A_2^{-1}, A_2^{-1})\mathbf{I}_{\mu_2 > 0} \qquad \textbf{(24)}$$

where $A_2 \equiv \sigma_2^{-2}\sum_{t=1}^{T}\mathbf{I}_{2t} + \kappa$, $B \equiv \sigma_2^{-2}\sum_{t=1}^{T} X_t \mathbf{I}_{2t}$, and

iv Model Selection

When estimating these models we invariably have a number of specifications with which we would like to ascertain as to which is more appropriate. We would like to ask whether in fact a Markov mixture representation is better than a simple unconditional model which has only a single regime in describing the data. Further, if the Markov mixture models are superior, is the model which imposes a more informative structure, more adequate than one which only uses uninformative priors. The issue of model selection is addressed by seeking to establish which model has he highest posterior probability that the data is generated by that model. Several methods have been developed that seek to estimate a model's marginal likelihood, *Pr*(Data|Model), using MCMC sampling techniques including Carlin and Chib (1995) and Kass and Raftery (1995). The approach adopted in this paper is the procedure developed by Chib (1995) which computes the marginal likelihood of the model using reduced MCMC sampling schemes when the full conditionals for the parameters are available in closed form. This approach is based on the following identity which can be easily constructed using Bayes Rule:

$$p(X \mid Model) = f(X \mid \theta)p(\theta) / p(\theta \mid X) \qquad \textbf{(25)}$$

Taking logs, an estimate of $p(X|Model)$ can be expressed as

$$\log p(X \mid Model) = \log f(X \mid \hat{\theta}) + \log p(\hat{\theta}) - \log p(\hat{\theta} \mid X) \qquad \textbf{(26)}$$

where $p(\hat{\theta} \mid X)$ is an estimate of the posterior distribution and $\hat{\theta}$ is chosen to be a point of high posterior density (typically the posterior mean) so as to maximize the accuracy of this approximation. Chib (1995) develops a straightforward Gibbs sampling algorithm from which the posterior distribution estimates can be obtained. Given the Markov mixture representation of the model, the marginal likelihood estimate of (24) can be easily constructed. For a review of these methods and their use in model selection, see Han and Carlin (2001)

When conducting Bayesian analysis of mixture models, parameter estimation can be complicated by the inability of the Markov chain to generate parameter iterates that belong solely to a single mixture component. This so-called label switching problem generally arises when taking a Bayesian approach to parameter estimation within mixture models. The problem has been identified by several authors, including Diebolt and Robert (1994), Richardson and Green (1997), and Fruhwirth-Schnatter (2001) and Celeux et al (2000). The problem arises due to the fact the likelihood and hence posterior distribution of the model parameters under diffuse priors are symmetric and hence invariant under relabelling of the mixture components. The MCMC sampling thus produces posterior distributions that are multi-modal and highly symmetric, rendering useless inference methods that summarise the parameters by their marginal distributions (e.g. by computing the posterior mean and mode). Several attempts have been made to remove label switching, the most popular being those which impose artificial identifiability constraints (see Richardson and Green, 1997). Yet this approach does not always provide a satisfactory solution particularly when there may be no prior knowledge as to how to label the parameters. The most promising approach however has been that developed by Stephens (2000) which attempts to relabel the iterates for each parameter by selecting the relabelling that minimises the posterior expected loss for a certain class of loss functions. An online algorithm has been adopted in this study, which attempts to relabel the parameters following each sweep of the Gibbs Sampler. All results reported in this study have been successfully relabelled using this algorithm. For details, the reader is referred to Stephens (2000).

v A Markov mixture model of the Conditioning Agents

Although we are testing for multiple regimes in the term premia, implicit in the inequality tests is the assumption that the regimes and hence information content are driven by the conditioning agents. Although the conditioning agents selected have been drawn from economic theory, for robustness it is prudent to check the distributional properties of said agents.

The evidence reported by BRSW which suggests that the LPH is only weakly violated, is fundamentally related to the choice of conditioning agents. Specifically, the examination of a possible violation of the LPH was based on conditioning on the shape of the yield curve, namely when it is negative or inverted. It would be interesting to consider whether such conditioning agents employed by BRSW represent a distinct information state during these periods when the yield curve is inverted, or whether such events are more anomalous which are unable to justify their use as a conditioning agents. Looking at this from another angle, BRSW found that when the slope of the yield curve is negative, there is an increased probability that the LPH is violated, although weak. Therefore, the information is contained, not in the term premia but in the shape of the yield curve. This seems to suggest that there are two regimes present in the slope of the yield curve – one positive and one negative. Thus, although they may be economically valid agents, the negative regime must be statistically significant in order to justify the yield curve as a conditioning agent. If this is not the case, tests of the LPH based upon such agents are invalid.

In order to investigate this and in line with previous studies on interest rates, we apply the Markov mixture model to the conditioning agent used in the LPH tests; the slope of the yield curve and investigate whether the we are able to uncover a significant second regime where the yield curve is characterised to be negative. Specifically we test whether the conditioning agent, I_t, (slope of the yield curve) can be characterised as a mixture process driven by the latent state variable S_t governed by (13)

$$E_t[I_{t+1}] = \alpha_t = \begin{cases} \alpha_1 \text{ if } S_t = 1 \\ \alpha_2 \text{ if } S_t = 2. \end{cases} \quad \textbf{(27)}$$

$$Var_t[I_{t+1}] = \delta_t = \begin{cases} \delta_1 \text{ if } S_t = 1 \\ \delta_2 \text{ if } S_t = 2. \end{cases} \quad \textbf{(28)}$$

Given this Markov process, I_t evolves according to

$$I_{t+1} = \alpha(S_t) + \delta(S_t)\upsilon_t \quad \textbf{(29)}$$

where

υ_t is a Gaussian random variable distributed $N(0,1)$,

$\alpha(S_t) = \alpha_i$ when $S_t = i$, and $\delta(S_t) = \delta_i$ when $S_t = i$.

Thus, similar to the multiple regime models for the term premium, we model the conditioning agent as a state dependent process with time varying mean and volatility. To do this we adopt the MCMC techniques derived in earlier sections in order to estimate the parameters specified in (13), (27) and (28), and the latent state variable driving the data generating process. We further utilise these Bayesian techniques for model selection in order to investigate the presence of possible negative states in the conditioning agent through the use of informative and uninformative priors.

II Data

The focus of our research is the ex ante term premium on government bonds therefore a time series of term premiums for a variety of maturity dates needed to be constructed. The term premium for a bond is the holding period return from t to t+1 less the risk free rate over the same period. Therefore it was necessary to source each element.

Treasury Fixed Coupon Tender Results containing all Australian Commonwealth bond tenders by maturity and coupon amount from 1991 to 2001 was provided by the Australian Office of Financial Management. After removal of duplicate bonds, 33 remained in the Commonwealth treasury sample over the 10 year period.

Reuters[*] provided the clean prices, yields, settlement date and accrued interest for all bonds in the sample. Government Bonds in Australia trade at the gross price (clean price plus accrued interest) with coupons paid every 6 months. Hence the holding period return on bonds is:

$$r_{t,t+1}(\tau) = \ln\left(\frac{GP_{t+1} + c}{GP_t}\right) \quad \textbf{(30)}$$

where

$r_{t,t+1}(\tau)$ is the holding period return from t to t+1 on a bond with maturity τ,

GP_t is the Gross Price at t, and

c is the coupon amount paid to the holder of the bond at t+1.

Holding period returns for maturities of 1, 3, 7 and 10 were constructed from the above data set resulting in 2540 data points for each of the four maturities for the period 12 December 91 to 10 December 01.

The risk free rate used in similar studies was the 90 day BAB rate. However, as the holding period, in this analysis, was daily it was more appropriate to use a guaranteed

[*] Thank you to Cushla Edwards at Reuters for her considerable help in accessing this data

return for the holding period and as such the overnight cash rate was used as the risk free rate $y_{t,t+1}$.

Recall that the inequality analysis on the holding premiums required the construction of an information set. Consistent with BRSW we considered the yield curve as an appropriate conditioning agent. Daily yield curve data for 3, 5 and 10 year maturities was sourced from Reuters for the period 1991 to 2001. The yield curve series was then transformed into purely positive conditioning agent sets.

III Results

Initial inspection of the holding period returns showed that both mean returns and standard deviations were increasing with time to maturity as reported in Table I, panel A. The mean annual returns ranged from 6.18% for bonds with a 1 year maturity to 9.38% for bonds with a 10 year maturity. The standard deviations ranged from 1.48% to 8.78% respectively. However, our focus here is on the term premium, which is net of the risk free rate. The average term premium on bonds with 1 year to maturity is 0.4% whilst the premium on bonds with 10 years to maturity is 3.45%. Construction of conditional mean returns, however, produced a very different account.

Figure 1 graphs the annualised unconditional term premium and reflects the LPH in that the premium is an increasing function of time. However when we condition the returns on one of the information sets the result is reversed. That is, it seems that the conditional returns are a negative function of time to maturity. Panels B to C graph the conditional risk premium conditioned on lagged instruments. In all occasions the average returns were negative however the most striking result is in Panel C. When conditioned on a downward sloping yield curve with a negative change then not only are the returns all negative but the returns seem to be a decreasing function of maturity. This is exactly opposite to the expectations of the LPH.

Table I also presents the correlation structures of both the holding period returns and the yield curve. As expected there is a higher correlation of holding period returns between near to maturity bonds. The correlation between 10 and 7 year bonds was .9501 whilst the correlation between 1 and 10 year bonds was only .478. Similarly, the yields showed cross correlations ranged from .673 between 5 year yields and the overnight cash rate to .9905 between 5 and 3 year yields. Therefore it is important that any formal test of the returns on the yields accounts for the correlation across maturities.

There are two points to note at this time. Firstly, we are only graphing *ex post* conditional returns at this point so whilst these results appear compelling they do not constitute a full test of the LPH. Secondly, the instances of non-monotonic, downward sloping, and downward sloping and negative change term structures are very few (3.1%, 1.8% and 1.0% respectively). Therefore, whether these conditioning agents can be used as a trading rule is unclear.

In the first formal test of the LPH we applied uninformative conditioning agents as used in the graphs discussed above. Results for these tests are provided in Table II. They show that the mean values of the differenced premiums are often negative. Specifically, when conditioned on downward sloping and negative change yield curves, the annualised mean premium of the 3 year over 1 year bonds is 3.56%, the 7 less 3 is –3.20% and 10 less 7 is –5.07%. However, when tested formally the LPH as a system, the results are not statistically significant. This is consistent with the LPH. Importantly as we condition on a finer information set[†] the significance improves from p values of 58% to 35% to 21%. However, we were unable to identify an economically sound conditioning agent that was able to reject the LPH.

We then applied the informative agents to see whether the magnitude of the yield curve would result in more significant returns. The annualised mean premia are reported in Table III. The change in mean values seems quite substantial with the annualised premium reaching –36.975% when conditioned on downward sloping and negative change yield curves. However, as with the uninformative instruments, when formally tested we are unable to reject the LPH. Although resulting p values were lower on each of the uninformative sets they were still insignificant overall. The lowest p value was 17.09% when the returns were conditioned on downward sloping and negative term structures.

We also tested the data over a shorter period from 1995 to 2001 and although not presented, the results were consistent. We were unable to resort to sub period analysis because of the small frequency of non-monotonic term structures during the ten year

[†] The broadest information set was non-monotonic yield curves. Downward sloping yield curves is a subset of monotonic and downward sloping with a negative change is a finer selection again.

period. In fact there was a period from January 1995 to August 1999 where the yield curve was always upward sloping.

From the results of the inequality analysis it seems that we have identified a driving factor in the returns, however it is not statistically significant. Nevertheless, we have certainly highlighted the possibility of two 'states of the world' in holding period returns. However we were unable to fully identify an economically significant conditioning agent. The above results consider the term premia from an economic perspective by utilising a set of conditioning agents drawn from economic theory. It is possible that not all possible conditioning agents were considered for these tests. Therefore, for robustness, we now analyse the term premia from a purely statistical perspective by employing a Bayesian regime switching model. This second prong of the analysis therefore directly compliments the inequality analysis.

The two regime Markov mixture model was run on the holding period return using the difference between the 10 year and 1 year returns. We seek to establish whether models describing two regimes in the term premium are better than an unconditional model which supposes only one state prevails. We compare the results of a single regime model to the results obtained from the two regime mixture models based on uninformative and informative priors. The results are summarised in Table IV. For each model considered, the parameter estimates represent the posterior means of the marginal densities generated from the Gibbs sampler. The Gibbs sampler is run for 11,000 iterations, with the first 1000 discarded; the following 10,000 iterations used to compute the sample means. The parameter estimates have also numerical standard errors computed using Newey and West's (1986) approach to adjust for heteroskedasticity and autocorrelation. We also report for each model the log marginal likelihoods using Chib's (1995) reduced sampler.

The results suggest that there is overwhelming evidence in favour of two regimes. Both Markov mixture representations display significantly larger log marginal likelihoods than the single regime model. It is interesting to check whether the informative priors were an appropriate modelling strategy. The close log marginal likelihoods (10,357 for the uninformative model and 10,480 for the informative model) suggest both models are adequate representations of the data. This is in contrast to the single regime log likelihood of 9,819. This is further reflected by the highly similar estimates for all parameters across both dual regime specifications. However, the larger marginal likelihood and the smaller numerical standard errors for all the parameters under the informative model suggest that the use of informative priors is preferred. Both Markov mixture representations however highlight the presence two regimes in the expected term premium; one which is positive and one negative. Focussing on Panel C, using the informative prior specification, these term premiums approximately equalled 13.00% p.a. for regime 1 and -18.48% p.a. for regime 2. The variance under the regime when expected returns are negative is generally smaller than the variance in the state when the expected term premium is positive. This suggests there is generally more "noise" in the observed holding period return in those periods where the term premium is expected to be positive, than in those periods when the term premium is expected to be negative. The transition probabilities from both specifications also suggest that the duration of the positive regime tends to be much longer than those periods characterised by negative regime. From Panel C, the persistence of regime 1 displayed by transition probability p_{11} equalling 68.01% compared to the persistence of regime 2 reflected by transition probability p_{22} equalling 34.18% suggests that the duration of regime 1 is twice as long as the duration of regime 2.

As noted earlier, the apparent lack of rejection of the LPH found both by BRSW and in the inequality tests conducted in this study were dependent on the selection of conditioning agents; namely the term structure of interest rates. The following set of results examines the information content of this agent. Specifically we test whether the observed negative slope of the yield curve constitutes a distinct regime. For this set of analyses we take the slope of the yield curve to be the difference between the yield on the 10 year bond and the yield on the 3 year bond. Similar to the approach taken above we compare the results of a single regime model to the results obtained from the two regime mixture models; one based on uninformative priors and the other based on informative priors which restrict one regime to have a positive mean, and second regime to have a negative mean. Table V summarises these results. The results from the Gibbs sampler suggest that a multiple regime representation of the data seems to be more appropriate than a single regime. The log marginal likelihood for the unconditional single regime model in Panel A equals 10,489, which is exceeded by the log marginal likelihoods of both mixture models in Panels B and C (11,779 for the model with uninformative priors model and 11,194 for the model with informative state dependent priors). Of more importance are the differences between the two mixture models. Unlike the results for the holding period return spread, the results of the yield curve spread in Table V show that the unrestricted mixture model represented in Panel B is superior to the restricted model in Panel C as evidenced by the higher marginal likelihood. Furthermore, we find that the two conditional means of the two regimes in Panel B are both positive, with values of 0.4873% p.a. and 1.0961% p.a. respectively. When examining the parameter estimates of the restricted model using informative priors in Table V, Panel C, we see that even when restricting one regime to have negative mean, its value is essentially zero (-0.0008%). These results clearly suggest the data is unable to support the presence of a distinct regime characterised by negative mean in the slope of the yield curve. Therefore, although we were able to uncover regimes with positive and negative means in the holding period return spread were unable to identify a negative regime in the yield curve spread.

We have approached the analysis of the LPH from two angles. First, we identified possible conditioning agents that would identify states of the world where the LPH was violated. This *ex ante* analysis was unable to detect any statistically significant negative states. We then conducted an *ex post* multiple regime test and determined that the data was characterised by two regimes of opposite sign. In other words, the LPH may be violated in some states of the world but the conditioning agent we selected in the inequality test was unable to uncover those states. For robustness, we then tested for positive and negative regimes in the yield curve spread which was our chosen conditioning agent. We found that the yield curve spread exhibited two regimes but they were both positive, providing confirmation that our choice of conditioning agent was inadequate.

IV Conclusions

This paper tests the LPH using returns on Australian Government Bonds. Using a lagged information set we provide a direct test of the LPH. Although initial indications suggested the series was in violation of the LPH, formal *ex ante* tests incorporating the cross correlations between maturities were unable to reject the hypothesis. In addition we conducted multiple regime analysis on the holding period returns. We sought to investigate whether a second regime existed in the term premium characterised by a negative conditional mean. The results from this *ex post* analysis suggest means of positive and negative value in the holding period returns. It is important to note that although a negative regime was uncovered, no conditioning agent was employed in the identification of the regimes. Hence prediction of subsequent positive and negative term premia cannot be made using this model. Therefore the choice of conditioning agent is paramount in the identification of violations in the LPH.

We then investigated more closely the conditioning agents adopted in the direct tests of the LPH. We are unable to identify statistically significant periods where the slope of the yield curve is negative. In the direct tests of the LPH we relied on the shape of the yield curve, namely periods when it is negative, to identify states where the LPH was violated. If the shape of the yield curve cannot characterised by positive and negative regimes, then this suggests that our choice of conditioning agent is inadequate. Therefore, the next step in this line of research will be to identify both economic and statistically significant conditioning agents which would improve the validity of tests of the LPH.

Table I Descriptive Statistics of Holding Period return and Yield Curve Data

Panel A reports the annualised values for the sample of daily holding period returns of Australian Government bonds of differing maturities (1,3,7 and 10 years) for the period 12 Dec 1991 to 10 Dec 2001. The correlation structure is also reported. Bond price data was sourced from Reuters using a complete list of Bond tenders provided by the Australian Office of Financial Management. Daily holding period returns were calculated for each bond and then the nearest maturity to 1, 3, 7 and 10 years was selected to form a series of returns of differing maturities. Panel B reports the sample of zero yield curve data and correlation structure for the overnight cash rate, 90 Bank Accepted Bill and 3, 5 and 10 year bonds for the period 12 Dec 1991 to 10 Dec 2001. Yield curve data was sourced from Reuters and reported as annual values.

Panel A

Bond Holding Period Returns	**1 year**	**3 year**	**7 year**	**10 year**
Annualised Mean return	6.18%	7.57%	8.89%	9.38%
Annualised standard deviation	1.48%	3.56%	6.84%	8.78%
Return Correlations	**1 year**	**3 year**	**7 year**	**10 year**
1 year	1			
3 year	0.653135	1		
7 year	0.51122	0.806414	1	
10 year	0.478556	0.787605	0.95016	1

Panel B

Yield Curve	**Overnight cash rate**	**90 Day BAB rate**	**3 year**	**5 year**	**10 year**
Annualised Mean return	5.756%	5.882%	6.582%	6.910%	7.343%
Annualised std deviation	1.019%	1.073%	1.449%	1.480%	1.572%
Yield Curve	**Overnight cash rate**	**90 Day BAB rate**	**3 year**	**5 year**	**10 year**
Overnight cash rate	1				
90 Day BAB	0.962161	1			
3 year	0.725099	0.824666	1		
5 year	0.673367	0.768936	0.990506	1	
10year	0.655237	0.730797	0.970697	0.990499	1

Table II Direct Tests of the LPH using lagged uninformative instruments.

The sample of daily holding period returns consisted of Australian Government bonds of differing maturities (1,3,7 and 10 years). Term premiums were calculated as holding period return for each maturity in excess of the daily cash rate (tests were also run with the 90 day BAB yield and the results were consistent). The statistic W is a joint test of multiple inequality restrictions corresponding to lagged uninformative information sets. The estimators,$\theta\mu Zt+$, 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).

Uninformative Instruments	**Non-monotonic**	**Downward sloping**	**Downward sloping and Negative change**
Probability of state	3.1%	1.8%	1.0%
3 yr less 1yr			
Mean $\theta_{\mu Zt+}$	0.875%	0.675%	-3.56%
(standard error)	(0.029)	(0.039)	(0.0305)
7 yr less 3yr			
Mean $\theta_{\mu Zt+}$	-0.950%	-2.50%	-3.25%
(standard error)	(0.055)	(0.0700)	(0.057)
10yr less 7yr			
Mean $\theta_{\mu Zt+}$	-1.000%	-5.925%	-5.200%
(standard error)	(0.0502)	(0.0815)	(0.06275)
Multiple inequality restrictions statistic W	0.03918	0.53084	1.452266
(p-value)	0.586335	0.352155	0.209736

Table III Direct Tests of the LPH using lagged informative instruments.

The sample of daily holding period returns consisted of Australian Government bonds of differing maturities (1,3,7 and 10 years). Term premiums were calculated as holding period return for each maturity in excess of the daily cash rate (tests were also run with the 90 day BAB yield and the results were consistent). The statistic W is a joint test of multiple inequality restrictions corresponding to lagged informative information sets. The estimators,$\theta\mu Zt+$, represent the annualized 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).

Informative Instruments	**Nonmonotonic Yield Curve**	**Downward sloping**	**Downward sloping and Negative change**
Probability of state	3.1%	1.8%	1.0%
3 yr less 1yr			
Mean $\theta_{\mu Zt+}$	-2.45%	0.600%	-16.875%
(standard error)	(0.04225)	(0.02825)	(0.16275)
7 yr less 3yr			
Mean $\theta_{\mu Zt+}$	-12.375%	-10.750%	-36.975%
(standard error)	(0.1175)	(0.11675)	(0.37725)
10yr less 7yr			
Mean $\theta_{\mu Zt+}$	-10.950%	-9.375%	-31.850%
(standard error)	(0.08775)	(0.1000)	(0.31375)
Multiple inequality restrictions statistic W	1.561724	0.993393	1.101856
(p-value)	0.182055	0.296686	0.170955

Table IV Bayesian Analysis of holding period return spread using Markov mixture models

The holding period return spread is the difference between the holding period returns on 10 year bonds less the holding period returns on 1 year bonds. The table reports the estimation results from the Gibbs sampling scheme. Panel A reports the mean (μ) and variance (σ^2) estimates for the unconditional, single regime model. These estimates are the Bayesian posterior means of the generated marginal densities computed after running 10,000 iterations after a suitable initial period. Newey and West (1986) numerical standard errors are also reported for these estimates. Model significance is measured using Chib's (1995) approach to construct the model's log marginal likelihood. Panel B presents the results for the 2 regime model that do not use informative priors and includes estimates for the transition probabilities, p_{11} and p_{22}. Panel C presents the results for the two regime model using truncated normals as the informative priors.

Panel A: Unconditional Model

	μ_1	σ^2_1
Estimates	2.94%	0.0062
(standard errors)	(.0000327)	(.000000238)
Log Marginal Likelihood	9818.5	

Panel B: Markov Mixture 2 Regime Model Uninformative Priors

	μ_1	μ_2	σ^2_1	σ^2_2	p_{11}	p_{22}
Estimates	12.66%	-19.93%	0.0140	0.0027	70.41%	32.03%
(standard errors)	(.00011)	(.0003781)	(.00000230)	(.0000122)	(.0083)	(.0080)
Log Marginal Likelihood	10357					

Panel C: Markov Mixture 2 Regime Model Informative Priors

	μ_1	μ_2	σ^2_1	σ^2_2	p_{11}	p_{22}
Estimates	13.00%	-18.48%	0.0026	0.0134	68.01%	34.18%
(standard errors)	(.0000861)	(0.0003254)	(.00000204)	(.0000102)	(0.0063)	(0.006)
Log Marginal Likelihood	10480					

Table V Bayesian Analysis of yield spread using Markov mixture models

The yield spread is the difference between the yield on 10 year bonds less the yield on 3 year bonds. The table reports the estimation results from the Gibbs sampling scheme. Panel A reports the mean (μ) and variance (σ^2) estimates for the unconditional, single regime model. These estimates are the Bayesian posterior means of the generated marginal densities computed after running 10,000 iterations after a suitable initial period. Newey and West (1986) numerical standard errors are also reported for these estimates. Model significance is measured using Chib's (1995) approach to construct the model's log marginal likelihood. Panel B presents the results for the 2 regime model that do not use informative priors and includes estimates for the transition probabilities, p_{11} and p_{22}. Panel C presents the results for the two regime model using truncated normals as the informative priors.

Panel A: Unconditional Model

	μ_1	σ^2_1
Estimates	.76%	.0000146
(standard errors)	(.00000113)	(.000000007)
Log Marginal Likelihood	10489	

Panel B: Markov Mixture 2 Regime Model Uninformative Priors

	μ_1	μ_2	σ^2_1	σ^2_2	p_{11}	p_{22}
Estimates	.4873%	1.0961%	.000005297	.000005664	96.66%	95.8%
(standard errors)	(.0000486)	(.0000685)	(.0000001)	(.0000003)	(.0119)	(.0093)
Log Marginal Likelihood	11779					

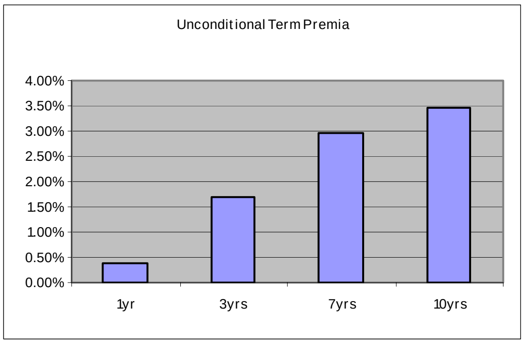
Panel C: Markov Mixture 2 Regime Model Informative Priors

	μ_1	μ_2	σ^2_1	σ^2_2	p_{11}	p_{22}
Estimates	.8569%	-.0008%	.00000995	.0000056	98.28%	90.09%
(standard errors)	(.0000053)	(.00000017)	(.000000008)	(.00000015)	(.0003)	(.0013)
Log Marginal Likelihood	11194					

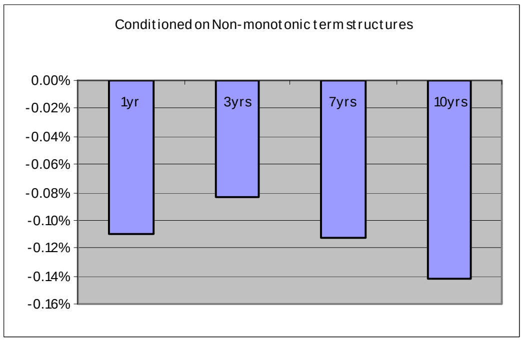
Figure 1 Conditional risk premiums

This Figure reports the annualised risk premiums on 1, 3, 7 and 10 year returns conditioned on lagged instruments. Specifically, Panel A shows a graph of the unconditional risk premium, which is the average of all daily risk premia in the sample. The average is then annualised assuming daily compounding. The conditional risk premium averages are calculated as the average of all risk premia at t conditioned on information at time t-1. (The risk premia was included in the sample if the lagged conditioning information was equal to 1).

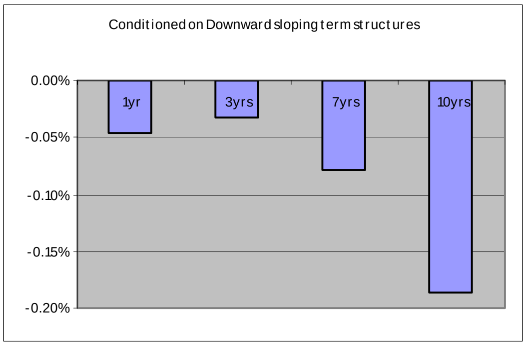
Panel A



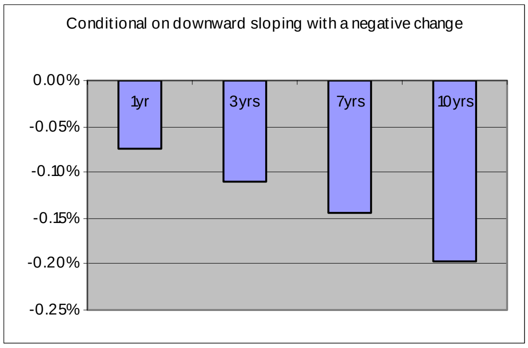
Panel B



Panel C



Panel D



References

Albert, J. H., and Chib, S, (1993), "Bayes Inference via Gibbs Sampling of Autoregressive Time Series Subject to Markov Mean and Variance Shifts", *Journal of Business & Economic Statistics*, 11, 1-15.

Alles, Lakshman, (1995), "Time varying risk premium and the predictive power of the Australian term structure of interest rates", *Accounting and Finance*, 35, 77-97

Bhar, Ramaprasad, (1996), "Modelling Australian bank bill rates: A Kalman Filter Approach", *Accounting and Finance*, 36,1-14

Boudoukh, J., M Richardson, T Smith and R Whitelaw, (1999), "Ex Ante Bond Returns and the Liquidity Preference Hypothesis", *Journal of Finance*, 54, No 3, 1153-1167.

Brailsford, Timothy J and K Maheswaran, (1998), "The Dynamics of the Short term interest rate", *Australian Journal of Management*, 23, 213-234

Carlin, B. P., and Chib, S. (1995), "Bayesian Model Choice via Markov Chain Monte Carlo Method", *Journal of Royal Statistical Society*, Ser.B, 57, *3*, 473-84.

Carter, C. K and R Kohn, (1994), "On Gibbs sampling for state space models", *Biometrika*, 81, 3, 541-53.

Casella, G and E I George, (1992) "Explaining the Gibbs Sampler", *The American Statistician*, Vol 46, 167 - 190

Celeux, G., Hurn, M. and Robert, C.P. (2000) Computational and Inferential Difficulties with mixture posterior distributions, *Journal of the American Statistical Association* , 95, 957-970.

Chib, S., (1995), "Marginal Likelihood From the Gibbs Output", *Journal of the American Statistical Association*, 90, 1313-21.

Chib, S., and Greenberg, E, (1995), "Hierarchical analysis of SUR models with extension to correlated serial errors and time varying parameter models", *Journal of Econometrics*,

Tanner, M. A. (1996), *Tools for Statistical Inference*, 3rd ed., New York: Springer., 339-60.

Chib, S and E Greenberg, (1996), "Markov chain monte carlo simulations in econometrics" *Econometric Theory*, Vol 12, 409-431

Diebolt, J., and Robert, C. P, (1994), "Estimation of finite Mixture Distributions trough Bayesian sampling", *Journal of Royal Statistical Society*, 56, *2*, 363-375.

Fama, Eugene F, (1984),"Term Premiums in Bonds" *Journal of Financial Economics*, 13, 529-546

Fama, Eugene F, (1986), "Term Premiums and default premiums in money markets" *Journal of Financial Economics*, 17, 175-196

Fama, Eugene F and Robert R Bliss (1987) “The Information in long-maturity forward rates”, *American Economic Review*, 77, 680-692

Fama, Eugene F and Kenneth R French, (1989)“Business conditions and expected returns on stocks and bonds” *Journal of Financial Economics*, 25, 23-49

Fruhwirth-Schnatter, S, (2001), “Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models”, *Journal of the American Statistical Association,* 96, 194-209.

Gelfand, A.E. and Smith, A.F.,. (1990) “Sampling Based Approaches to Calculating Marginal Densities”, *Journal of the American Statistical Association*, 85, 398-409.

Gray, Stephen F, (1996a), “Modeling the conditional distribution of interest rates as a regime-switching process”, *Journal of Financial Economics*, 42, 27–62.

Gray, Stephen F, (1996b), “Regime Switching in Australian Short term interest rates”, *Accounting and Finance*, Vol 36 Iss 1, 65-88.

Gray, Stephen F. and Sirimon Treepongkaruna, (2002), “On the Robustness of Short-term Interest Rate Models,” *Accounting and Finance*, 43, 87-121.

Hamilton, J.D, (1988) “Rational Expectations Econometric Analysis of changes in Regime: An Investigation of the Term Structure of Interest Rates”, *Journal Economic Dynamics and Control*, 12, 385-423.

Hamilton, J. D, (1989) “Analysis of Time Series Subject to Changes in Regime”, *Journal of Econometrics*, 45, 39-70.

Hamilton, J. D, (1994) *Time Series Analysis*, Princeton, New Jersey: Princeton University Press.

Han, C., and Carlin, B. P, (2001) “Markov Chain Monte Carlo Methods for Computing Bayes Factors: A Comparative Review”, *Journal of the American Statistical Association*, 96, *455*, 1122-33.

Hansen, Lars P and Kenneth J Singleton, (1982) “Generalized instrumental variables estimation of nonlinear rational expectations models”, *Econometrica* 50, 1269-1286 (with corrections in Econometrica, 52 267-268)

Heaney, Richard, (1994) “Predictive power of the term structure in Australia in the late 1980’s: A note” *Accounting and Finance*, 34, 37-47

Hicks, John R, (1946) *Value and Capital* ,Oxford University Press, London

Kass, R. E., and Raftery, A. E, (1995) “Bayes Factors”, *Journal of the American Statistical Association*, 90, *430*, 773-793.

Kessel, Reuben A., (1965) “The cyclical behaviour of the term structure of interest rates”, *NBER occasional working paper* no 91.

Klemkosky, Rober C and Eugene A Pilotte, (1992), “Time-varying term premiums on US Treasury bills and bonds, *Journal of Monetary Economics*, 30, 87 -106

McCulloch, J Huston, (1987), “The monotonicity of the Term Premium: A closer look” *Journal of Financial Economics*, 18, 185–192.

Newey, Whitney K and Kenneth D West, (1987), “A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix”, *Econometrica* 55, 703-708.

Richardson, Matthew, Paul Richardson and Tom Smith, (1992), “Time-varying term premiums on US Treasury bills and bonds”, *Journal of Monetary Economics*, 30, 87-106.

Richardson, S., and Green, P. J, (1997), “ Bayesian Analysis of Mixtures With An Unknown Number of Components” *Journal of the Royal Statistical Society*, Ser. B, 59, 731-758.

Stambaugh, Robert F, (1988) The information in forward rate: Implications for models of the term structure”, *Journal of Financial Economics*, 21, 41-70

Stephens, M., (2000) “Dealing with label switching in mixture models”, *Journal of Royal Statistical Society: Ser B*, 62, *4*, 795-809.

Tanner, M A, (1996) *Tools for Statistical Inference*, 3rd ed., Springer, New York

Walsh, Kathleen D (2004), “Is the Ex Ante Premium always Positive? Evidence and Analysis from Australia, *UNSW Working Paper* 2004 – 14.