IMF Working Paper

Exploration of the Brazilian Term Structure in a Hidden Markov Framework

Richard Munclinger

INTERNATIONAL MONETARY FUND

IMF Working Paper

Monetary and Capital Markets Department

Exploration of the Brazilian Term Structure in a Hidden Markov Framework

Prepared by Richard Munclinger[*]

Authorized for distribution by Carlos Medeiros

January 2011

Abstract

We apply a hidden Markov model of the term structure to modeling the Brazilian swap rate curve. We examine the model's characteristics and its performance in describing the cross-sectional and time-series dynamics of the term structure. Two regimes are identified, a high level and a high volatility regime and a low level and low volatility regime. Both regimes are persistent and are explained by the level and the slope of the term structure. The model is estimated using a Bayesian MCM algorithm that produces consistent standard errors and a reliable method for testing the differences between the model parameters.

This Working Paper should not be reported as representing the views of the IMF.
The views expressed in this Working Paper are those of the author(s) and do not necessarily represent those of the IMF or IMF policy. Working Papers describe research in progress by the author(s) and are published to elicit comments and to further debate.

Contents

Boxes

I. INTRODUCTION

Most term structure models assume that interest rate series are stationary. However, the fact that the dynamics of interest rates and macroeconomic variables vary over time has been documented in a number of regime switching studies based on the work of Hamilton (1989,1990). This behavior implies the existence of non-stationarity that is often attributed to changes in the business cycle. Another cause of nonstationarity are structural changes that may occur over short and long horizons. An example of a sudden structural change is the US Federal Reserve "monetary experiment" that took place from 1979 to 1982. The stabilization of the economic environment in Brazil since 2004, and the election of president Luiz Inácio Lula da Silva, is an example of a gradual shift which has effected the level and volatility of interest rates permanently.

In this paper, we exploit the hidden Markov model of the term structure proposed by Day, Singleton, and Yang (2007) to examine the dynamics of the Brazilian term structure. We outline the benefits of this approach and draw lessons for future applications. Emerging markets' term structures are usually more volatile and include extreme yield curve outcomes which makes the analysis interesting. In addition, data on term structure in emerging markets are typically limited in the cross-sectional and the time-series dimensions. These limitations may have important implications on the applicability of hidden Markov models. We also propose a Bayesian Markov Chain Monte Carlo (MCMC) algorithm to estimate affine hidden Markov models of the term structure. A step by step guideline to applying this algorithm is provided. MCMC allows for the specification of the posterior densities of the parameters, in contrast to DSY who used a maximum likelihood approach and report asymptotic standard errors. The difference between asymptotic errors and actual errors may be significant especially if only a short time-series is available.

The core finding is that the Brazilian term structure exhibits regime switching behavior. We identify two regimes in the data; a high level, slope and volatility regime (Regime 1), and a low level, slope, and volatility regime (Regime 2). Regime 1 encapsulates the exchange rate crisis from 1998 to 1999 and the presidential election in October 2002. Regime 2 is characterized by the most recent period of stable monetary and fiscal policy. We observe that regimes are highly persistent and that regime changes can not be explained by inflation, GDP growth or other macroeconomic variables. All transition probability information is reflected in the slope and the curvature of the yield curve. The market price of level risk is relatively high in Regime 2; a result that is attributable to the relatively low volatility of the term structure factors.

The hidden Markov model overperforms a single regime ATSM benchmark for all estimated measures of fit, including those that penalize the model for complexity. The improvement originates from both the time-series and cross sectional dimensions, although time-series fit remains relatively poor. In-sample fit does not, however, imply superior forecasting performance. One-step-ahead forecasting tests were not performed due to the complexity involved in re-estimating the models.

Controlling for non-stationarity in time series behavior is important for a number of reasons.

Firstly, models that do not control for changes in levels and volatility are mis-specified. For example, Bollerslev, Chou and Kroner (1992) found a unit root in interest rate volatility which implies explosive variances and invalidates many equilibrium term structure models. An introduction of a regime process mitigates this problem. Secondly, both in-sample and out of sample fit can be improved by controlling for changes in regime. Gray (1996) documents this for interest rates, Engel and Kim (1996) for exchange rates and So, Lam and Li (1998) for stock index volatility. Thirdly, pricing and risk management of fixed income portfolios is dependent on the correct specification of interest rate densities. Numerous deviations from the usual assumptions can be captured by introducing a hidden Markov process.[1] Recognition of these benefits has prompted the development of numerous term structure models which incorporate hidden Markov behavior.

Landen (2000) assumed that the mean and variance of the short interest rate process depends on a Marked Point Process (MPP) modulated by a Markov chain.[2] In the most general case, this implies that the long term mean, speed of mean reversion and volatility are regime dependent. Landen derives a closed form solution for bond prices under the Equivalent Martingale Measure (EMM), given a semi-affine term structure for the short rate.

Bansal and Zhou (2002) present an affine model in which both the market price of risk and the short rate parameters undergo regime switching. The regime transitions are governed by a two-state Markov chain, both regimes and factors are assumed to be observable, transition probabilities are modeled as a constant and the market price of regime switching risk is equal to zero. One and two factor versions of the model are benchmarked against one-factor, two-factor and three-factor Cox, Ingersoll and Ross (CIR) models and a three-factor affine model. Specification tests suggest that the two-factor regime switching model is superior to all tested alternatives. Bansal and Zhou (2002) conclude that regimes are dependent primarily on the volatility, but not on mean reversion. Surprisingly, the parameter that governs risk premium is higher in the low volatility regime. A similar study by Bansal, Tauchen and Zhou (2004) corroborates these results and illustrates that the regime switching model can account for the predictability in the yield curve and the transition dynamics of yields.

Wu and Zeng (2003) present a continuous time treatment of risk-free regime switching term structure based closely on Landen (2000). They derive the pricing kernel and a pseudo closed form (up to a system of differential equations) solution for bond prices. Dai, Singleton, and Yang (2007) (DSY) develop a three-factor, two-regime arbitrage free, affine regime switching model in discrete time that is a special case of the Wu and Zeng (2003) model. They specify functional forms for the market prices of risk, restrict factor dynamics and derive a closed-form expression for a zero-coupon bond price. They allow for state-dependence in probabilities and regime shifts under the real measure, but not under the risk neutral measure. They examine the market prices of risk and conclude that there is a significant difference between states and that regime switching risk appears to be priced. Likelihood tests reject the

---

[1]Regime switching and hidden Markov processes are identical, with the former terminology being popular in economics, and the latter in finance literature.

[2]This model is a generalization of both Hansen and Poulsen (2000) and Lee and Naik (1998).

constant regime transition probabilities model in favor of the state-dependent model.

These models characterize the term structure relatively well, but have several important deficiencies. Firstly, most hidden Markov term structure models have an affine structure, and they therefore share some of it's shortcomings. In particular, estimation of Affine Term Structure Models (ATSMs) is difficult due to the non-linear arbitrage-free restrictions which are imposed on the dynamics of the yield curve. The introduction of a hidden Markov process exacerbates this problem. Additionally, ATSMs do not guarantee the positivity of the yields in either discrete or continuous time. Secondly, the number of regimes is not known a priori and model comparison is difficult due to the fact that models with different numbers of regime are not nested. Thirdly, regimes may or may not be predictable; a fact that can have an adverse impact on forecasting performance.

This work continues with a short review of single regime affine term structure models. A discussion of the regime switching model of Dai, Singleton and Yang (2007) follows in Section III. Section IV summarizes the Bayesian MCMC estimation methodology, including a short review of the Gibbs and Metropolis-Hastings algorithms. Section V outlines the term structure and macroeconomic data used in this study and the discussion of the main results is presented in Section VI. Section VII concludes.

II.   REVIEW OF AFFINE TERM STRUCTURE MODELS

Affine term structure models are popular in academic literature because they allow closed form solutions and are flexible in describing yields. Dai and Singleton (2000) review this class of models and characterize them into several subsets. Only one subset for ATSMs, denoted by $\mathbb{A}_0(3)$ by Dai and Singleton (2000), are relevant for the DSY hidden Markov model of term structure. More complex versions of ATSMs do not have closed form solution in this setting.

ATSMs assume that the yields are a linear function of latent term structure factors. The ATSMs of the $\mathbb{A}_0(3)$ class are homoscedastic and follow a three dimensional Gaussian diffusion[3]. They are characterized in discrete time,[4] under the real world probability measure, by

$$r_t = a + \mathbf{b}X_t \tag{1}$$

$$X_t = X_{t-1} + \mathbf{K}(\boldsymbol{\theta} - X_{t-1})\Delta t + \boldsymbol{\Sigma}\boldsymbol{\epsilon}_t \tag{2}$$

$$\Psi_t = \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\lambda}_0 + \boldsymbol{\Lambda}_1 X_t\right) \tag{3}$$

---

[3]Theoretically, any number of factors can be used to describe the dynamics of the term structure. Three factors are most often used because they explain over 90% of yield variance and have an appealing interprettation as the level, slope and curvature of the yield curve.

[4]Some readers will be more familiar with equivalent expressions in continuous time. Discrete time treatment was chosen here because of its intuitive appeal to a broader audience. Furthermore, all practical applications are performed in discrete time and would require a discretization of continuous-time results.

where r_t is the short rate of interest, X_t is a three-dimensional latent term structure factor vector and ϵ_t is a Gaussian error term. $\mathbf{K}$ is the mean reversion parameter, $\boldsymbol{\theta}$ is the long term mean of X_t and $\boldsymbol{\Sigma}$ is a Cholesky decomposition of the covariance matrix. The functional form of the market price of risk, Ψ_t, is assumed to be given by expression (3). The market price of risk is necessary as a link between the real world probability measure and the risk neutral probability measure under which pricing takes place. $\boldsymbol{\Lambda}_1$, $\boldsymbol{\lambda}_0$, a and $\mathbf{b}$ are parameters. Δt is the time period over which the model was discretized and must be consistent with the cross-sectional frequency at which the data are observed. Without loss of generality, we can assume that Δt is one.[5]

If the price of a zero coupon bond at time t maturing in τ periods is given by

$$P(t, \tau) = e^{A(\tau) + B(\tau)X_t} \tag{4}$$

then the coefficients $A(\tau)$ and $B(\tau)$ are given by the following expressions.

$$A(\tau + 1) = A(\tau) + B(\tau)(\mathbf{K}\boldsymbol{\theta} - \boldsymbol{\lambda}_0) + \frac{1}{2}B(\tau)'\boldsymbol{\Sigma}\boldsymbol{\Sigma}'B(\tau) - a \tag{5}$$

$$B(\tau + 1) = B(\tau) - B(\tau)(\mathbf{K} - \boldsymbol{\Lambda}_1) - \mathbf{b} \tag{6}$$

This result is a solution to a system of difference equations. The process to obtain $A(\tau)$ and $B(\tau)$, given the form of the bond price in expression (4) is summarized in Ang and Piazzesi (2003).

In the $\mathbb{A}_0(3)$ subclass of ATSM models conditional second moments are constant and therefore term structure factors can become negative. This characteristic is counter-factual and forces the distribution of the interest rates to be Gaussian and homoscedastic. Additionally, the positivity of the interest rate is not guaranteed, although a lower bound of zero can be placed on the interest rates when simulating from Gaussian ATSMs. The inclusion of a hidden Markov process induces both heteroscedasticity and excess kurtosis; characteristics which are necessary to describe the interest rate process correctly. The sign of the unconditional correlation between state variables is unconstrained in ATSMs. The flexibility afforded by this characteristic is desirable and is not shared by non-Gaussian ATSMs although positivity and negative correlations between interest rate factors is not simultaneously possible in ATSMs.

A. Normalization

The Gaussian Homoscedastic ATSM given by expressions (1) to (6) is globally unidentified because it is not invariant with respect to a linear transformation. In other words, for any invertible matrix $\mathbf{H}$, there is a function $g(\mathbf{H}, \tau)$ such that the equations in (2) and (4) can be written as

$$P(t, \tau) = e^{A(g(\mathbf{H},\tau)) + B(g(\mathbf{H},\tau))X_t} \tag{7}$$

---

[5]Normal font indicates a scalar, capitalized or bold Greek letters indicate a vector and all matrices are denoted by bold capitalized font.

$$\tilde{X}_t = \tilde{X}_{t-1} + \tilde{\mathbf{K}}(\tilde{\boldsymbol{\theta}} - \tilde{X}_{t-1})\Delta t + \tilde{\boldsymbol{\Sigma}}\boldsymbol{\epsilon}_t \tag{8}$$

where $\tilde{X}(t) = \mathbf{H}X(t)$, $\tilde{\mathbf{K}} = \mathbf{H}\mathbf{K}\mathbf{H}^{-1}$, $\tilde{\boldsymbol{\theta}} = \mathbf{H}\boldsymbol{\theta}$ and $\tilde{\boldsymbol{\Sigma}} = \mathbf{H}\boldsymbol{\Sigma}$. There are several restrictions that will make the model identifiable. We follow Dai and Singleton (2000) in choosing the following normalizing restrictions (assuming X_t is a (3×1) vector).

$$\mathbf{K} = \begin{pmatrix} k_{11} & 0 & 0 \\ k_{21} & k_{22} & 0 \\ k_{31} & k_{32} & k_{33} \end{pmatrix}$$

$$\boldsymbol{\theta} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Additionally, the diagonal elements of $\mathbf{K}$ are positive, ensuring that the factors X_t are stationary.

III. HIDDEN MARKOV MODEL

In this section, we examine the model of Dai, Singleton and Yang (2007) which differs from the single regime model in a number of important characteristics. DSY (i) include a regime switching process which modulates between different term structure dynamics; (ii) assume that regime switching risk is priced; and (iii) include a price of regime switching parameter in the pricing kernel. Due to the presence of the state process and the additional market price of risk, the restrictions on the dynamics of the term structure in the hidden Markov model differ from those in (5) and (6).

Dai, Singleton, and Yang (2007) assume that conditional on the regime, the term structure follows Gaussian ATSM structure. In particular, under the real world measure $\mathbb{P}$,

$$r_t = a(s_t) + \mathbf{b}X_t \tag{9}$$

$$X_t = X_{t-1} + \mathbf{K}(s_t)(\boldsymbol{\theta}(s_t) - X_{t-1})\Delta t + \boldsymbol{\Sigma}(s_t)\epsilon_t \tag{10}$$

and the market price of interest rate risk is given by

$$\Psi_t(s_t) = \boldsymbol{\Sigma}(s_t)^{-1}(\boldsymbol{\lambda}_0(s_t) + \boldsymbol{\Lambda}_1(s_t)X_t) \tag{11}$$

The specification in (9) and (10) are identical to those in (1) and (2) except for the dependency on the state s_t. The process s_t (with state space q) is assumed to be governed by a Markov chain with a transition probability matrix $\boldsymbol{\Pi}^{\mathbb{P}}(t)$. Under the real world measure $\mathbb{P}$, $\boldsymbol{\Pi}^{\mathbb{P}}(t)$ can

be time dependent. In order to obtain a close form solution, however, the transition probability matrix $\mathbf{\Pi}^{\mathbb{Q}}$ must be homogeneous under the measure $\mathbb{Q}$.[6] This is achieved by the following parametrization of the market price of regime switching risk.

$$\Gamma^{i,j}(t) = log\left(\frac{\pi_{i,j}^{\mathbb{P}}(t)}{\pi_{i,j}^{\mathbb{Q}}}\right), \forall i, j \tag{12}$$

$\pi_{i,j}^{\mathbb{P}}(t)$ and $\pi_{i,j}^{\mathbb{Q}}$ are the (i, j) elements of $\mathbf{\Pi}^{\mathbb{P}}(t)$ and $\mathbf{\Pi}^{\mathbb{Q}}$ respectively and represent the transition from state i to state j. Under these conditions, and assuming that the price of a zero coupon bond is given by

$$P(s_t, t, \tau) = e^{A(s_t,\tau)+B(\tau)X_t} \tag{13}$$

Dai, Singleton, and Yang (2007) show that

$$A(s_t, \tau + 1) = B(\tau)(\mathbf{K}(s_t)\boldsymbol{\theta}(s_t) - \boldsymbol{\lambda}_0(s_t)) + log\left(\sum_{j=0}^{S} \pi_{i,j}^{\mathbb{Q}} e^{A(j,\tau)}\right) +$$
$$\frac{1}{2}B(\tau)\mathbf{\Sigma}(s_t)\mathbf{\Sigma}(s_t)'B(\tau)' - a(s_t) \tag{14}$$

$$B(\tau + 1) = -b + B(\tau)(1 - \mathbf{K}(s_t) - \mathbf{\Lambda}_1(s_t)) \tag{15}$$

with initial conditions $A(s_t, 0) = 0$ and $B(0) = 0$. To ensure that $B(\tau)$ is independent of s_t, we assume that $(1 - \mathbf{K}(s_t) - \boldsymbol{\lambda}_1(s_t))$ is regime invariant.

A. Normalization

Normalizations, as was the case in the single regime model, are necessary for the identification of the model. In a two regime case, DSY assume that in Regime 1, $\mathbf{\Sigma}$ is the identity matrix, $\mathbf{K}$ is a lower triangular matrix and $\boldsymbol{\theta}$ is a vector of zeros. Independence between factors is also necessary in Regime 2 which is achieved by assuming that $\mathbf{\Sigma}$ is a diagonal matrix.

IV. ESTIMATION

Dai, Singleton, and Yang (2007) outline a Maximum Likelihood (ML) method of obtaining the parameters of the model. In this paper, we take a Bayesian approach to estimation. The reason for this is that using Bayesian MCMC allows us to generate entire parameter distributions and therefore obtain consistent standard error estimates. Testing of equality between parameters is

---

[6]This is a technical condition that is required to ensure that the transition probability can be taken out of the expectation under the risk neutral measure $\mathbb{Q}$. It is a restriction that is couterfactual, but which will have small effect on model fit. The reason for this is that the model is estimated under the measure $\mathbb{P}$, which allows for heterogeneous transitions. Additionally, this restriction only effects cross-sectional fit which is excellent in these models. The dynamics of X_t under measure $\mathbb{P}$ do allow for state-dependent intercept.

also possible directly, which enables us to make inferences about the differences between coefficients in alternate regimes. Small sample issues are acute in our case due to the lack of reliable Brazilian term structure data before 1998. This lack of data, coupled with the relatively large number of parameters to be estimated, makes the use of MCMC necessary to obtain consistent standard errors. We begin this section by developing an econometric model that will be used to estimate the term structure models in Sections II and III. We briefly summarize the MCMC algorithm used for estimation and provide a step by step guide to applying it.

A. Data Likelihood

The solution given by expressions (14) and (15) identifies the restrictions on the shape of the term structure necessary to ensure arbitrage free pricing. In this section, we link the model to observable yields and specify a likelihood function. We focus the discussion on the model in Section III.

A yield at time t with a maturity τ denoted by $r(t, \tau)$ is given by

$$r(t, \tau) = -\frac{ln\left(P(s_t, t, \tau)\right)}{\tau} \tag{16}$$

where $P(s_t, t, \tau)$ is given by expression (13). This implies that

$$r(t, \tau) = -\frac{A(s_t, \tau) + B(\tau)X_t}{\tau} \tag{17}$$

for each yield with maturity τ. We combine all yields into a $(m \times 1)$ vector R_t (where m is the total number of maturities available at each time t) and all coefficients $\frac{A(s_t, \tau)}{\tau}$ and $\frac{B(\tau)}{\tau}$ into a vector $A(s_t)$ and matrix B.

$$R_t = A(s_t) + \mathbf{B}X_t \tag{18}$$

Assuming a three-dimensional latent factor vector X_t, as is standard in ATSMs, the dimensions of $\frac{A(s_t, \tau)}{\tau}$ and $\frac{B(\tau)}{\tau}$ are $(m \times 1)$ and $(m \times 3)$ respectively.

This specification for yields is rather naive since it assumes that all yields are modeled exactly. Error can be introduced into (18) in one of two ways. Firstly, one can assume that all yields are matched with error. In this case (18) is given by

$$R_t = A(s_t) + \mathbf{B}X_t + \epsilon_t \tag{19}$$

and the model could be estimated using a Kalman Filter. Alternatively, we can assume that only a subset, $m - n$, yields are matched with error and that n yields are modeled without error, where n is the number of latent factors X_t.[7]

The disadvantage of fitting some yields exactly is that it imposes a restriction on the factors X_t which could lead to inferior fit of the term structure as a whole. The advantage of this

---

[7]The choice of yields that are fitted exactly and those that are modeled with error is arbitrary in theory. In practice, it is advantageous to select yields that span the entire maturity spectrum.

approach is that the term structure is arbitrage free at n maturities and that the yields can be inverted to obtain the latent factors X_t. This significantly simplifies estimation which is why this approach is preferred in practice and is employed here.

We separate R_t, $A(s_t)$ and $\mathbf{B}$ into vectors and matrices corresponding to yields matched without error ($\hat{R}_t$, $\hat{A}(s_t)$ and $\hat{B}$) and those corresponding to yields matched with error ($\tilde{R}_t$, $\tilde{A}(s_t)$ and $\tilde{B}$). The factors X_t are obtained by inverting the yields.

$$X_t = \hat{\mathbf{B}}^{-1} \left(\hat{R}_t - \hat{A}(s_t) \right) \tag{20}$$

The yields matched with error are given by

$$\tilde{R}_t = \tilde{A}(s_t) - \tilde{\mathbf{B}}\hat{\mathbf{B}}^{-1}\hat{A}(s_t) + \tilde{\mathbf{B}}\hat{\mathbf{B}}^{-1}\hat{R}_t + \mathbf{\Omega}(s_t)e_t \tag{21}$$

where $\mathbf{\Omega}(s_t)$ is the Cholesky decomposition of the error covariance matrix and e_t is a standard normal error vector. A more complicated expression can be obtained for $\hat{R}_t$ by substituting for the dynamics of X_t given in (10) and inverting the lagged yields.

$$\hat{R}_t = \boldsymbol{\mu}_{\hat{R}} + \hat{\mathbf{B}}\mathbf{\Sigma}(s_t)\epsilon_t \tag{22}$$

$$\boldsymbol{\mu}_t = \hat{A}(s_t) + \hat{\mathbf{B}} \left(\mathbf{K}(s_t)\boldsymbol{\theta}(s_t) - (1 - \mathbf{K}(s_t))\hat{\mathbf{B}}^{-1}\hat{A}(s_{t-1}) \right) + \hat{\mathbf{B}}(1 - \mathbf{K}(s_t))\hat{\mathbf{B}}^{-1}\hat{R}_{t-1}$$

ϵ_t is also standard normal and is assumed to be uncorrelated with e_t. The likelihood, conditional on the regime vector $S = \{s_1, \ldots, s_T\}$ and the parameters of the model, is given by

$$\mathbb{L}(\mathbf{R}|S, \ldots) = \prod_{t=1}^{T} N \left(\tilde{A}(s_t) - \tilde{\mathbf{B}}\hat{\mathbf{B}}^{-1}\hat{A}(s_t) + \tilde{\mathbf{B}}\hat{\mathbf{B}}^{-1}\hat{R}_t, \mathbf{\Omega}(s_t)\mathbf{\Omega}(s_t)' \right) N \left(\boldsymbol{\mu}_t, \hat{\mathbf{B}}\mathbf{\Sigma}(s_t)\mathbf{\Sigma}(s_t)'\hat{\mathbf{B}}' \right) \tag{23}$$

where $\mathbf{R}$ is a matrix of stacked yields.

B. The State Process

The likelihood function derived above is conditional on the state vector S. In order to obtain the likelihood for the entire model a specification for S must be given. We assume that, under the real world probability measure $\mathbb{P}$, each component of S, s_t, is a categorically distributed random variable, with a transition probability matrix $\mathbf{\Pi}^{\mathbb{P}}(t)$, and with $\mathbf{\Pi}_i^{\mathbb{P}}(t)$ being the ith row of $\mathbf{\Pi}^{\mathbb{P}}(t)$.

$$p \left(s_{t+1}|S^-, \ldots \right) \sim Cat \left(\mathbf{\Pi}_{s_t}^{\mathbb{P}}(t) \right) \tag{24}$$

S^- denotes, in the most general case,[8] the entire vector S not including s_{t+1}. Under $\mathbb{P}$, the transition matrix is heterogeneous. We assume that each element of $\mathbf{\Pi}^{\mathbb{P}}(t)$, $\pi_{i,j}^{\mathbb{P}}(t)$ depends on

---

[8]In this case, we assume that the process is Markovian and therefore only the previous state, s_t, is relevant for the determination of s_{t+1}.

a linear index through a Logit transformation.[9]

$$Logit\left(\pi^{\mathbb{P}}_{s_t,s_{t+1}}(t)\right) = \boldsymbol{\gamma}^{s_t,s_{t+1}} M_t \tag{25}$$

M_t is a $(k \times 1)$ vector of some known macroeconomic or term structure variables. Only $q - 1$ probabilities need to be estimated because each row of $\boldsymbol{\Pi}^{\mathbb{P}}(t)$ must sum to unity. For notational parsimony we stack the column vectors $\boldsymbol{\gamma}^{s_t,s_{t-1}}$ into a $(kq(q-1) \times 1)$ vector Γ.

Under the measure $\mathbb{Q}$, s_t depends on a homogeneous probability matrix $\boldsymbol{\Pi}^{\mathbb{Q}}$ (with $\boldsymbol{\Pi}^{\mathbb{Q}}_i$ being the ith row of $\boldsymbol{\Pi}^{\mathbb{Q}}$).

$$p^{\mathbb{Q}}\left(s_{t+1}|S^-,\dots\right) \sim Cat\left(\boldsymbol{\Pi}^{\mathbb{Q}}_{s_t}\right) \tag{26}$$

We assume that $\boldsymbol{\Pi}^{\mathbb{Q}}_i$ is distributed according to the Dirichlet distribution,

$$p^{\mathbb{Q}}\left(\boldsymbol{\Pi}^{\mathbb{Q}}_{s_t}|S,\dots\right) \sim Dirichlet\left(\alpha_{s_t,1},\dots,\alpha_{s_t,q}\right) \tag{27}$$

with

$$\alpha_{s_t,s_{t+1}} = \frac{1}{T}\sum_{t=1}^{T} 1(s_t = i, s_{t+1} = j)) \tag{28}$$

where $1(\cdot)$ denotes an indicator function.

C. Bayesian Estimation

In the previous two sections, we have derived the likelihood for the vector of yields and have assumed a hierarchical probability model for the state process. This section outlines the Bayesian MCMC algorithm that can be utilized to estimate the parameters of the model. The goal is to obtain the full posterior distribution of the model parameters, which is given by the Bayes rule[10] as

$$p\left(\boldsymbol{\Theta}, S, \boldsymbol{\Omega}, \Gamma|\mathbf{R}\right) \propto p\left(\mathbf{R}|\boldsymbol{\Theta}, S, \boldsymbol{\Omega}, \Gamma\right) p\left(\boldsymbol{\Theta}, S, \boldsymbol{\Omega}, \Gamma\right) \tag{29}$$

---

[9]The Logit transformation is achieved via the Logit function which is given by

$$\pi^{\mathbb{P}}_{i,j}(t) = \frac{e^{-\boldsymbol{\gamma}^{i,j} M_t}}{\sum_{j=1}^{S} e^{-\boldsymbol{\gamma}^{i,j} M_t}}$$

where a single $\boldsymbol{\gamma}^{i,j}$ is normalized to zero since $\sum_{j=1}^{S} \pi^{\mathbb{P}}_{i,j}(t) = 1$.

[10]Most readers will be familiar with another version of the Bayes rule which is as follows.

$$p\left(\boldsymbol{\Theta}, S, \boldsymbol{\Omega}, \Gamma|\mathbf{R}\right) = \frac{p\left(\mathbf{R}, \boldsymbol{\Theta}, S, \boldsymbol{\Omega}, \Gamma\right)}{p\left(\mathbf{R}\right)}$$

Conditioning on the unknown parameters in the right hand side yields

$$p\left(\boldsymbol{\Theta}, S, \boldsymbol{\Omega}, \Gamma|\mathbf{R}\right) = \frac{p\left(\mathbf{R}|\boldsymbol{\Theta}, S, \boldsymbol{\Omega}, \Gamma\right) p\left(\boldsymbol{\Theta}, S, \boldsymbol{\Omega}, \Gamma\right)}{p\left(\mathbf{R}\right)}$$

Expression (29) is obtained by dropping the denominator, which is a constant with respect to the unknown parameters.

where $\boldsymbol{\Theta} = \{a(s_t), \mathbf{b}, \mathbf{K}(s_t), \theta(s_t), \boldsymbol{\Sigma}(s_t), \boldsymbol{\lambda}_0(s_t), \boldsymbol{\Lambda}_1(s_t)\}$ represents all term structure parameters. $p(\mathbf{R}|\boldsymbol{\Theta}, S, \boldsymbol{\Omega}, \boldsymbol{\Gamma})$ is the likelihood function in expression (23) and $p(\boldsymbol{\Theta}, S, \boldsymbol{\Omega}, \boldsymbol{\Gamma})$ is called the prior distribution. Applying the Bayes rule consecutively on the prior distribution in expression (29) results in the following.

$$p(\boldsymbol{\Theta}, S, \boldsymbol{\Omega}, \boldsymbol{\Gamma}) \propto p(\boldsymbol{\Theta}|S, \boldsymbol{\Omega}, \boldsymbol{\Gamma}) \, p(\boldsymbol{\Omega}|S, \boldsymbol{\Gamma}) \, p(S|\boldsymbol{\Gamma}) \, p(\boldsymbol{\Gamma}) \tag{30}$$

Box 1. Gibbs Sampler

For two random variables W and Y, which have a joint density $p(w, y)$, a Gibbs sampler can generate a Markov chain $(W(t), Y(t))$ that will converge to the draw from $p(w, y)$. This is useful when the form of $p(w, y)$ can not be recognized but the kernels of the conditional distributions $p(w|y)$ and $p(y|w)$ are known. The Markov chain $(W(t), Y(t))$ can be generated according to the following algorithm.

1. With a starting value, $W(0) = w(0)$, generate from the full conditional

$$Y(t) \sim p(y(t)|w(t-1))$$

2. Take previously generated value of $Y(t) = y(t)$ and generate $W(t)$ from the following

$$W(t) \sim p(w(t)|y(t))$$

$p(y(t)|w(t))$ and $p(w(t)|y(t))$ are conditional distributions which are ideally of a known form and can be used to generate random values easily.

This algorithm can be generalized to more than two variables. After an initial burn-in phase, the generated random numbers can be recorded and statistics based on the random draw can be calculated. This approach is very flexible and, as long as full conditionals for all variables can be obtained, is universally applicable. However, depending on the number of parameters and the shape of the likelihood, mixing can be slow and convergence may be difficult to achieve. In some cases (in term structure models for example) the full conditionals themselves are not of known form and the Metropolis-Hastings algorithm is used within the Gibbs sampler.

Prior specification

The form of the components of (30) is either given by the term structure model or is assumed by the modeler. We stack all coefficients in $\boldsymbol{\Theta}$ into a single vector and assume that its prior is

given by a diffuse multivariate normal distribution.[11]

$$p\left(\boldsymbol{\Theta}|S, \boldsymbol{\Omega}, \Gamma\right) = N(\mu_{\boldsymbol{\Theta}}, \boldsymbol{\sigma}_{\boldsymbol{\Theta}}) \tag{31}$$

The covariance matrix $\boldsymbol{\Omega}\boldsymbol{\Omega}'$ is assumed to have an independent Inverse-Wishart prior with an inverse scale matrix $\mathbf{G}$ and g degrees of freedom and

$$p\left(\boldsymbol{\Omega}\boldsymbol{\Omega}'|S, \boldsymbol{\Omega}, \Gamma\right) = InvWishart(G, g) \tag{32}$$

$p\left(\Gamma\right)$ is a multivariate normal density.

$$p\left(\Gamma\right) = N(\mu_{\Gamma}, \boldsymbol{\sigma}_{\Gamma}) \tag{33}$$

The form of $p\left(S|\Gamma\right)$ is given by a combination of expressions (24) and (25).

Full posterior densities

Given the set of prior distributions, the posterior density (29) is known up to a constant of integration. It is theoretically possible to obtain samples from this density directly via the Metropolis-Hastings algorithm. This approach is inefficient due to the size of the unknown parameter vector. It is more productive to obtain full conditional distributions for each parameter vector and then sample from each density separately using the Gibbs algorithm. In this case we obtain the full conditional for $\boldsymbol{\Theta}$, S, $\boldsymbol{\Omega}$ and Γ.[12] The Gibbs sampler and the Metropolis-Hastings algorithm are outlined briefly in Boxes 1 and 2.

The full conditional for $\boldsymbol{\Theta}$ is given by

$$p(\boldsymbol{\Theta}|\mathbf{R}, S, \boldsymbol{\Omega}, \Gamma) \propto \left(\prod_{t=1}^{T} p\left(\hat{R}_t|\boldsymbol{\Theta}, S, \boldsymbol{\Omega}, \Gamma\right) p\left(\tilde{R}_t|\boldsymbol{\Theta}, S, \boldsymbol{\Omega}, \Gamma\right)\right) p\left(\boldsymbol{\Theta}\right) \tag{34}$$

where

$$p\left(\hat{R}_t|\boldsymbol{\Theta}, S, \boldsymbol{\Omega}, \Gamma\right) = N\left(\boldsymbol{\mu}_{\hat{R}}, \hat{B}\boldsymbol{\Sigma}(s_t)\boldsymbol{\Sigma}(s_t)'\hat{B}'\right) \tag{35}$$

and

$$p\left(\tilde{R}_t|\boldsymbol{\Theta}, S, \boldsymbol{\Omega}, \Gamma\right) = N\left(\tilde{A}(s_t) - \tilde{B}\hat{B}^{-1}\hat{A}(s_t) + \tilde{B}\hat{B}^{-1}\hat{R}_t, \boldsymbol{\Omega}(s_t)\boldsymbol{\Omega}(s_t)'\right) \tag{36}$$

The prior $p\left(\boldsymbol{\Theta}\right)$ is given by (31). Although the likelihood and priors are (log)normal, the parameters of the term structure enter the means and variances in a non-linear manner. As a

---

[11]The normal priors for the diagonal matrix $\boldsymbol{\Theta}$ are appropriate because the covariance matrix of the factors X_t is given by $\boldsymbol{\Theta}\boldsymbol{\Theta}'$. The diagonal of this covariance matrix will be strictly positive with each element being the square of the corresponding element of $\boldsymbol{\Theta}$ and all off-diagonal elements are zero.

[12]Full conditional distributions can be produced by conditioning the posterior density in (29) on all coefficients except for the one of interest. For instance, the full conditional of $\boldsymbol{\Theta}$ is obtained by conditioning on S, $\boldsymbol{\Omega}$ and Γ, and the interest rate matrix $\mathbf{R}$. All expressions that do not contain $\boldsymbol{\Theta}$ can be dropped and are subsumed in the constant of integration.

$$p\left(\boldsymbol{\Theta}|S, \boldsymbol{\Omega}, \Gamma, \mathbf{R}\right) \propto p\left(\mathbf{R}|\boldsymbol{\Theta}, S, \boldsymbol{\Omega}, \Gamma\right) p\left(\boldsymbol{\Theta}\right)$$

The first expression on the right hand side is the likelihood and $p\left(\boldsymbol{\Theta}\right)$ is the prior distribution given by (31).

result, the kernel of the posterior density can not be recognized and the Metropolis-Hastings algorithm (within the Gibbs algorithm) is used to obtain samples from (34).

Likewise, the kernel for $\mathbf{\Gamma}$ is unknown, due to the logistic structure on the heterogeneous transition probabilities. The Metropolis-Hastings algorithm is again employed in conjunction with the following full conditional distribution,

$$p\left(\mathbf{\Gamma}|S,\mathbf{\Gamma},\mathbf{M}\right) \propto \prod_{i=1}^{q}\prod_{j=1}^{q}\prod_{t=1}^{T}\left(\frac{e^{\gamma^{i,j}M_t}}{\sum_{i=1}^{S}e^{\gamma^{i,j}M_t}}\right)^{1(s_t=j,s_{t-1}=i)}p\left(\mathbf{\Gamma}\right) \tag{37}$$

where $p\left(\mathbf{\Gamma}\right)$ is given by (33).

Box 2. Metropolis-Hastings Algorithm

The purpose of this algorithm is to generate random numbers from an unknown distribution $f(\cdot)$ by making use of a candidate density, $g(y,w)$ which has a known kernel. The ratio

$$\frac{f(y)}{g(y,w)}$$

has to be known up to constant independent of w. A Markov chain (W(t)), which has a steady state density $f(\cdot)$, can be produced by the following translation

1. Given an initial value $W(0) = w(0)$, generate from the candidate density

$$Y(t) \sim g(y(t)|w(t-1))$$

2. Take

$$W(t+1) = \begin{cases} Y(t) \text{ with probability } p(w(t),Y(t)) \\ w(t) \text{ with probability } 1 - p(w(t),Y(t)) \end{cases}$$

where

$$p(w,y) = \min\left(\frac{f(y)}{f(w)}\frac{g(w|y)}{g(y|w)}\right)$$

$p(w,y)$ is referred to as the acceptance probability. The algorithm is independent of normalization constants and accepts values such that the ratio $f(y)/g(y,w)$ increases from its previous value.

A number of variations of this algorithm exist. For their specifications, please refer to Robert and Casella (2004). The choice of the proposal density is extremely important for the efficiency of the algorithm. Any density that envelops the candidate can in theory be chosen, but the convergence properties of the algorithm will depend heavily on this choice. A popular algorithm is the Random Walk (RW) algorithm, which uses a Gaussian proposal density.

Each element of the state vector S can be generated recursively from a categorical distribution with a parameter vector $\mathbf{h}$, with each element h_i given by

$$\frac{p\left(s_t = i | s_{t-1}, s_{t+1}, \Theta, \Omega, R_t, \Pi^{\mathbb{P}}(t), \Pi^{\mathbb{P}}(t+1)\right)}{\sum_{j=1}^{q} p\left(s_t = j | s_{t-1}, s_{t+1}, \Theta, \Omega, R_t, \Pi^{\mathbb{P}}(t), \Pi^{\mathbb{P}}(t+1)\right)} \tag{38}$$

where

$$p\left(s_t = i | s_{t-1}, s_{t+1}, \Theta, \Omega, R_t, \Pi^{\mathbb{P}}(t), \Pi^{\mathbb{P}}(t+1)\right) = \pi^{\mathbb{P}}_{s_{t-1}, s_t}(t) p\left(\tilde{\mathbf{r}}_t | \hat{\mathbf{r}}_t, \Theta, s_t, s_{t-1}\right)$$
$$p\left(\hat{\mathbf{r}}_t | \hat{\mathbf{r}}_{t-1}, \Theta, s_t\right) \pi^{\mathbb{P}}_{s_t, s_{t+1}}(t) \tag{39}$$

Finally, the kernel of the full conditional of the covariance matrix $\Omega(s_t)\Omega(s_t)'$ can be identified, providing that the prior distribution is Inverse Wishart (given by expression (32)). The resulting full conditional distribution is also Inverse Wishart, characterized by

$$(\Omega(i)\Omega(i)'|\mathbf{R}, \Theta, S) \sim InvWishart\left(\mathbf{G} + \sum_{t=1}^{T} 1(s_t = i)\mathbf{E}'\mathbf{E}, k + T_i\right) \tag{40}$$

where $\mathbf{E}$ is the vector of stacked error terms $e_t = 1(s_t = i)(R_t - \tilde{A}(s_t) - \tilde{\mathbf{B}}X_t)$ and $T_i = \sum_{t=1}^{T} 1(s_t = i)$. $\Omega(s_t)$ can then be obtained by using Cholesky decomposition on the covariance matrix $\Omega(s_t)\Omega(s_t)'$.

Summary of the MCMC algorithm

Given a set of initial values, the steps of the Metropolis-Hastings within Gibbs algorithm used to estimate this model are as follows:

1. Generate the covariance matrix $\Omega(s_t)\Omega(s_t)'$ for each state using (40).

2. Use a single step of the Metropolis Hastings algorithm to generate all parameters in Θ via (34), (35) and (36).

3. Produce Γ via expression (37) and obtain the heterogeneous transition probabilities $\Pi^{\mathbb{P}}(t)$ for all t.

4. Generate the vector S recursively from (38) and (39).

5. Obtain the homogeneous transition matrix $\Pi^{\mathbb{Q}}$ via expressions (26), (27) and (28).[13]

6. Return to step 1.

At each step, the randomly generated parameters are recorded. After convergence is achieved, the burn-in sample is discarded and a sufficient sample from the joint posterior density is obtained. The number of iterations necessary to obtain a reliable sample will depend on the autocorrelation properties of the algorithm.

---

[13]Given the state vector S, we can evaluate $\alpha_{i,j}$ for all i, j and generate the homogeneous probability matrix $\Pi^{\mathbb{Q}}$ which appears in expression (14). This is the only step of the estimation algorithm that takes place under the measure $\mathbb{Q}$. There is no effect on the state process because S is identical under both measures.

V. DATA

The Brazilian term structure is characterized by Di-Pre (Deposito Interbanco) swap rates. The floating rate of the swap is set daily as the average of the one day interbank deposit rate. The swap pays a single payment at maturity. The swap rates are available monthly for 6, 12, 18, 24, 30, and 36 month maturities from January 1998 to May 2007. There are several advantages in using swaps. Swaps are not subject to repo specials and are constant maturity by nature and not by construction.[14] Table 1 summarizes some key characteristics of the swap data.

Table 1. Brazilian Swap Rates Summary

	Central moments and median					Autocorrelations	
	Mean	Median	Stdev	Skew	Kurt	Lag 1	Lag 2
6 mth	20.9433	19.2600	6.2510	1.6984	4.5881	0.8706	0.6956
12 mth	21.6156	19.3800	6.7365	1.3453	2.8871	0.8809	0.7315
18 mth	22.2440	19.5650	7.2997	1.1699	1.9330	0.8917	0.7572
24 mth	22.6781	19.9200	7.6969	1.0636	1.3445	0.8974	0.7703
30 mth	23.0327	20.4800	7.9486	0.9644	0.8773	0.8980	0.7710
36 mth	23.3138	20.9050	8.1618	0.9040	0.5818	0.8991	0.7738

This table summarizes the term structure data for Brazil which was collected at monthly intervals from January 1998 to May 2007. The 6, 12, 18, 24, 30 and 36 month Brazilian rates are the Di-Pre (Deposito Interbanco) swap rates obtained from the Brazilian Treasury. These rates are based on a single payment at maturity.

Paradoxically, long term swap rates are more volatile than short rates, possibly due to liquidity distortions at the long end of the yield curve. Long term rates are more persistent and have lower kurtosis, facts which are consistent with studies of US yields. The average yield curve over the period is relatively flat.

Macroeconomic determinants of regime transitions

Macroeconomic variables are not usually included in term structure studies. The latent term structure factors, X_t, have been shown to describe a large proportion of yield variation. More recently, however, Ang and Piazzesi (2003) and Diebold, Rudebusch and Auroba (2004) have shown that the inclusion of inflation, real activity and capacity utilization does improve time series fit and potentially forecasting performance. Although it is possible to include macroeconomic variables in affine hidden Markov models of the term structure directly, it is impractical, given the number of extra parameters that this requires. Macroeconomic variables can, however, describe the intensity of regime transitions. Regime changes have been most commonly linked to the business cycle fluctuations, which can be partially predicted by certain leading indicators. These, together with the term structure factors X_t, can be included in M_t (in expression (25)). Macroeconomic variables would therefore affect the term structure, although only indirectly, through s_t.

A large number of leading indicators can potentially be exploited to explain regime transitions.

---

[14]In developing markets swaps are also more liquid than treasury securities. This is not necessarily true in developing markets where swaps rates may include liquidity distortions.

The number of variables included in the model, however, are restricted by data availability. Only 114 monthly observations are on hand to estimate Γ. M_t includes a vector of ones (allowing for a non-zero intercept) and three term structure variables bringing the dimensions of Γ to eight, for a two-regime model, even before the inclusion of macroeconomic variables. As a result, only two variables were selected to explain regime changes in addition to the latent term structure factors; GDP growth and inflation.[15] Table 2 presents the key statistics for GDP and inflation. Both are stationary as reflected by the Dickey-Fuller p-value.

Table 2. Macroeconomic Variables Summary

	Central moments and median					Autocorrelations		Stationarity
	Mean	Median	Stdev	Skew	Kurt	Lag 1	Lag 2	ADF test
%ΔGDP(Reais)	0.0096	0.0082	0.0380	0.1257	-0.6497	-0.0267	-0.1622	0.0100
Inflation	0.0055	0.0047	0.0049	1.8897	6.6624	0.6553	0.3837	0.0208

This table summarizes the macroeconomic data for Brazil which was collected at monthly intervals from January 1998 to May 2007. The last column presents the p-values for the augmented Dickey-Fuller test. Both inflation and GDP were obtained from the IFS database.

VI. RESULTS

We estimate the single regime ATSM presented in Section II (denoted **ATSM** model) and a two-regime hidden Markov model discussed in Section III (denoted **HMM** model). The **ATSM** model is included for comparison purposes and will not be discussed in detail since its characteristics have been studied extensively in previous literature.

A. Hidden Markov Process

The principal finding of this study is that the Brazilian term structure exhibits hidden Markov behavior. Two regimes were identified in the data; a high level, slope and volatility regime (Regime 1), and a low level, slope and volatility regime (Regime 2). Figure 1 depicts the Brazilian term structure and the estimated hidden Markov process (shown as the mean of the posterior density for each t). Figure 2 illustrates the differences in volatilities, depicted as first differences, between the regimes, and Figure 3 links the estimated hidden Markov process to the slope of the term structure.

Regime 1 captures the end of the Asian crisis, the consequences of Russian default, the Brazilian currency crisis in 1998 and 1999, as well as the period (from April 2001 to August 2003) encompassing the Brazilian election in October 2002. Regime 2 is identified from early

---

[15]Several other variables were available such as the level and the slope of the US term structure, the Reais-Dollar exchange rate, a Dow Jones commodities and Bovespa index returns, foreign currency reserves, current account and net Brazilian public debt. Inclusion of these variables (results not shown) did not materially affect the final conclusions.

2000 to mid-2001 and again from mid-2003 to the end of the sample. The most recent period is characterized by stable monetary and fiscal policy and decreasing yields for all maturities.

Figure 1. 6-month Swap Rates and the Regime Process from Jan. 1998 to May 2007

Source: Fund staff estimates.
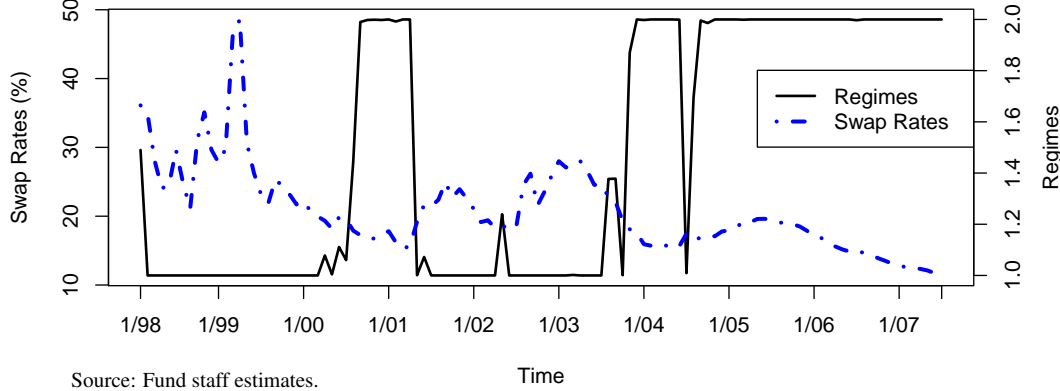
Both regimes are highly persistent, as evidenced by expression (41) that shows the mean and the 95 percent credible interval of the homogenous transition probability matrix $\Pi^{\mathbb{Q}}$. This result is consistent with other studies of regime switching models. The half-life, or the average duration of Regime 1 is 8.81 months and the half-life of Regime 2 is approximately 13.12 months. The practical implication of this is that once the system is in a particular regime, the probability of transitioning from this regime to another is low.

$$\Pi^{\mathbb{Q}} = \begin{bmatrix} 0.9244\% & 0.0756\% \\ (0.8800; 0.9444) & (0.0555; 0.1200) \\ 0.0515\% & 0.9485\% \\ (0.0312; 0.0857) & (0.9153; 0.9686) \end{bmatrix} \tag{41}$$

Figure 2. Maturity Spread and the Regime Process from Jan. 1998 to May 2007

Source: Fund staff estimates.
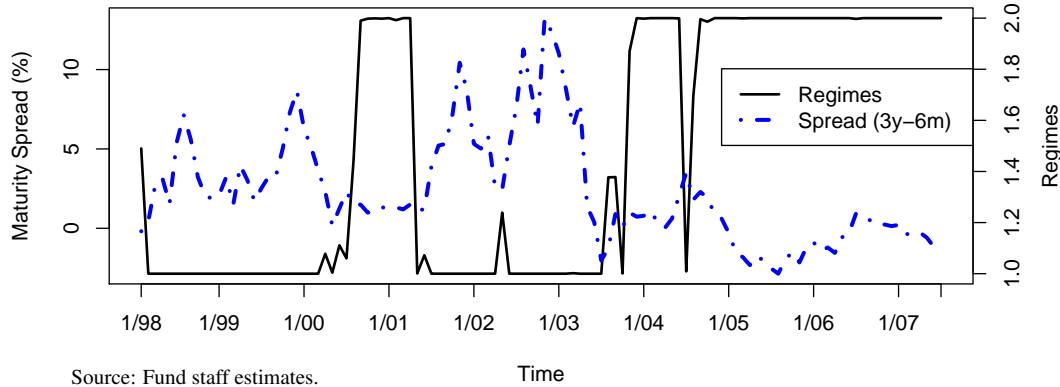
Regime changes are dependent on the level and the slope of the term structure. Table 3 shows the coefficients of the term structure factors, GDP growth and inflation. The coefficients of the macroeconomic variables are not significant in contrast to the coefficients of the level and the slope of the term structure. The level is proxied by the second term structure factor $-X[2]$, the slope by third factor $-X[3]$ and the curvature by $X[1]$.
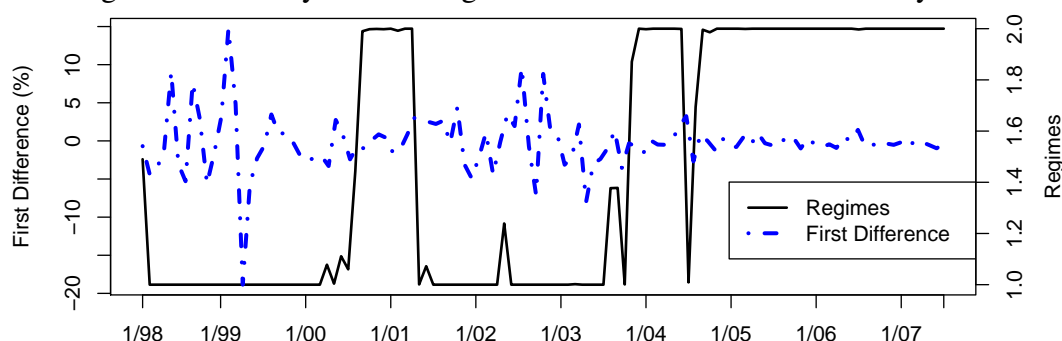
Table 3. Heterogeneous Probability Parameters

	$\gamma^{R1,R1}$	$\gamma^{R2,R2}$
Const.	-1.2419 (0.33)	**9.7479** (0.97)
	(-10.3537;9.3619)	(0.8887; 21.0374)
Curvature	-2.146173 (0.35)	2.094288 (0.74)
	(-9.5669;3.3073)	(-2.3839; 8.0721)
Level	$-\mathbf{7.1689}$ (0.01)	-1.1862 (0.40)
	(-16.2493;-1.0735)	(-7.8938;4.9628)
Slope	$-\mathbf{4.2520}$ (0.02)	$-\mathbf{8.0350}$ (0.01)
	(-11.8828;-0.3385)	(-15.9164;-1.6171)
ΔGDP	-0.1717 (0.50)	-6.8869 (0.29)
	(-23.8065; 23.3050)	(-28.5787;14.5967)
Inflation	0.2209 (0.50)	0.3904 (0.52)
	(-21.3753;23.2929)	(-21.7693;22.9547)

This table presents the summary of the heterogeneous probability parameters ($\pi^{\mathbb{P}}_{s_t,s_{t+1}}(t)$). $R1$ and $R2$ represent Regime 1 and Regime 2 respectively. The probability of regime transitions was hypothesized to be a function of term structure factors X_t, GDP and inflation. For each coefficient, a mean of the posterior distribution, the 95 percent credible interval (in brackets below the estimate), and the probability that the estimate was greater than zero ($P(x>0)$ based on the posterior distributions) is shown. Those variables that significantly differ from zero at the one-sided 5 percent level are emphasized in bold.

The interpretation of the coefficients in Table 3 is not clear-cut. The Logistic function is defined such that the coefficients are inversely related to probability. The likelihood of remaining in Regime 1 is positively related to the level and the slope of the term structure. This finding is consistent with intuition. The probability of staying in regime 2 is likewise related to slope, but not the level of the term structure. This result is surprising, given the fact that high levels and steep term structure characterize Regime 1. The comparison of the coefficients is difficult, however, because we model the conditional transition probabilities. In contrast, Probit or Logit regressions model the probability of being in a particular state, unconditional on the proceeding regime. A test using the posterior densities of the parameters revealed that there are no statistically relevant differences between the parameters. This is due to the large variation in the estimates that is caused by a combination of a small data sample and small explanatory power of the exogenous variables.

Figure 3. Volatility and the Regime Process from Jan. 1998 to May 2007

Source: Fund staff estimates.

Figure 4 graphs the heterogeneous probability of transitioning from Regime 1 to Regime 2 ($P[1,2]$) and the probability of switching from Regime 2 to Regime 1 ($P[2,1]$). In-sample, the model is capable of explaining the regime changes rather well. The regimes and probability

coincide (offset) in the top (bottom) figure. In several instances, the probability is a leading indicator of regime change. Despite the relatively good in-sample explanatory power of the model, the large variances of the parameters and the high persistence in the regimes would make accurate prediction of future regime transitions difficult.

Figure 4. Heterogeneous Probability

Source: Fund staff estimates.
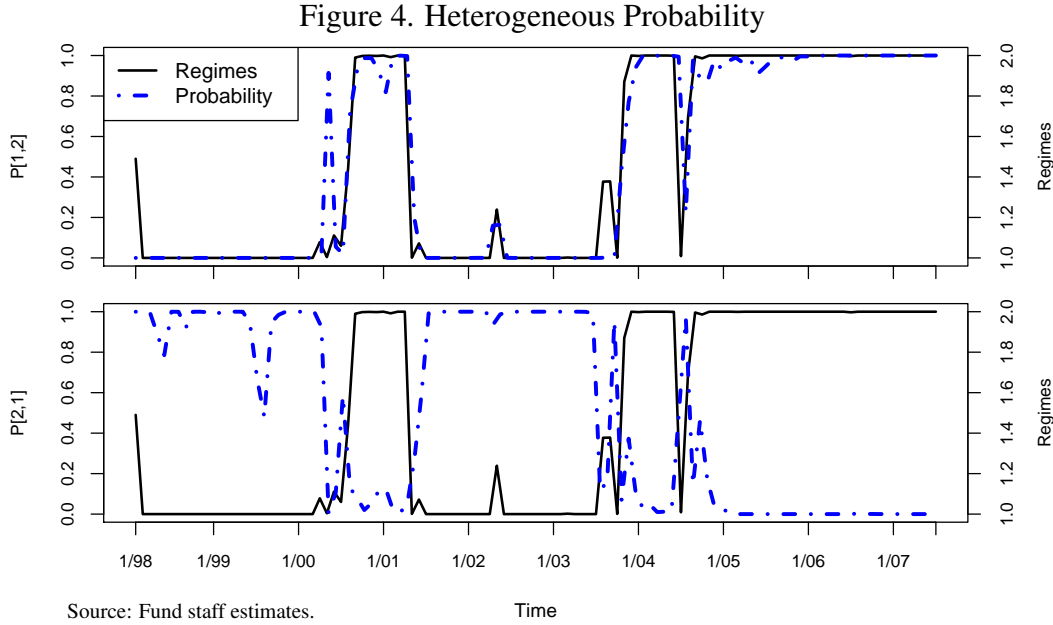
B. Term Structure Characteristics

Table 4 displays the estimates of the single regime **ATSM** model and Table 5 presents the coefficients of the **HMM** model. The number to the right of the coefficient represents the probability that the coefficient is greater than zero.[16] Bold coefficients are significantly greater, or less then zero (depending on the coefficient sign), at the 5 percent significance level. In order to identify the differences between the regimes, we also tested for equality of the parameters. One asterisk indicates that the parameter is greater in Regime 1 and two asterisks that it is greater in Regime 2 at the 5 percent (one sided) level.

The coefficients concur with the regime themes outlined in Figures 1, 2, and 3. The residual variance of X_t, $\Sigma(s_t)\Sigma(s_t)'$ as well as the residual variance of the rates modeled with error, $\Omega(s_t)\Omega(s_t)'$, are greater in Regime 1 than in Regime 2. Error variances in both cases are smaller than the error variances in the **ATSM** model, indicating a better over-all fit to the data.

---

[16]

This was calculated as the mean of a indicator function which took the value of one if the coefficient value was greater than zero or zero otherwise. Probability greater than 97.5 percent therefore indicates that the variables is positive at the 5 percent two sided significance level. Probability smaller than 2.5 percent indicates that the parameter is significant and negative.

Table 4. **ATSM** Model Parameters

a	10.3893		
b	**0.9437** (1) (0.7772;1.1885)	**−1.1049** (0) (-1.3047;-0.8801)	**−0.2136** (0) (-0.2389;-0.1085)
K	**0.2950** (1) (0.1808;0.4198)	0	0
	0.0224 (0.67) (-0.0868;0.1284)	**0.1070** (1) (0.0507;0.2065)	0
	−0.2236 (0) (-0.3582;-0.0947)	**−0.2040** (0) (-0.3035;-0.0957)	**0.1649** (1) (0.0917;0.2467)
$\boldsymbol{\lambda}_0$	0	0	**11.3225** (0) (10.3605; 13.2405)
$\boldsymbol{\Lambda}_1$	**−0.2202** (0) (-0.3769;-0.0536)	**−0.7107** (0) (-0.8304;-0.5897)	**−0.0679** (0) (-0.12167;-0.0211)
	-0.0388 (0.33) (-0.1853;0.0904)	-0.0263 (0.45) (-0.2145; 0.1105)	**0.1422** (1) (0.1121; 0.1678)
	−0.4630 (0) (-0.6675;-0.2675)	**−0.5897** (0) (-0.7510;-0.3204)	**0.1606** (1) (0.0691;0.2598)
$\boldsymbol{\Omega\Omega'}$	**0.1378** (1) (0.1022;0.1866)	**0.0989** (1) (0.0699;0.1387)	0.0132 (0.93) (-0.0049;0.0311)
		0.1116 (1) (0.0837;0.1491)	**0.0347** (1) (0.0193;0.0527)
			0.0628 (1) (0.0462;0.0852)

This table presents the summary of the term structure parameters for the **ATSM** model. The normalizing restriction were imposed so that θ is a null vector and $\boldsymbol{\Sigma}$ is an identity matrix. **K** is lower triangular with all eigenvalues greater than zero. Overidentifying restrictions were also imposed on the vector $\boldsymbol{\lambda}_0$ and a was fixed at the mean of the short rate. For each coefficient a mean of the posterior distribution, a 95 percent credible interval (in brackets below the estimate) and the probability that the estimate was greater than zero ($P(x > 0)$ based on the posterior distributions) is shown (right of the estimate). Those variables that are significantly different from zero at one-sided 5 percent are emphasized in bold.

X_t dynamics are stationary in both regimes in the **HMM** model. All elements of the diagonal of **K** in Regime 1 are greater than zero and the eigenvalues of $1 - \mathbf{K}$ in Regime 2 are all less than one. Stationarity in Regime 1 and in the **ATSM** model were imposed for identification purposes. Direct comparison of the mean reversion is difficult because of the mixed size of the diagonal and off-diagonal elements of $\mathbb{K}(1)$ and $\mathbb{K}(2)$. To benchmark the mean reversion across regimes and models we compare the real portions of the eigenvalues of $\mathbb{K}$.

$$Eigen(\mathbf{K^{HMM}}(1)) = [0.3759, 0.1637, 0.1213]$$
$$Eigen(\mathbf{K^{HMM}}(2)) = [0.3576, 0.0531, 0.0034] \tag{42}$$
$$Eigen(\mathbf{K^{ATSM}}) = [0.2950, 0.1070, 0.1649]$$

Expression (42) shows that the mean reversion in Regime 1 is greater than in Regime 2. The comparison with the **ATSM** is inconclusive.

Figures 5 and 6 depict the term structure factors X_t for the **ATSM** and **HMM** model respectively. The factors are correlated to yield curve characteristics and can be interpreted as the level ($X[2]$), slope ($X[3]$) and curvature ($X[1]$). $X[1]$ is only moderately correlated to curvature, as defined by a butterfly consisting of a two short positions in 2-year rates and long position in both the 1-year and 3-year rates.

The factors of the **ATSM** and the **HMM** models are highly correlated. The level of the factors is the most conspicuous disparity between the models, most noticeably in the case of $X[2]$ and $X[3]$. The long term mean parameter in Regime 1 ($\boldsymbol{\theta}(1)$) in the **HMM** model was normalized

Table 5. **HMM** Model Parameters

	Regime Invariant Parameters		
a	10.3893		
b	**1.1132** (1)	**−1.4149** (0)	**−0.1453** (0)
	(0.9336145;1.311132)	(-1.62039;-1.209320)	(-0.1940596;-0.1049315)

	Regime 1 Parameters		
K(1)	**0.3759**(1)	0**	0
	(0.2220;0.5257)		
	-0.0182 (0.432)	**0.1637** (1)	0
	(-0.1858;0.1372)	(0.0643;0.2830)	
	−0.3033** (0)	**−0.1183**** (0.13)	**0.1213*** (1)
	(-0.4882;-0.1322)	(-0.2852;0.04877)	(0.03586;0.2052)
λ$_0$(1)	0	0	**8.9326*** (1)
			(8.0477;9.8567)
Λ$_1$(1)	**0.2989**** (1)	**0.4208*** (1)	**0.1786** (1)
	(0.1422;0.4719)	(0.3429;0.4923)	(0.1463;0.2192)
	0.0008 (0.49)	**−0.1400** (0.02)	**−0.1393** (0)
	(-0.15642;0.1585)	(-0.2644; -0.0276)	(-0.1670; -0.1109)
	0.3828* (0.998)	**0.3672** (1)	**−0.2363** (0)
	(0.1960;0.5829)	(0.1718;0.5637)	(-0.3531;-0.1114)
Ω(1)**Ω**(1)′	**0.1203*** (1)	**0.0794*** (1)	0.0134 (0.90)
	(0.0960;0.1501)	(0.0591;0.1059)	(-0.0034;0.0316)
		0.1002* (1)	**0.03*** (1)
		(0.0785;0.1283)	(0.0191;0.0515)
			0.0847* (1)
			(0.0660;0.1075)

	Regime 2 Parameters		
θ(2)	0.0713 (0.36)	0.7737 (0.81)	-0.3744 (0.43)
	(-6.9318;5.1837)	(-3.6822;6.4435)	(-10.5536;9.6446)
Diag(**Σ**(2)**Σ**(2)′)	**0.0345*** (1)	**0.0331*** (1)	**0.0338*** (1)
	(0.0221;0.0513)	(0.0256;0.0426)	(0.0237;0.0459)
K(2)	**0.1878** (1)	**0.1900**** (1)	0.0105 (0.63)
	(0.0845;0.2748)	(0.1124;0.2611)	(-0.0318;0.0543)
	0.1484 (0.985)	**0.1982** (0.992)	-0.0227 (0.34)
	(0.0412;0.2661)	(0.0763;0.3227)	(-0.0986;0.0489)
	0.0925**(0.92)	0.0756** (0.84)	**0.0281*** (0.72)
	(-0.0161;0.2208)	(-0.0465;0.1919)	(-0.0282;0.0994)
Λ$_1$(2)	**0.4870**** (1)	**0.2309*** (1)	**0.1682** (1)
	(0.3740;0.6132)	(0.1760;0.2998)	(0.1159;0.2195)
	−0.1658 (0.01)	**−0.1744** (0.01)	**−0.1166** (0.01)
	(-0.2970;-0.0498)	(-0.3005; -0.0524)	(-0.1921; -0.0267)
	-0.0129** (0.49)	**0.1733** (1)	**−0.1430** (0)
	(-0.1970;0.1316)	(0.0717;0.2933)	(-0.2731;-0.0656)
Ω(2)**Ω**(2)′	**0.0140*** (1)	**0.0046*** (1)	-0.0013(0.13)
	(0.0109;0.0175)	(0.0024; 0.0073)	(-0.0034;0.0006)
		0.0142* (1)	-0.0011* (0.18)
		(0.0113;0.0176)	(-0.0031;0.0008)
			0.0108* (1)
			(0.0086;0.0135)

This table presents the summary of the term structure parameters for the bxHMM model. There are regime invariant parameters a and b, as well as regime dependent parameters. The normalizing restriction were imposed so that $\theta(1)$ is a null vector and $\Sigma(1)$ is an identity matrix. **K** is lower triangular with all eigenvalues greater then zero. Overidentifying restrictions were also imposed on the vector $\lambda_0(1)$ and a was fixed at the mean of the short rate. Additionally, $\lambda_0(2)$ was assumed to be a null vector. For each coefficient a mean of the posterior distribution and a 95 percent credible interval (in brackets below the estimate) is presented. The probability that the estimate was greater than zero, $P(x > 0)$ based on the posterior distributions is shown in brackets next to each coefficient. Those variables that are significantly different from zero at the one-sided 5 percent significance level are emphasized in bold. A star ($*$) signifies that at the one sided 5 percent significance level the coefficient in Regime 1 is greater than the coefficient in Regime 2. Two stars ($**$) indicate that parameter in Regime 2 is greater than in Regime 1.

Figure 5. Latent Factor X_t for the **ATSM** Model

Source: Fund staff estimates.
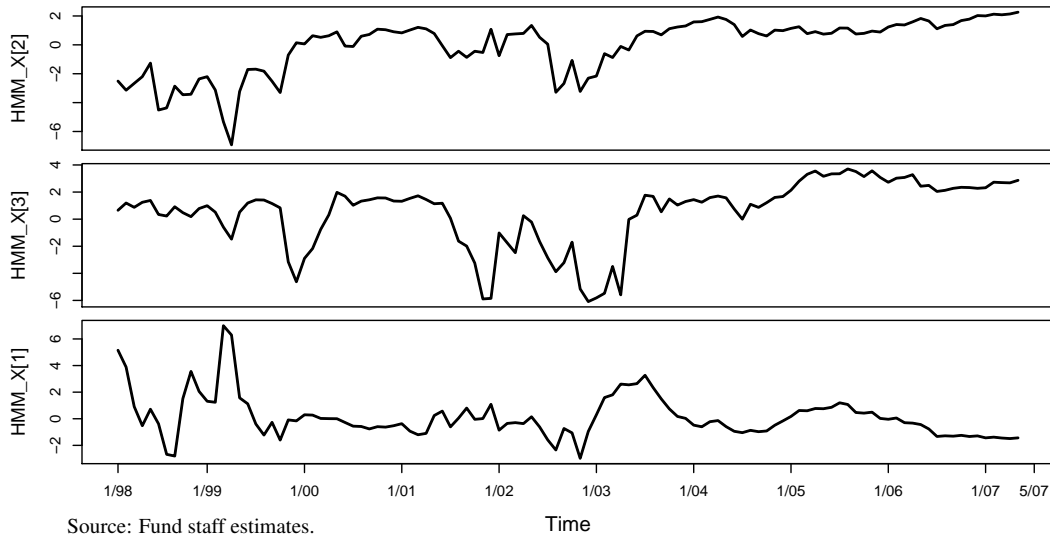
to a null vector and therefore, mean of X_t in Regime 1 should be zero. This result only holds asymptotically and there is no guarantee that the sample mean of X_t will equal its long term mean without restricting the model.[17]

Figure 6. Latent Factor X_t for the **HMM** Model

Source: Fund staff estimates.
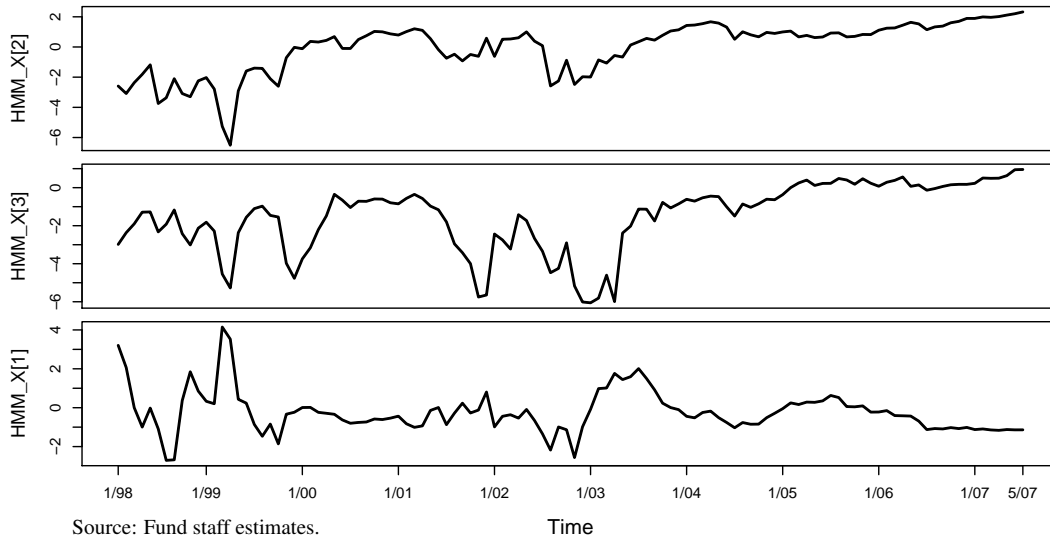
The implication of this is that the asymptotic level of the term structure factors in Regime 1 is different from the estimated factors. This is not a problem in-sample, but is troubling if the model is used for forecasting. Forecasting a few periods ahead would cause the the term structure in Regime 1 to revert to a level that is significantly lower than the mean of the

---

[17]The estimation algorithm does not require the mean of X_t to be zero explicitly. Including a function of the mean X_t as a penalty term in the likelihood function does ensure that the mean is zero, but leads to severe deterioration in fit.

historical data. Although this may be consistent with the long run expectations of the Brazilian swap rate movements, it does suggest that the characteristics of Regime 1 are not completely described by the model. One reason for this could be that the overidentifying restrictions imposed by Dai, Singleton and Yang (2007) are not appropriate. We tested this hypothesis by relaxing the restrictions on the value of a, λ_0 and λ_1 and found that the non-zero mean is persistent. An alternative explanation could be that two regimes are insufficient to describe the extreme behavior of the Brazilian term structure in the beginning of the sample.

C.   Market Prices of Risk

Figures 7 and 8 portray the market prices of risk for the **ATSM** model and the **HMM** model respectively. The market prices of risk are given by expressions (3) and (11) and can best be understood as the excess return on the bond per unit of risk.

Figure 7. Market Prices of Factor Risk for the **ATSM** Model



Source: Fund staff estimates.
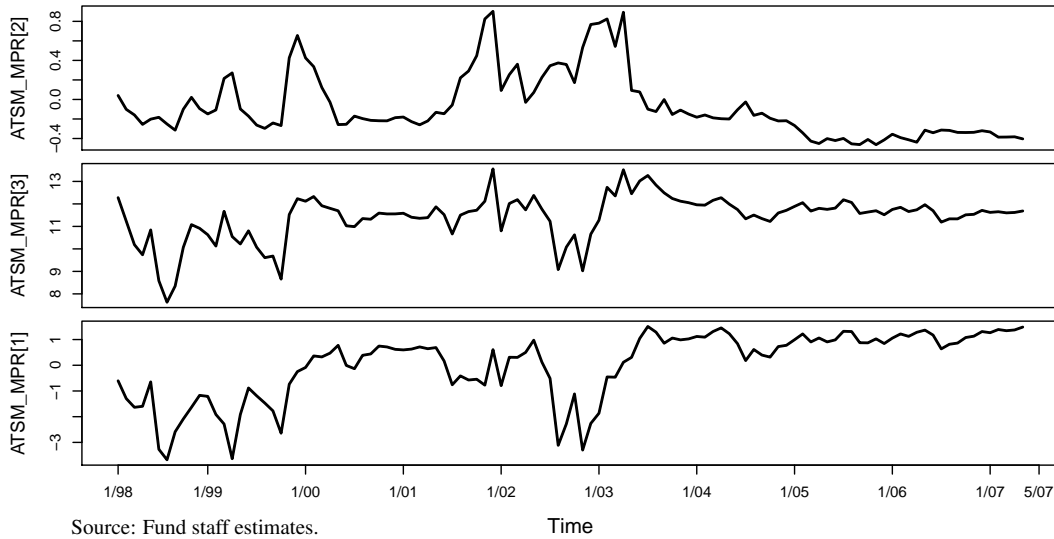
The market price of risk in regime 2 is more volatile for all factors and is higher for the level factor. This is consistent with Bansal and Zhou (2002) who find that the risk premium is higher in the low volatility regime. The differences between the volatilities in the regimes dominate the differences in the levels of the latent term structure factors. This could be an indication that the interest rate level is not commensurate with the relatively small risk inherent in the term structure in Regime 2. This is possibly because the recent stable macroeconomic policies have not been fully incorporated in interest rate expectations.

The market price of regime switching risk appears to be priced and is calculated via expression (12). The price is negative when the real world probability of regime change is lower than the probability under the risk neutral measure. The price of transitioning from a high volatility regime to a low volatility regime is lower than the reverse. At the end of the sample, the risk of switching to regime 2 is presumably relatively low.

Figure 8. Market Prices of Factor Risk for the **HMM** Model

Source: Fund staff estimates.
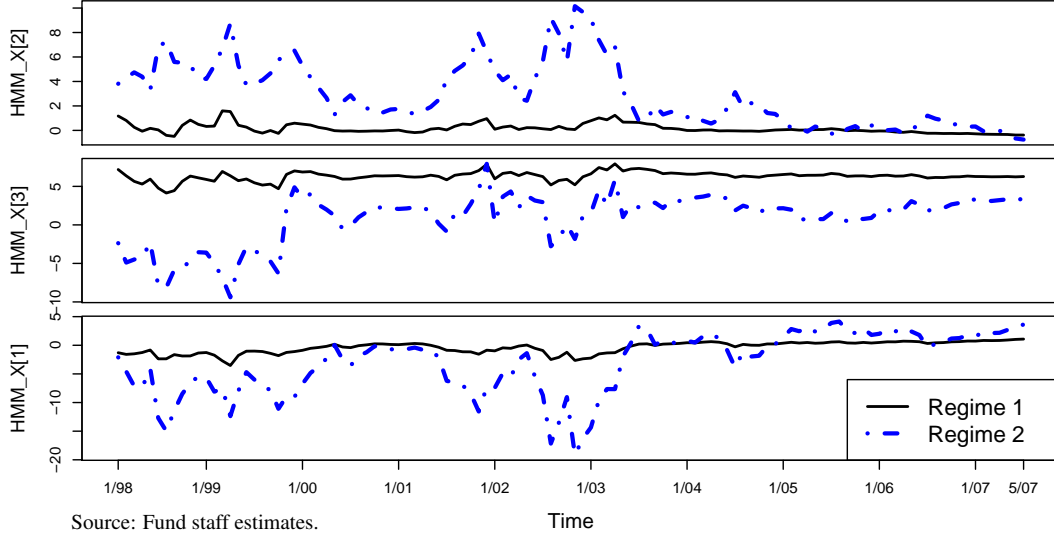
D.   Model Fit

The fact that the fit of the **HMM** is superior to the **ATSM** model should not be surprising considering the results in the previous sections. The difference in likelihoods could be attributable solely to the number of parameters in the **HMM** model. In order to control for complexity, the Deviance Information Criterion (DIC), the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) were calculated to compare the models.

Figure 9. Market Prices of Regime Switching Risk for the **HMM** Model

Source: Fund staff estimates.
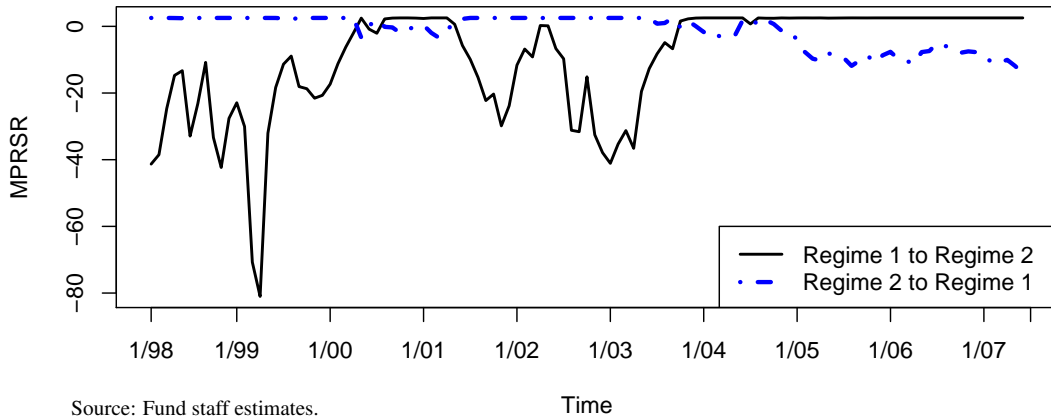
The DIC, AIC and BIC penalize the model for the number of free parameters, but in the hidden Markov models the number of parameters is not immediately obvious.[18] We calculate the effective number of parameters (denoted k) via expression (43) and thereafter calculate the

---

[18]The problem is that each element of the state vector S could be taken as a separate parameter, adding a further T parameters to the total. This would be a gross over-estimate however, given the persistence in the regimes.

values of the fit statistics which are presented in Table 6.

$$k = \bar{D}(\mathbf{\Theta}) - D(\bar{\mathbf{\Theta}}) \tag{43}$$

$D(\cdot)$ is the deviance and is defined as $-2log(P(\mathbf{\Theta}|\mathbf{V}))$ where $P(\mathbf{\Theta}|\mathbf{V})$ is the log likelihood function. $\bar{D}(\mathbf{\Theta})$ refers to the mean deviance and $D(\bar{\mathbf{\Theta}})$ to the deviance of the posterior parameter means. AIC is given by $2k - 2ln(L)$, BIC as $kln(n) - 2ln(L)$ and DIC is calculated via $k + \bar{D}(\mathbf{\Theta})$.

Furthermore, Table 6 separates the likelihood into time-series and cross-sectional fit. Time series fit refers to the ability of the model to capture the dynamics of interest rates over time. This is calculated simply as the portion of the likelihood corresponding to expression (22). Cross sectional fit corresponds to equation (21) and measures the ability of the model to explain the span of yields at a particular point in time.

Table 6. Comparative Measures of Fit

| | ATSM | | | HMM | | |
	Total	Time-Series	Cross-Section	Total	Time-Series	Cross-Section
DIC	1361.781	-	-	891.240	-	-
AIC	1378.471	-	-	922.600	-	-
BIC	1442.181	-	-	1042.857	-	-
Mean(ln(L))	-672.545	-657.589	-14.956	-429.940	-518.085	88.144
97.5%(ln(L))	-678.108	-662.301	-20.539	-445.525	-528.955	76.520
2.5%(ln(L))	-667.663	-653.385	-9.728	-415.835	-509.334	98.698

This table presents comparative statistics of fit for the two models. The Total column presents the statistics for the entire model. DIC, AIC and BIC penalize overall fit by the number of parameters used, which are calculated via expression (43). The effective number of parameters were 16.7 and 31.4 for the ATSM and HMM models respectively. Mean(log(L)) refers to the mean of the log likelihood and is a raw measure of fit that does not adjust for the complexity of the model. 97.5 percent and 2.5 percent credible intervals for the log likelihood (log(L)) are also shown. The table also separates the overall likelihood into time-series and cross-sectional components. The time-series component refers to the ability of the model to fit the dynamics of the yields matched without error. The cross-sectional column identifies the fit of the yields matched with error.

In all cases, the **HMM** model exhibits a superior in-sample fit. The improvement in the overall likelihood is due to improvement in both the time-series and the cross sectional dimensions. Both models are better at characterizing contemporaneous rates than fitting time-series dynamics. ATSMs, in general, have been shown to have inadequate forecasting performance. Cross-sectional fit is superior because of the high correlation between rates of different maturities.

E. A Note on Convergence

Due to the small sample available and because of the highly non-linear nature of the ATSMs, convergence for the **HMM** model was difficult to achieve. The samples from the full posterior density exhibit high autocorrelation as a result of slow mixing in the Metropolis-Hastings within Gibbs algorithm. Consequently, a large number of iterations are necessary to verify convergence and a high degree of thinning was required to obtain a reliable sample. Over 1

million iterations were run to achieve convergence and a sample of 200,000 was collected. Figure 10 displays the sample from the full posterior densities of selected parameters. The convergence properties of the **ATSM** model were much faster due to the absence of the hidden Markov process.

The complex restrictions on the term structure dynamics necessitates the use of the Metropolis-Hastings algorithm. The performance of this algorithm relies on the appropriate choice of the proposal density. The Random Walk Metropolis-Hastings algorithm (RW) that uses a multivariate normal proposal density was utilized in this case. The specification of the covariance matrix is crucial in the RW algorithm. A very small (large) covariance will result in high (low) acceptance rates, but small (large) step sizes. A popular option (Robert and Casella (2000)) is to use the inverse of the Hessian at the Maximum Likelihood (ML) estimate. This approach can be computationally expensive however, since it requires the calculation of the ML. In term structure models, this can be complicated and time consuming. An alternative approach (and the one used in this study) obtains the covariance matrix by calculating the covariances of the initial sample of the parameters.The initial sample is based on a naïve covariance matrix where the diagonal elements of the covariance matrix are identical and the off-diagonal elements are zero. Roberts and Rosendal (2001) review optimal scaling literature and suggest an acceptance rate of 23.4 percent for a RW algorithm. The size of the covariance matrix in this application was optimized so that the resulting acceptance rates were between 20 percent and 25 percent.

VII. CONCLUSION

We review Gaussian affine term structure models and a hidden Markov model of the term structure developed by Dai, Singleton and Yang (2007) and apply them to modeling the Brazilian term structure. This is the first study that applies hidden Markov models to an emerging market term structure and illustrates the challenges posed by small samples and relatively extreme term structure outcomes. We also develop a Bayesian MCMC algorithm to estimate hidden Markov models of the term structure. The methodology is easy to apply and produces consistent standard errors.

The application of these models showed that the dynamics of the Brazilian term structure has undergone material shifts. A high level, slope and volatility regime (Regime 1), and a low level, slope and volatility regime (Regime 2) are identified in the Brazilian term structure. Regime 1 is also characterized by a higher mean reversion, and factor and error variances. The currency crisis from 1998 to 1999 is the main driver of Regime 1, although the period spanning the Brazilian election of 2002 is also a factor. Regime 2 is primarily characterized by the stable monetary and fiscal policy in recent years. The underlying drivers of the regimes may be both short and long term. The Brazilian economy has undergone a significant structural changes over the past decade leading to greater stability and lower interest rates. This long term trend is reflected by the incidence of Regime 1 at the beginning of the sample and Regime 2 at the end. This would suggest that only the dynamics under Regime 1 are relevant in the future. Regime

Figure 10. Convergence of Selected Parameters
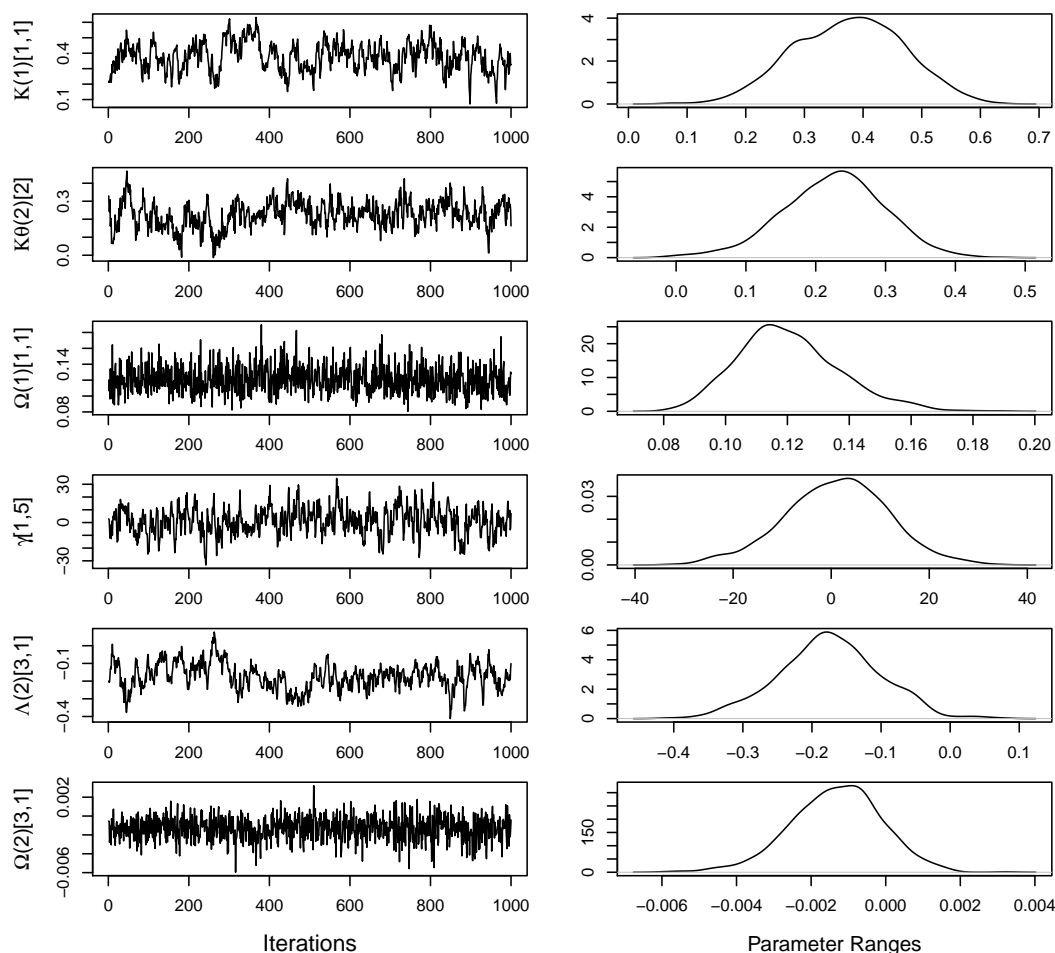
Source: Fund staff estimates.

2 is, however, also driven by short term uncertainty, as indicated by its occurrence around the time of the 2002 elections. The possibility of more volatile yields should be incorporated into forecasting or in making policy decisions, such as those involved in public debt management.

The hidden Markov model overperforms the single factor ATSM model using all available measures of in-sample fit, including those that penalize for the complexity of the model. The improvement in fit is attributable to both time-series and cross-sectional dimensions. Time series fit remains poor, which is consistent with other affine models of interest rates. Forecasting performance of the models remains undetermined since one step ahead forecasts are not practical in this framework.

We also find that, interestingly, macroeconomic variables do not have any power to explain regime shifts conditional on the inclusion of term structure factors. This result implies that all macroeconomic information is incorporated in the latent term structure factors. This is inconsistent with some recent studies of the term structure (Ang and Piazzesi (2003) and Diebold, Rudebusch and Auroba (2006)) that show that incorporating macroeconomic variables directly in the term structure model improves in-sample fit and forecasting

performance. Our results are not directly comparable, since macroeconomic variables were not used as term structure factors. Future studies could focus on disentangling the effect of regime shifts and direct macro-dependencies.

In contrast, term structure factors do explain state transitions, but their explanatory power is weak. The estimated regimes are highly persistent, implying that forecasting regime changes using these models is difficult. In the light of this fact, and the general poor performance of affine models of the term structure in forecasting interest rates in general, the use of hidden Markov models should be restricted to situations where the distributional characteristics of the model are of paramount importance. This qualification excludes forecasting and pricing, for which arbitrage free models of the term structure are far more appropriate, but encompasses risk management and fixed income management including public debt management. Additionally, hidden Markov models identify data driven states that can be used to determine scenario baselines in scenario analysis.

Future research should focus on developing restrictions that will allow simplified estimation of hidden Markov models. Benchmarking their forecasting performance to more popular models such as the dynamic Nelson and Siegel models is necessary. The ability of these models to generate realistic scenarios for interest rate outcomes is also worth investigating.

REFERENCES

Ang, A. and M. Piazzesi, 2003, "A no-arbitrage vector autoregression of the term structure dynamics with macroeconomic and latent variables," *Journal of Monetary Economics*, 50, 745-787.

Bansal, R., and H. Zhou, 2002, "Term structure of interest rates with regime shifts," *Journal of Finance*, 57, 1997-2043.

Bansal, R., G. Tauchen and H. Zhou, 2004, "Regime-shifts, risk premiums in the term structure, and the business cycle," *Journal of Business and Economic Statistics*, 22, 396-409.

Dai, Q., J. Singleton and W. Yang, 2007, "Regime shifts in a dynamic term structure model of US treasury bond yields," *Review of Financial Studies*, 20, 1669-1706.

Diebold, F. X., G. D. Rudebusch and B. Aruoba, 2006, "The macroeconomy and the yield curve: A dynamic latent factor approach," *Journal of Econometrics*, 131, 309-338.

Engel C. and C. Kim, 1996, "The long run US/UK real exchange rate," Working Paper, National Bureau of Economic Research.

Gray, S. F., 1996, "Modeling the conditional distribution of interest rates as a regime-switching process," *Journal of Financial Economics*, 42, 27-62.

Hamilton, J. D., 1989, "A new approach to the economic analysis of non-stationary time series and the business cycle," *Econometrica*, 57, 357-384.

Hamilton, J. D., 1990, "Analysis of time series subject to changes in regime," *Journal of Econometrics*, 45, 39-70.

Kim, C., 1994, "Unobserved component time series models with Markov-switching heteroscedasticity: Changes in regime and the link between inflation rates and inflation uncertainty," *Journal of Business and Economic Statistics*, 11, 341-349.

Kim, C., 1996, "Predicting business cycle phases with indexes of leading and coincident indicators: A multivariate regime-shift approach," *Journal of Monetary Economics*, 12, 1-27.

Landen, C., 2000, "Bond pricing in a hidden Markov model of the short rate," *Finance and Stochastics*, 4, 371-389.

Nelson, C. R. and A. F. Siegel, 1987, "Parsimonious modeling of yield curves," *The Journal of Business*, 60, 473-489.

Robert, C. P. and G. Casella, 2004, "Monte Carlo Statistical Methods," Springer Texts in Statistics.

So, M. K. P., K. Lam and W. K. Li, 1998, "A stochastic volatility model with Markov switching," *Journal of Business and Economic Statistics*, 16, 244-253.

Wu, S. and Y. Zeng, 2003, "Affine regime-switching models for interest rate term structure," *Contemporary Mathematics*, 351, 375-386