Regime-Switching Univariate Diffusion Models of the Short-Term Interest Rate

Seung-Moon Choi*

University of Wisconsin-Madison
Department of Economics

First Draft: September 15, 2004

Abstract

This article proposes very general regime-switching univariate diffusion models and estimates these using short-term interest rates. There is a large literature supporting the existence of regime changes in the evolution of interest rates. In particular, these papers report strong evidence of high volatility in U.S. short-term interest rates during some periods of historic events. My study extends the existing models in four directions. First, I use a very general parametrization which nests most of the earlier specifications. Keeping the same constant elasticity of variance diffusion function of Conley et al. (1997), I add a third order term to the drift function. Second, all parameters of the drift and the volatility components are subject to regime shifts. Third, the transition probabilities governing the dynamics of the regime variable vary with the interest rates. Fourth and finally, when constructing the likelihood function by using the algorithm established by Hamilton (1989), I use a very accurate approximate transition density function of the diffusion process obtained by the method of Aït-Sahalia (2002b). The maximum likelihood estimates are calculated using the weekly series of U.S. 3-month treasury bill rates. For comparison, the regime-switching Vasicek and CIR models are also estimated. Finally, I derive a system of partial differential equations for the bond price of our model, which can be solved numerically since the analytic solution is not available.

The estimation results reveal that there is strong evidence for the existence of two regimes, for the time varying transition matrix, and for the high persistence of both regimes. The volatility, but not the drift, is estimated accurately and plays a key role in explaining the dynamics of the interest rates. The volatility of one regime is about three times that of the other regime over the support of interest rates. Based on the inferred probability of the process being in a regime, I can classify the sample period into high and low volatility states quite distinctively. The likelihood-based test apparently rejects the other two models in favor of our model. This implies that misspecification can result in misleading outcomes particularly regarding the volatility and transition probabilities of the regime index.

*I am grateful to Bruce Hansen and Dennis Kristensen for helpful comments and suggestions.
1 Introduction

There is a large literature supporting the existence of regime changes in the evolution of interest rates (e.g., Hamilton (1988), Cai (1994), Gray (1996), Garcia and Perron (1996), and Ang and Bekaert (2002a)). In particular, these papers report strong evidence of high volatility in U.S. short-term interest rates during the episodes of the 1973 and 1979 OPEC oil crises, the 1979-82 Federal Reserve Monetary Experiment, and the 1987 stock market crash. Most of the literature focus on discrete time models, such as autoregressive (AR) and autoregressive conditional heteroskedasticity (ARCH) specifications in order to discover the regime-switching features of interest rates.

On the other hand, the short-term interest rate is often modeled as a continuous time diffusion process. Many earlier models of term structure of interest rates are based on the simplifying assumption that changes in all yields are driven by a single underlying, random factor. Studies of one-factor diffusion model of the instantaneous rate include Vasicek (1977), Cox, Ingersoll, and Ross (1985) (hereafter CIR), Chan, Karolyi, Longstaff, and Sanders (1992) (hereafter CKLS), Aït-Sahalia (1996b), Conley, Hansen, Luttmer, and Scheinkman (1997) (hereafter CHLS), Stanton (1997), Pritsker (1998), Durham (2002), Jones (2003), and Kristensen (2003), naming a few. Litterman and Scheinkman (1991) and Chapman and Pearson (2001) document that three factors (level, slope and curvature) can fully capture most of the variability of interest rates. In fact, roughly 90 percent of the variation in U.S. yields changes can be explained by the first factor, which is typically identified with the short rate of interest. Recent extensive articles on the univariate diffusion models of the short-term interest rate can be interpreted as focusing on the dominant first factor and attempting to investigate the relationship between the level of interest rates and their expected changes and volatilities in detail.

However, empirical evidence proves that additional factors are needed to fit the data better (e.g. Dai and Singleton (2000), Ahn, Dittmar, and Gallant (2001), Andersen, Benzoni, and und (2003)). My regime-switching univariate diffusion model is a two-factor model. One is the continuously evolving short rates and the other is the regime variable changing discretely, taking finite number of values. While there have been considerable works in which the instantaneous rate follows a univariate stochastic differential equation (SDE), to my knowledge, there are only a few papers that introduce the regime shifts into the continuous time model to study the dynamics of short-term interest rates. Naik and Lee (1998) analyze the regime-switching Vasicek model in which only the volatility is subject to regime shifts. They use the known transition density function of the Vasicek process to carry out the maximum likelihood (ML) estimation. Three long-term bond yields data sets are inverted with the bond price formula they derived to get the short rate and estimate the model. Using the Markov Chain Monte Carlo (MCMC) method Liechty and Roberts (2001) estimate a regime-switching Gaussian diffusion model where only the drift component is regime-dependent. The Regime-switching CIR model is studied by Driffl, Kenc, Sola, and Spagnolo (2004) (hereafter DKSS). They
approximate the transition density function by discretizing the CIR model and find the ML estimates of the parameters. Note that applying ML estimation to a discretization of the stochastic differential equation yields biased and inefficient estimates (Lo (1988)). Landén (2000) and Wu and Zeng (2003) discuss bond pricing of a univariate diffusion models in the regime-switching context. Some studies have examined discretized versions of the regime-switching diffusion models (e.g. Ang and Bekaert (2002b) (Vasicek model), Bansal and Zhow (2002) (three factor CIR model where all factors are independent using the efficient method of moments (EMM) method) and Dai, Singleton, and Yang (2003) (hereafter DSY) (three factor Gaussian model)).

This paper extends in four directions the aforementioned research on the continuous time regime-switching diffusion models of the short-term interest rate. First, to let the model decide which fits the data well, I use a general diffusion model which nests most of the existing one factor diffusion models. Keeping the same constant elasticity of variance (CEV) diffusion function of CHLS, I add a third order term to the second order polynomial form of the drift function to allow more flexibility. Second, all parameters of the drift and the volatility components are subject to regime shifts. Third, the transition probabilities of the Markov chain which govern the dynamics of the regime index vary with interest rates. The Normal cumulative distribution and a logistic functions of interest rates are used to ensure that the transition probability is between zero and one. Finally, highly accurate transition density function of the diffusion process is used when constructing the likelihood function by applying the recursive algorithm developed by Hamilton (1989). Therefore, my study does not suffer from discretization bias. The true transition probability density function of a diffusion process is, in general, not known with the exception of a few cases such as Vasicek, CIR and Black ans Scholes (1973). However, Aït-Sahalia (2002a) establishes a theory to approximate the true density function of an arbitrary univariate diffusion process in closed form. Exploiting his theory, I generate the approximate density function and conduct ML estimation.

The regime variable follows a continuous time two-state Markov chain. As usual in regime-switching literature, the two states are labeled as low (L) and high (H) regimes depending on the magnitude of the variance of the interest rate in each state. Markov switching models are useful in modeling occasional but sudden regime shifts. My conjecture is that the drift and the volatility of the two states have very different features. The regime L is expected to have relatively lower variance and lower mean reversion than the regime H. Whether interest rates have a tendency to return to a long run mean and if so, the strength is much stronger at high or low levels are important but still unresolved questions in univariate models of the short-term interest rate. We anticipate that the drift of the state H, typically characterized by high interest rates and high volatility, show this nonlinear property. The volatility function is expected to explain well documented heteroskedastic behavior of the conditional variance of interest rates.

Five different models are estimated depending on the number of regimes, the specification for the initial state distribution and the parametrization of the transition matrix. Then, for the purpose of comparison
to the previous articles, the regime-switching Vasicek and CIR models are estimated. Notice that these are contained in the general model. Weekly series of the U.S. 3-month treasury bill rates which spans from January 8, 1971 through December 26, 2003 are used to estimate the model.

The usual test statistics cannot be applied to test the existence of the second regime since parameters associated with the second state are unidentified under the null of one regime. There are some researchers (e.g., Davis (1987) and Hansen (1992, 1996)) who address to this problem but, generally, testing for multiple regimes is not an easy task, as in the case for my model. The enormous increase in the likelihood value when moving from a single-regime model to a two-regime model does however provide a certain amount of evidence for the existence of the additional regime. The log likelihood test (LRT) can be safely applied to compare the regime-switching models. The specification of the initial state probability and the functional form for the transition probability of the Markov chain do not matter much. But the null of the time-invariant transition probability of the regime variable is strongly rejected in favor of the time-varying transition matrix. Moreover, both regimes are greatly persistent in all models. When it comes to the performance of classifying regimes, both state dependent and independent transition matrices result in quite similar outcomes. In addition to those high volatility periods mentioned above, early 1970s, several periods between the end of 1991 and the beginning of 1994, 1998 Long Term Capital Management crisis, the September 11, 2001 attacks, the stock market downturn of 2002, and the 2003 invasion of Iraq are also identified as the economy being in the high variance regime according to the smoothed probabilities. While the high volatility regime occurs mostly at higher interest rates, the low variance state prevails at lower interest rates.

It is not the drift term but the diffusion function that plays a crucial role in explaining the evolution of the short rates for each regime. The drift functions of all regime-switching models are insignificantly estimated and I cannot even find the evidence against zero drift in any models. Chapman and Pearson (2000) and Bandi and Phillips (2003) provide explanations why the drift term is inaccurately estimated. In contrast, the volatility terms of two states are not only highly significant but also very different from each other for all models. The volatility of the regime $H$ is three times that of the regime $L$ over all observed interest rates. Therefore, two different regimes are clearly confirmed in terms of distinct variance by the data. The effect of forcing a single-regime model to fit a two-regime process is to underestimate (overestimate) the variance at high (low) interest rates. Although the drift is not significantly different from zero in any regime, the stationarity of the process in each regime is induced by increasing volatility (CHLS). Not only the level effect but also the different variance and the high persistency of each state give a good explanation for the volatility clustering. Despite the difficulty of testing the hypothesis of one regime against two regimes, the substantially different variance and transition probability in two regimes strongly indicate the presence of a second regime.

The estimation results of the regime-switching Vasicek and CIR models exhibit very similar patterns to those of the general model across different assumptions on the number of regimes, the initial state probability
and the transition probability of the regime index. There are some interesting differences between my model and these two cases. My model gives a significantly better fit than both the Vasicek and CIR specifications. Given this and the indistinguishableness of the drift from zero, I conclude that the CIR and Vasicek regime-switching models do a poor job in matching volatility by underestimating variance. The fitted Vasicek model shows that the transition probability of the economy being in the same state in the next period is a decreasing (increasing) function of state variable in regime $L (H)$. This result is consistent with the one found in Ang and Bekaert (2002) who analyze a discrete version of Vasicek regime-switching model. This implies that misspecification can result in misleading outcomes, in particular, about the volatility and transition probabilities of the regime index. Qualitatively, the regime classifications are similar for all cases.

Another merit of a regime-switching diffusion model lies in bond pricing. The evolution of interest rates over time plays a key role in valuing and hedging fixed income securities, options and other derivatives and also in designing macroeconomic policies. In the one-factor model of term structure, all yields on the bonds with different maturities are perfectly correlated since they are calculated based on one factor. This perfect correlation is easily contradicted by the empirical evidence. The other problem is that the shape of the yield curve that can be obtained by the one-factor model is severely restricted. Because the regime index enters highly nonlinear way, the regime variable is capable of resolving the perfect correlation problem as well as producing many different shapes of term structure observed in the real world. As shown by Landén (2000) in her calibration study, even a simple Vasicek regime-switching model, in which only the constant term of the drift is subject to regime shifts, can generate a rich family of shapes of the yield curve. My estimation results suggest that a more general model is needed to avoid misspecification of the dynamics of interest rates. Hence, I derive a system of the partial differential equations of the bond price for my general model and it can be solved numerically using a finite difference method since the analytical solution to it is not available.

The remainder of this paper is organized as follows. The next section discusses the data and the model. Hamilton algorithm and the way to obtain approximate transition density function of the diffusion process are presented in Section 3. The results for my model and comparisons are contained in Section 4. Section 5 describes bond pricing of my model and Section 6 draws together our conclusions. Appendix I presents assumptions needed for the existence and uniqueness of the solution to the stochastic differential equation and for obtaining the approximate transition density function. Parameter restrictions to satisfy the assumptions are derived in Section II. The closed-form transition density functions of my diffusion model are given in the final section of the Appendix.
2 Data and Model

2.1 Data and Motivation

Weekly three month Treasury Bill secondary market rates, from January 8, 1971 to December 26, 2003, are taken from the H-15 Federal Reserve Statistical Release (Selected Interest Rate Series) for the estimation. These interest rates are averages of 7 calendar days ending on Friday except when unavailable due to a holiday, in which case the Thursday rate is used. The data set is annualized using a 360-day year and has 1731 observations. The daily rates are the averages of the bid rates quoted on a bank discount basis by a sample of primary dealers who report to the Federal Reserve Bank of New York. The rates reported are based on quotes at the official close of the U.S. government securities market for each business day. The descriptive statistics of the data series are given in Table 1.

Plots of the data and weekly changes in the interest rates illustrate that the interest rate behaves quite differently in different time periods. It is clear from Figure 1 that large moves follow large moves and small moves follow small moves, which is a stylized phenomenon in many asset returns and is termed volatility clustering. Some periods of clustered large movements coincide with those during which significant historic events took place. Studying Figure 1, I conjecture that there will be at least two regimes in terms of the variability of interest rates. This motivates us to analyze the diffusion model in the context of regime-switching.

My data set covers periods during which a number of events caused strong shifts in the behaviour of the yield curve of interest rates. The 1973 energy crisis occurred due to the onset of an oil embargo by OPEC and oil prices continued increasing until 1975. The effect of the oil crisis lingered through the 1970s. The Vietnam War began in 1961 and ended in 1975. In the wake of 1979 Iranian Revolution, the 1979 oil crisis occurred but did not last long. During 1979-82, the Federal Reserve conducted "Monetary Experiment" by directing its policy away from targeting federal fund rate toward targeting non-borrowed reserve. During 1984, the Federal Reserve temporarily tightened its monetary policy, which caused an increase in interest rates. In late 1984 and throughout 1985, the interest rate decreased because of a monetary easing (DSY). On October 19, 1987, the Dow Jones Industrial Average fell 22.6%, the largest one-day decline in recorded stock market history. By mid-1988 the stock market had recovered, and the U.S. economy was largely unaffected by the crash.

A number of other events may also have caused high volatility of interest rates between January 1971 and December 2003. From late 1960s, U.S. economic growth was slowing down, which began to become apparent in the early 1970s. This was an era of stagflation - the simultaneous occurrence of recession and inflation. Additionally, the Gulf War started in August 1990 and it ended in early 1991. On December 31, 1991, the Union of Soviet Socialist Republics officially ceased to exist. On November 3, 1991 Bill Clinton
defeated George H. W. Bush in the U.S. presidential election. A decade long of the longest peacetime economic expansion in U.S. history began in March 1991. At the end of September 1998, the Long Term Capital Management fund had lost substantial amounts of investors’ equity capital. It is usually claimed that the September 11, 2001 attacks had immediate and far-ranging economic effects. In October 2001, the United States invaded Afghanistan as part of its "War on Terrorism" campaign. During the stock market downturn of 2002, there were sharp drops in stock prices in the United States and Europe. Finally, the 2003 invasion of Iraq began on March 19, 2003.¹

2.2 Model

The following univariate diffusion model has been estimated by CHLS, Durham (2002), Choi (2002) and Jones (2003) using different estimation methods.

\[
dr_t = (\alpha_{-1}r_t^{-1} + \alpha_0 + \alpha_1r_t + \alpha_2r_t^2)dt + \beta r_t^\rho dW_t
\]

where \( W_t \) is a standard Brownian motion. If both \( \alpha_{-1} \) and \( \alpha_2 \) are zero then the model collapsed to the CKLS model with a linear drift term where the strength of mean reversion is the same for all levels of the short rate. In this case, \( \alpha_1 \) determines the rate at which \( r_t \) returns to its long-run average value \(-\frac{\alpha_0}{\alpha_1}\). It seems more reasonable that mean reversion is stronger for low or high levels of \( r_t \). This drift specification of (1), is designed to capture the property of very little mean reversion while interest rate values remain in the middle part of its domain, and strong mean reversion at either end of the domain. In the model (1), the speed of returning \( r_t \) to its long run-average value is equal to the first derivative of the drift with respect to interest rate, \(-\frac{\alpha_{-1}}{r_t} + \alpha_1 + 2\alpha_2r_t\). Therefore, the model (1) can have different mean reversion speeds at different interest rates. The diffusion component of (1) represents the conditional volatility of the change in short rates as a function of the interest rate, which accounts for the heteroskedastic behavior of the changes in \( r_t \). Here, \( \rho \) is the elasticity of the volatility with respect to \( r_t \) and it measures the sensitivity of the volatility of the interest rate with respect to interest rate level. Nonlinearity in the volatility may guarantee stationarity of \( r_t \) which is called volatility-induced stationarity by CHLS. The remaining volatility after accounting for the level effect is attributed to the parameter \( \beta \). The drift and the volatility terms of the model (1) are general enough to encompasses most of the single-factor models in literature such as Vasicek (\( \alpha_{-1} = \alpha_2 = \rho = 0 \)), CIR (\( \alpha_{-1} = \alpha_2 = 0 \) and \( \rho = 1/2 \)) and CKLS (\( \alpha_{-1} = \alpha_2 = 0 \)) among others.

We generalize the model (1) in two ways. One way is to add the third order term to the drift function to see if specifying more flexible drift can help answer the question of whether the interest rate reverts to its constant long-run mean, and if so, how strong it is at different level of rates. The other way is to allow

¹Most of these U.S. history except the one cited separately are obtained from the encyclopedia Wikipedia, http://en.wikipedia.org/wiki/Main_Page.
the parameters to change to incorporate the possible regimes shifts in the movements of the data. Following most of the previous works on regime-switching models, I assume that there are two regimes. My model assumes that the dynamics of short rate is described by a continuous time regime-switching Markov process:

\[ dr_t = \mu (r_t, s_t; \theta) \, dt + \sigma (r_t, s_t; \theta) \, dW_t \]  

where \( \mu (r_t, s_t; \theta) = \alpha_{-1} s_t r_t^{-1} + \alpha_0 s_t + \alpha_1 s_t r_t + \alpha_2 s_t r_t^2 + \alpha_3 s_t r_t^3 \), \( \sigma (r_t, s_t; \theta) = \beta_{s_t} r_t^{\rho_{s_t}} \) and the regime index \( s_t \) follows a continuous time first order Markov chain with two states. Motivated by the behavior of the change in interest rates, I let low (L) and high (H) volatility regimes characterize two different economic environments. Thus, the short rate can evolve in a regime with a low or a high volatility at any given moment. At all times, the true state of the economy is unobservable to econometricians and it has to be inferred. The Markov chain model is useful in capturing not only a permanent regime change but also a short-lived regime shift of the process.

The regime-switching framework allows the conditional mean and variance of the change in the interest rates to be different depending on the state of the economy. So, the regime-switching model can help explain some features of data or unresolved issues of short-term interest rate models. First, the nonlinearity of drift is an inconclusive issue in the literature of univariate diffusion model for the short rate. Durham (2002) and Choi (2002) adopted the ML estimation method to estimate and compare various models using three different sets of interest rate data. In Durham (2002), tests based on likelihood constructed by simulation reveal that a constant term is more preferable for the drift. Choi (2002) approximates true but unknown transition density function analytically and concludes that no drift model cannot be rejected in favor of a model with more flexible drift function. There are also mixed results regarding the nonlinear behavior of the drift in nonparametric or semiparametric papers. Aït-Sahalia (1996a) and Stanton (1997) find evidence of nonlinearity in the drift. On the other hand, Pritsker (1998) criticizes that the specification test developed by Aït-Sahalia (1996b) rejects true models too frequently. He argues that Aït-Sahalia (1996b) treats the data as if it were independently and identically distributed even if interest rate data is highly persistent, which is an important factor in small samples. Chapman and Pearson (2000) show that the methods proposed by Aït-Sahalia (1996b) and Stanton (1997) have a tendency to find evidence of nonlinearity in the drift even though the true drift is linear. This bias is due to the fact that there is only a few extreme data, which make the estimates associated with the strength of mean reversion imprecise. Using the nonparametric generalized likelihood ratio test, Fan and Zhang (2003) cannot find strong evidence against linear drift. Recently, Kristensen (2003) reports very strong mean reversion of the drift by nonparametrically estimate while keeping the diffusion term parametrized. Using the MCMC method, Jones (2003) concludes that the role of the prior distribution is critical in finding nonlinear drift. In particular, the drift of the regime H is expected to show nonlinear behavior. Second, the regime-switching specification is able to explain the volatility clustering better than a single-regime model. When interest rates are persistent, the dependence of
the volatility on the level of interest rate (so called “level effect”) can give rise to volatility clustering during periods of high interest rates. In addition, the strong persistence of the regime and distinct volatilities across two regimes are capable of clustering the periods of high volatility.

Two factors describe the behavior of interest rates in my model. While the continuously evolving short rates catch frequently arriving information, the regime variable changes discretely, capturing infrequent but sudden regime-shift, which can last for long or short time periods. It is the regime index that has the capacity to resolve the problems of single-factor models that the yields with different time to maturity are perfectly correlated and the shapes of term structure are quite restricted.

2.2.1 Transition matrix of the continuous time Markov chain with two states

The continuous time two state Markov chain with the conservative infinitesimal matrix

$$Q = \begin{pmatrix} q_{LL} & q_{HL} \\ q_{HL} & q_{HH} \end{pmatrix} = \begin{pmatrix} -q_{HL} & q_{HL} \\ q_{HL} & -q_{HL} \end{pmatrix}$$

is assumed to govern the shifts between two regimes. The intensity parameter $q_{ij}$ is the rate of the probability at which the process switches from the state $i$ to the state $j$ as time goes to zero and $q_{ij} > 0$ for $i \neq j$. The corresponding transition matrix is

$$P^\Delta = \frac{1}{q_{HL} + q_{LL}} \begin{pmatrix} q_{HL} + q_{HL}e^{-\Delta(q_{HL}+q_{HH})} & q_{HL}(1 - e^{-\Delta(q_{HL}+q_{HH})}) \\ q_{HL}(1 - e^{-\Delta(q_{HL}+q_{HH})}) & q_{HL} + q_{HL}e^{-\Delta(q_{HL}+q_{HH})} \end{pmatrix}$$

where $\Delta$ is the time interval between two consecutive observations. Hence, the transition probabilities depend not only on the intensity but also on $\Delta$. Furthermore, $(\pi_L, \pi_H) = \left(\frac{q_{HL}}{q_{HL}+q_{LL}}, \frac{q_{HL}}{q_{HL}+q_{LL}}\right)$ are the unconditional probabilities that the above Markov chain will be in state $L$ and $H$, respectively at any time. $\Delta$ is small enough to say that at most one regime shift can occur every seventh day. Since $\Delta$ is constant, i.e. $\frac{1}{52}$ in my data set, for convenience’s sake, we can reparametrize $P^\Delta$ as

$$P = \begin{pmatrix} p_{LL} & p_{HL} \\ p_{HL} & p_{HH} \end{pmatrix}$$

(3)

The transition probabilities, $p_{ij} = P(s_t = j|s_{t-1} = i), i, j = L, H$ are of greater interest than the intensity parameters, so I will focus on $P$ rather than $Q$. $q_{ij}$‘s can be recovered by

$$q_{HL} = -\frac{(p_{LL} - 1)}{(p_{LL} + p_{HH} - 2)} \ln \left(\frac{p_{LL} + p_{HH} - 1}{\Delta}\right) \text{ and } q_{HL} = -\frac{(p_{HH} - 1)}{(p_{LL} + p_{HH} - 2)} \ln \left(\frac{p_{LL} + p_{HH} - 1}{\Delta}\right)$$

 Representing stationary distribution of the Markov chain in terms of $p_{LL}$ and $p_{HH}$, we get

$$(\pi_L, \pi_H) = \left(\frac{1 - p_{HH}}{2 - p_{LL} - p_{HH}}, \frac{1 - p_{LL}}{2 - p_{LL} - p_{HH}}\right)$$

(4)
which is, not surprisingly, the same as the unconditional probabilities of a discrete time Markov chain with transition matrix (3).

I model the transition probabilities, that is, the intensity parameters as time-varying by making them depend on the state variable \( r_t \). Any parameterizations of state variable can be used as long as the transition probabilities are bounded between zero and one. Therefore, we can allow the probability of switching from one regime to the other to vary with the interest rate by setting

\[
\begin{align*}
  p_{LL} &= P(s_t = L|s_{t-1} = L) = F(c_L + d_L r_{t-1}) \\
  p_{LH} &= P(s_t = H|s_{t-1} = L) = 1 - F(c_L + d_L r_{t-1}) \\
  p_{HH} &= P(s_t = H|s_{t-1} = H) = F(c_H + d_H r_{t-1}) \\
  p_{HL} &= P(s_t = L|s_{t-1} = H) = 1 - F(c_L + d_L r_{t-1}).
\end{align*}
\]

Following the literature, both the cumulative Normal distribution function, \( \Phi(x) \) (Gray (1996)) and a logistic function, \( \frac{e^x}{1+e^x} \) (Diebold, Lee, and Weinbach (1994) and DSY) are used for \( F(x) \). If \( d_i \) is positive (negative), then the probability of the economy staying in the same regime \( i \) in the next period increases (decreases) as interest rate increases. Moreover, when \( d_L = d_H = 0 \), the transition probabilities become constant and the time-varying transition matrix is reduced to the time-invariant transition matrix.

### 2.2.2 Initial state distribution

When constructing the likelihood function of the regime-switching models, I treat the unconditional probability of the initial state, \( P(s_1 = L) \) as an additional parameter, \( p \) to be estimated. For the model of time-invariant transition matrix, the unconditional probability (4) is also considered for the initial state probability.

According to the number of regimes, the starting probability and the transition probability matrix specifications, I will estimate five different models: R1 (single-regime), R2-1 (two regimes, time-invariant transition matrix, unconditional probability for the probability of the initial state), R2-2 (two regimes, time-invariant transition matrix, additional parameter for the probability of the initial state), R2TVTP-1 (two regimes, time-varying transition matrix with standard Normal cumulative density function, additional parameter for the probability of the initial state), and R2TVTP-2 (two regimes, time-varying transition matrix with a logistic function, additional parameter for the probability of the initial state)
3 Maximum Likelihood Estimation Method

3.1 Transition Density Function

In each regime, the interest rate follows a stochastic differential equation (SDE) which is a Markov process. And if we know the transition density function of the diffusion process, the log-likelihood function of the regime-switching Markov process, $r_t$ can be calculated by using the algorithm developed by Hamilton (1989) as we will see below. As is often the case with most diffusion processes with a few exceptions of Vasicek, CIR and Black and Scholes (1973) models, the explicit true transition density function of my diffusion model is unknown. Aït-Sahalia (2002a) suggests an approximation of the true but unknown transition density function. This approximation of an arbitrary univariate diffusion process seems to work very well.

Within each regime $r_t$ follows (2) without regime variable. For this section, we can ignore the regime index of the model (2) because we merely need to find the transition density function $p(r_{t+\Delta}|r_t, s_t)$ for later use in constructing the likelihood function by employing the Hamilton algorithm. As a first step to obtain the closed-form approximate density function, we need to transform the process $r_t$ into a unit diffusion process $Y_t$ by using the equation below

$$Y_t \equiv \gamma (r_t; \theta) = \int^{r_t} \frac{1}{\sigma (u; \theta)} du.$$ (5)

Through this transformation, we can standardize $r_t$ to the following unit diffusion process $Y_t$ by applying Ito's Lemma.

$$dY_t = \mu_Y (Y_t; \theta) dt + dW_t$$

where

$$\mu_Y (y; \theta) = \frac{\mu (\gamma^{-1} (y; \theta); \theta)}{\sigma (\gamma^{-1} (y; \theta); \theta)} - \frac{1}{2} \frac{\partial \sigma}{\partial x} (\gamma^{-1} (y; \theta); \theta)$$ (6)

The first transformation makes the tail behavior of $p_Y$ close to Gaussian by controlling the tail behavior of the process $r_t$. However, $p_Y$ gets peaked around the conditional value $y_0$ when the time difference between two consecutive observations, $\Delta$ becomes small. Thus, we need the second transformation from $Y$ to $Z$

$$z \equiv \frac{Y - y_0}{\sqrt{\Delta}}$$

This makes the transition density of $Z_t, p_Z$ very close to the standard Normal and we can Hermite-expand $p_Z$ around it as follows:

$$p_Z^{(j)} (\Delta, z|y_0; \theta) \equiv \phi (z) \sum_{j=0}^{J} \eta_Z^{(j)} (\Delta, y_0; \theta) H_j (z)$$ (7)

where $\phi (z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$, standard normal density function. Hermite polynomials $H_j (z)$'s are orthogonal base and defined by $H_j (z) \equiv \phi (z)^{-1} \frac{d^j}{dz^j} \phi (z)$, for example $H_0 (z) = 1, H_1 (z) = -z, H_2 (z) = z^2 - 1,
That is to say, we have to calculate

Using the orthogonality of Hermite polynomials we can obtain the coefficients $\eta^{(j)}(\Delta, y_0; \theta)$ of Hermite expansion as follows:

\[
\eta^{(j)}(\Delta, y_0; \theta) = (1/j!) \int_{-\infty}^{\infty} H_j(z) p_{zY}(\Delta, z|y_0; \theta) \, dz
\]

\[
= (1/j!) \int_{-\infty}^{\infty} H_j(z) \sqrt{\Delta} p_Y(\Delta, \sqrt{\Delta} z + y_0|y_0; \theta) \, dz
\]

\[
= (1/j!) \int_{-\infty}^{\infty} H_j \left( \frac{y-y_0}{\sqrt{\Delta}} \right) p_Y(\Delta, y|y_0; \theta) \, dz
\]

\[
= (1/j!) E \left[ H_j \left( \frac{Y_s - y_0}{\sqrt{\Delta}} \right) \mid Y_s = y_0; \theta \right]
\]

The conditional expectation can be approximated up to any order by Taylor expansion using the infinitesimal operator,\(^2\) $A_Y$ of the diffusion $Y_t$ defined by

\[
A_Y \circ f(\Delta, y, y_0; \theta) = \mu_Y(y; \theta) \frac{\partial f(\Delta, y, y_0; \theta)}{\partial y} + \frac{1}{2} \frac{\partial^2 f(\Delta, y, y_0; \theta)}{\partial y^2}
\]

Hence, for any infinitely differentiable function $f$

\[
E[f(\Delta, Y_t, Y_s) \mid Y_s = y_0; \theta] = \sum_{i=0}^{I} \frac{\Delta^i}{i!} A^i_Y \circ f(y, y_0; \theta)|_{y=y_0} + O(\Delta^{I+1})
\]

In this way, $p_Y$ is approximated as

\[
p_Y^{(J, I)}(\Delta, y, t|y_0, s; \theta) = \Delta^{-\frac{J}{2}} \phi \left( \frac{y-y_0}{\Delta} \right) \left\{ \sum_{j=0}^{J} \frac{\Delta^j}{j!} A^j_Y \circ H_j \left( \frac{y-y_0}{\sqrt{\Delta}} \right) \bigg|_{y=y_0} \right\} H_j \left( \frac{y-y_0}{\sqrt{\Delta}} \right)
\]

That is to say, we have to calculate $\eta^{(j)}(\Delta, y_0; \theta)$'s up to $J$'th order and truncate (7) at $J$ in order to find the approximation, $p_Y^{(J, I)}$. Alternatively, we can rearrange all terms in (8) according to increasing powers of

\(^2\)In general, if univariate diffusion process $X_t$ follows

\[
X_t = \mu(X_t, t; \theta) \, dt + \sigma(X_t, t; \theta) \, dW_t
\]

then

\[
\lim_{\Delta \to 0} \frac{U_{\Delta}g(X_t, t) - U_0g(X_t, t)}{\Delta} = \lim_{\Delta \to 0} \frac{E[g(X_t, t) \mid X_s = x_0] - g(x_0, s)}{\Delta} = A_X \circ g(x, t)
\]

\[
= \frac{\partial g(x, t)}{\partial t} + \mu(x, t; \theta) \frac{\partial g(x, t)}{\partial x} + \frac{1}{2} \sigma(x, t; \theta)^2 \frac{\partial^2 g(x, t)}{\partial x^2} \bigg|_{x=x_0, t=s}
\]

11
Δ. The resultant K-th order expansion of the approximate density function to \( p_Y \) is denoted as \( p_Y^{(K)} \) and given by

\[
p_Y^{(K)} (\Delta, y|y_0; \theta) = \Delta^{-\frac{k}{2}} \phi \left( \frac{y - y_0}{\Delta} \right) \exp \left( \int_{y_0}^{y} \mu_Y (\omega; \theta) \, d\omega \right) \sum_{k=0}^{K} c_Y^{(k)} (y|y_0; \theta) \frac{\Delta^k}{k!}
\]

where \( c_Y^{(0)} (y|y_0; \theta) = 1 \). The other coefficients are determined recursively by

\[
c_Y^{(k)} (y|y_0; \theta) = k (y - y_0)^{-k} \int_{y_0}^{y} (\omega - y_0)^{k-1} \left\{ \lambda_Y (\omega; \theta) c_Y^{(k-1)} (\omega|y_0; \theta) + \frac{1}{2} \frac{\partial^2 c_Y^{(k-1)} (\omega|y_0; \theta)}{\partial \omega^2} \right\} \, d\omega
\]

with \( \lambda_Y (y; \theta) = -\frac{1}{2} \left[ \mu_Y (y; \theta)^2 + \frac{\partial \mu_Y (y; \theta)}{\partial y} \right] \). As a matter of fact, \( K = 1 \) or \( 2 \) at most gives very precise density function with the usual values of \( \Delta \) for data sets which researchers use in practice (Aït-Sahalia (1999)). In this paper, I use a weekly data set so \( \Delta = 1/52 \) and the approximate density function with \( K = 1 \).

For computing the likelihood of my model, it is convenient to make \( Y_t \) nonnegative by adopting minus the integral in equation (5) for \( \rho > 1 \). Depending on the value of the elasticity, \( \rho \) I end up with different transformations and domains of the variable \( Y_t \), \( D_Y \) summarized in (10). Note that \( D_r = (0, +\infty) \).

\[ Y_t \equiv \gamma (r_t; \theta) = \begin{cases} 
\frac{1}{\beta} r_t & \text{if } \rho = 0 \\
\frac{r_t^{1-\rho}}{\beta(1-\rho)} & \text{if } 0 < \rho < 1 \\
r_t & \text{if } \rho = 1 \\
\frac{r_t^{\rho-1}}{\beta(\rho-1)} & \text{if } 1 < \rho 
\end{cases} \]

By Ito’s formula we get the following unit diffusion processes

\[
dY_t = \mu_Y (Y_t; \theta) \, dt + dW_t \quad \text{and} \quad D_Y = (0, +\infty) \quad \text{if } \rho = 0
\]

\[
dY_t = \mu_Y (Y_t; \theta) \, dt + dW_t \quad \text{and} \quad D_Y = (0, +\infty) \quad \text{if } 0 < \rho < 1
\]

\[
dY_t = \mu_Y (Y_t; \theta) \, dt + dW_t \quad \text{and} \quad D_Y = (-\infty, +\infty) \quad \text{if } \rho = 1
\]

\[
dY_t = \mu_Y (Y_t; \theta) \, dt - dW_t \quad \text{and} \quad D_Y = (0, +\infty) \quad \text{if } 1 < \rho
\]

So, \( \sigma_Y^2 (Y_t; \theta) = 1 \) but \( \mu_Y (Y_t; \theta) \)'s are different and obtained by using (6) in the appendix for all cases.

Replacing \( \mu_Y (y; \theta) \) in (9) with the one in (11) and calculating \( c_Y^{(k)} (y|y_0; \theta) \)'s, the transition density function \( p_Y^{(K)} (y_{t+\Delta}|y_t; \theta) \) can be found for all cases. Once we have \( p_Y^{(K)} \), we can recover \( p^{(K)} (r_{t+\Delta}|r_t; \theta) \) by the change of variable from \( Y \) to \( r \) using Jacobian.

\[
p^{(K)} (r_{t+\Delta}|r_t; \theta) = \frac{\beta r_t^\rho \, p_Y^{(K)} (\gamma (r_{t+\Delta}; \theta) | \gamma (r_t; \theta); \theta)}{\beta r_t^\rho}
\]

Conditions for the existence and the uniqueness of a solution to the SDE and for obtaining the sequence of approximate density function are presented in appendix I. In order for my model to satisfy those conditions, certain restrictions on parameters have to be imposed. Construction of the approximation to the true density function does not hinge upon the stationarity of the process. Usually, stationarity is assumed for the purpose.
of maximum likelihood estimation. As we will see below, however, the resultant parameter space I get after imposing the constraints on my model ensure stationarity of the process. In appendix III, the explicit first order density expression for \( p_Y(y_{t+\Delta}|y_t; \theta) \) and restrictions needed for the parameter space \( \Theta \) for all cases in (11) are given. A list of parameter restrictions for many nested models of (1) and their approximate density expansions are found in Choi (2002).

Let us consider some features of the resultant parameter space of my model. First, the elasticity of volatility, \( \rho \) is nonnegative due to Assumption 2 and it cannot exceed 2 to satisfy the condition \( \lim_{y \to 0} \lambda_Y(y; \theta) < +\infty \) in Assumption 3. This result is, of course, weaker than usual linear growth conditions, which can be found in Karatzas and Shreve (1991), for the existence of a solution to a general stochastic differential equation. Second, for all cases of (11) except the two cases \( \rho = 0, 2 \), the left boundary and right boundary conditions for \( \mu_Y(y; \theta) \) require , which also guarantee stationarity of the process (Aït-Sahalia (1996b)). For \( \rho = 0 \) and \( \rho = 2 \), we need to impose \( \alpha_3 \leq 0 \) only and \( \alpha_{-1} > 0 \) and \( \alpha_3 < 0 \), respectively. Appendix II demonstrates how I achieve these restrictions on the parameter space.

### 3.2 Hamilton Algorithm

Unlike Hamilton (1989), the parameters of conditional density function \( r_{t+\Delta} \) depend not on the current regime \( s_{t+\Delta} \) but on the previous regime \( s_t \). The same timing convention was employed by DSY and DKSS. As \( \Delta \) goes to zero, both specifications are equivalent\(^3\). As a matter of convenience, let’s define a new state variable \( s^*_t \) as follows:

\[
\begin{align*}
    s^*_t &= 1 \text{ if } s_{t-\Delta} = L \text{ and } s_t = L \\
    s^*_t &= 2 \text{ if } s_{t-\Delta} = L \text{ and } s_t = H \\
    s^*_t &= 3 \text{ if } s_{t-\Delta} = H \text{ and } s_t = L \\
    s^*_t &= 4 \text{ if } s_{t-\Delta} = H \text{ and } s_t = H 
\end{align*}
\]

Then \( s^*_t \) follows a four-state continuous time Markov chain with the following transition matrix

\[
P^* = \begin{pmatrix}
    p_{LL} & 0 & p_{LL} & 0 \\
    p_{LH} & 0 & p_{LH} & 0 \\
    0 & p_{HL} & 0 & p_{HL} \\
    0 & p_{HH} & 0 & p_{HH}
\end{pmatrix}
\]

\(^3\)In practice, either convention leads to almost the same likelihood value and the estimation results.
Let
\[ \hat{\xi}_{t+\Delta|t} = \begin{pmatrix} P(s_{t+\Delta}^* = 1|I_t; \theta) \\ P(s_{t+\Delta}^* = 2|I_t; \theta) \\ P(s_{t+\Delta}^* = 3|I_t; \theta) \\ P(s_{t+\Delta}^* = 4|I_t; \theta) \end{pmatrix} \]
and \( \eta_{t+\Delta} = \begin{pmatrix} p(r_{t+\Delta}|s_{t+\Delta}^* = 1, I_t; \theta) \\ p(r_{t+\Delta}|s_{t+\Delta}^* = 2, I_t; \theta) \\ p(r_{t+\Delta}|s_{t+\Delta}^* = 3, I_t; \theta) \\ p(r_{t+\Delta}|s_{t+\Delta}^* = 4, I_t; \theta) \end{pmatrix} \)
where \( I_t = \{ r_r | r \leq t \} \) is the information set which consists of data through time \( t \). At step \( t+\Delta \), the input of the algorithm is \( \hat{\xi}_{t+\Delta|t} \) whose elements are the inferences about the value of \( s_{t+\Delta}^* \) based on \( I_t \) and knowledge of the population parameter vector \( \theta \). \( \eta_{t+\Delta} \) contains the conditional density functions \( p(r_{t+\Delta}|r_t, s_t = L; \theta) \) by exploiting the Markovian property of the diffusion process, \( r_t \) and the timing assumption that the parameters of conditional density function \( r_{t+\Delta} \) depend on \( s_t \). Recall that the transition density function \( p(r_{t+\Delta}|r_t, s_t = L; \theta) \) is not known in general, so we have to replace this with the approximate density function obtained in Section 3.1.

The optimal inference about \( s_{t+\Delta}^* \) and \( \eta_{t+\Delta} \) can be computed by the element by element multiplication of \( \hat{\xi}_{t+\Delta|t} \) and \( \eta_{t+\Delta} \), that is
\[ \hat{\xi}_{t+\Delta|t} \odot \eta_{t+\Delta} = \begin{pmatrix} P(s_{t+\Delta}^* = 1|I_t; \theta) p(r_{t+\Delta}|s_{t+\Delta}^* = 1, I_t; \theta) \\ P(s_{t+\Delta}^* = 2|I_t; \theta) p(r_{t+\Delta}|s_{t+\Delta}^* = 2, I_t; \theta) \\ P(s_{t+\Delta}^* = 3|I_t; \theta) p(r_{t+\Delta}|s_{t+\Delta}^* = 3, I_t; \theta) \\ P(s_{t+\Delta}^* = 4|I_t; \theta) p(r_{t+\Delta}|s_{t+\Delta}^* = 4, I_t; \theta) \end{pmatrix} \]
Then, summing \( \hat{\xi}_{t+\Delta|t} \odot \eta_{t+\Delta} \) over the variable \( s_{t+\Delta}^* \), we get the transition density function
\[ p(r_{t+\Delta}|I_t; \theta) = \sum_{j=1}^{4} p(r_{t+\Delta}, s_{t+\Delta}^* = j|I_t; \theta). \tag{14} \]
The optimal inference about \( s_{t+\Delta}^* \) denoted as \( \hat{s}_{t+\Delta|t+\Delta} \) is calculated by
\[ \hat{s}_{t+\Delta|t+\Delta} = \frac{\hat{\xi}_{t+\Delta|t} \odot \eta_{t+\Delta}}{p(r_{t+\Delta}|I_t; \theta)} \]
and the input of the next step, \( \hat{\xi}_{t+2\Delta|t+\Delta} \) is updated by premultiplying the transition matrix \( P^* \) by \( \hat{\xi}_{t+\Delta|t+\Delta} \).

By Bayes’ rule, the log-likelihood function is of the form
\[ l_T(\theta) = \frac{1}{n} \sum_{i=1}^{n} \ln [ p(r_{i\Delta}|I_{i-1\Delta}; \theta) ] \tag{16} \]
where \( r_t \)'s are observed discretely at dates \( \{ t = i\Delta | i = 0, 1, \cdots, n \} \).
The initial state probabilities are \( (\pi_1, \pi_2, \pi_3, \pi_4) = (p_{LL}, p_{LH}, p_{HL}, p_{HH} (1-p), p_{HH} (1-p)) \) when \( P(s_1 = L) = p \). Given these starting values, iterating the above procedure and collecting the conditional density function (14) of each step, we can maximize the log-likelihood function (16) to find the ML estimates. 

The estimation results do not tell us exactly which state the economy was in at each time point. We can only infer the probability of the economy being in each regime based on the observations of interest rates. The \( i \)-th entry of (15), known as the filtered probability, is the inferred probability of \( s_t^* \) being \( i \) using information up to time \( t \). The inference can be made using all information available. I also calculate the smoothed probability of the economy being in each state at time \( t \), \( P(s_t^* = i|I_T; \theta) \), where \( T = n\Delta \) is the last time point of the data. The state of \( s_t \) rather than \( s_t^* \) is of more interest and it is apparent from (13) that 

\[
P(s_t = L|I_t; \theta) = P(s_t^* = 1|I_t; \theta) + P(s_t^* = 3|I_t; \theta)
\]

\[
P(s_t = H|I_t; \theta) = P(s_t^* = 2|I_t; \theta) + P(s_t^* = 4|I_t; \theta)
\]

\( P(s_t = L|I_T; \theta) \) and \( P(s_t = H|I_T; \theta) \) can be found in the same way. When the transition probability varies with interest rates, the above recursive procedure does not change and we only need to replace the elements of \( P^* \) with the corresponding time dependent probabilities.

## 4 Estimation Results

### 4.1 Main Models

Table 2 provides ML estimates of parameters with standard errors in parentheses for the five different models I consider. The standard errors are calculated by taking the inverse of the sum of the outer product of the score functions evaluated at the ML estimates for each observation. The first column of Table 2 presents ML estimation results for the single-regime model, R1. The outcomes of the ML estimation for the four regime-switching models, depending on the specification of the transition probability and the starting value of the algorithm, R2-1 (time-constant transition matrix, unconditional probability for the starting value), R2-2 (time-invariant transition matrix, additional parameter for the starting value), R2TVTP-1 (time-varying transition matrix with the Normal cumulative density function, additional parameter for the starting value), and R2TVTP-2 (time-varying transition matrix with a logistic function, additional parameter for the starting value), are reported in turn in the rest of the columns of Table 2. We imposed \( \alpha_3 = 0 \) for the model R1 and \( p = 0 \) for the regime-switching models whose initial state probability is an unknown parameter because these parameters converge to zero when maximizing the likelihood function. The asterisk by the parameter estimate

\(^4\)I use Matlab built in optimizer fmincon to find the optimum of the objective function.
indicates that it is different from zero at the 5% significance level. In the last three rows of Table 2, Wald statistics are calculated to jointly test if each of the corresponding parameter of the two regimes are equal for three cases: the drift \((\alpha_{-1L} = \alpha_{-1H}, \alpha_{0L} = \alpha_{0H}, \alpha_{1L} = \alpha_{1H}, \alpha_{2L} = \alpha_{2H} \text{ and } \alpha_{3L} = \alpha_{3H})\), diffusion functions \((\beta_L = \beta_H \text{ and } \rho_L = \rho_H)\) and the transition probabilities \((P(s_{t+\Delta} = L|s_t = L) = P(s_{t+\Delta} = H|s_t = H))\), separately.

Traditional test statistics cannot be used to test whether there is one regime or two regimes because the parameters related to the second state of the economy are not identified under the null of no regime-switching. For example, given \(P(s_1 = L) = 1\), if the null hypothesis is \(p_{LL} = 1\), then \(\alpha_{-1H}, \alpha_{0H}, \alpha_{1H}, \alpha_{2H}, \alpha_{3H}, \beta_H, \rho_H\) and \(p_{HH}\) are unidentified under the null. In this case, standard asymptotic distribution theory cannot be applied and the usual test statistics are not \(\chi^2\) distributed. Hansen (1992, 1996) proposed a standardized likelihood ratio statistic to conduct the test in such situations. The idea is to concentrate the identified parameters out of the likelihood function and take the supremum of the concentrated likelihood function over the possible unidentified parameters after standardizing it. The optimization is done by using a grid search across plausible values for the nuisance parameters. In practice, concentrating out the likelihood function for each grid point is computationally extremely expensive unless the model is very simple. In my model, however, there are eight parameters not identified under the null of a single-regime. We cannot even find convincing bounds for the nuisance parameters, not to mention being unable to do a computationally very intensive grid search over the eight dimensional parameter space. Therefore, I will resort to the ordinary likelihood ratio test (LRT) to compare models with different number of states as Hamilton and Susmel (1996) and Gray (1996) do. The huge increase in the log-likelihood value, when we move from the model R1 to the model R2-1 in Table 2, implies an LRT statistic of 758.190. Ignoring the fact that the LRT statistic does not have \(\chi^2(8)\) distribution under the null of one regime, the model R1 is rejected in favor of the model R2-1 even at the 0.5% significance level because the corresponding critical value is 23.59. Comparisons of the model R1 to other two-regime models also reveal strong evidence for the existence of the second regime.

On the other hand, LRT can be conducted safely to compare regime-switching models. The model R2-2 nests R2-1 parametrization and two regime-switching models with time-varying transition probabilities encompass the specification R2-2. In the case of comparing two time-constant transition probability models, R2-1 and R2-2, the LRT cannot reject the hypothesis of the model R2-1 since the LRT statistic, which is \(\chi^2_1\) distributed, is 2.518 and its p-value is 0.113. Thus, this shows that there is not much difference between adding a parameter and using unconditional probabilities for the initial state distribution. The difference in the maximized likelihood between the models R2TVTP-1 and R2TVTP-2 is negligible. This indicates that the results of the regime-switching models with state-dependent transition matrix are not affected much by using different functional forms for the transition probabilities of the Markov chain. Comparing the models R2-2 and R2TVTP-2 (R2TVTP-1), the LRT statistic, which has \(\chi^2_2\) distribution under the null of constant transition probabilities, is 18.074 (18.368) and the corresponding p-value is 0.00012 (0.00010). So, there is a
strong evidence, in terms of a likelihood based test, that the transition matrix of the Markov chain is much better described by the state-dependent probabilities than by the time constant probabilities.

The estimates for the coefficients of the state variable in time-varying transition probabilities, $d_L$ and $d_H$ are positive and significant at any usual level. Hence, the probability of the process staying in the same regime in a week decreases (increases) as the magnitude of interest rates falls (increases). However, the estimates themselves are not very informative for the size of transition probability. The first top two panels of Figure 2 respectively depict the transition probabilities of shifting from the state $L$ to the state $L$ and from the regime $H$ to the regime $H$ evaluated at the ML estimates for the model R2TVTP-2 over the support of the data. It is apparent from the graphs that the probability of remaining in the same regime in a week is very high for both regimes at all interest rates in the data, which proves the high persistency of each state. That is to say, the probabilities of changing regimes are very low over the sample. One difference between the two panels is that the probability of staying in the regime $H$ gets close to 0.8 when the interest rate falls to 0.02, whereas the likelihood of being in the regime $L$ is above 0.9 for all observed interest rates. This 0.1 difference in the probabilities creates a big discrepancy in persistency. For instance, let $p_{LL} = 0.9$ and $p_{HH} = 0.8$, such that the transition matrix is

$$P = \begin{pmatrix} 0.9 & 0.2 \\ 0.1 & 0.8 \end{pmatrix}$$

then

$$P^4 \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0.7467 \\ 0.2533 \end{pmatrix} \text{ and } P^4 \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0.5066 \\ 0.4934 \end{pmatrix}.$$  

In four periods, the probability of remaining in the state $L$ is still above 0.5 but that of staying in the state $H$ falls below 0.5. As a matter of fact, once we begin in the state $L$, the likelihood of shifting to the state $H$ is always less than 1/3 in finite periods since the unconditional probabilities for this example are $(\pi_L, \pi_H) = (2/3, 1/3)$. The bottom two panels of Figure 2 plot $P(s_{t+\Delta} = L|s_t = L)$ and $P(s_{t+\Delta} = H|s_t = H)$ computed using ML estimates over the time series data. Each regime is very persistent over the sample. The model R2TVTP-1 has almost the same results as R2TVTP-2. In the cases of two-regime models with time-invariant transition probability, the estimates of $p_{LL}$ and $p_{HH}$ are very big and highly significant in both cases. Therefore, it is very likely that each regime remains in the same state in the next period. According to Wald test statistics in Table 2, for all two-regime models, $P(s_{t+\Delta} = L|s_t = L)$ and $P(s_{t+\Delta} = H|s_t = H)$ are significantly different from each other except in the case R2TVTP-2. This is just because $c_H$ is estimated with low precision but $d_H$ is much more important and significantly estimated than $c_H$.

We retrieve two kinds of inferred probabilities of each regime conditional on the information which prevails at time $t$ (filtered probability) and at time $T$ (smoothed probability). Figure 3 draws the filtered (top) and the smoothed (bottom) probabilities of the regime variable being in the regime $H$ over the sample for the model R2TVTP-2. Again, the inferred probabilities for the model R2TVTP-1 are virtually identical.
to those for the model R2TVTP-2. The state of an observation at time $t$ is classified as regime $i$ if the smoothed probability, $P(s_t = i | I_T) > 0.5$, following the convention in practice.

Some of the important incidents in U.S. history can be connected with most of the high variance periods. The U.S. economy suffered from high inflation rates from late 1960s until 1983. A high volatility regime in the beginning of my sample seems to be related with high inflation. We could not find any historic occasion for this except that there was stagflation during the early 1970s. The 1973-1975 period and subsequent states of $H$ coincide with the OPEC oil crisis and its lingering effects through the 1970s. In addition, the Vietnam War occurred between 1961 and 1975. The 1979-1982 high volatility regime corresponds to the period of what is known as the Federal Reserve Monetary Experiment. The nine successive weeks of state $H$ in 1987 is linked to the 1987 stock market crash. The beginning of some of the occasional high variance states from the end of 1991 to 1997 match the starting point of the decade long bull market. It is also coincident with the break up of the Soviet Union at the end of 1991. However I could not identify any special event for these high volatility periods. The seven consecutive weeks of regime $H$ in late 1990s coincide with the LTCM crisis in September 1998. There are four spikes in 2000s. The first one is associated with the September 11, 2001 attacks. The second spike has to do with the stock market downturn of 2002. The last two sharp increases match the outbreak of the 2003 invasion of Iraq.

The regime $L$ happens in the range of $[0.0084, 0.1065]$. However, the high volatility states occur during the periods of both high and low interest rates. The maximum (minimum) interest rate for the regime $H$ is 0.1676 (0.0092). In the earlier literature, the high volatility regime is mostly characterized only by high interest rates.

The division of regimes for the two models with time-constant transition probability is very close to each other. The filtered and smoothed probabilities of the specification R2-2 are illustrated in Figure 4. As is clear from Figure 3 and 4, regime classifications for both of the time-constant and time-varying cases are quite similar. One striking difference between those two is that only the model R2-2 assigns high volatility regime to the period of 1985. DSY provides an explanation for this period. During 1984, the Federal Reserve temporarily tightened monetary policy, which caused an increase in interest rates. In late 1984 throughout 1985, the interest rate decreased because of a monetary easing.

The wide spread belief that the regime $H$ is closely related to recessions can be checked in Figure 3 and 4. There are five NBER recession periods\(^5\) (shaded area) between January 1971 and December 2003 and four of them are identified as being in the regime $H$. There are also some periods of the state $H$ during expansion.

The estimated transition matrix of the model R2-2 and the sample mean of the transition probabilities

\(^5\)The official business cycles dates of the NBER can be found in www.nber.org/cycles.html.
evaluated at the ML estimates of the model R2TVTP-2 are

\[
\begin{pmatrix}
0.981 & 0.047 \\
0.019 & 0.953
\end{pmatrix}
\quad \text{and} \quad
\begin{pmatrix}
0.979 & 0.065 \\
0.021 & 0.935
\end{pmatrix}
\]

Correspondingly, the unconditional probabilities are

\[
\begin{pmatrix}
0.712 \\
0.288
\end{pmatrix}
\quad \text{and} \quad
\begin{pmatrix}
0.759 \\
0.241
\end{pmatrix}
\]

which implies that the economy was in regime \( L \) 71.2\% (75.9\%) of the time for the model R2-1 (R2TVTP-2). Overall, the economy spent more time in the regime \( L \) than in the state \( H \) and regime \( L \) is the prevailing regime.

\[
\begin{pmatrix}
0.704 \\
0.296
\end{pmatrix}
\quad \text{and} \quad
\begin{pmatrix}
0.711 \\
0.289
\end{pmatrix}
\]

are the sample mean of the smoothed probabilities of the economy being in each regime for these two models. Calculating the ratio of the economy being ascribed to the states \( L \) and \( H \) for each model according to the smoothed probabilities, we get respectively

\[
\begin{pmatrix}
0.710 \\
0.290
\end{pmatrix}
\quad \text{and} \quad
\begin{pmatrix}
0.724 \\
0.276
\end{pmatrix}
\]

The above three kinds of figures regarding the state \( L \) (equivalently \( H \)) are close across two models as well as different sorts of rates.

A new metric, Regime Classification Measure (RCM) suggested by Ang and Bekaert (2002a) is computed to compare the performance of different regime-switching models in identifying the regime over the sample. Define the RCM statistic for \( R \) regime case as

\[
RCM (R) = 100R \frac{1}{n} \sum_{t=1}^{n} \left( \prod_{i=1}^{R} p_{i,t} \right)
\]

where \( p_{i,t} = P(s_t = i|I_T) \), or alternatively any inferred probability can be used. The role of the constant 100\( R^R \) is to normalize the statistic to be between 0 and 100. A model that does a good (bad) job of distinguishing between regimes will make an inference about \( s_t \) being in a regime close to 1 or 0 (1/2). Therefore, based on the inferred probabilities of being in a particular regime, the closer the RCM value is to 0, the more perfect the regime classification of a model is. In my case, \( R = 2 \) and Table 2 presents the RCM for each one of the two-regime models. RCM decreases, but only slightly when we move from the second column (model R1) to the last column (model R2TVTP-2) of Table 2. Based on RCM, the benefit from using time-dependent transition probability is negligible. In terms of the allocation of regimes, state-dependence of the transition probability appears to be of little importance.
Using the ML estimates provided in Table 2, I drew the drift and the volatility terms over the region of the observed interest rates and the approximate conditional density functions given \( r_0 = 0.05 \) obtained in Section 3.1 for the models R1 and R2TVTP-2 in Figure 5. The 95% confidence band is obtained by the delta method\(^6\). The corresponding graphs of the models R2-1, R2-2 and R2TVTP-1 are virtually the same as those of the model R2TVTP-2. Thus, I omit them and compare only the cases R1 and R2TVTP-2.

All parameters of the drift term in the model R1 are significantly different from zero at the 0.05 level of significance. In contrast, none of the parameters of the drift function for the regime-switching models are significant at any usual significance level. The estimates in themselves, however, do not offer much information about the properties of the drift until we graph it. My more flexible drift component is expected to reveal the nonlinearity of the drift, particularly in the regime \( H \). So we can see very little mean reversion for moderate levels of interest rates in the regime \( L \) and a fairly high mean reverting behavior in the extremes of the data in the regime \( H \). Unfortunately, examination of the data does not provide evidence supporting this nonlinearity even for the model R1. Panels A and B of Figure 2 respectively depicts the drift function and its 95% confidence band of the models R1 and R2TVTP-2 using the ML estimates over the range of the interest rates in my data. The drift cannot be distinguished from zero in either case because the 95% confidence band includes the \( x \)-axis. Equivalently, restricting the drift to be zero in the model R2TVTP-2, the maximized likelihood of this constrained model is not much improved.\(^7\) Similarly, inaccurate estimation of the drift is obtained for the rest of regime-switching models. Therefore, I cannot find evidence of short rate’s reversion to a long-run level for any level of interest rates in either regime. This appears to be due to the lack of high and low interest rate data as pointed out by Chapman and Pearson (2000). Wald test cannot reject the null that the drift parameters of the two regimes are identical in any regime-shift models.

Unlike the drift, I can estimate not only all parameters determining the diffusion function but also the volatility term with very high precision for each model I consider. The entire estimates of the parameters related to the volatility part are significant at the level of 0.01. Moreover, the differential variances clearly separate two regimes in that the standard deviation of one regime is about three times that of the other regime and the confidence bands of the volatility functions of the two regimes do not overlap with each other in panel D of Figure 2. Not surprisingly, the Wald test also provides strong evidence against the hypothesis that the volatility parameters are equal across two regimes in all models. My classification of regimes into low variance state \( L \) and high variance state \( H \) is confirmed by the data.

Let us take a closer look at the results for volatility. While the estimate for \( \rho \) is 1.116 in the model R1,\(^6\) I also calculated the confidence band using bootstrap and the results are practically identical to those reported here.

\(^7\)The log-likelihood value of the model R2TVTP-2 with zero drift at the optimum is 8957.400, which implies that no drift R2TVTP-2 model cannot be rejected in favor of the model R2TVTP-2. Notice that the LRT statistic to test the no drift hypothesis is 9.726 and its p-value equals 0.465.
\( \hat{\rho}_L \) and \( \hat{\rho}_H \) are respectively 0.968 and 0.973 in all regime-switching models. Hence, most of the volatility differentiation between the two regimes is attributed to the coefficient, \( \beta \). Failing to account for the regime-shift in the data, single-regime model seems to overestimate the sensitivity of volatility to interest rates. These results are very similar to those in Bliss and Smith (1998) who reexamined CKLS after taking the temporary structural shift period of the Federal Reserve Experiment into account. Using the Chow-type test, they found strong evidence of structural break and the GMM estimate of the elasticity dropped from 1.5 of CKLS to 0.948. \( \beta \) also decreased from 1.30 to 0.27, which is closer to my estimates \( \hat{\beta}_L = 0.11 \) and \( \hat{\beta}_H = 0.33 \).

The bottom two panels of Figure 5 provide the estimated approximate transition density functions of the SDEs for the models R1 and R2TVTP-2 conditional on 0.05. At around the given rate, the density function of the state \( L \) is highly peaked but that of the regime \( H \) is flattened out.

Although there are some high volatility periods at low interest rates, the regime \( L \) is dominant at observations less than 0.1065, which is the maximum interest rate of the state \( L \) for the model R2TVTP-2. At high interest rates, the regime \( H \) prevails. What the single-regime model does can be understood as mixing two regimes across the support of the interest rate. Graphically, forcing a single-regime model to fit a two-regime process has an effect of averaging two regimes. As a result, the single-regime model overestimates (underestimates) the variance of the process at low (high) interest rates. The effect of the distribution of the regime \( H \) is to level the concentrated regime \( L \) distribution in mixing two regimes.

The estimation results imply that the specification of the volatility function rather than that of drift function plays a key role in modeling the dynamics of interest rates in each regime. Bandi and Phillips (2003) present an explanation why the diffusion term is estimated more precisely than the drift term. Although I cannot find the evidence against a non-zero drift within each regime, the diffusion processes of both regimes are still stationary, which is induced by increasing volatility. My result corresponds to CHLS’s one special case of volatility induced stationarity where drift is non-negative constant and the volatility elasticity exceeds one half. Each regime is very persistent and the process is more likely generated from the regime \( H \) as interest rate increases. Then at high interest rates, because the drift is zero, the process can explode and show random walk behavior, however, the increased volatility prevents this. Not only the level effect but also the high persistency of each regime and distinct variances of the two regimes can explain the volatility clustering observed in my data set.

### 4.2 Model Comparisons

For the purpose of comparison, I estimate the same five kinds of models presented in Section 2.2 using two different parametrizations for the SDE studied by Naik and Lee (1998) (Vasicek) and DKSS (CIR). The former assumes that only the constant volatility parameter is regime-dependent. They use bond pricing
formula which they derived to invert long term yields of monthly data to estimate the parameters. The latter
uses quarterly data and discretizes the diffusion process to get the approximate transition density function.
Both of them assume a time-constant transition matrix. Although the true transition density functions of
Vasicek and CIR models are known, I will use the first order approximations to them for compatibility with
my model. In fact, the first order approximate density function is remarkably close to the true density
function, as mentioned in Section 3.1.

Table 3 and 4 summarize the same aspects of estimation results as Table 2 for the Vasicek and CIR
models, respectively. Within each table, we see very similar patterns to those in Table 2 across different
assumptions on the number of regimes, the initial state probability and the transition probability. The
likelihood of a two-regime case is much greater than that of a single-regime model. Adding another parameter
for the distribution of the initial state makes little difference in all outcomes compared to using unconditional
probability. Again, the additional parameter goes to zero and I restrict it to zero while maximizing the
likelihood function. The null of the time-constant transition matrix is strongly rejected by LRT in favor
of the state-dependent transition probability. Using different functions for the transition probability gives
rise to practically the same outcome. While I get very accurate estimates for the coefficients of the state
variable, \( d_L \) and \( d_H \) in the case of the Vasicek model, only \( d_H \) is significant in the CIR model. Regarding
the regime classification, overall periods of high volatility regime identified by the two models are virtually
the same, even if the RCM of the CIR case is a little less than that of the Vasicek model.

As before, the drift term is indistinguishable from zero and all drift parameters are not significant except
\( c_0H \) of the CIR-R2TVTP-1 model. In fact, the different covariance structure of the drift parameters of the
regime \( H \) causes the high Wald statistic for testing difference in drift parameters of the two regimes in the
CIR-R2TVTP-1 model. Diffusion parameters and functions of two regimes are significantly different from
each other in all models. We can see similar graphical illustrations of these results to Figure 5 so I omit
them here.

Comparing these two models with my model, we can find some striking differences between them. First
of all, the LRTs very strongly rejects the hypothesis of the Vasicek and CIR specifications. For instance, the
LRT statistic (p-value) to make a comparison the models R2TVTP-2 and CIR-R2TVTP-2 is 86.592 (0.000)
and, R2TVTP-2 and Vasicek-R2TVTP-2 is 324.794 (0.000). The second difference arises in time-dependent
transition probabilities of the Vasicek model. In the other two cases, both states are highly persistent for all
observed interest rates and the probability of the process remaining in the same state increases as interest
rate increases. But Vasicek model reveals that the probability of the economy staying in the same regime \( L \)
(\( H \)) in the next period decreases (increases) as interest rate increases. In Figure 6, the transition probabilities
of the economy staying in the same regime for the model Vasicek-R2TVTP-2 are plotted. Comparing these
to the top two panels of Figure 2, we can see the apparent difference. This result is consistent with And
and Bekaert (2002) who investigate a discrete time regime-switching Vasicek model. Given the superiority
of my model to the Vasicek case in terms of the likelihood based test, their results may have come from misspecification. Third, the indistinguishableness of the drift from zero for all models and the LRT results imply that the volatility is underestimated for all interest rates in these two parametrizations as is clear from comparing Figure 5 and Figure 7, where the diffusion functions of the CIR-R2TVTP-2 and Vasicek-R2TVTP-2 models are drawn. Here, I focus solely on the model implied volatility because drift is not an important factor. Finally, although the RCM value of CIR and Vasicek cases are much less than that of my model, qualitatively, the identified high variance regime periods are about the same. One difference is that here the smoothed probabilities do not find much evidence of the regime $H$ in the early 1990s and 2000s.

5 Bond Pricing

The evolution of interest rates play an important role in pricing and hedging portfolios of fixed-income derivative securities. There has been a good deal of papers on term structure of interest rates based on the simplifying assumption that changes in yields with all maturities are driven by a single underlying, random factor. However, empirical evidence suggests that a multi-factor model explains the term structure better than a one factor model. Moreover, a one factor model suffers from a perfect correlation between the yields with different time to maturity. This is due to the assumption that the term structure is derived by only one factor. The yields are not perfectly correlated in practice even though there are strong correlations between them. In addition, the possible shapes of term structure one can obtain from the usual one factor model are very restrictive. Thus, people are moving towards multi-factor term structure models. Multi-factor models can solve the problems of one-factor models. Short term interest rates are used as a proxy for the first factor. However, the unobservedness of the additional factors makes the estimation more challenging and difficult.

The regime-switching univariate diffusion model for term structure is a two factor model. One factor is the continuously evolving short rates and the other factor is the regime variable changing discretely, taking finite (often two) number of values. Furthermore, the estimation results in the previous section also highlight the importance of using a more general diffusion model in explaining the movement of interest rates. In this section, I derive a system of partial differential equations which the bond price of my model with time-constant transition probability must satisfy because the closed-form solution to it is unavailable. We assume time-constant transition probability of the Markov chain.

Among other equivalent methods of derivation of the bond price, here, I follow the equilibrium bond pricing approach using stochastic discount factor which is also called the pricing kernel. DKSS follows the same approach to derive the bond price of the regime-switching CIR model. I assume that the stochastic discount factor, $\Lambda_t$, follows the stochastic differential equation

$$d\Lambda_t = -r_t \Lambda_t dt - \Lambda_t \lambda (r_t) dW_t$$

(17)
where $\lambda(r_t)$ is the market price of the risk of the continuous change in short rate and $W_t$ is the same Brownian motion as the one in (2). We assume that investors do not hedge the regime-switching risk. The market price of the continuous changes in interest rates are assumed to be the same in both regimes. Under the standard economic assumptions and (2), the price of a zero coupon bond at time $t$ maturing at time $T$, $P(t,T)$ is the function of short rate, $r_t$ and $s_t$, $P(t,r_t,s_t,T)$.

Since $P(T,T) = 1$, $P(t,T) = E_t \left( \frac{1}{X_t} \right)$ implies that $P(t,T) \Lambda_t$ is a martingale.

$$E_t \left( d \{ P(t,T) \Lambda_t \} \right) = 0$$

with a slight abuse of notation. From Ito’s formula we get

$$dP = \mu_P(t,T) P(t,T) dt + \sigma_P(t,T) P(t,T) dW_t$$

with

$$\mu_P(t,T) P(t,T) = \mu(r_t,s_t;\theta) \frac{\partial P}{\partial r} + \frac{1}{2} \sigma^2(r_t,s_t;\theta) \frac{\partial^2 P}{\partial r^2} + \frac{\partial P}{\partial H} + \sum_{i=1}^{2} \sum_{j=1}^{2} q_{ij} \Delta P$$

and

$$\sigma_P(t,T) P(t,T) = \sigma(r_t,s_t;\theta) \frac{\partial P}{\partial r}.$$

where $\Delta P$ is the change in $P$ when $s_t$ shifts. Recall that $\mu(r_t,s_t;\theta) = \alpha_{-1s} r_t^{-1} + \alpha_{0s} + \alpha_{1s} r_t + \alpha_{2s} r_t^2 + \alpha_{3s} r_t^3$ and $\sigma(r_t,s_t;\theta) = \beta_{s} r_t^{\rho_{st}}$.

Using the Ito’s Lemma

$$E_t \left( d \{ P(t,T) \Lambda_t \} \right) = E_t \left( dP(t,T) \Lambda_t \right) + E_t \left( dP(t,T) d\Lambda_t \right) + E_t \left( dP(t,T) d\Lambda_t \right)$$

into which we plug (17) and (18). Setting (19) equal zero and assuming $\lambda(r_t) = \lambda$ gives the following system of partial differential equations for the bond price:

$$\begin{cases} 
(\mu(r_t,s_t = L;\theta) - \lambda \sigma(r_t,s_t = L;\theta)) \frac{\partial P_L}{\partial r} + \frac{1}{2} \sigma^2(r_t,s_t = L;\theta) \frac{\partial^2 P_L}{\partial r^2} + \frac{\partial P_L}{\partial s} + \sum_{i=L,H} q_{Li} \Delta P_L - r P_L = 0 \\
(\mu(r_t,s_t = H;\theta) - \lambda \sigma(r_t,s_t = H;\theta)) \frac{\partial P_H}{\partial r} + \frac{1}{2} \sigma^2(r_t,s_t = H;\theta) \frac{\partial^2 P_H}{\partial r^2} + \frac{\partial P_H}{\partial s} + \sum_{i=L,H} q_{Hi} \Delta P_H - r P_H = 0 
\end{cases}$$

which can be solved numerically using finite difference method. Note that $P_i = P(t,r_t,s_t = i,T)$ where $i = L$ or $H$.

Once we find the price $P(t,T)$ of the discount bond, the zero coupon yield is given by

$$Y(t,r_t,s_t,T) = -\frac{\ln(P(t,r_t,s_t,T))}{T-t}$$

After solving the bond pricing problem numerically for the two dimensional grid in spot rate and time $(r_t,t)$ space, we need to interpolate the bond pricing function to get the full yield curve.
6 Conclusion

We develop and estimate a continuous time univariate regime-switching diffusion model of the short-term interest rate. The diffusion specification is general enough to encompass most existing models in literature. The estimation results reveal that there is strong evidence of the existence of two regimes. The specification of the time-varying transition matrix is preferred by the LRT and both regimes are highly persistent. Within each regime, the volatility rather than the drift is important in explaining the dynamics of interest rates. The volatility function of the regime $H$ is three times that of the regime $L$ and, of course, the two volatility functions are significantly different from each other. Based on the inferred probability, I can classify the sample period into high and low volatility states quite distinctively. Also, I could identify some of historic events that are very likely to cause high volatile interest rates as the regime $H$. Comparing my model with the two simpler models used in earlier papers, the likelihood based test apparently rejects the other two models in favor of my model. This implies that having misspecified model fit the data can result in misleading outcomes, in particular, about the volatility and time-varying transition probability.

A system of partial differential equations for the bond price of my model has been derived and it can be solved numerically since an analytic solution is not available. I will further look into the performance of the different specifications in terms of matching term structure. Using the estimation results, the term structure at different points in time can be obtained and compared with the actual prices to evaluate different models.

There have been growing studies on multi-factor term structure models. As mentioned before, multi-factor models can resolve some problems from which one factor models suffer. In addition, empirical evidence supports that additional factors are needed to fit the data better (e.g. Dai and Singleton (2000), Ahn, Dittmar, and Gallant (2001), Andersen, Benzoni, and Lund (2003)). To my knowledge, there is no written paper on continuous time multivariate regime-switching diffusion models for term structure. A few researchers (Bansal and Zhou (2002) and DSY) study discrete versions of simple multivariate diffusion models taking regime-shift into account. I will continue to work on continuous time multivariate regime-switching diffusion model.

Appendix

Appendix I. Assumptions

Assumption 1 (Smoothness of the coefficients): The functions $\mu(x; \theta)$ and $\sigma(x; \theta)$ are infinitely differentiable in $x$ in $D_X$, and twice continuously differentiable in $\theta$ in the parameter space $\Theta \subset R^d$.

Assumption 2 (Nondegeneracy of the diffusion): If $D_X = (0, +\infty)$, I allow for the possible local degeneracy of $\sigma$ at $x = 0$: If $\sigma(0; \theta) = 0$, then there exist constants $\xi_0, \omega \geq 0, \delta \geq 0$ such that $\sigma(x; \theta) \geq \omega x^\delta$ for all $0 < x < \xi_0$ and $\theta \in \Theta$. Away from 0, $\sigma$ is nondegenerate; that is, for each $\xi > 0$, there exists a constant $c_\xi$ such that $\sigma(x; \theta) \geq c_\xi > 0$ for all $x \in [\xi + \infty)$ and $\theta \in \Theta$. 
Assumption 3 (Boundary behavior): For all \( \theta \in \Theta, \mu_{\lambda}(y; \theta), \partial \mu_{\lambda}(y; \theta) / \partial y, \) and \( \partial^2 \mu_{\lambda}(y; \theta) / \partial y^2 \) have at most exponential growth near the infinity boundaries and \( \lim_{y \to -\infty} \gamma \lambda_{\lambda}(y; \theta) < +\infty \).

1. Left Boundary:
   i. If \( y = 0^+ \), there exist constants \( \epsilon_0, \kappa, \alpha \) such that for all \( 0 < y \leq \epsilon_0 \) and \( \theta \in \Theta, \mu_{\lambda}(y; \theta) \geq \kappa y^{-\alpha} \) where either \( \alpha > 1 \) and \( \kappa > 0 \), or \( \alpha = 1 \) and \( \kappa \geq 1 \).
   ii. If \( y = -\infty \), there exist constants \( E_0 > 0 \) and \( K > 0 \) such that for all \( y \geq E_0 \) and \( \theta \in \Theta, \mu_{\lambda}(y; \theta) \leq K y \).

2. Right Boundary: If \( y = +\infty \), there exist constants \( E_0 > 0 \) and \( K > 0 \) such that for all \( y \geq E_0 \) and \( \theta \in \Theta, \mu_{\lambda}(y; \theta) \leq K y \).

Assumption 4 (Strengthening of Assumption 2 in the limiting case where \( \alpha = 1 \) and the diffusion is degenerate at 0): Recall the constant \( \delta \) in Assumption 2(2), and the constants \( \alpha \) and \( \kappa \) in Assumption 3(1.i). If \( \alpha = 1 \), then either \( \delta \geq 1 \) with no restriction on \( \kappa \), or \( \kappa \geq 2\delta/(1-\delta) \) if \( 0 < \rho < 1 \). If \( \alpha > 1 \), no restriction is required.

See Alá-Habalia (2002a) for more about these assumptions. Assumption 1-3 are needed for the existence and uniqueness of a solution to a univariate stochastic differential equation. Assumption 4 is only for the purpose of maximizing the log-likelihood function.

Appendix II. Proof of parameter restrictions

In this appendix, I will show that why I need to impose restrictions on parameter space given in the following section to satisfy Assumptions 1-4. I will show only the case, \( \rho > 1 \) other cases can be similarly shown.

Assumption 1: This holds obviously.

Assumption 2: Since \( \sigma(0; \theta) = 0 \), in order for \( \beta_0 \rho \geq \omega x^\delta \) to be true for all \( 0 < x < \xi_0 \) and \( \theta \in \Theta \) for some \( \xi_0 > 0, \omega > 0 \) and \( \delta \geq 0 \), we should have \( \beta \geq 0 \). We need to have \( \beta > 0 \) and \( \rho > 0 \) for \( \beta_0 \rho \geq c_\xi \) to be satisfied for some \( c_\xi > 0 \) and for all \( x \in [\xi, \infty) \) and \( \theta \in \Theta \).

Assumption 3: Exponential growth rate conditions of \( \mu_{\lambda}(y; \theta), \partial \mu_{\lambda}(y; \theta) / \partial y, \) and \( \partial^2 \mu_{\lambda}(y; \theta) / \partial y^2 \) near the infinity boundary for all \( \theta \in \Theta \) are trivially true. The upper bound \( \lim_{y \to -\infty} \gamma \lambda_{\lambda}(y; \theta) < +\infty \). does not restrict \( \lambda_{\lambda}(y; \theta) \) from going to \( -\infty \) near the boundaries. Note that \( \lambda_{\lambda}(y; \theta) = -\frac{1}{2} \left\{ \alpha_1^2 \beta \frac{\partial^2}{\partial y^2}(\rho - 1)^{\frac{2(\rho + 1)}{\rho + 1}} y^{\frac{2(\rho + 1)}{\rho + 1}} + (\alpha_2^2 + 2\alpha_1\alpha_1) \beta \frac{\partial}{\partial y}(\rho - 1)^{\frac{2(\rho - 1)}{\rho + 1}} y^{\frac{2(\rho - 1)}{\rho + 1}} - \frac{\rho \theta}{\rho - 1} \right\}

\( \times (\rho - 1)^2 y^2 + (\alpha_2^2 + 2\alpha_1\alpha_1) \beta \frac{\partial^2}{\partial y^2}(\rho - 1)^{\frac{2(\rho - 1)}{\rho + 1}} y^{\frac{2(\rho - 1)}{\rho + 1}} - \frac{\rho \theta}{\rho - 1} \left\{ (\rho - 1)^{\frac{2\rho^2}{\rho - 1}} y^{\frac{2\rho^2}{\rho - 1}} - \alpha_1 \frac{\partial}{\partial y}(\rho - 1)^{\frac{2\rho^2}{\rho + 1}} y^{\frac{2\rho^2}{\rho + 1}} \right\}

\times y^{\frac{1}{\rho - 1}} + 2(\alpha_2^2 + 2\alpha_1\alpha_1) \beta \frac{\partial}{\partial y}(\rho - 1)^{\frac{2\rho^2}{\rho + 1}} y^{\frac{2\rho^2}{\rho + 1}} - \alpha_1 (\rho - 1)^{\frac{2\rho^2}{\rho + 1}} y^{\frac{2\rho^2}{\rho + 1}} - \alpha_2 \beta \frac{\partial^2}{\partial y^2}(\rho - 1)^{\frac{2\rho^2}{\rho + 1}} y^{\frac{2\rho^2}{\rho + 1}} + \alpha_2^2 \beta \frac{\partial}{\partial y}(\rho - 1)^{\frac{2\rho^2}{\rho + 1}} y^{\frac{2\rho^2}{\rho + 1}} - \alpha_2 (\rho - 1)^{\frac{2\rho^2}{\rho + 1}} y^{\frac{2\rho^2}{\rho + 1}} - \alpha_3 \beta \frac{\partial^2}{\partial y^2}(\rho - 1)^{\frac{2\rho^2}{\rho + 1}} y^{\frac{2\rho^2}{\rho + 1}} + \alpha_3^2 \beta \frac{\partial}{\partial y}(\rho - 1)^{\frac{2\rho^2}{\rho + 1}} y^{\frac{2\rho^2}{\rho + 1}} - \alpha_3 (\rho - 1)^{\frac{2\rho^2}{\rho + 1}} y^{\frac{2\rho^2}{\rho + 1}} \}

\( \begin{align*}
\end{align*} \)
and \( D_Y = (0 + \infty) \) when \( \rho > 1 \).

Therefore, \( \lim_{y \to 0^+} \lambda_Y(y; \theta) = \lim_{y \to 0^+} -\frac{1}{2} \left\{ \alpha_2^2 \beta^{-\frac{2}{\rho-1}} (\rho - 1) - \frac{2(\rho-2)}{2(\rho-1)} y \frac{2(\rho-2)}{2(\rho-1)} + \frac{4(\rho-2)}{4(\rho-1)} y^{-2} + 2\alpha_1 \alpha_2 \beta^{-\frac{1}{\rho-1}} \right\} \times (\rho - 1) y^\frac{2}{\rho-1} - \alpha_1 (2\rho - 1) - 2\alpha_2 \beta^{-\frac{3}{\rho-1}} (\rho - 1) y^\frac{2}{\rho-1} + \alpha_3 \beta^{-\frac{1}{\rho-1}} (\rho - 1) y^\frac{2(\rho-3)}{2(\rho-1)} - 2\alpha_3 \beta^{-\frac{1}{\rho-1}} (\rho - 1) y^\frac{2p-3}{2p-1} - \alpha_3 \beta^{-\frac{1}{\rho-1}} (2\rho - 3) y^\frac{2p-5}{2p-1} y^\frac{2}{\rho-1} \right\} \}

When \( 1 < \rho < 2 \), \( y^\frac{2p-3}{2p-1} \) dominates because \( 2(\rho-3) < \min \left\{ \frac{2(p-2)}{2p-1}, -2, \frac{2p-3}{2p-1}, \frac{2p-5}{2p-1}, \frac{2p-7}{2p-1} \right\} \) on \( 1 < \rho < 2 \), so \( \lim_{y \to 0^+} \lambda_y = +\infty \) holds for \( 1 < \rho < 2 \).

If \( 2 < \rho \), \( y^{-2} \) dominates. Thus, we need \( -\frac{\rho(\rho-2)}{4(\rho-1)} y^\frac{2}{\rho-1} \geq 0 \), which is true only if \( 0 \leq \rho \leq 2 \) for \( \lim_{y \to 0^+} \lambda_y = +\infty \) to be satisfied. Therefore, \( \rho \) can not be greater than 2. Now, let see the case, \( \lim_{y \to \infty} \lambda_y < +\infty \). Since

\[
\lim_{y \to \infty} \lambda_y = \lim_{y \to \infty} -\frac{1}{2} \left\{ \alpha_2^2 \beta^{-\frac{2}{\rho-1}} (\rho - 1) y^\frac{2(\rho-1)}{2(\rho-1)} + (\alpha_0^2 + 2\alpha_0 \alpha_2 + \alpha_2 \alpha_3) (\rho - 1)^2 y^2 + (\alpha_2 + 2\alpha_1 \alpha_3) \beta^{-\frac{1}{\rho-1}} \times (\rho - 1) y^\frac{2(\rho-2)}{2(\rho-1)} + 2\alpha_1 \alpha_3 \beta^{-\frac{1}{\rho-1}} (\rho - 1) y^\frac{2p-3}{2p-1} + 2 (\alpha_1 \alpha_2 + \alpha_0 \alpha_1) \beta^{-\frac{1}{\rho-1}} (\rho - 1) y^\frac{2p-3}{2p-1} \right\} \times y^\frac{2p-3}{2p-1} - \alpha_1 (2\rho - 1) \right\} \}

and \( 2(p+1) \rho(p-1) > \min \left\{ \frac{2p+1}{p-1}, \frac{2p+1}{p-1}, \frac{2p+1}{p-1}, \frac{2p-3}{p-1}, \frac{2p-3}{p-1} \right\} \) \( y^\frac{2p+1}{p-1} \) dominates and

\[
\lim_{y \to \infty} \lambda_y = +\infty \) is satisfied.

Left Boundary

\[
\mu_y(y; \theta) = -\alpha_1 \beta^{-\frac{2}{\rho-1}} (\rho - 1) y^\frac{2(p+1)}{p-1} - \alpha_0 \beta^{-\frac{1}{\rho-1}} (\rho - 1) y^\frac{2p-3}{2p-1} - \alpha_1 (\rho - 1) y - \alpha_2 \beta^{-\frac{1}{\rho-1}} (\rho - 1) y^\frac{2p-3}{2p-1} - \alpha_3 \beta^{-\frac{1}{\rho-1}} (\rho - 1) y^\frac{2p-3}{2p-1} + \frac{\rho}{2(p-1)} y^{-1} \geq \kappa y^{-\alpha}. \]

Hence, for small enough \( \varepsilon_0 \), \( 0 < \varepsilon_0 \mu_y(y; \theta) \approx -\alpha_3 \beta^{-\frac{2}{\rho-1}} (\rho - 1) y^\frac{2p-3}{2p-1} \geq \kappa y^{-\alpha} \) to hold for all \( 0 < y \leq \varepsilon_0 \) and for some \( \varepsilon_0 > 0 \), \( \alpha > 1 \) and \( \kappa > 0 \) we need \( \alpha_3 < 0 \).

Right Boundary

For large enough \( y \), \( \mu_y(y; \theta) \approx -\alpha_1 \beta^{-\frac{2}{\rho-1}} (\rho - 1) y^\frac{2p+1}{p-1} - \alpha_0 \beta^{-\frac{1}{\rho-1}} (\rho - 1) y^\frac{2p-3}{2p-1} - \alpha_1 (\rho - 1) y \).

For \( \mu_y(y; \theta) \leq K y \) to hold for all \( y \geq E_0 \) and \( \theta \in \Theta \) and for some \( E_0 > 0 \) and \( K > 0 \) we need \( \alpha_1 > 0 \) because \( y^\frac{2p+1}{p-1} \) dominates.

Assumption 4 Therefore no restriction is required in this case because \( \alpha > 1 \).
Appendix III. Density Expansion

The approximate transition probability density function of \( Y_t \) for the case \( K = 1 \) is given when \( Y_t \) is \( y_0 \) at \( t \) as a function of \( y \) at \( t + \Delta \) for four cases, \( \rho = 0, 0 < \rho < 1, \rho = 1, \) and \( 1 < \rho < 2 \).

This table contains the closed-form transition density approximation to \( p_Y \) and the restrictions on \( \theta \) corresponding to the case \( \rho = 0 \) for all models we consider. In this case \( Y_t = \gamma(X_t; \theta) = \frac{\lambda}{m} \) and \( \mu_Y(y; \theta) = \alpha_{-1} y^{-2} + \alpha_0 y^{-1} + \alpha_1 y + \alpha_2 y^2 + \alpha_3 y^3, (\alpha_3 \leq 0) \). Then we can get

\[
\hat{p}_y^{(1)}(\Delta, y|y_0; \theta) = \hat{p}_y^{(0)}(\Delta, y|y_0; \theta) \{1 + c_1 (y|y_0; \theta) \Delta\}
\]

where

\[
\hat{p}_y^{(0)}(\Delta, y|y_0; \theta) = \frac{1}{\sqrt{\Delta}} \left( \frac{y - y_0}{\sqrt{\Delta}} \right) \exp \left\{ \alpha_{-1} y^{-2} (\ln y - \ln y_0) + \alpha_0 y^{-1} (y - y_0) + \frac{1}{2} \alpha_1 (y^2 - y_0^2) \right\}
+ \frac{1}{3} \alpha_2 y (y^3 - y_0^3) + \frac{1}{4} \alpha_3 y^2 (y^4 - y_0^4)
\]

and if \( y \neq y_0 \)

\[
c_1 (y|y_0; \theta) = -\frac{1}{2(y - y_0)} \left\{ - (\alpha_{-1}^2 y^{-4} - \alpha_{-1}^2 y^{-2}) (y^{-1} - y_0^{-1}) + 2 \alpha_{-1} \alpha_0 \beta^{-3} (\ln y - \ln y_0) \right\}
+ \left\{ (\alpha_0^2 + 2 \alpha_{-1} \alpha_1) \beta^{-2} + \alpha_1 \right\} (y - y_0) + (\alpha_{-1} \alpha_2 \beta^{-1} + \alpha_0 \alpha_1 \beta^{-1} + \alpha_2 \beta) (y^2 - y_0^2)
+ \frac{1}{3} (\alpha_2^2 + 2 \alpha_0 \alpha_2 + 2 \alpha_{-1} \alpha_3 + 3 \alpha_3 \beta^2) (y^3 - y_0^3) + \frac{1}{2} (\alpha_1 \alpha_2 + \alpha_0 \alpha_3) \beta (y^4 - y_0^4)
+ \frac{1}{3} (\alpha_2^2 + 2 \alpha_1 \alpha_3) \beta^2 (y^5 - y_0^5) + \frac{1}{3} \alpha_2 \alpha_3 \beta^3 (y^6 - y_0^6)
\]

otherwise

\[
c_1 (y|y_0; \theta) = -\frac{1}{2} \left\{ (\alpha_{-1}^2 \beta^{-4} - \alpha_{-1} \beta^{-2}) y_0^{-2} + 2 \alpha_{-1} \alpha_0 \beta^{-3} y_0^{-1} + (\alpha_0^2 + 2 \alpha_{-1} \alpha_1) \beta^{-2}
+ \alpha_1 + 2 (\alpha_{-1} \alpha_2 \beta^{-1} + \alpha_0 \alpha_1 \beta^{-1} + \alpha_2 \beta) y_0 + (\alpha_2^2 + 2 \alpha_0 \alpha_2 + 2 \alpha_{-1} \alpha_3 + 3 \alpha_3 \beta^2) y_0^2
+ 2 (\alpha_1 \alpha_2 + \alpha_0 \alpha_3) \beta y_0^3 + (\alpha_2^2 + 2 \alpha_1 \alpha_3) \beta^2 y_0^4 + 2 \alpha_2 \alpha_3 \beta^3 y_0^5 + \alpha_2^2 \beta^4 y_0^6 \right\}
\]
This table contains the closed-form transition density approximation to $p_Y$ and the restrictions on $\theta$ corresponding to the case $0 < \rho < 1$ for all models we consider. In this case $Y_t = \gamma (X_t; \theta) = \frac{X_t^{1-\rho}}{1-\rho}$ and $\mu_Y (y; \theta) = \alpha_{-1}\beta^{\frac{2-\rho}{1-\rho}} (1-\rho)^{\frac{1}{1-\rho}} y^{\frac{1}{1-\rho}} + \alpha_0 \beta^{\frac{2-\rho}{1-\rho}} y^{\frac{1-\rho}{1-\rho}} + \alpha_1 (1-\rho) y^{\frac{1-\rho}{1-\rho}} + \alpha_2 \beta^{\frac{2-\rho}{1-\rho}} (1-\rho)^{\frac{1}{1-\rho}} y^{\frac{1}{1-\rho}} - \frac{\rho(\rho-2)}{2(1-\rho)^2} y^2 + \alpha_3 \beta^{\frac{2-\rho}{1-\rho}} (1-\rho)^{\frac{1}{1-\rho}} y^{\frac{1}{1-\rho}} .(\alpha_{-1} > 0 \text{ and } \alpha_3 \leq 0)

Then we can get

$$\tilde{p}_Y^{(1)} (\Delta, y|y_0; \theta) = \tilde{p}_Y^{(0)} (\Delta, y|y_0; \theta) \{1 + c_1 (y|y_0; \theta) \Delta\}$$

where

$$\tilde{p}_Y^{(0)} (\Delta, y|y_0; \theta) = \frac{1}{\sqrt{\Delta}} \left(\frac{y - y_0}{\sqrt{\Delta}}\right) \exp \left\{ -\alpha_{-1}\beta^{\frac{2-\rho}{1-\rho}} \frac{1}{2\rho} (1-\rho)^{\frac{1}{1-\rho}} \left( y^{\frac{1}{1-\rho}} - y_0^{\frac{1}{1-\rho}} \right) + \frac{\alpha_0}{1-2\rho} (1-\rho)^{\frac{1}{1-\rho}} \left( y^{\frac{1}{1-\rho}} - y_0^{\frac{1}{1-\rho}} \right) + \frac{1}{2} \alpha_1 (1-\rho) \left( y^2 - y_0^2 \right) + \alpha_2 \beta^{\frac{2-\rho}{1-\rho}} \frac{1}{3-2\rho} (1-\rho)^{\frac{1}{1-\rho}} \left( y^{\frac{1}{1-\rho}} - y_0^{\frac{1}{1-\rho}} \right) - \frac{\rho(\rho-2)}{2(1-\rho)^2} \left( \ln y - \ln y_0 \right) \right\}$$

and if $y \neq y_0$

$$c_1 (y|y_0; \theta) = -\frac{1}{2 (y - y_0)} \left\{ \alpha_{-1}\beta^{\frac{2-\rho}{1-\rho}} \frac{1}{1+3\rho} (1-\rho)^{\frac{1}{1-\rho}} \left( y^{\frac{1}{1-\rho}} - y_0^{\frac{1}{1-\rho}} \right) + \alpha_0 \beta^{\frac{2-\rho}{1-\rho}} \frac{1}{1-2\rho} (1-\rho)^{\frac{1}{1-\rho}} \left( y^{\frac{1}{1-\rho}} - y_0^{\frac{1}{1-\rho}} \right) + \frac{1}{3} (\alpha_1^2 + 2\alpha_0 \alpha_2 + 2\alpha_{-1} \alpha_1) (1-\rho)^{\frac{1}{1-\rho}} (y^3 - y_0^3) + \frac{\rho(\rho-2)}{4 (1-\rho)^2} \left( y - y_0 \right) \right\}$$

otherwise

$$c_1 (y|y_0; \theta) = -\frac{1}{2} \left\{ \alpha_{-1}\beta^{\frac{2-\rho}{1-\rho}} (1-\rho)^{\frac{1}{1-\rho}} y^{\frac{1}{1-\rho}} - y_0^{\frac{1}{1-\rho}} + \alpha_0 \beta^{\frac{2-\rho}{1-\rho}} (1-\rho)^{\frac{1}{1-\rho}} y^{\frac{1}{1-\rho}} - y_0^{\frac{1}{1-\rho}} + \alpha_1 (1-2\rho) \left( y - y_0 \right) + \alpha_2 \beta^{\frac{2-\rho}{1-\rho}} (1-\rho)^{\frac{1}{1-\rho}} y^{\frac{1}{1-\rho}} - y_0^{\frac{1}{1-\rho}} + \frac{3}{2} \alpha_3 \beta^{\frac{2-\rho}{1-\rho}} (1-\rho)^{\frac{1}{1-\rho}} y^{\frac{1}{1-\rho}} - y_0^{\frac{1}{1-\rho}} \right\}$$
This table contains the closed-form transition density approximation to $p_Y$ and the restrictions on $\theta$ corresponding to the case $\rho = 1$ for all models we consider. In this case $Y = \gamma (X_1; \theta) = \frac{1}{\theta} \ln X_2$ and $\mu_Y (y; \theta) = \alpha_1 \beta^{-1} e^{-2\beta y} + \alpha_0 \beta^{-1} e^{-\beta y} + \alpha_1 \beta^{-1} + \alpha_2 \beta^{-1} e^{\beta y} + \alpha_3 \beta^{-1} e^{2\beta y} - \frac{1}{2} \beta$. ($\alpha_1 > 0$ and $\alpha_3 \leq 0$). Then we can get

$$p_Y^{(1)} (\Delta, y|y_0; \theta) = p_Y^{(0)} (\Delta, y|y_0; \theta) \{1 + c_1 (y|y_0; \theta) \Delta\}$$

where

$$p_Y^{(0)} (\Delta, y|y_0; \theta) = \frac{1}{\sqrt{\Delta}} \phi \left( \frac{y - y_0}{\sqrt{\Delta}} \right) \exp \left\{ -\frac{1}{2} \alpha_1 \beta^{-2} \left( e^{-2\beta y} - e^{-2\beta y_0} \right) - \alpha_0 \beta^{-2} \left( e^{-\beta y} - e^{-\beta y_0} \right) + \alpha_1 \beta^{-2} (y - y_0) + \frac{1}{2} \alpha_3 \beta^{-2} \left( e^{2\beta y} - e^{2\beta y_0} \right) \right\}$$

and if $y \neq y_0$

$$c_1 (y|y_0; \theta) = -\frac{1}{2 (y - y_0)} \left\{ -\frac{1}{4} \alpha_2 \beta^{-3} \left( e^{-4\beta y} - e^{-4\beta y_0} \right) + \frac{1}{2} \left( \alpha_2 \beta^{-3} + 2 \alpha_0 \alpha_1 \beta^{-3} - 3 \alpha_0 \beta^{-1} \right) \left( e^{-2\beta y} - e^{-2\beta y_0} \right) \right. \left. + \left( \alpha_2 \beta^{-3} + 3 \alpha_0 \beta^{-1} + 2 \alpha_0 \alpha_2 \beta^{-3} - \alpha_1 + \alpha_1 \alpha_3 \beta^{-2} \right) \left( e^{2\beta y} - e^{2\beta y_0} \right) \right. \right.$$  

$$\left. + \left( \alpha_2 \beta^{-3} + 3 \alpha_0 \beta^{-1} + 2 \alpha_0 \alpha_2 \beta^{-3} - \alpha_1 + \alpha_1 \alpha_3 \beta^{-2} \right) \left( e^{2\beta y} - e^{2\beta y_0} \right) \right.$$  

$$\left. + 2 \left( \alpha_2 \beta^{-3} \left( e^{3\beta y} - e^{3\beta y_0} \right) + \frac{1}{4} \alpha_2 \beta^{-3} \left( e^{4\beta y} - e^{4\beta y_0} \right) \right) \right\}$$

otherwise

$$c_1 (y|y_0; \theta) = -\frac{1}{2} \left( \alpha_1 \beta^{-2} e^{-4\beta y_0} + (\alpha_2 \beta^{-2} + 2 \alpha_0 \alpha_1 \beta^{-2} - 3 \alpha_0) e^{-2\beta y_0} + 2 \alpha_0 \alpha_0 \beta^{-2} e^{-3\beta y_0} \right) \left( e^{-2\beta y} - e^{-2\beta y_0} \right) + \left( \alpha_2 \beta^{-2} + \frac{1}{4} \beta^2 + 2 \alpha_0 \alpha_2 \beta^{-2} - \alpha_1 + 2 \alpha_0 \alpha_3 \beta^{-2} \right) \left( e^{2\beta y} - e^{2\beta y_0} \right) + \left( \alpha_2 \beta^{-2} + 3 \alpha_0 + 2 \alpha_0 \alpha_3 \beta^{-2} \right) e^{3\beta y_0}$$

$$+ 2 \left( \alpha_2 \beta^{-2} + 2 \alpha_0 \alpha_2 \beta^{-2} - \alpha_0 \beta^2 + 2 \alpha_0 \alpha_2 \beta^{-2} e^{3\beta y_0} + 2 \alpha_2 \alpha_3 \beta^{-2} e^{3\beta y_0} + \frac{2}{3} \alpha_2 \beta^{-3} e^{4\beta y_0} \right)$$
This table contains the closed-form transition density approximation to \( p_Y \) and the restrictions on \( \theta \) corresponding to the case 

\[ 1 < \rho < 2, \ (\rho \neq \frac{1}{2}, \frac{3}{2}, \frac{5}{2}) \] for all models we consider. In this case \( Y_t = \gamma (X_t; \theta) = \frac{X_t}{\gamma (X_{t-1}; \theta)} \) and 

\[ p_Y (y; \theta) = -\alpha_{-1} \beta^{\rho - 1} (\rho - 1) \frac{2\rho - 3}{2 \rho - 1} y^{\rho - 1} - \alpha_0 \beta^{\rho - 1} (\rho - 1) \frac{2\rho - 3}{2 \rho - 1} y^{\rho - 1} - \alpha_1 (\rho - 1) \frac{2\rho - 3}{2 \rho - 1} y^{\rho - 1} + \frac{2 \rho - 3}{2 \rho - 1} y^{2 \rho - 3} - \alpha_3 \beta^{\rho - 1} (\rho - 1) \frac{2\rho - 3}{2 \rho - 1} y^{2 \rho - 3}. \] 

(\( \alpha_{-1} > 0 \) and \( \alpha_3 \leq 0 \))

Then we can get 

\[ p_Y^{(1)} (\Delta, y|y_0; \theta) = p_Y^{(0)} (\Delta, y|y_0; \theta) \{1 + c_1 (y|y_0; \theta) \Delta \} \]

where

\[ p_Y^{(0)} (\Delta, y|y_0; \theta) = \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left( \frac{y - y_0 \Delta}{\sqrt{\Delta}} \right)^2 \right\} \]

and if \( y \neq y_0 \)

\[ c_1 (y|y_0; \theta) = \frac{1}{3} \left\{ \alpha_2 + 2\alpha_{-1} \alpha_1 \right\} \beta^{\rho - 1} \left( \rho - 1 \right) \frac{2\rho - 3}{2 \rho - 1} \left( \frac{y - y_0}{\sqrt{\Delta}} \right)^{\rho - 1} \]

\[ + \frac{2 \alpha_{-1} \alpha_0 \beta^{\rho - 1}}{3 \rho - 4} \left( \rho - 1 \right) \frac{2\rho - 3}{2 \rho - 1} \left( \frac{y - y_0}{\sqrt{\Delta}} \right)^{\rho - 1} \]

\[ \times (\rho - 1)^2 y_0^2 + \frac{(\alpha_2 + 2\alpha_{-1} \alpha_3) \beta^{\rho - 1}}{3 \rho - 4} \left( \rho - 1 \right) \frac{2\rho - 3}{2 \rho - 1} \left( \frac{y - y_0}{\sqrt{\Delta}} \right)^{\rho - 1} \]

otherwise

\[ c_1 (y|y_0; \theta) = \frac{1}{2} \left\{ \alpha_2 + 2\alpha_{-1} \alpha_1 \beta^{\rho - 1} (\rho - 1) \frac{2\rho - 3}{2 \rho - 1} y^{\rho - 1} + \frac{(\alpha_2 + 2\alpha_{-1} \alpha_3) \beta^{\rho - 1}}{3 \rho - 4} (\rho - 1) \frac{2\rho - 3}{2 \rho - 1} y^{\rho - 1} \right\} \]
References


34


Descriptive statistics for the weekly 3-month U.S. T-Bill Yields from 01/08/71 to 12/26/03 are computed. Here, skewness coefficient $= \mu_3/\sigma^3$ and degree of excess $= \mu_4/\sigma^4 - 3$ are respectively normalized measure of the asymmetry and the thickness of the tails of the distribution relative to standard Normal distribution. Note that $\mu_i = E[(r_t - \mu)^i]$ and $\sigma^2 = E[(r_t - \mu_t)^2]$.

<table>
<thead>
<tr>
<th>Observations</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std. Dev</th>
<th>Skewness Coef</th>
<th>Degree of Excess</th>
</tr>
</thead>
<tbody>
<tr>
<td>1721</td>
<td>0.0084</td>
<td>0.1676</td>
<td>0.0621</td>
<td>0.0287</td>
<td>0.9241</td>
<td>1.3574</td>
</tr>
</tbody>
</table>
Table 2: Maximum Likelihood Estimation Results of Main Models

<table>
<thead>
<tr>
<th>Model Regime</th>
<th>R1</th>
<th>R2-1</th>
<th>R2-2</th>
<th>R2TVTP-1</th>
<th>R2TVTP-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P(s_1 = L) )</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>( P(s_{t+1} = i</td>
<td>s_t = i) ) with ( i = L ) and ( H )</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>8572.872</td>
<td>8951.967</td>
<td>8953.226</td>
<td>8962.410</td>
<td>8962.263</td>
</tr>
<tr>
<td>Parameters</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>( \alpha_{1L} )</td>
<td>0.000189* (0.0000825)</td>
<td>0.0000757 (0.000144)</td>
<td>0.0000891 (0.000125)</td>
<td>0.0000106 (0.000178)</td>
<td>0.0000102 (0.000163)</td>
</tr>
<tr>
<td>( \alpha_{1H} )</td>
<td>0.000418 (0.000191)</td>
<td>0.000340 (0.000539)</td>
<td>0.000390 (0.000649)</td>
<td>0.000376 (0.000633)</td>
<td></td>
</tr>
<tr>
<td>( \alpha_{0L} )</td>
<td>-0.0110 (0.0217)</td>
<td>-0.0127 (0.0191)</td>
<td>-0.0153 (0.0289)</td>
<td>-0.0148 (0.0263)</td>
<td></td>
</tr>
<tr>
<td>( \alpha_{0H} )</td>
<td>-0.0627 (0.0723)</td>
<td>-0.0542 (0.0790)</td>
<td>-0.0636 (0.0937)</td>
<td>-0.0620 (0.0912)</td>
<td></td>
</tr>
<tr>
<td>( \alpha_{1L} )</td>
<td>0.401 (0.913)</td>
<td>0.456 (0.814)</td>
<td>0.538 (1.389)</td>
<td>0.521 (1.250)</td>
<td></td>
</tr>
<tr>
<td>( \alpha_{1H} )</td>
<td>1.396 (2.822)</td>
<td>1.208 (3.098)</td>
<td>1.488 (3.585)</td>
<td>1.440 (3.492)</td>
<td></td>
</tr>
<tr>
<td>( \alpha_{2L} )</td>
<td>-3.076 (13.962)</td>
<td>-3.503 (12.677)</td>
<td>-4.412 (24.591)</td>
<td>-4.298 (21.947)</td>
<td></td>
</tr>
<tr>
<td>( \alpha_{2H} )</td>
<td>-8.276 (38.044)</td>
<td>-7.250 (42.014)</td>
<td>-8.322 (47.605)</td>
<td>-8.116 (46.449)</td>
<td></td>
</tr>
<tr>
<td>( \alpha_{3L} )</td>
<td>-0.0546 (67.902)</td>
<td>-0.314 (62.904)</td>
<td>-0.454 (139.831)</td>
<td>-0.0372 (123.820)</td>
<td></td>
</tr>
<tr>
<td>( \alpha_{3H} )</td>
<td>-0.118 (155.997)</td>
<td>-0.476 (173.138)</td>
<td>-4.092 (193.329)</td>
<td>-3.166 (188.916)</td>
<td></td>
</tr>
<tr>
<td>( \beta_{L} )</td>
<td>0.300* (0.0113)</td>
<td>0.107* (0.0344)</td>
<td>0.107* (0.0122)</td>
<td>0.108* (0.0148)</td>
<td>0.108* (0.0136)</td>
</tr>
<tr>
<td>( \beta_{H} )</td>
<td>0.328* (0.000)</td>
<td>0.329* (0.0143)</td>
<td>0.332* (0.0476)</td>
<td>0.331* (0.0464)</td>
<td></td>
</tr>
<tr>
<td>( \rho_{L} )</td>
<td>1.116* (0.0126)</td>
<td>0.968* (0.0439)</td>
<td>0.968* (0.0456)</td>
<td>0.968* (0.0436)</td>
<td></td>
</tr>
<tr>
<td>( \rho_{H} )</td>
<td>0.973* (0.0417)</td>
<td>0.973* (0.0468)</td>
<td>0.973* (0.0527)</td>
<td>0.973* (0.0515)</td>
<td></td>
</tr>
<tr>
<td>( p_{LL} )</td>
<td>—</td>
<td>0.980* (0.00527)</td>
<td>0.981* (0.00485)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>( p_{HH} )</td>
<td>—</td>
<td>0.954* (0.0113)</td>
<td>0.953* (0.00973)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>( c_{L} )</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>1.353* (0.284)</td>
<td>2.169* (0.653)</td>
</tr>
<tr>
<td>( c_{H} )</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.544 (4.441)</td>
<td>0.691 (11.051)</td>
</tr>
<tr>
<td>( d_{L} )</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>12.954* (0.313)</td>
<td>32.706* (0.585)</td>
</tr>
<tr>
<td>( d_{H} )</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>18.198* (5.016)</td>
<td>38.510* (10.298)</td>
</tr>
<tr>
<td>( p )</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Wald Test Statistics (p-value) to test ( \beta_{L} = \beta_{H} )</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

Maximum likelihood estimates with standard errors in parentheses for the five different models with the general diffusion specification are presented in this table. The standard errors are calculated using the mean of the outer product of the score functions evaluated at ML estimates. According to the number of regimes, the starting probability and the transition probability matrix specifications, the five models are: R1 (single regime), R2-1 (two regimes, time invariant transition matrix, unconditional probability for the probability of initial state), R2-2 (two regimes, time invariant transition matrix, additional parameter for the probability of initial state), R2TVTP-1 (two regimes, time varying transition matrix with Normal cumulative density function, additional parameter for the probability of initial state), and R2TVTP-2 (two regimes, time varying transition matrix with logistic function, additional parameter for the probability of initial state). In addition, the maximized likelihood values, RCMs are provided. Wald test statistic and its p-value are also calculated to jointly test if the parameters of the drift and diffusion functions and transition probabilities, separately, of two regimes are equal.
fusion functions and transition probabilities, separately, of two regimes are equal.

Additional parameters for the probability of initial state. In addition, the maximized likelihood values, RCMs are provided. Wald test statistic and its p-value

Maximum likelihood estimates with standard errors in parentheses for the parameters of the drift and diffusion functions and transition probabilities, separately, of two regimes are equal.

### Table 3: Maximum Likelihood Estimation Results CIR Models

<table>
<thead>
<tr>
<th>Model</th>
<th>R1</th>
<th>R2-1</th>
<th>R2-2</th>
<th>R2TVTP-1</th>
<th>R2TVTP-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regime</td>
<td>Single Regime</td>
<td>Two Regimes</td>
<td>Time Invariant Transition Matrix</td>
<td>Time Varying Transition Matrix</td>
<td>p</td>
</tr>
<tr>
<td>P(s1 = L)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>P(s_{t+1} = i</td>
<td>s_1 = i)</td>
<td>—</td>
<td>—</td>
<td>p_{ii}</td>
<td>p_{ii}</td>
</tr>
<tr>
<td>with i = L and H</td>
<td>—</td>
<td>—</td>
<td>Φ(c_i + d_i r_{t-1})</td>
<td>Φ(c_i + d_i r_{t-1})</td>
<td></td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>8359.901</td>
<td>8908.169</td>
<td>8909.647</td>
<td>8919.147</td>
<td>8918.967</td>
</tr>
<tr>
<td>Parameters</td>
<td>Maximum Likelihood Estimates (standard error)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α_{0L}</td>
<td>0.00584 (0.00553)</td>
<td>0.00109 (0.00323)</td>
<td>0.00109 (0.00322)</td>
<td>0.00108 (0.00323)</td>
<td>0.00108 (0.00254)</td>
</tr>
<tr>
<td>α_{0H}</td>
<td>0.0291 (0.0232)</td>
<td>0.0292 (0.0233)</td>
<td>0.0264 (0.0115)</td>
<td>0.0264 (0.0275)</td>
<td></td>
</tr>
<tr>
<td>α_{1L}</td>
<td>-0.104 (0.0692)</td>
<td>-0.0108 (0.0604)</td>
<td>-0.0108 (0.0601)</td>
<td>-0.00875 (0.0608)</td>
<td>-0.00900 (0.0459)</td>
</tr>
<tr>
<td>α_{1H}</td>
<td>-0.414 (0.270)</td>
<td>-0.415 (0.271)</td>
<td>-0.384 (0.229)</td>
<td>-0.385 (0.305)</td>
<td></td>
</tr>
<tr>
<td>β_L</td>
<td>0.0569* (0.000385)</td>
<td>0.0296* (0.000567)</td>
<td>0.0296* (0.000563)</td>
<td>0.0293* (0.000569)</td>
<td>0.0293* (0.000380)</td>
</tr>
<tr>
<td>β_H</td>
<td>0.108* (0.00291)</td>
<td>0.108* (0.00292)</td>
<td>0.108* (0.00292)</td>
<td>0.109* (0.00288)</td>
<td></td>
</tr>
<tr>
<td>P_{LL}</td>
<td>—</td>
<td>0.989* (0.00321)</td>
<td>0.989* (0.00312)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>P_{HH}</td>
<td>—</td>
<td>0.964* (0.00853)</td>
<td>0.962* (0.00885)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>p</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>2.152* (0.298)</td>
<td>4.127* (0.262)</td>
</tr>
<tr>
<td>Drift Function</td>
<td>—</td>
<td>2.199 (0.333)</td>
<td>2.203 (0.332)</td>
<td>16.930 (0.000)</td>
<td>1.738 (0.419)</td>
</tr>
<tr>
<td>Diffusion Function</td>
<td>—</td>
<td>720.722 (0.000)</td>
<td>890.979 (0.000)</td>
<td>731.670 (0.000)</td>
<td>723.874 (0.000)</td>
</tr>
<tr>
<td>Transition Probability</td>
<td>—</td>
<td>8.591 (0.063)</td>
<td>9.342 (0.002)</td>
<td>53.909 (0.000)</td>
<td>119.594 (0.000)</td>
</tr>
<tr>
<td>RCM</td>
<td>—</td>
<td>7.657</td>
<td>7.614</td>
<td>8.790</td>
<td>8.689</td>
</tr>
</tbody>
</table>

Maximum likelihood estimates with standard errors in parentheses for the five different models with the general diffusion specification are presented in this table. The standard errors are calculated using sample mean of the outer product of the score functions evaluated at the ML estimates. According to the number of regimes, the starting probability and the transition probability matrix specifications, the five models are: R1 (single regime), R2-1 (two regimes, time invariant transition matrix, unconditional probability for the probability of initial state), R2-2 (two regimes, time invariant transition matrix, additional parameter for the probability of initial state), R2TVTP-1 (two regimes, time varying transition matrix with Normal cumulative density function, additional parameter for the probability of initial state), and R2TVTP-2 (two regimes, time varying transition matrix with logistic function, additional parameter for the probability of initial state). In addition, the maximized likelihood values, RCMs are provided. Wald test statistic and its p-value are also calculated to jointly test if the parameters of the drift and diffusion functions and transition probabilities, separately, of two regimes are equal.
Table 4: Maximum Likelihood Estimation Results of Vasicek Models

<table>
<thead>
<tr>
<th>Model Regime</th>
<th>R1</th>
<th>R2-1</th>
<th>R2-2</th>
<th>R2TVTP-1</th>
<th>R2TVTP-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time Invariant Transition Matrix</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( P(s_1 = L) )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( P(s_{t+1} = i</td>
<td>s_t = i) )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>with ( i = L ) and ( H )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>7970.945</td>
<td>8783.087</td>
<td>8784.638</td>
<td>8800.077</td>
<td>8799.866</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Maximum Likelihood Estimates (standard error)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha_{0l} )</td>
<td>0.00712 (0.00753)</td>
</tr>
<tr>
<td>( \alpha_{0H} )</td>
<td>0.0251 (0.0468)</td>
</tr>
<tr>
<td>( \alpha_{1l} )</td>
<td>-0.127 (0.0672)</td>
</tr>
<tr>
<td>( \alpha_{1H} )</td>
<td>-0.358 (0.402)</td>
</tr>
<tr>
<td>( \beta_L )</td>
<td>0.0154* (0.000692)</td>
</tr>
<tr>
<td>( \beta_H )</td>
<td>0.0350* (0.000883)</td>
</tr>
</tbody>
</table>

| \( p_{LL} \) | — | 0.988* (0.00317) | 0.988* (0.00306) | — | — |
| \( p_{HH} \) | — | 0.958* (0.00929) | 0.955* (0.00962) | — | — |
| \( c_L \) | — | — | — | 2.763* (0.332) | 5.671* (0.321) |
| \( c_H \) | — | — | — | -0.498 (0.303) | -1.188* (0.382) |
| \( d_H \) | — | — | — | 24.418* (4.869) | 47.971* (6.168) |
| \( p \) | — | — | 0 | 0 | 0 |

Wald Test Statistics (p-value) to test \( \theta_L = \theta_H \)

| Drift Function | — | 3.557 (0.169) | 3.540 (0.170) | 3.990 (0.136) | 3.590 (0.166) |
| Diffusion Function | — | 969.076 (0.000) | 963.893 (0.000) | 1070.866 (0.000) | 955.274 (0.000) |
| Transition Probability | — | 11.453 (0.000) | 12.520 (0.000) | 119.970 (0.000) | 372.883 (0.000) |
| RCM | — | 8.263 | 8.213 | 10.237 | 9.971 |

Maximum likelihood estimates with standard errors in parentheses for the five different models with the general diffusion specification are presented in this table. The standard errors are calculated using sample mean of the outer product of the score functions evaluated at the ML estimates. According to the number of regimes, the starting probability and the transition probability matrix specifications, the five models are: R1 (single regime), R2-1 (two regimes, time invariant transition matrix, unconditional probability for the probability of initial state), R2-2 (two regimes, time invariant transition matrix, additional parameter for the probability of initial state), R2TVTP-1 (two regimes, time varying transition matrix with Normal cumulative density function, additional parameter for the probability of initial state), and R2TVTP-2 (two regimes, time varying transition matrix with logistic function, additional parameter for the probability of initial state). In addition, the maximized likelihood values, RCMs are provided. Wald test statistic and its p-value are also calculated to jointly test if the parameters of the drift and diffusion functions and transition probabilities, separately, of two regimes are equal.
Weekly 3-Month U.S. Treasury Bill Rates (Panel A) and their weekly changes (Panel B) are displayed for the period from January 8, 1971 to December 26, 2003.
Panel A and B depict the transition probabilities of shifting from the state $L$ to the state $L$, and from the regime $H$ to the regime $H$ evaluated at the ML estimates over the support of the data, respectively, for the regime-switching model R2TVTP-2, whose regime-shifting probability is a logistic function of state variable. Panel C and D draw the same probabilities calculated at each observed interest rate over the time series data.
For the model R2TVTP-2 (logistic function for the transition probability), two time series plots of the filtered and smoothed probabilities are plotted in panel A and B, respectively. The former is the inferred probability that the economy was in the regime $H$ based on information available at time $t$, $P(s_t = H | I_t; \theta)$ and the latter is the inferred probability that the economy was in the regime $H$ using information available at time $T$, $P(s_t = H | I_T; \theta)$. We also overlay the NBER recessions periods (shaded area).
For the model R2-2 (time invariant transition probability), two time series plot of the filtered and smoothed probabilities are plotted in panel A and B, respectively. The former is the inferred probability that the economy was in the regime $H$ based on information available at time $t$, $P(s_t = H | I_t; \theta)$ and the latter is the inferred probability that the economy was in the regime $H$ using information available at time $T$, $P(s_t = H | I_T; \theta)$. We also overlay the NBER recessions periods (shaded area).
Using the ML estimates, the drift (first row) and the volatility (second row) terms over the region of the observed interest rates and the approximate conditional density functions (last row) given $x_0 = 0.05$ are drawn for the models R1 and R2TVTP-2. The 95% confidence band is obtained by delta method.
Panel A and B depict the transition probabilities of shifting from the state $L$ to the state $L$, and from the regime $H$ to the regime $H$ evaluated at the ML estimates over the support of the data, respectively, for the regime-switching model Vasicek-R2TVTP-2, whose regime-shifting probability is a logistic function of state variable.
Using the ML estimates, the volatility terms over the region of the observed interest rates are drawn for the models Models CIR-R2TVTP-2 and Vasicek-R2TVTP-2. The 95% confidence band is obtained by delta method.