Nonparametric Specification Tests of Discrete Time Spot Interest Rate Models in China

Yongmiao Hong
Department of Economics and Department of Statistical Science, Cornell University, Ithaca, NY 14850
Department of Economics, Tsinghua University, Beijing, China 100084(yh20@cornell.edu)

Hai Lin
Department of Finance, Xiamen University, Xiamen, China 361005(cfc@xmu.edu.cn)

Abstract

Understanding the dynamics of Chinese spot rates is very important for Chinese asset/derivatives pricing, risk management and interest rate liberalization. We examine a wide variety of popular spot rate models in China, including the single-factor diffusion models, GARCH models, Markov regime-switching models and jump-diffusion models. We describe the historical development of Chinese interest rate policy and Chinese bond markets, and fit these models to daily data of Chinese 7-day repo rates from July 22, 1996 to August 26, 2004. The estimation and specification testing suggest that introducing GARCH, Regime-switching and jump effect substantially improve the goodness of fit of the spot rate models. Regime-switching models and jump-diffusion models can effectively capture the excess kurtosis and heavy-tails of spot rate. Introducing the mean reversion also improves the in sample fit. However, unlike the findings of Hong, Li and Zhao (2004) which showed no evidence of Ait-Sahalia’s (1996) nonlinear drift contribution in USA, we find obvious evidence of nonlinear drift in China. By introducing a dummy variable for the sub-period 1996-1998, we find that the level, volatility of spot rates and the probability of jumps were significantly higher during that period. However, the sensitivity of interest rate change on interest rate level (level effect) became stronger after 1999. Another interesting observation from the estimation is that the jump-diffusion models seem to have the ability of capturing the sudden jump of interest rates according to Chinese IPO actions.

To compare different models further, we adopt the nonparametric model specification test recently proposed by Hong and Li (2005). The result is surprisingly similar to Hong and Li (2005): Although the GARCH, Regime-switching and jump-diffusion models improve the performance of spot rate models a lot, they are all overwhelmingly rejected at any reasonable confidence level. This suggests that although we have made great progress in modeling the dynamics of Chinese spot interest rate, we have not got an exact model yet.

Key Words: Spot Rate Models, Single-Diffusion Models, GARCH Models, Markov Regime-Switching Models, Jump-Diffusion Models, Nonparametric Specification Tests, Maximum Likelihood Estimation, Generalized residuals, Probability Integral Transform, Q-Stats.

JEL Classification: E4,G0,G12,

1. Introduction

Term Structure, namely the curve of spot risk free interest rates for different maturities, plays a fundamental role in modern economics and finance, especially in derivatives pricing and risk management. For example, to price a derivative, the expected payoff in the risk neutral world should be discounted by the correspondent risk free rate. To price fixed income securities, one should forecast the change of interest rate since their payoffs are based on the dynamics of interest rates.

The research on Chinese term structure of interest rate has other deeper theoretical and practical influence on the development of Chinese financial markets. Generally speaking, in such an emerging and market oriented market as China, the term structure, will play a role similar to FED fund rates in USA, and is fundamental in developing other financial markets, such as bond market and derivatives markets.

First, it gives a solid theoretical foundation for Chinese asset pricing. The interest rates provide the pricing benchmark not only for bonds (including government and corporate bonds), but also for many other derivatives, especially the fixed-income securities. For example, the price of convertible bond depends seriously on the exact estimation of interest rate dynamics. To price such assets and derivative, we have to well understand the probability distribution of interest rate dynamics.

Secondly, it helps to develop Chinese financial markets. In a complete market, the assets are priced reasonably and the arbitrage opportunities do not exist. A good estimation of Chinese interest rate dynamics reveals useful information on whether the asset prices are reasonable or not, provides some theoretical guidelines to reduce arbitrage opportunities and develop financial markets.

Thirdly, it provides some benchmark support for Chinese interest rate liberalization. Interest rate liberalization is the orientation of Chinese interest rate reform. How to choose the benchmark rate is a key job during the process of interest rate liberalization. Such benchmark rate should be based on the expectation of the investors, and have the ability of reflecting such expectations.

Fourthly, since the central bank controls the dynamics of short term interest rate to affect the long term interest rate by the transmission mechanism of term structure, understanding the term structure of interest rate provides some policy suggestions to choose suitable interest rate policy and pilot the investment.

Fifthly but not finally, it is also particularly important to model Chinese interest rates for investors. For example, by knowing the dynamics of interest rates in China, investors can well predict interest rate movements and take efficient hedging strategies.

Among the interest rates of different maturities on the term structure, the instantaneous spot rate is the state variable that determines the evolution of the term structure by the expectation hypothesis or liquidity premium hypothesis. For instance, according to expectation hypothesis, the long term interest rate equals the expected value of instantaneous spot rates integral during the correspondent future period. By liquidity premium hypothesis, the long term interest rate is equivalent to the expected value of instantaneous spot rates integral during the correspondent future period plus a liquidity premium.

Due to importance of instantaneous spot rate in the term structure, the academic has opened a separate research to model their dynamics in continuous time. Many spot rate models, such as Vasicek (1977) model and Cox, Ingersoll and Ross (CIR, 1985) model have been proposed and over the last decade, a vast body of literature has been developed to estimate and test these models.(see. e.g. Chan et al. (CKLS, 1992), Bali (1999) and Chapman and Pearson (2001) , Dai and Singleton (2003) for a survey of empirical literature.)

When studying the interest rates dynamics in USA, the academia first proposed many one factor-diffusion models (e.g. Vasicek (1977), CIR (1985), CKLS (1992)). The research suggested mean reversion of interest rate change and showed clear evidence of level effect. The marginal contribution of nonlinear drift by Ait-Sahalia (1996) is, however, ambiguous. Some researches (e.g. Ait-Sahalia (1996), Stanton (1997), Ahn and Gao (1999)) showed evidence of nonlinear drift while some others (e.g. Chapman and Pearson (2000), Pritsker (1998), Hong, Li and Zhao (2004)) doubted it.

Despite the popularity of one factor-diffusion models, they failed to explain some important stylized facts of interest rate behavior, such as non-normality, excess kurtosis and volatility clustering. In order to capture such characters of interest rates dynamics in USA, many other models were introduced, for example, stochastic volatility models(e.g. Anderson and Lund (1997), Gallant and Tauchen (1998))/GARCH models(e.g. Brenner, Harjes and Kroner (1996)), Markov Regime-switching models(e.g. Bansal and Zhou (2001), Gray (1996), Ball and Torous (1998), Ang and Bekaert (2002), Sanders and Unal (1988), Li and Xu (2002)) and jump-diffusion models(e.g. Baz and Das (1996) and Das (2002)). We may find some important facts when reviewing such models in USA.

First, it is important to model conditional heteroscedasticity through stochastic volatility/GARCH effect. Stochastic volatility/GARCH models significantly improve the in sample fit of one-factor diffusion models. Secondly, Regime switching and jumps help capture volatility clustering and especially the excess kurtosis and heavy-tails of the interest rate. Thirdly, once stochastic volatility/GARCH, regime-switching, or jump effects were introduced, the importance of modeling mean reversion in drift term decreased a lot. It was more important to correctly model the diffusion function than the drift function in fitting interest rate data (e.g. Bali (2001) and Durham (2003)). Introducing additional specification in drift term had little influence on in sample fit (e.g. Durham (2003)).

Therefore, the spot rates dynamics has been well investigated in USA, but not so in China. There has been no pioneering work on modeling the Chinese spot rates. This is mainly due to the relatively short history of Chinese bond markets, the strong regulation of Chinese interest rates, and the research focus on Chinese stock market. The main purpose of this paper is thus to investigate the statistical properties of Chinese spot rates. In particular, we are interested in whether Chinese spot rates share similar statistical features with USA spot rates.

In all, despite the importance of Chinese spot rate research and the progress that has been made in modeling and testing interest rate dynamics in developed countries such as USA, relatively little attention has been paid to China. Therefore, we have no idea that whether Chinese spot rates share similar statistical features with USA spot rates. Moreover, there is no guarantee that a model that well fits historical data of other countries will also perform well in China.

This paper examines a wide variety of popular spot rates models in China, including the single-factor diffusion models, GARCH model, Markov regime-switching models and jump-diffusion models. Our paper thus contributes to the literature by three aspects. First, we study and compare a wide variety of popular spot rates models rather than focus on one specific class. Second, our paper contributes to the relative few specification tests of discretized spot rate models by applying the nonparametric test procedures recently proposed by Hong and Li (2005). It also provides some econometric backgrounds and comparisons for different interest rate dynamics. Third, we provide the first comprehensive empirical analysis of a wide variety of popular spot rate models in China.

In Section 2, we would like to trace the historical development of Chinese term structure and Chinese bond markets. In section 3, we introduce and discuss the specification tests procedures recently proposed by Hong and Li (2005) in details. In Section 4, we introduce a wide variety of popular spot rate models. In Section 5, we describe data, estimation method, and the in-sample performance of each model. In Section 6, we subject each model to the specification tests evaluation, and we conclude in section 7.

2. Historical Development of Chinese Term Structure.

2.1 History of Chinese Term Structure

Chinese term structure had a relatively short history for about 20 years. In early 1980s, the marketlization degree of Chinese economy was relatively low, and the adjustment of macroeconomic was mainly determined by fiscal policy. The term structure during that period was almost blank except the regulated saving rates. With the transformation of Chinese economy institution, a lot of financial markets have been developed to meet the market demand since late 1980s and particularly in 1990s. In current status, the interest rate is gradually becoming an important tool in economic adjustment, risk management and asset pricing. However, due to the short history of Chinese market economy and the main focus on development of stock market, Chinese bond market and correspondent term structure are now quite incomplete. Moreover, they are artificially and unreasonably segmented.

For short terms, there are mainly two markets: Chinese inter-bank borrowing market and bond repurchase market. Chinese inter-bank borrowing markets began in 1980s at different places and were united into one on Jan, 1996. After Mar 1, 1996, a framework of two-level inter-bank borrowing market was proposed in China. The first level is the headquarters of 15 commercial banks and 35 financing centers, while the second level includes the correspondent bank branches and other non-bank financial organizations. The market has a uniform inter-bank borrowing rate in China, named as “CHIBOR”. In 1996, the upper limit of CHIBOR was cancelled to reflect the true demand and supply condition. The CHIBOR mainly consisted of short term interest rates, for example, 1 day and 7 day. The longest maturity is 4 month. In 2000, the percentage of 1 day and 7 day inter-bank borrowing in total inter-bank borrowing was 71.4%. Therefore, the CHIBOR mainly determines Chinese short term interest rate.

Chinese bond repurchase began in 1991 at some stock exchanges, for example, Shanghai Stock Exchange, Wuhan Stock Trading Center, Tianjin Stock Trading Center, and STAQ system (the later three trading centers were closed later). In 1997, to prevent banks from stock markets, The People’s Bank of China ordered all the commercial banks to quit from the bond repurchase on stock exchanges and opened another bond repurchase sub-market at inter-bank market. Therefore, there are two independent and segmented bond repurchase markets in China, the OTC market at inter-bank markets and the electronic trading market at stock exchanges. These two markets are artificially segmented with different interest rates for the same maturity.

Compared with inter-bank borrowing, the members taking part in the repurchase trading are more extensive. The repurchase is the mortgaged borrowing, whose credit risk is fewer than credit borrowing. As a result, the bond repurchase market is more active (after 1999, the trading volume of repurchase was much higher than that of inter-bank borrowing) and the interest rate is more stable, making it more representative as Chinese short term interest rate.

For long terms, Chinese government issued a lot of coupon bonds with maturity more than 8 years, which became an independent long term bond market. Most of the bonds in this market have terms more than 5 years, and reflect Chinese long term interest rate level. Similar to short term interest rate, there are two independent markets for the same bond, the OTC bond market at inter-bank market and the electronic trading market at stock exchange. We may often find the price for the same bond is different at different markets, which implies a possible pure arbitrage opportunity if the markets were opened to all and short sell was permitted.

The interest rates of middle maturities were decided and controlled rigorously by Chinese central bank. Such interest rates are not market oriented. They do not change every day to reflect the market but keep unchanged for a relatively long period. They only change when Chinese government uses them to adjust the macroeconomic situations, resulting in pure jump of saving rates. When reviewing the evolution of Chinese interest rate policy, we observed many interesting behavior which may be researched further in modern economics framework. For example, in order to encourage the saving of public, the government proposed a lot of favorable terms, one of which was to adopt "the interest rate protection" for the depositors. If the regulated saving rates increased during the period of saving, the saving rate changed and increased to the new higher level. However, if the interest rate decreased during this period, the depositing rate would not change. Therefore, the depositors could benefit from the increase of interest rates while have no loss when the interest rates decreased. Such protection was in fact an interest rate derivative-Guaranteed Interest Rate Option. Another similar protection was "Real Interest Rate Subsidy", which was mainly to protect the depositors from loss by inflation. In late 1980s and beginning 1990s, the inflation in China was so serious that the real saving rates became minus. The central government decided to subsidize the loss of saving from inflation, resulting in the same real interest rate as that before the inflation. Such protection was in fact another option-Guaranteed Real Interest Rate Option. In all, it is mainly planned mechanism that decides Chinese saving rates for middle maturities.

Therefore, the repurchase/inter-bank borrowing markets reflected the short term interest rate, while the long term bond market reflected long term interest rate. In the markets, the interest rates change every day. However, the mid-term interest rates are seriously regulated and controlled by the central bank through administrative mechanism.

2.2 Shortcomings of Chinese Term Structure and Recent Reforms

Despite the rapid progress the Chinese government has made to reform the interest rate policy and develop the financial markets, Chinese term structure is still far from complete. As we can see from historical introduction, there are two main deficiencies for Chinese current term structure. First, there exist two independent markets that have similar functions and trade the same products, inter-bank OTC market and exchange electronic market. The repo terms and bonds traded on two markets are almost the same. However, since they are artificially divided, it is very common to find that the same bond has different price at these two markets, which results in a totally different term structure between inter-bank market and exchange market. Therefore, although many kinds of estimation methods (e.g. spline approximation by McCulloch (1971)) can be applied to derive the market-implied term structure, there exists no uniform Chinese term structure.

Secondly, the deposit rates in China are rigorously regulated by the central bank. They have no ability to reflect the market demand and supply conditions of money. Violating such regulation to absorb deposit by higher rate is a serious financial crime in China. Therefore, there is a large gap between the regulated deposit rates and the market oriented interest rates, and many serious problems and arbitrage opportunities may arise. For example, if the deposit rate is lower than the market implied rate for the same term, some large investors would lend money from the banks to invest on bond market to construct an arbitrage portfolio.

Generally speaking, Chinese term structure is quite dispersed and composed of many different markets and decision mechanism. The bond markets are artificially divided into inter-bank market and exchange market, while the mid-term deposit rates are seriously regulated. Such problems will give more and more negative impact on development of Chinese economy and financial markets. For example, it is almost impossible to develop derivatives market without a uniformed market oriented term structure.

Fortunately, recognizing the importance of a uniform term structure, Chinese government has recently proposed many reforms. In order to unite the inter-bank market and exchange market, the government issued some bonds at both inter-bank market and exchange markets. Some eligible security companies and trust companies were permitted to enter the inter-bank market to take part in the issuing. At secondly markets, the central bank proposed the construction of market maker system in 2001, permitting some eligible banks to be the bid-ask quoter which had similar function as market maker.

Secondly, in order to construct a completely market-oriented term structure, gradually decrease the function of deposit rates in modern economy and thus push the liberalization process of Chinese interest rate, the government began to issue bonds systematically gradually from long term to short term. With the issuing and trading of bonds with different terms, a systemized bond market will be constructed to provide a robust benchmark for pricing and hedging. On the other hand, Chinese government will gradually decrease the use of regulated deposit rates and liberalize them in the end. Moreover, in order to develop financial market and apply the fundamental functions of bond markets, the government is now considering and proposing many other financial instruments, such as Bond Futures, Stock Index Futures and Monetary Market Fund (MMF). In all, although the current Chinese term structure in China is far incomplete, its development is in stable process.

3. Nonparametric Evaluation Method

When using discretely-sampled data to estimate the above mentioned continuous time models, it can result in inconsistent parameter estimates (Lo 1998). Therefore, the last decade has also seen the development of a large and still growing literature on estimation and testing of continuous-time methods, including the nonparametric method of Ait-Sahalia (1996), Stanton (1997) and Jiang and Knight (1997), the simulated method of moments of Duffie and Singleton (1993), the efficient method of moments (EMM) of Gallant and Tauchen (1996, 2001), the generalized method of moments of Hansen and Scheinkman (1995), and the maximum likelihood methods of Lo (1998) and Ait-Sahalia (2002). Asymptotic properties of these estimators have been well established and related inference procedures have been developed.

In contrast to the rapid development of estimation methods, there is relatively few literature on specification analysis for continuous time models. In a pioneering paper, Ait-Sahalia (1996) develops probably the first nonparametric test for stationary time-homogeneous one-factor diffusion models. Ait-Sahalia (1996) compares a model marginal density estimator with a nonparametric kernel density estimator based on a discretely sampled data. In an application to Eurodollar interest rates, Ait-Sahalia (1996) rejects all existing one-factor linear drift models using asymptotic theory and finds that the principal source of rejection of existing models is the strong nonlinearity of the drift. However, Pritsker (1998) documents that the size performance of Ait-Sahalia's (1996) test appears inadequate even for rather large samples: it requires 2,755 years of daily interest rate data generated by an empirically relevant Vasicek (1977) model to attain the accuracy of a kernel density estimator implied by its asymptotic distribution. The main reasons, as pointed out in Pritsker (1998), are the highly persistent dependence of the interest rate data and the slow convergence of nonparametric estimators. The asymptotic distribution of Ait-Sahalia's (1996) test statistic remains the same whether the sample is independent and identically distributed (i.i.d.) or persistently dependent. The level of persistence, however, severely affects the finite sample distribution.

In a related context, Gallant and Tauchen (1996) propose a minimum chi-square specification test for continuous-time models using EMM. Among other things, the greatest appeal of the EMM approach is that it applies to a wide range of stationary continuous-time processes, including both one-factor and multi-factor diffusion processes. In addition to the minimum chi-square test for over-identifying restrictions, the EMM approach also provides a spectrum of constructive individual t-statistics that are informative in revealing possible sources of model misspecification. However, Tauchen (1997) points out that it may have no power against certain alternatives, because the semi-nonparametric score function can have zero expectation under a mis-specified model distribution. Thus, strictly speaking, one still cannot conclude that a continuous-time model is correctly specified even when the minimum chi-square EMM test statistic is insignificant.

Hong and Li (2005) recently propose two new nonparametric transition density-based specification tests for continuous-time models. These tests share the appealing features of both Ait-Sahalia (1996) and Gallant and Tauchen (1996) nonparametric approaches, and have many additional nice properties. First, unlike Ait-Sahalia (1996) marginal density-based test, the tests are based on the transition density, which captures the full dynamics of a continuous-time process. Second, to achieve robustness to persistent dependence, the data is transformed via a dynamic probability integral transform using the model transition density, which is well known in statistics (e.g., Rosenblatt 1952) and is more recently used to evaluate out-of-sample density forecasts in discrete-time analysis (e.g., Diebold, Gunther and Tay 1998, Hong, Li and Zhao 2004). The transformed sequence is i.i.d. U [0, 1] under correct model specification, irrespective of the dependence structure of the original data. Third, to eliminate the well-known "boundary bias" of kernel estimators as documented in Chapman and Pearson (2000), a boundary-modified kernel is introduced. Fourth, to reduce the impact of parameter estimation uncertainty, a test based on the Hellinger metric is proposed. Fifth, the regularity conditions for asymptotic analysis are based on the model transition density rather than the stochastic differential equation of the underlying process. As a consequence, the tests are applicable to a vast variety of continuous-time and discrete-time dynamic models, such as GARCH/stochastic volatility models, regime-switching models, jump-diffusion models, and multi-factor diffusion models.

This paper uses the nonparametric tests proposed recently by Hong and Li (2005) to evaluate different spot rate models. Assuming the underlying process $\{X_t\}$ follows the following data generating process:

$$dX_t = \mu_0(X_t, t)dt + \sigma_0(X_t, t)dW_t \qquad (3.1)$$

where $\mu_0(X_t,t)$ and $\sigma_0(X_t,t)$ are the drift and diffusion functions respectively, and W_t is a standard Brownian motion. Let $p_0(x,t\mid y,s)$ be the true transition density of the diffusion process $\{X_t\}$, that is the conditional density of $X_t=x$ given $X_s=y, s<t$. For a given pair of drift and diffusion models $\mu(X_t,t,\theta)$ and $\sigma(X_t,t,\theta)$, a certain family of transition densities $\{p(x,t\mid y,s,\theta)\}$ is characterized. If a model is correctly specified, there exists some $\theta_0\in\Theta$ satisfying $\{p(x,t\mid y,s,\theta_0)=p_0(x,t\mid y,s)\}$ almost everywhere for some $\theta_0\in\Theta$. To test such a hypothesis, Hong and Li (2005) first transform the discretized data $\{X_{\tau\Delta}\}_{\tau=1}^n$ via a probability integral transform and define this discrete transformed sequence by

$$Z_\tau(\theta)\equiv\int_{-\infty}^{X_{\tau\Delta}} p[x,\tau\Delta\mid X_{(\tau-1)\Delta},(\tau-1)\Delta,\theta]dx, \tau=1,2,\ldots,n \qquad (3.2)$$

if the model is correctly specified, the exists some $\theta_0\in\Theta$ such that $p[x,\tau\Delta\mid X_{(\tau-1)\Delta},(\tau-1)\Delta,\theta_0)] = p_0[x,\tau\Delta\mid X_{(\tau-1)\Delta},(\tau-1)\Delta)]$ almost surely for all $\Delta>0$. Consequently, the transformed series $\{Z_\tau\equiv Z_\tau(\theta_0)\}_{\tau=1}^n$ is i.i.d. U[0,1] under correct specification. We call $\{Z_\tau(\theta)\}_{\tau=1}^n$ "generalized residuals" of the model $\{p(x,t\mid y,s,\theta)\}$. Here, i.i.d. U[0,1] property captures two important aspects of model specification; i.i.d. characterizes the correct specification of model dynamics, and U[0,1] characterizes correct specification of the model marginal distribution.

The test that whether $\{Z_\tau(\theta)\}_{\tau=1}^n$ follows i.i.d. U[0,1] is not a trivial task, because it is a joint hypothesis. The well-known Kolmogorov-Smirnov test checks U[0,1] under the i.i.d. assumption rather than test i.i.d. and U[0,1] jointly. It would miss the alternatives whose marginal distribution is uniform but not i.i.d. To make such joint hypothesis tests, Hong and Li (2005) develop two nonparametric tests of i.i.d.U[0,1] by comparing a kernel estimator $\hat{g}_j(z_1,z_2)$ for the joint density $g_j(z_1,z_2)$ of $\{Z_\tau,Z_{\tau-j}\}$ with unity, the product of two U[0,1] densities.

The kernel joint density estimator is for any integer $j>0$,

$$\hat{g}_j(z_1,z_2)=(n-j)^{-1}\sum_{\tau=j+1}^{n}K_h(z_1,\hat{Z}_\tau)K_h(z_1,\hat{Z}_{\tau-j}) \qquad (3.3)$$

where $K_h(x,y)=\begin{cases} h^{-1}k(\frac{x-y}{h})/\int_{-(x/h)}^{1}k(u)du, x\in[0,h) \\ h^{-1}k(\frac{x-y}{h}), x\in[h,1-h] \\ h^{-1}k(\frac{x-y}{h})/\int_{-1}^{(1-x)/h}k(u)du, x\in(1-h,1] \end{cases}$,

and the kernel $k(.)$ is a bounded symmetric probability density with support [-1,1] so that $\int_{-1}^{1}k(u)du=1, \int_{-1}^{1}uk(u)du=0,$ and $\int_{-1}^{1}u^2k(u)du<\infty$. One choice is the quartic kernel:

$$k(u)=\frac{15}{16}(1-u^2)\sum_{i=1}^{n}X_i^2 1(|u|\leq 1) \qquad (3.4)$$

where $1(|u|\leq 1))$ is the indicator function, taking value 1 if $|u|\leq 1$ and value zero otherwise.

$\hat{Z}_\tau=Z_\tau(\hat{\theta})$, and $\hat{\theta}$ is a $\sqrt{n}$-consistent estimator for θ_0. Like Scott (1992), $h=\hat{S}_Z n^{-1/6}$, where $\hat{S}_Z$ is the sample standard deviation of $\{Z_\tau\}_{\tau=1}^{n}$.

The first tests is based on a quadratic form between $\hat{g}_j(z_1,z_2)$ and 1, the product of two U[0,1] densities,

$$\hat{M}_1(j)\equiv\int_0^1\int_0^1[\hat{g}_j(z_1,z_2)-1]^2dz_1dz_2\,, \qquad (3.5)$$

and the first test statistic is a properly centered and scaled version of $\hat{M}_1(j)$:

$$\hat{Q}(j)\equiv[(n-j)h\hat{M}_1(j)-A_h^0]/V_0^{1/2}\,, \qquad (3.6)$$

where the non-stochastic centering and scale factors

$$A_h^0\equiv(h^{-1}-2)\int_{-1}^{1}k^2(u)du+2\int_0^1\int_{-1}^{b}k_b^2(u)dudb\,, \qquad (3.7)$$

$$V_0\equiv 2[\int_{-1}^{1}[\int_{-1}^{1}k(u+v)k(v)dv]^2du]^2\,, \qquad (3.8)$$

and $k_b(.)\equiv k(.)/\int_{-1}^{b}k(v)dv$.

Under correct model specification, Hong and Li (2004, Theorem 1) has shown that

$$\hat{Q}(j)\to N(0,1) \text{ in distribution.}$$

And under model misspecification,

$$\hat{Q}(j)\to\infty \text{ in probability}$$

whenever $\{Z_\tau, Z_{\tau-j}\}$ are not independent or U[0,1].(Hong and Li (2005), Theorem 3).

The quadratic form test $\hat{Q}(j)$, though convenient and quite accurate when the true parameter

θ_0 were known, might be adversely affected by any imprecise estimate for $\hat{\theta}$ in finite samples. To alleviate this problem, Hong and Li (2005) propose a second test based on the square Hellinger metric,

$$\hat{M}_1(j) \equiv \int_0^1 \int_0^1 [\sqrt{\hat{g}_j(z_1, z_2)} - 1]^2 dz_1 dz_2 \qquad (3.9)$$

which is a quadratic form between $\sqrt{\hat{g}_j(z_1, z_2)}$ and $\sqrt{1 \bullet 1} = 1$. The associated test statistic is

$$\hat{H}(j) \equiv [4(n-j)h\hat{M}_2(j) - A_h^0]/V_0^{1/2}, \qquad (3.10)$$

where A_h^0 and V_0 are as in (3.7) and (3.8). Under correct model specification, this test has the same asymptotic distribution as $\hat{Q}(j)$ and is asymptotically equivalent to $\hat{Q}(j)$ in the sense that $\hat{Q}(j) - \hat{H}(j) \to 0$ in probability. Under model misspecification, we also have $\hat{H}(j) \to \infty$ as $n \to \infty$ whenever $\{Z_\tau, Z_{\tau-j}\}$ are not independent or U[0,1].

We summarize the omnibus evaluation procedures following Hong and Li (2005): (i) estimate the parameters of discrete spot rate models using maximum likelihood estimation(MLE) method to yield a $\sqrt{n}$ -consistent estimator $\hat{\theta}$;(ii) compute the model generalized residual$\{\hat{Z}_\tau = Z_\tau(\hat{\theta})\}_{\tau=1}^n$, where $Z_\tau(\theta)$ is given in (3.2); (iii) compute the boundary-modified kernel joint density estimator $\hat{g}_j(z_1, z_2)$ in (3.3) for a pre-specified lag j, using a kernel in (3.4) and the bandwidth $h = \hat{S}_Z n^{-1/6}$, where $\hat{S}_Z$ is the sample standard deviation of the model generalized residual $\{\hat{Z}_\tau\}_{\tau=1}^n$; (iv) compute the test statistics $\hat{Q}(j)$ in (3.6) and $\hat{H}(j)$ in (3.10); (v) compare the value of $\hat{Q}(j)$ or $\hat{H}(j)$ with the upper-tailed N(0,1) critical value C_α at level α (e.g., $C_{0.05} = 1.645$). The upper-tailed rather than two sided N(0,1) critical values is suitable since negative values $\hat{Q}(j)$ and $\hat{H}(j)$ occur only under correct model specification when n is sufficiently large. Both of $\hat{Q}(j)$ and $\hat{H}(j)$ diverge to $+\infty$ when $\{Z_\tau, Z_{\tau-j}\}$ are not independent or $U[0,1]$ under model specification, granting the tests asymptotic unit power.

4. Spot Rate Models

We apply Hong and Li (2005) procedure to evaluate the in sample performance of a variety of populate spot rate models, including single-factor diffusion, GARCH, regime-switching, and jump-diffusion models. These models are now discussed in detail first.

4.1 Single-Factor Diffusion Model

One popular and important class of spot rate models is the single-factor diffusion model, which have been widely used in modern finance and fixed-income securities pricing. For many single-diffusion models, i.e. Vasicek model and CIR model, the prices of discounted bonds can reach a closed form solution, which gives a lot of convenience in pricing other interest rate derivatives.

Specifically, the spot rate is assumed to follow single-factor diffusion in continuous time,

$$dr_t = \mu(r_t, \theta)dt + \sigma(r_t, \theta)dW_t, \qquad (4.1)$$

where $\mu(r_t, \theta)$ and $\sigma(r_t, \theta)$ are the drift and diffusion functions, W_t is a standard Brownian motion. For diffusion models, $\mu(r_t, \theta)$ and $\sigma(r_t, \theta)$ completely determine the model transition density, which in turn captures the full dynamic of r_t.

Our paper evaluates a variety of single-factor diffusion models which are nested by Ait-Sahalia (1996) nonlinear drift model,

$$\mu(r_t, \theta) = \alpha_{-1} r_{t-1}^{-1} + \alpha_0 + \alpha_1 r_{t-1} + \alpha_2 r_{t-1}^2, \;\; \sigma(r_t, \theta) = \sigma r_{t-1}^{\rho},$$

In discrete approximation, the change of spot rate follows:

$$\begin{cases} \Delta r_t = \alpha_{-1} r_{t-1}^{-1} + \alpha_0 + \alpha_1 r_{t-1} + \alpha_2 r_{t-1}^2 + \sigma r_{t-1}^{\rho} z_t \\ \{z_t\} \sim iid.N(0,1) \end{cases} \qquad (4.2)$$

where $\Delta r_t = r_t - r_{t-1}$. Similar to Stanton (1997) and Das (2002), the discretization bias for daily data that we use in this article is not so significant. To consider different model specification, in (4.2), we allow the drift to have zero, linear, and nonlinear specification and allow the diffusion to be a constant or to depend on the interest rate level, which is referred to as the "level effect". The volatility specification in which the elasticity parameter ρ is estimated from the data is called the *constant elasticity variance* (CEV). The detailed single-diffusion models examined in our paper are listed in Table 1(a). The nonparametric specification tests and statistical value for the spot rate r_t is given from the model implied transition density $p(r_t, t \mid I_{t-1}, \hat{\theta})$, where $\hat{\theta}$ is a parameter estimator using the maximum likelihood estimate (MLE) method.

4.2 GARCH Models

Despite the convenience and popularity of single-diffusion models, many other studies (e.g. Brenner, Harjes and Kroner (1996), Anderson and Lund (1997)) have demonstrated that the single-factor diffusions failed to capture the well-known persistent volatility clustering of financial returns including interest rates. Brenner et al. (1996) introduced various GARCH models for volatility and showed that the GARCH models significantly outperformed single-factor diffusion models.

To evaluate the importance of GARCH effect in modeling spot rate, we examine six GARCH models as listed in Table 1(b), including three different drift specifications (zero, linear and nonlinear) and two volatility specifications (pure GARCH and combined CEV-GARCH). These models are nested by the following specification:

$$\begin{cases} \Delta r_t = \alpha_{-1} r_{t-1}^{-1} + \alpha_0 + \alpha_1 r_{t-1} + \alpha_2 r_{t-1}^2 + \sigma r_{t-1}^{\rho} \sqrt{h_t} z_t, \\ h_t = \beta_0 + h_{t-1}(\beta_2 + \beta_1 r_{t-2}^{2\rho} z_{t-1}^2), \\ \{z_t\} \sim iid.N(0,1) \end{cases} \tag{4.3}$$

Various GARCH models allow us to examine the contribution of drift term in modeling spot rate in the presence of GARCH and CEV, and to examine the marginal improvement of GARCH with respect to CEV. For identification, we set $\sigma = 1$ in all GARCH models.

4.3 Markov Regime-Switching Models

Due to the change of monetary policy, the business cycle and other macroeconomic conditions, the dynamics of interest rate change over time. The most popular approach to test this change is the Markov regime-switching model which was proposed by Hamilton (1989) to test the business cycle. Such a method has been widely used by Bansal and Zhou (2001), Gray (1996), Ball and Torous (1998), Ang and Bekaert (2002), Sanders and Unal (1988), Li and Xu (2002) and many others. Following most studies, we examine a class of two-regime models for the spot rates, where the latent state variable s_t follows a two state, first order Markov chain. We refer to the regime in which s_t =1 (2) as the first (second) regime. Following Ang and Bakaert (2002), the transition probability of $\{s_t\}$ is assumed to depend on the one-lagged spot rate level,

$$\Pr(s_t = l \mid s_{t-1} = l) = \frac{1}{1 + \exp(-c_l - d_l r_{t-1})}, \quad l = 1, 2 \tag{4.4}$$

Table 1(c) lists a variety of regime-switching models, all of which are nested by the following specification:

$$\begin{cases} \Delta r_t = \alpha_{-1}(s_t) r_{t-1}^{-1} + \alpha_0(s_t) + \alpha_1(s_t) r_{t-1} + \alpha_2(s_t) r_{t-1}^2 + \sigma(s_t) r_{t-1}^{\rho} \sqrt{h_t} z_t \\ h_t = \beta_0 + \beta_1 E[e(r_{t-2}, s_{t-1} \mid r_{t-2}, s_{t-2})]^2 + \beta_2 h_{t-1} \\ e(r_{t-2}, s_{t-1} \mid r_{t-2}, s_{t-2}) = [\Delta r_{t-1} - E(\Delta r_{t-1} \mid r_{t-2}, s_{t-1})] / \sigma(s_{t-1}) \\ \{z_t\} \sim iid.N(0,1) \end{cases} \tag{4.5}$$

Following above mentioned models, we consider three specifications of drift term: zero, linear and nonlinear drift, and three specifications of diffusion term: CEV, GARCH and CEV-GARCH. Therefore, we have a total number of nine regime-switching models. Different from Gary (1996), we use the same GARCH specification across different regimes. And unlike many previous studies that set the elasticity parameter to be 0.5, we allow it to be regime-dependent and estimate it from the data. Similarly, for identification, we set the diffusion constant $\sigma(s_t)$ =1 for $s_t = 1$ in the regime-switching models with GARCH effect.

It can be easily shown that the conditional density of the interest rate r_t in a regime switching model is

$$p(\Delta r_t \mid I_{t-1}) = \sum_{l=1}^{2} p(\Delta r_t \mid s_t = l, I_{t-1}) p(s_t = l \mid I_{t-1}), \qquad (4.6)$$

where the ex ante probability that the data are generated from regime l at time t, $p(s_t = l \mid I_{t-1})$, can be computed using Bayes' s rule via a recursive procedure described by Halmiton (1989). Therefore, the conditional density of regime-switching models is a mixture of two normal distributions, which have great flexibility in modeling skewness, kurtosis and heavy tails.

4.4 Jump-Diffusion Models

Various economic shocks, news announcement and the actions of central bank on financial markets, will undoubtedly influence the spot rate in a sudden way, resulting in the jump of interest rate. Baz and Das (1996) discussed the estimation of jump-diffusion model by likelihood function method. Das (2002) and Johannes (2003), among others, have shown that diffusion models including those with stochastic volatility cannot explain the excessive leptokurtosis exhibited by the changes of spot rates. Jump-diffusion models are a convenient way to generate such data.

Quite different from other countries, the Chinese IPO at the stock market has a large impact on the sudden change of spot rate in repurchase market, especially before 1999. The main reason behind this surprising phenomenon is the serious under-pricing of IPO stock in China. As a result, on the first listed day, the price of new issuing stock will increase more than 100%. Before 1999, the return from competing to buy IPO Stock at primary market and selling it immediately on the first listed day could be as high as 100%. Then when there is an IPO at the primary market, large institution investors will demand a large amount of money for a few days at a high rate, i.e. 20%, which results in a sudden jump of repo rate from a low level to a high level. After IPO, the spot rate falls immediately. This can be shown by the after mentioned figure of Chinese 7 days repo rate (fig.1). This is quite different from the behavior of interest rates in other countries, where the change of interest rate is relatively stable.

Similarly, we consider a class of discretized jump-diffusion models listed in Table 1(d). We consider zero, linear and nonlinear drift specifications. For volatility, we consider CEV, GARCH and combined CEV-GARCH specifications. These nine models are nested by the following specification:

$$\begin{cases} \Delta r_t = \alpha_{-1} r_{t-1}^{-1} + \alpha_0 + \alpha_1 r_{t-1} + \alpha_2 r_{t-1}^2 + \sigma r_{t-1}^{\rho} \sqrt{h_t} z_t + J(\psi, \gamma^2) \pi(q_t), \\ h_t = \beta_0 + \beta_1 [r_{t-1} - E(r_{t-1} \mid r_{t-2})]^2 + \beta_2 h_{t-1}, \\ \{z_t\} \sim iid.N(0,1), \\ \{\pi(q_t)\} \sim iid.Bernoulli(q_t), \\ J \sim N(\psi, \gamma^2), \end{cases} \qquad (4.7)$$

where J is the jump size and q_t is the jump probability with

$$q_t = \frac{1}{1 + \exp(-c - d r_{t-1})} \qquad (4.8)$$

The conditional density of above mentioned jump-diffusion models can be written as

$$
\begin{aligned}
&p(\Delta r_{t-1} \mid r_{t-1}) \\
&= (1-q_t)\frac{1}{\sqrt{2\pi\sigma^2(r_{t-1})}}\exp\{\frac{-[\Delta r_t-\mu(r_{t-1})]^2}{2\sigma^2(r_{t-1})}\} \\
&+q_t\frac{1}{\sqrt{2\pi[\sigma^2(r_{t-1})+\gamma^2]}}\exp\{\frac{-[\Delta r_t-\mu(r_{t-1})-\psi]^2}{2[\sigma^2(r_{t-1})+\gamma^2]}\}
\end{aligned} \qquad (4.9)
$$

where $\mu(r_{t-1})$ and $\sigma^2(r_{t-1})$ are the conditional mean and variance of diffusion specification in (4.7). For example, for linear drift CEV jump-diffusion model (linear drift JD CEV model), $\mu(r_{t-1})=\alpha_0+\alpha_1 r_{t-1}$, $\sigma^2(r_{t-1})=\sigma r_{t-1}^{2\rho}$. Similar to regime-switching models, the conditional density of jump-diffusion models is also a mixture of two normal distributions. However, the specifications of regime-switching models are more sophisticated. In (4.5), all drift parameters are regime dependent, whereas in (4.7) only the intercept is different in conditional mean and variance.

[Insert Table 1.]

5. Model Estimation and In-Sample Performance

5.1 Data and Estimation Method

In modeling the dynamics of spot rate, yields on different short term debts are used as proxies, such as 1 month T-bill rates used by Gray (1996) and Chan, Karolyi, Longstaff and Sanders (CKLS, 1992) and Hong, Li and Zhao (2004), the 3 month T-bill rates used by Stanton (1997) and Anderson and Lund (1997), the 7 day Eurodollar rates used by Ait-Sahalia (1996) and Hong and Li (2002) and the Fed funds rates used by Conley, Hansen, Luttmer and Scheinkman (1997) and Das (2002). In this paper we use the repo rate of Chinese bond repurchase market. As we can see in the discussion of section 2, the repo rate is more representative than CHIBOR as short term interest rate. We use the daily data of 7-day repo rate from July 22, 1996 to August 26, 2004, with a total of 1954 observations. However, because of the influence of holiday on the repurchase time, the original data do not represent the true 7 day repo rates. For instance, one 7-day repurchase buyer will generally repurchase the bond at pre-specified price in 7 days. However, if in 7 days the market is closed due to holiday and other reasons, the repurchase is delayed to the next immediate opening day, while the repurchase price and total interest keep unchanged. Thus, the investor could use the fund for more than 7 days while only paying 7 day interest. Since this information is public, the 7 day repo rate will increase to counteract the delay of repurchase and interest payout. Therefore, to study the dynamic behavior of exact 7 day repo rates, we have to transform the original data to eliminate such effect. The transformation method is written by:

$$\tilde{r}=\frac{r\times 7}{t}$$

where $\tilde{r}$ is exact 7 repo rate after transformation, r is the 7 day repo rate listed on exchange, t is the true repurchase term. It is easily shown that if $t=7$, $\tilde{r}=r$, the marked 7 day repo rate is the exact 7 day repo rate. However, if $t>7$, $\tilde{r}<r$, the marked 7 day repo rate is higher than the exact 7 day repo rate.

Figure 1 plots the level and change series of the transformed daily 7-day repo rates, as well as their histograms. There is obvious persistent volatility clustering, and in general, the volatility is higher at higher level of interest rates (before 1999). There is a systematic change on repo rate behavior after 1999. The marginal distribution of the interest rate level is skewed to the right, with a long right tail. Most daily changes in the interest rate level are very small, with sudden jump. And quite different from the data in other countries such as USA in which the interest rates change stably with some trends, the repo rates in China show more jump behaviors, with small rate changes following sudden jump. Similar to other spot rate data, daily changes of repo rates also exhibit a high peak around 0. Such stylized facts imply excess kurtosis, and has motivated the use of jump models in the literature (Das (2002), Johannes (2003)).

[Insert fig.1]

In estimation, the optimization algorithm is the well-known BHHH with STEPBT for step length calculation and is implemented via the constrained optimization code in GAUSS windows Version 6.0. The optimization tolerance level is set such that the gradients of parameters are less than or equal to 10^{-6}. Tables 2-5 report parameters for four classes of spot rate models. To account for the different period before 1999, we introduce dummy variables to the drift, volatility, and elasticity parameters, that is, $\alpha_D, \sigma_D, \rho_D$ are 0 after 1999.

5.2 In-Sample Performance

To compare the performance of different classes, we discuss their estimations in details first. Table 2 reports the parameter estimates with estimated robust standard errors and log-likelihood values for discretized single-factor diffusion models. The estimate of the drift parameters of Vaiscek, CIR and CKLS models all suggest mean-reversion in the conditional mean, with a long run mean estimated around 2.7% (estimate of $-\alpha_0 / \alpha_1$). For other models such as random walk and nonlinear drift models, some drift parameters are not significant. For Dot model, the parameters are significant but the log-likelihood is the smallest. This is consistent with the estimation result in USA. A comparison of the pure CEV, CKLS and nonlinear drift models of Ait-Sahalia (1996) indicates the marginal contribution of nonlinear drift was obvious, although much less than the linear drift. By adding the linear drift term in pure CEV, the log likelihood increases from 5680.69 to 5881.61. And by adding the nonlinear drift term, the log likelihood increase to 5995.76. This is different from Hong, Li and Zhao (2004) which found that the marginal contribution of nonlinear drift is small in USA. On the other hand, there is also a clear evidence of level effect; all estimates of the elasticity parameter are significant. But unlike the previous studies (e.g. CKLS) which found and estimated elasticity parameter close to 1.5 and Hong, Li and Zhao (2003) whose estimate was about 0.25, our estimate is about 0.5, which is consistent with the CIR model of square root rule. Our estimates, among other studies such as Brenner et al. (1996), Anderson and Lund (1997), Bliss and Simth (1998), confirm that the elasticity parameter is very sensitive to the choice of interest rate data, data frequency, sample periods and specifications of volatility function. The estimates of dummy variable between 1996 and 1998 suggest that both drift and volatility behave quite differently during that period. For Vasicek, CIR, CKLS and nonlinear drift models that have higher log likelihood value, the drift dummy α_D is significantly positive, suggesting a higher interest rate behavior before 1999. For Vasicek and CIR models, the volatility dummy σ_D is significantly positive. For Pure CEV, CKLS and nonlinear drift models, the level dummy ρ_D is significantly negative. Therefore, the volatility between 1996 and 1998 is significantly higher. However, the sensitivity of interest rate change on interest rate level becomes stronger after 1999. There may be two main reasons for such interesting observations. On the one hand, the borrowing and lending of short term money was mainly through inter-bank before 1999. The trading of repurchase market was not active. However, since 1999, the trading of repurchase market has increased a lot and exceeded the inter-bank market. Repurchase market replaced the inter-bank market as the dominant market of short term financing for large institution investors. The short term financing of such large institution investors was more influenced by the correspondent interest rate level. Therefore, the sensitivity of interest rate change on interest rate level increased. On the other hand, Chinese SEC (Security Regulation Commission) proposed a lot of reforms on IPO mechanism and made serious regulation on the flowing of bank money into stock market after 1999. The degree of IPO under-pricing decreased gradually (in current status, there have been some stocks whose prices of first listed day decreased to below IPO prices). Therefore, the demand of large amount of money for the pure arbitrage at Chinese first stock market decreased, resulting in the decrease of interest rate volatility and jump possibility. The dependence of interest rate volatility on interest rate level thus increased.

[Insert Table 2.]

Estimation results of GARCH models are listed in Table 3. The comparison of log-likelihood value with discretized single-factor diffusion models show that adding the GARCH effect significantly improves the in-sample fit of singe-factor diffusion models. The log-likelihood increases from less than 6000 to more than 6200. In Table 3, all estimates of GARCH parameters are significant. The sum of GARCH parameter estimates, $\hat{\beta}_1 + \hat{\beta}_2$, is slightly larger than 1 without the level effect. When considering the level effect, $\hat{\beta}_1 + \hat{\beta}_2$ increases to some degree. However, it is still possible that the spot rate model is strictly stationary (see Nelson (1991) for more detail discussion). The level effect in GARCH model is also significant with an estimate about 0.3, which is smaller than that of single-factor diffusion (0.5). The specification of conditional variance affects the estimate of drift parameters. However, the drift parameters are still significant under GARCH model, suggesting a mean-reversion trend. This is different from the estimation results in USA where mean reversion decreased immediately after introducing GARCH effect. Most of dummy variables α_D, σ_D are significantly positive, suggesting a higher interest rate level and higher volatility before 1999. Furthermore, the specification of conditional mean and variance affects the estimate of dummy variables a lot. For no drift GARCH-CEV and linear drift GARCH-CEV models, α_D is insignificant and ρ_D is significantly negative. However, After adding the nonlinear drift specification in GARCH-CEV models, the dummy variables α_D, ρ_D become significantly positive, suggesting higher interest rate level and more dependence on interest rate level during 1996-1998. This is quite different from single-factor diffusion models for which the dummy variable ρ_D is significantly negative. Among all GARCH models considered, the model with nonlinear drift and level effect has the best in sample performance. The marginal contribution of nonlinear drift is important.

[Insert Table 3.]

Parameter estimation results of Markov regime-switching models are listed in Table 4. The results show that spot rate behaves quite differently between regimes. For models of linear drift, both regimes show mean reversion, with one higher and one lower long-run mean. However, the specification of conditional variance affects the long run mean a lot. For linear drift CEV model, the higher long run mean is 6.73% and the lower long run mean is 2.76%. For linear drift GARCH model, the higher long run mean is 0.82% and the lower long run mean is 0.22%. Considering both CEV and GARCH give higher long run mean 1.56% and lower long run mean 0.25%. All the GARCH parameters are significant, and the sum of parameters $\hat{\beta}_1 + \hat{\beta}_2$ are slightly (much) larger than 1 without (with) level effect. The level effect in two regimes is also significant with or without GARCH effect, with one regime about 0.5 and the other about 1.5. Such estimation results of level effect parameters are higher and more stable than the estimation results in USA. The specification of conditional variance affects the estimate of volatility a lot. For CEV models, the volatility of one regime is about 8 times of the other. For GARCH models, it is about 5 times. However, for CEV-GARCH models, the relationship is quite unstable, depending on the specification of drift function. For no drift function, it is 2 times, but for linear and nonlinear drift function, it is 6 times and 3 time respectively. However, the ratio keeps quite stable in USA. The relationship between volatility and level effect is also quite unstable. For CEV models and no drift CEV-GARCH model, higher volatility is related to higher level effect, i.e. the regime with higher volatility has higher dependence on interest rate level. However, for linear drift CEV-GARCH and nonlinear drift CEV-GARCH models, higher volatility is related to lower level effect. Compared with GARCH models, the Markov regime-switching models have much higher log-likelihood, suggesting the improvement of in sample fit by considering the regime-switching effect. This is different from the estimation result in USA. Among all Markov regime-switching models considered, the models with level effect performed better than GARCH effect. This is also in contrast with the estimation results in USA. Considering both level effect and GARCH effect has no much help on improving the in sample fit. The models with nonlinear drift have the largest log-likelihood, although some parameters are insignificant.

[Insert Table 4.]

Table 5 lists the parameter estimation results of discretized jump-diffusion models. The mean reversion is still significant, with long-run mean about 2.34%. All GARCH parameters are significant, with sum smaller than 1. The GARCH parameters are smaller than pure GARCH models, confirming the part explanation of volatility clustering by jump. Without GARCH effects, the level effect is more than 1.5.However, if the GARCH effects are considered, the level effect decreases to less than 0.2 immediately. Thus GARCH specifications help capture the volatility clustering of interest rates.

Different from Markov regime-switching models, the parameters of transition probability for jumps are overwhelmingly significant under GARCH and CEV-GARCH specification. The specification of conditional mean and conditional variance both affect the estimate of jump size. For CEV models, the jump size is high at about 1%. With the GARCH effects, the jump size decreases a lot to some insignificant value. Considering both CEV and GARCH effect also gives similar results. This is different from the estimation results of USA spot rates. It is reasonable to give such a result since the interest rate not only jump up, but also jump down, which offsets to a large degree. The volatility parameters in all specifications keep stable at about 2.3%. The dummy variables α_D, σ_D are significant in all models, suggesting a higher interest rate level and a higher volatility before 1999. Without GARCH effect, the ρ_D before 1999 is significant. However, under GARCH specification, it is insignificant, suggesting the functions of GARCH in capture interest rate clustering. Both the dummy variables for jump probability before 1999 are significant, suggesting a different jump behavior before 1999. Similar to Markov regime-switching models, the models with CEV perform a little better than those of GARCH specification. This is in contrast with the estimation results in USA. Considering both CEV and GARCH has no help on improving the in sample fit. Again, the models with nonlinear drift have the largest log-likelihood, although some parameters are insignificant.

[Insert Table 5.]

An interesting question for testing the jump-diffusion models in China is the relationship between conditional jump probability and the large spot rate change, which is mainly brought by the huge capital demand for arbitrage at IPO market by large investors. It is impossible to result in such large increase of spot rate by other factors. To check this, we compute the conditional jump probability and compare it with the change of spot rate in China. Figure 2 figures out the conditional probability of linear drift CEV model as an example (the conditional probabilities of other models are similar). It is intuitively shown that there is a systematic difference before 1999 and after that. The conditional jump probability is much higher before 1999, reflecting more sudden interest rate change. This is quite reasonable since before 1999, the under-pricing of IPO stocks was so serious that large investors borrowed money from repo market at an unreasonable high rate to arbitrage at IPO first market. After 1999, such phenomena decreased gradually since the SEC in China reformed the IPO mechanism. However, there still exists some profit for arbitrage at IPO market, resulting in fewer sudden jump in spot rate after 1999. Another interesting observation is that there is a straight relationship between the sudden increase of spot rate in China and the conditional jump probability. After 1999, the figure of difference for 7-day repo rate series and that of conditional jump probability is quite similar. Therefore, the jump-diffusion model in China help to capture the large change of spot rate resulted by IPO actions. This is just an intuitive explanation, and the exact relationship is for future research.

To sum up, our in-sample discussion reveals some important stylized facts for the spot rates in China:

(1) The importance of modeling mean reversion is significant. Although some parameters are insignificant, considering the nonlinear drift can improve the in-sample fit. Furthermore, the specification of conditional mean can affect the estimate results of other parameters such as volatility and level effect. The contribution of Ait-Sahalia’s (1996) nonlinear drift is evident. These are different from the estimation results in USA.

(2) It is important to model conditional heteroscedasticity through GARCH or level effect. Considering both GARCH effect and level effect have no help on improving in sample fit. Quite different from the estimation results in USA, the models with level effects perform better than the models with GARCH effect. Table 6 lists some important differences between the estimation results of Chinese spot rates and USA spot rates.

(3) Regime switching and jump help capture volatility clustering and especially the excess kurtosis and heavy-tails of interest rates.

(4) Interest rates behave quite differently during the period of 1996-1998. The level/ volatility of interest rates and the probability of jumps seem significantly higher during this period. However, the dependence of the interest rate volatility on the interest rate level becomes stronger after 1999.

(5) There is an intuitively straight relationship between the jump behavior of interest rates and IPO impact in China. This seems to be captured by the jump-diffusion models.

6. Nonparametric Specification Tests.

In this section, to do the specification tests, we follow the test procedures of Hong and Li (2005) and compute the relevant $\hat{Q}(j)$ for $j=1,5,10$ for each class of spot rate models, which is listed in Table 7 (the results of $\hat{H}(j)$ tests are quite similar).

Table 7(a) reports the $\hat{Q}(j)$ test statistics as function of lag order $j=1,5,10$ for the singe-diffusion models. As shown in the table 7(a), the $\hat{Q}(j)$ statistics for the eight models range from 287.81 to 14471.64. Compared with the upper tailed N(0,1) critical value (e.g. 2.33 at the 1% level), the large $\hat{Q}(j)$ statistics are overwhelmingly significant, suggesting that all eight models are severely mis-specified at any reasonable significance level. The lognormal model performs the worst among the eight models, with $\hat{Q}(j)$ values increasing with lag to more than 14000. This is different from the comparison results of likelihood value in which the Dot model has the smallest likelihood value. The CIR model dramatically reduce the $\hat{Q}(j)$ of Vasicek and the goodness of fit is further improved by nonlinear drift model. Introducing the level effect in CIR increases the errors and suggests that the level effect is unimportant for modeling interest rate dynamics. This is quite different from the comparison of likelihood value in which CKLS model has larger likelihood value than CIR model. Consistent with Ait-Sahalia (1996) which found evidence of nonlinear drift, the nonlinear drift CEV model performs the best. In all, the extremely large test statistics for all the eight single-factor diffusion models indicate that none of them can adequately capture the interest rate dynamics.

Table 7(b) reports the $\hat{Q}(j)$ test statistics as function of lag order $j=1,5,10$ for the GARCH models. As shown in the Table 6(b), the $\hat{Q}(j)$ statistics for the six models range from 126.01 to

505.39. Compared with the stats of single-factor diffusion models, the GARCH models significantly reduce the stats, showing evidence of GARCH effect in modeling spot rates. However, compared with the upper tailed N(0,1) critical value (e.g. 2.33 at the 1% level), the large $\hat{Q}(j)$ statistics are also overwhelmingly significant, suggesting that all six models are mis-specified at a reasonable significance level. The no drift model performs the worst among the six models, with $\hat{Q}(j)$ values more than 500. Introducing the linear drift decreases the $\hat{Q}(j)$ value a lot and improves the fit of models furthermore. However, considering the level effect has little help on improving the performance. This suggests that only GARCH or level effect is necessary to model the spot rates. This is quite different from the comparison of likelihood value. Moreover, different from the important marginal contribution of nonlinear drift in single-factor diffusion models, the $\hat{Q}(j)$ values of nonlinear drift models are larger than linear models, suggesting that they perform worse than the linear drift models.

Table 7(c) reports the $\hat{Q}(j)$ test statistics as function of lag order $j=1,5,10$ for the Markov regime-switching models. As shown in the Table 7(c), the $\hat{Q}(j)$ statistics for the nine models range from 11.66 to 53.10. Compared with the stats of single-factor diffusion models and GARCH models, the $\hat{Q}(j)$ stats of regime-switching models decrease tremendously, suggesting that the interest rates behave quite differently at different times. However, compared with the upper tailed N(0,1) critical value (e.g. 2.33 at the 1% level), the large $\hat{Q}(j)$ statistics are still overwhelmingly significant, and all the nine specifications are refused at reasonable significance level. The no drift model performs the worst among the nine models, with $\hat{Q}(j)$ values more than 50. Introducing the linear drift decreases the $\hat{Q}(j)$ value a lot and improves the fit of models. The nonlinear drift models have some help on improving the performance, however it is much less than linear drift. The joint introduction of level and GARCH effect also improve the performance of models a little. The models with level effect (linear drift and nonlinear drift) have smaller $\hat{Q}(j)$ value than the models with GARCH effect, suggesting that the level effect is better for capturing the volatility clustering than GARCH models. This is consistent with the observation from likelihood value comparisons.

Table 7(d) reports the $\hat{Q}(j)$ test statistics as function of lag order $j=1,5,10$ for the jump-diffusion models. As shown in the Table 7(d), the $\hat{Q}(j)$ statistics for the nine models range from 18.11 to 51.75, which is quite similar as that of Markov regime-switching Models. Therefore, regime-switching models and jump-diffusion models are two efficient alternative ways to capture the excess kurtosis of interest rates. Compared with the stats of single-factor diffusion models and GARCH models, the $\hat{Q}(j)$ stats of jump-diffusion models decrease tremendously too, suggesting that the interest rates have some significant jumps. However, compared with the upper tailed N(0,1) critical value (e.g. 2.33 at the 1% level), the large $\hat{Q}(j)$ statistics are all overwhelmingly significant, and all the nine specifications are refused at reasonable significance level. The no drift model performs the worst among the nine models, with $\hat{Q}(j)$ values more than 50. Introducing the linear drift decreases the $\hat{Q}(j)$ value a lot and improves the fit of models. There is a small marginal improvement (for CEV models) when the nonlinear drift is introduced. Under the conditions of no drift and linear drift, the $\hat{Q}(j)$ values of CEV models are not smaller than GARCH models, although they have larger likelihood value. The joint introductions of level and GARCH effect have no help in improving the performance.

[Insert Table 7]

To sum up, our specification tests also reveal some important findings in modeling Chinese spot rates. Some of these results are similar to the comparison results of likelihood value, while others are quite different.

(1) Although introducing the GARCH/level effect, regime-switching effect and jump effect can improve the performance of a variety of popular spot rate models a lot in China, they are all refused by the $\hat{Q}(j)$ tests at any reasonable significance level, suggesting that they are still grossly mis-specified. There is a very long way to go before obtaining an adequate specification form for Chinese interest rate dynamics.

(2) Although both the level and GARCH effect can help capture the volatility clustering of interest rate, the joint introduction is not necessary.

(3) The introduction of linear drift improves the performance a lot, while the marginal contribution of nonlinear drift is much smaller. For the single-factor diffusion models, the marginal contribution of nonlinear drift is obvious. However, if the GARCH, regime-switching or jump effect are introduced, the evidence becomes small. GARCH, regime-switching and jump models help capture the nonlinear effect in interest rate dynamics.

6. Conclusion

Using discrete data in China, we have applied the nonparametric specification tests of Hong and Li (2005) to a variety of popular spot rate models estimated by MLE method. The models considered are discretized single-factor diffusion models, GARCH models, Markov Regime-Switching Models and Jump-Diffusion Models. Introducing GARCH significantly improves the in sample fit. Regime-Switching and Jump effects help capturing volatility clustering and especially the excess kurtosis and heavy-tails of interest rates. Moreover, compared with the stylized facts in USA, the contribution of nonlinear drift is more evident, and the elasticity of level effect is much higher.

Although the traditional log-likelihood comparison showed the importance of GARCH, regime-switching and jump in modeling the interest rate dynamics, the modern nonparametric specification test found that the models with GARCH effect, regime-switching effect and jump effect are still grossly mis-specified, while much more slightly than single-factor diffusion. All these imply that there is still a long way before we reach the adequate specification for interest rate dynamics. This is for future research. Another specific and interesting question for future research is to test the relationship between the IPO actions at stock market and the interest rate jump which can be detected by jump-diffusion models.

Reference:

Ahn, D., and B. Gao (1999), "A Parametric Nonlinear Model of Term Structure Dynamics", *Review of Financial Studies* 12, 721-762.

Ait-Sahalia ,Y. (2002), "Maximum-Likelihood Estimation of Discretely Sampled Diffusions: A Closed-Form Approach," *Econometrica* 70, 223 – 262.

Ait-Sahalia, Y. (1996), "Testing Continuous-Time Models of the Spot Interest Rate," *Review of Financial Studies* 9, 385 – 426.

Anderson, T.G. and J. Lund (1997), "Estimating Continuous-Time Stochastic Volatility Models of the Short- Term Interest Rate", *Journal of Econometrics* 72, 343-377.

Ang, A., and G. Bekaert (2002), "Regime Switches in Interest Rates," *Journal of Business & Economic Statistics* 20, 163 – 182.

Bali, T.G.(1999), "An Empirical Comparison of Continuous Time Models of the Short Term Interest Rate", *The Journal of Futures Markets* 19, 777-197.

Ball, C., and W. Torous (1998), "Regime Shifts in Short-Term Riskless Interest Rates," unpublished manuscript, *University of California Los Angeles*.

Bansal, R., and H. Zhou (2001), "Term Structure of Interest Rate with Regime Shifts", *Working paper of Duke University.*

Baz, J. and S.R. Das (1996), "Analytical Approximation of the Term Structure for Jump-Diffusion Process: a Numerical Analysis", *Journal of Fixed Income* 6, 78-86.

Black, F. and M. Scholes (1973), "The pricing of option and corporate liabilities," *Journal of Political Economy* 81, 637-659.

Bliss, R., and D. Smith (1998), "The Elasticity of Interest Rate Volatility: Chan, Karolyi, Longstaff, and Sanders Revisited," unpublished manuscript, *Federal Reserve Bank of Atlanta.*

Brenner, R. J., R. H. Harjes, and K. B. Kroner (1996), "Another Look at Models of Short-Term Interest Rate", *Journal of Financial and Quantitative Analysis* 31, 95-107.

Chan, K. C., G. A.Karolyi, F.A.Longstaff, and A.B.Sanders (1992), "An Empirical Comparison of Alternative Models of the Short-Term Interest Rate", *Journal of Finance*, 47,1209-1227.

Chapman, D., and N. Pearson (2001), "Recent Advances in Estimating Models of the Term-Structure," *Financial Analysts Journal* 57, 77 – 95.

Conley, T. G., L. P. Hansen, E. G. J. Luttmer, and J. A. Scheinkman (1997), "Short-Term Interest Rates as Subordinated Diffusions," *Review of Financial Studies* 10, 525 – 578.

Corradi, V. (2000), "Reconsidering the Continuous-Time Limit of the GARCH(1, 1) Process," *Journal of Econometrics* 96, 145 – 153.

Cox, J. C., Ingersoll, J. E., and Ross, S. A. (1985), "A Theory of the Term Structure of Interest Rates," *Econometrica* 53, 385 – 407.

Dai, Q. , and K. Singleton (2003), "Term Structure Dynamics in Theory and Reality", *Review of Financial Studies* 16, 631-678.

Das, S. R. (2002), “The Surprise Element: Jumps in Interest Rates,” *Journal of Econometrics* 106, 27 − 65.
Diebold, F. X., T. A. Gunther, and A. S. Tay (1998), “Evaluating Density Forecasts With Applications to Financial Risk Management,” *International Economic Review* 39, 863 − 883.
Duffie, D., and K. Singleton (1993), “Simulated Moments Estimation of Markov Models of Asset Prices”, *Econometrica* 61, 929-952.
Gallant, A. R., and G. Tauchen (1998), “Reprojection Partially Observed Systems With Applications to Interest Rate Diffusions,” *Journal of the American Statistical Association* 93, 10 − 24.
Gallant, A.R. and G. Tauchen (1996), “Which Moments to Match?”, *Econometric Theory* 12, 657-681.
Gallant, R. and G. Tauchen (2001), “Efficient Method of Moments”, Working paper, *Department of Economics, University of North Carolina and Department of Economics, Duke University*.
Gray, S. (1996), “Modeling the Conditional Distribution of Interest Rates as a Regime-Switching Process,” *Journal of Financial Economics* 42, 27 − 62.
Hamilton, J. D. (1989), “A New Approach to the Econometric Analysis of Nonstationary Time Series and the Business Cycle,” *Econometrica* 57, 357 − 384.
Hansen, L.P. and J.A. Scheinkman (1995), “Back to the Future: Generating Moment Implications for Continuous Time Markov Processes”, *Econometrica* 63, 767-804.
Hong, Y., and H. Li(2002), ???.
Hong, Y., and H. Li (2004), “Nonparametric Specification Testing for Continuous-Time Models With Applications to Interest Rate Term Structures,” *Review of Financial Studies*, ???, ??? − ???.
Hong, Y., H. Li and F. Zhao (2004), “Out-of-Sample Performance of Discrete-Time Spot Interest Rate Models”, *Journal of Business & Economics Statistics*, forthcoming.
Houweling, P., J. Hoek, and F. Kleibergen (2001), “The Joint Estimation of Term Structure and Term Spread”, *Journal of Empirical Finance* 8, 297-323.
Jiang, G. and J. Knight (1997), “A Nonparametric Approach to the Estimation of Diffusion Processes with an application to a short-term interest rate model”, *Econometric Theory* 13, 615-645.
Johannes, M. (2003), “The Statistical and Economic Role of Jumps in Interest Rates,” *Journal of Finance*, ???, ??? − ???.
Li, H., and Y. Xu (2000), “Regime Shifts and Short Rate Dynamics,” unpublished manuscript, *Cornell University*.
Lo, A.W.(1988) “Maximum Likelihood Estimation of Generalized Ito Processes with Discretely Sampled Data”, *Econometric Theory 4*, 231-247.
Mcculloch, J.H.(1971),” Measuring the Term Structure of Interest Rates”, *Journal of Business* 44, 19-31.
Nelson, D. (1990), “ARCH Models as Diffusion Approximation,” *Journal of Econometrics* 45, 7 − 39.
Nelson, D. (1991), “Conditional Heteroskedasticity in Asset Returns: A New Approach,” *Econometrica* 59, 347 − 370.
Pritsker, M.(1998), “Nonparametric Density Estimation and Tests of Continuous Time Interest Rate Models”, *Review of Financial Studies* 11, 449-487.
Rosenblatt, M. (1952), “Remarks on a Multivariate Transformation,” *The Annals of Mathematical Statistics* 23, 470 − 472.
Sanders, A.B., and H. Unal (1988), “On the Intertemporal Behavior of the Short-Term Rate of

Interest", *Journal of Financial and Quantitative Analysis* 23, 417-423.
Scott, D.W. (1992), Multivariate Density Estimation, Theory, Practice, and Visualization, *JohnWiley and Sons,* New York.
Stanton, R. (1997), "A Nonparametric Model of Term Structure Dynamics and the Market Price of Interest Rate Risk", *Journal of Finance* 52, 1973-2002.
Tauchen, G. (1997), "New Minimum Chi-Square Methods in Empirical Finance", in: D. Kreps and K.Wallis, eds., *Advances in Econometrics: Seventh World Congress* (Cambridge University Press, Cambridge).
Vasicek, O. (1977), "An Equilibrium Characterization of the Term Structure," *Journal of Financial Economics* 5, 177 − 188.

Table 1. Spot Rate Models Considered for Evaluation

Model	Mean	Variance
	(a) Discretized single-factor diffusion models	
Random walk	α_0	σ
Log-normal	$\alpha_1 r_{t-1}$	σr_{t-1}
Dothan	0	σr_{t-1}
Pure CEV	0	σr_{t-1}^{ρ}
Vasicek	$\alpha_0 + \alpha_1 r_{t-1}$	σ
CIR	$\alpha_0 + \alpha_1 r_{t-1}$	$\sigma r_{t-1}^{0.5}$
CKLS	$\alpha_0 + \alpha_1 r_{t-1}$	σr_{t-1}^{ρ}
Nonlinear drift	$\alpha_{-1} / r_{t-1} + \alpha_0 + \alpha_1 r_{t-1} + \alpha_2 r_{t-1}^2$	σr_{t-1}^{ρ}
	(b) GARCH models	
No drift GARCH	0	$\sqrt{h_t}$
Linear drift GARCH	$\alpha_0 + \alpha_1 r_{t-1}$	$\sqrt{h_t}$
Nonlinear drift GARCH,	$\alpha_{-1} / r_{t-1} + \alpha_0 + \alpha_1 r_{t-1} + \alpha_2 r_{t-1}^2 + \sigma r_{t-1}^{\rho} z_t$	$\sqrt{h_t}$
No drift CEV-GARCH	0	$r_{t-1}^{\rho}\sqrt{h_t}$
Linear drift CEV-GARCH	$\alpha_0 + \alpha_1 r_{t-1}$	$r_{t-1}^{\rho}\sqrt{h_t}$
Nonlinear drift CEV GARCH	$\alpha_{-1} / r_{t-1} + \alpha_0 + \alpha_1 r_{t-1} + \alpha_2 r_{t-1}^2 + \sigma r_{t-1}^{\rho} z_t$	$r_{t-1}^{\rho}\sqrt{h_t}$
	(c)Markov regime-switching models	
No drift RS CEV	0	$\sigma(s_t) r_{t-1}^{\rho(s_t)}$
Linear drift RS CEV	$\alpha_0 + \alpha_1 r_{t-1}$	$\sigma(s_t) r_{t-1}^{\rho(s_t)}$
Nonlinear drift RS CEV,	$\alpha_{-1} / r_{t-1} + \alpha_0 + \alpha_1 r_{t-1} + \alpha_2 r_{t-1}^2 + \sigma r_{t-1}^{\rho} z_t$	$\sigma(s_t) r_{t-1}^{\rho(s_t)}$
No drift RS GARCH	0	$\sigma(s_t)\sqrt{h_t}$
Linear drift RS GARCH	$\alpha_0 + \alpha_1 r_{t-1}$	$\sigma(s_t)\sqrt{h_t}$
Nonlinear drift RS GARCH	$\alpha_{-1} / r_{t-1} + \alpha_0 + \alpha_1 r_{t-1} + \alpha_2 r_{t-1}^2 + \sigma r_{t-1}^{\rho} z_t$	$\sigma(s_t)\sqrt{h_t}$

No drift RS CEV GARCH	0	$\sigma(s_t)r_{t-1}^{\rho(s_t)}\sqrt{h_t}$
Linear drift RS CEV GARCH	$\alpha_0+\alpha_1 r_{t-1}$	$\sigma(s_t)r_{t-1}^{\rho(s_t)}\sqrt{h_t}$
Nonlinear drift RS CEV GARCH	$\alpha_{-1}/r_{t-1}+\alpha_0+\alpha_1 r_{t-1}+\alpha_2 r_{t-1}^2+\sigma r_{t-1}^{\rho}z_t$	$\sigma(s_t)r_{t-1}^{\rho(s_t)}\sqrt{h_t}$
	(d) Discretized jump-diffusion models	
No drift JD CEV	0	σr_{t-1}^{ρ}
Linear drift JD CEV	$\alpha_0+\alpha_1 r_{t-1}$	σr_{t-1}^{ρ}
Nonlinear drift JD CEV,	$\alpha_{-1}/r_{t-1}+\alpha_0+\alpha_1 r_{t-1}+\alpha_2 r_{t-1}^2+\sigma r_{t-1}^{\rho}z_t$	σr_{t-1}^{ρ}
No drift JD GARCH	0	$\sqrt{h_t}$
Linear drift JD GARCH	$\alpha_0+\alpha_1 r_{t-1}$	$\sqrt{h_t}$
Nonlinear drift JD GARCH	$\alpha_{-1}/r_{t-1}+\alpha_0+\alpha_1 r_{t-1}+\alpha_2 r_{t-1}^2+\sigma r_{t-1}^{\rho}z_t$	$\sqrt{h_t}$
No drift JD CEV GARCH	0	$r_{t-1}^{\rho}\sqrt{h_t}$
Linear drift JD CEV GARCH	$\alpha_0+\alpha_1 r_{t-1}$	$r_{t-1}^{\rho}\sqrt{h_t}$
Nonlinear drift JD CEV GARCH	$\alpha_{-1}/r_{t-1}+\alpha_0+\alpha_1 r_{t-1}+\alpha_2 r_{t-1}^2+\sigma r_{t-1}^{\rho}z_t$	$r_{t-1}^{\rho}\sqrt{h_t}$

NOTE: The eight discretized single-factor diffusion models are nested by the following specification: $\Delta r_t=\alpha_{-1}/r_{t-1}+\alpha_0+\alpha_1 r_{t-1}+\alpha_2 r_{t-1}^2+\sigma r_{t-1}^{\rho}z_t$, where $\{z_t\}\sim iid.N(0,1)$. The six GARCH models are nested by the following specifications: $\Delta r_t=\alpha_{-1}/r_{t-1}+\alpha_0+\alpha_1 r_{t-1}+\alpha_2 r_{t-1}^2+\sigma r_{t-1}^{\rho}\sqrt{h_t}z_t$, where $h_t=\beta_0+h_{t-1}(\beta_2+\beta_1 r_{t-2}^{2\rho}z_{t-1}^2)$ and $\{z_t\}\sim iid.N(0,1)$. The nine regime-switching models are nested by the following specification: $\Delta r_t=\alpha_{-1}(s_t)/r_{t-1}+\alpha_0(s_t)+\alpha_1(s_t)r_{t-1}+\alpha_2(s_t)r_{t-1}^2+\sigma(s_t)r_{t-1}^{\rho}\sqrt{h_t}z_t$,where $h_t=\beta_0+\beta_1 E\{e(r_{t-2},s_{t-1}\mid r_{t-2},s_{t-2})\}^2+\beta_2 h_{t-1}$, $e(r_{t-2},s_{t-1}\mid r_{t-2},s_{t-2})$, $=[\Delta r_{t-1}-E(\Delta r_{t-1}\mid r_{t-2},s_{t-1})]/\sigma(s_{t-1})$, $\{z_t\}\sim iid.N(0,1)$ and s_t follows a two state, first order Markov chain with transition probability $p(s_t=l\mid s_t=l)=[1+\exp(-c_l-d_l r_{t-1})]^{-1}$ for $l=1,2$. The nine discretized jump-diffusion models are nested by the following specification: $\Delta r_t=\alpha_{-1}/r_{t-1}+\alpha_0+\alpha_1 r_{t-1}+\alpha_2 r_{t-1}^2+\sigma r_{t-1}^{\rho}\sqrt{h_t}z_t+J(\psi,\gamma^2)\pi(q_t)$, where $h_t=\beta_0+\beta_1[r_{t-1}-E(r_{t-1}\mid r_{t-2})]^2+\beta_2 h_{t-1}$, $\{z_t\}\sim iid.N(0,1)$, $J\sim N(\psi,\gamma^2)$, and $\pi(q_t)$ is Bernoulli(q) with $q=[1+\exp(-c-dr_{t-1})]^{-1}$.

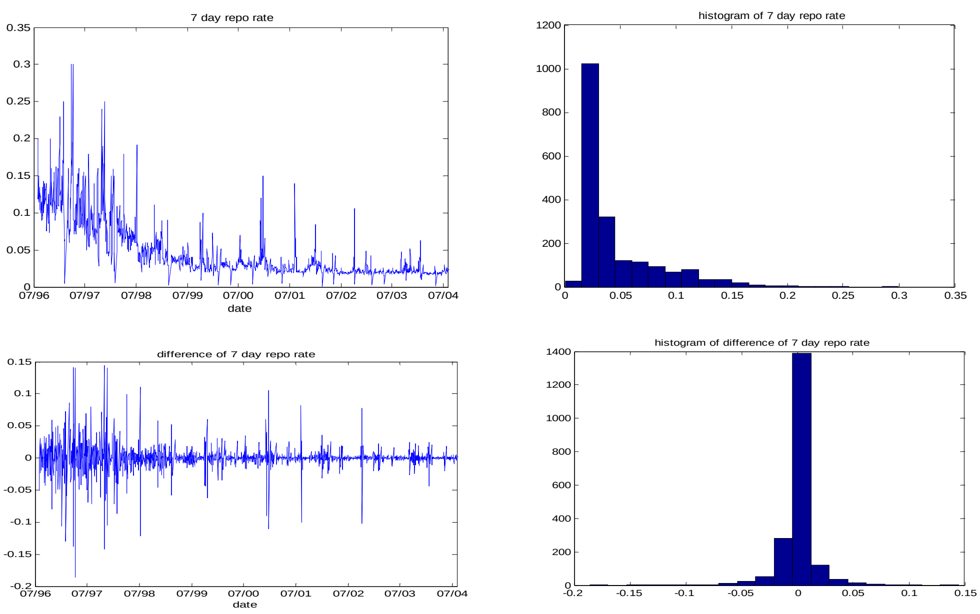

Figure 1. Daily 7-day repo rates between July 22, 1996 and August 26, 2004. This figure plots the level and

change series of daily data as well as their histograms.

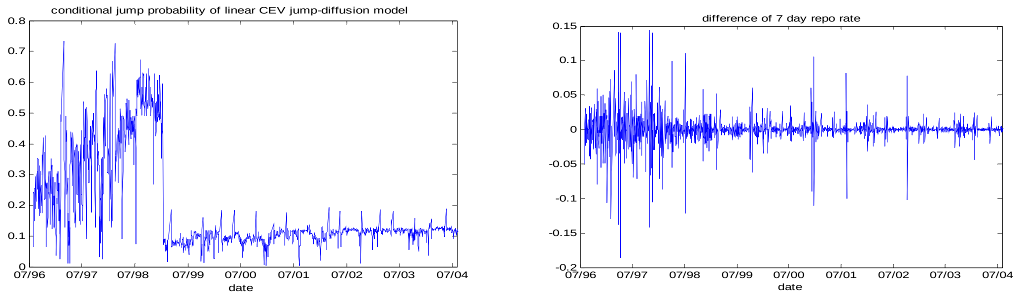

Figure2. Daily conditional jump probability for Linear Drift CEV Jump-Diffusion Model computed by $\hat{q}+\hat{q}_D=[1+\exp(-\hat{c}-\hat{c}_D-(\hat{d}+\hat{d}_D)r_{t-1})]^{-1}$ where the parameters are estimated from MLE method. The figure plots the conditional jump probability and difference of 7-day repo rates.

Table2. Parameter Estimates for the Single-Factor Diffusion Models

Parameters	RW	Log-normal	Dot	PCEV	Vasicek	CIR	CKLS	Nonlinear drift
α_{-1}								2.2E-5
								(2.00E-06)
α_0	-1.40E-05	01021			9.74E-03	0.013	0.0104	6.799E-03
	(2.81E-04)	(0.03072)			(5.34E-04)	(4.66E-04)	(5.22E-04)	(9.44E-04)
α_1					-0.3664	-0.45	-0.375	-0.2818
					(0.017696)	(0.01867)	(0.0193)	(0.04153)
α_2								-0.09359
								(0.2546)
σ	0.0104	1.1319	1.1365	0.0347	0.0092	0.0533	0.0358	0.0532
	(0.0002)	(0.0218)	(0.0219)	(0.0016)	(0.0002)	(0.001)	(0.0023)	(0.0036)
ρ				0.3434			0.4085	0.5177
				(0.0114)			(0.0154)	(0.0182)
α_D	-2.53E-04	-0.02841			0.02862	0.0247	0.0221	0.01803
	(1.21E-03)	(0.04871)			(1.599E-03)	(1.37E-03)	(1.55E-03)	(1.466E-03)
σ_D	0.0186	-0.2001	-0.2018		0.0182	0.042		
	(0.0009)	(0.0346)	(0.0347)		(0.0008)	(0.003)		
ρ_D				-0.2276			-0.2421	-0.2173
				(0.0148)			(0.0156)	(0.0140)
Log-likelihood	5525.11	3605.10	3597.75	5680.69	5722.50	5854.59	5881.61	5995.76

NOTE: The eight models are nested by the following specification: $\Delta r_t = \alpha_{-1}/r_{t-1} + (\alpha_0 + \alpha_D) + \alpha_1 r_{t-1} + \alpha_2 r_{t-1}^2 + (\sigma + \sigma_D) r_{t-1}^{\rho+\rho_D} z_t$, **where** $\{z_t\} \sim iid.N(0,1)$.

Table 3. Parameter Estimates for GARCH Models

Parameters	No drift	linear drift	Nonlinear drift	No drift CEV	linear drift CEV	Nonlinear drift CEV
α_{-1}			4.70E-05			4.3E-05
			(7.0E-06)			(4.00E-06)
α_0		2.734E-03	-3.401E-03		2.584E-03	-4.356E-03
		(3.67E-04)	(9.24E-05)		(3.49E-04)	(7.06E-04)
α_1		-0.1206	0.08445		-0.1082	0.1613
		(0.01702)	(0.03704)		(0.0156)	(0.0307)
α_2			-1.1289			-1.8788
			(0.2765)			(0.1988)
ρ				0.1588	0.1593	0.3078
				(0.0254)	(0.0321)	(0.0278)
β_0	1.994E-06	4.06E-06	1.826E-06	7.786E-06	1.5142E-05	1.4982E-05
	(3.22E-07)	(6.63E-07)	(3.42E-07)	(2.124E-06)	(4.423E-06)	(4.077E-06)
β_1	0.3968	1.0224	0.6296	1.1617	2.9405	3.744
	(0.0303)	(0.1203)	(0.0670)	(0.2207)	(0.7119)	(0.8011)
β_2	0.7120	0.4677	0.6416	0.7066	0.4552	0.7165
	(0.0125)	(0.0429)	(0.0247)	(0.0138)	(0.0358)	(0.0132)
α_D		-1.271E-03	-3.857E-03		-2.38E-4	0.01169
		(7.57E-04)	(8.20E-04)		(1.81E-03)	(1.53E-03)
σ_D	0.3620	0.1397	0.1039			
	(0.0576)	(0.0551)	(0.0504)			
ρ_D				-0.0708	-0.1526	0.0874
				(0.0174)	(0.0730)	(0.02)
Log-likelihood	6200.75	6279.06	6306.81	6214.98	6290.89	6349.94

NOTE: The six GARCH models are nested by the following specifications: $\Delta r_t = \alpha_{-1}/r_{t-1} + (\alpha_0 + \alpha_D) + \alpha_1 r_{t-1} + \alpha_2 r_{t-1}^2 + (1+\sigma_D) r_{t-1}^{\rho+\rho_D} \sqrt{h_t} z_t$, **where** $h_t = \beta_0 + h_{t-1}(\beta_2 + \beta_1 r_{t-2}^{2\rho} z_{t-1}^2)$ **and** $\{z_t\} \sim iid.N(0,1)$

Table 4. Parameter Estimates for Markov Regime-Switching Models

Parameters	No drift CEV	linear drift CEV	Nonlinear drift CEV	No drift GARCH	linear drift GARCH	Nonlinear drift GARCH	No drift CEV-GARCH	Linear drift CEV-GARCH	nonlinear drift CEV-GARCH
$\alpha_{-1}(1)$			2.04E-04			2.7E-05			5.00E-06
			(3.5E-05)			(2.8E-05)			(5.00E-06)
$\alpha_0(1)$		1.40E-3	-0.01521		6.73E-04	-5.08E-04		0.00106	2.123E-03
		(2.64E-04)	(3.379E-03)		(1.39E-04)	(3.03E-04)		(1.92E-04)	(3.53E-04)
$\alpha_1(1)$		-0.0508	0.3026		-0.08167	0.3911		-0.06787	-0.9640
		(0.01146)	(0.095)		(0.02446)	(0.0945)		(0.03122)	(0.1914)
$\alpha_2(1)$			-1.6047			-3.0240			11.9651
			(0.6995)			(0.5742)			(2.3512)
$\alpha_{-1}(2)$			2.2E-05			6.2E-05			2.47E-04
			(9.00E-06)			(1.20E-05)			(2.4E-05)
$\alpha_0(2)$		0.01911	-0.009521		1.5E-04	-2.18E-04		2.18E-04	-1.98E-03
		(2.079E-3)	(0.005431)		(2.50E-05)	(1.21E-04)		(2.80E-05)	(2E-04)
$\alpha_1(2)$		-0.2839	-0.01512		-0.0681	-0.04969		-0.08878	0.4468
		(0.0572)	(0.1920)		(0.0111)	(0.03787)		(0.01212)	(0.0475)
$\alpha_2(2)$			-1.6792			0.6761			-2.9128
			(1.0398)			(0.2973)			(0.2599)
$\rho(1)$	1.4795	1.4386	1.4947	0	0	0	0.4329	0.2712	0.5454
	(0.0407)	(0.0537)	(0.0381)				(0.0413)	(0.0351)	(0.0545)
$\rho(2)$	0.2419	0.4150	0.3518	0	0	0	0.7162	1.3156	1.3632
	(0.0375)	(0.0405)	(0.0682)				(0.0457)	(0.0541)	(0.0461)
$\sigma(1)$	0.6353	0.5353	06206	1	1	1	1	1	1
	(0.08816)	(0.1014)	(0.08143)						
			To be continued			To be continued			

σ (2)	0.06599	0.1077	0.0803	0.2008	0.2233	0.2102	0.4408	6.1191	3.545
	(9.96E-03)	(0.01619)	(0.02099)	(0.0103)	(0.0117	(0.0126)	(1.4866)	(1.1199)	(0.7657)
β_0				7.8E-05	6.578E-06	7.11E-06	3.80E-03	2.27E-03	1.216E-01
				(1.7E-05)	(1.372E-06)	(1.56E-06)	(1.24E-03)	(7.92E-04)	(5.54E-02)
β_1				0.0929	0.1132	0.1098	26.4987	6.4414	100.1910
				(0.0144)	(0.0170)	(0.0209)	(6.6496)	(2.6340)	(37.6058)
β_2				0.7857	0.7674	0.7676	0.2438	0.1506	0.1263
				(0.0169)	(0.0190)	(0.0237)	(0.0397)	(0.1419)	(0.1309)
c_1	-2.9052	-2.8861	-2.7275	-0.1056	-0.1996	0.4781	-2.2637	-1.0593	-1.3487
	(0.2234)	(0.2232)	(0.2937)	(0.2454)	(0.2504)	(0.4356)	(0.6635)	(0.3381)	(0.2912)
d_1	7.6123	8.0743	5.9252	-8.1197	-9.0335	-17.2068	116.76	19.0164	29.7506
	(5.0422)	(5.0622)	(8.6452)	(2.4242)	(2.8395)	(4.0973)	(27.84)	(7.2729)	(5.5602)
c_2	-0.8963	-0.5430	-0.5028	-3.0802	-3.1372	-3.1915	-2.6187	-3.396	-2.1102
	(0.2961)	(02949)	(0.2922)	(0.2261)	(0.2446)	(0.2638)	(0.1839)	(0.3297)	(0.1669)
d_2	8.7697	0.2587	-1.8408	27.1278	25.1505	30.0579	11.8228	37.90	-51.826
	(5.319)	(4.8145)	(5.3644)	(4.7078)	(5.4329)	(5.735)	(3.9583)	(10.2606)	(22.995)
Log-likelihood	6877.59	6928.47	6962.83	6835.64	6869.30	6929.35	6877.51	6935.50	6997.92

NOTE: The nine regime-switching models are nested by the following specification: $\Delta r_t = \alpha_{-1}(s_t)/r_{t-1} + \alpha_0(s_t) + \alpha_1(s_t) r_{t-1} + \alpha_2(s_t) r_{t-1}^2 + \sigma(s_t) r_{t-1}^{\rho} \sqrt{h_t} z_t$ **,where** $h_t = \beta_0 + \beta_1 E\{e(r_{t-2}, s_{t-1} | r_{t-2}, s_{t-2})\}^2 + \beta_2 h_{t-1}$, $e(r_{t-2}, s_{t-1} | r_{t-2}, s_{t-2})$, $= [\Delta r_{t-1} - E(\Delta r_{t-1} | r_{t-2}, s_{t-1})] / \sigma(s_{t-1})$, $\{z_t\} \sim iid.N(0,1)$ **and** S_t **follows a two-state, first order Markov chain with transition probability** $p(s_t = l | s_t = l) = [1 + \exp(-c_l - d_l r_{t-1})]^{-1}$ **for** $l = 1, 2$.

Table 5. Parameter Estimates for the Jump-Diffusion Models

Parameters	No drift CEV	linear drift CEV	Nonlinear drift CEV	No drift GARCH	linear drift GARCH	Nonlinear drift GARCH	No drift CEV-GARCH	Linear drift CEV-GARCH	nonlinear drift CEV-GARCH
α_{-1}			4.80E-05			4.6E-05			5.8E-05
			(1.20E-06)			(5.00E-06)			(6.00E-06)
α_0		3.043E-03	-4.25E-04		2.476E-03	-1.72E-03		2.546E-03	-2.668E-03
		(3.47E-04)	(1.256E-03)		(3.09E-04)	(3.15E-04)		(3.17E-04)	(7.46E-04)
α_1		-0.1298	-0.08296		-0.1097	1.59E-04		-0.1131	0.02135
		(0.01599)	(0.04243)		(0.01399)	(6.507E-03)		(0.01439)	(0.02960)
α_2			0.3124			-0.8377			-0.9654
			(0.3446)			(0.1536)			(0.2557)
ρ	1.9596	1.8560	1.9306				-0.0236	0.0506	0.1762
	(0.0642)	(0.0660)	(0.0624)				(0.0408)	(0.0413)	(0.0493)
σ	3.9957	2.5962	3.3771						
	(0.9343)	(0.6291)	(0.7733)						
β_0				2.68E-06	2.652E-06	1.152E-06	2.203E-06	3.89E-06	4.977E-06
				(2.93E-07)	(3.51E-07)	(1.83E-07)	(7.62E-07)	(1.4E-06)	(2.131E-06)
β_1				0.4163	0.3922	0.2514	0.3628	0.5908	0.7879
				(0.0508)	(0.0537)	(0.0364)	(0.1143)	(0.1882)	(0.2947)
β_2				0.0811	0.1043	0.4913	0.0827	0.1040	0.49
				(0.0301)	(0.0471)	(0.0393)	(0.0340)	(0.0454)	(0.0409)
c	0.6953	1.4108	1.4129	5.2362	4.8305	4.9892	5.2974	4.6979	4.6341
	(0.4924)	(0.441)	(0.4506)	(0.4391)	(0.4403)	(0.4909)	(0.4492)	(0.4551)	(0.4858)
			To be continued			To be continued			

d	62.5886	29.7391	28.7330	−119.84	−101.1267	−92.7187	−121.5257	−95.3798	−80.86
	(22.9778)	(19.4544)	(19.7320)	(14.66)	(15.3808)	(16.5926)	(14.9623)	(16.0539)	(16.8226)
ψ	0.0093	0.0102	0.0087	−0.0017	0.0022	0.0028	−0.0019	0.0024	0.0045
	(0.0015)	(0.0016)	(0.0015)	(0.0010)	(0.0014)	(0.0024)	(0.0011)	(0.0014)	(0.0023)
γ	0.0211	0.0208	0.0205	0.0234	0.0231	0.0260	0.0238	0.0231	0.0253
	(0.0012)	(0.0012)	(0.0015)	(0.0011)	(0.0011)	(0.0016)	(0.0011)	(0.0011)	(0.0016)
α_D		3.825E-03	2.818E-03		5.214E-03	5.178E-03		5.313E-03	5.077E-03
		(1.04E-03)	(6.63E-04)		(6.51E-04)	(6.89E-04)		(6.56E-04)	(7.50E-04)
ρ_D	0.3624	0.2862	0.3194				−0.0870	−0.0369	−0.0011
	(0.0526)	(0.0573)	(0.0491)				(0.0326)	(0.0352)	(0.0753)
σ_D				0.2699	0.1987	0.0818			
				(0.1049)	(0.1215)	(0.0907)			
c_D	−2.4901	−2.5047	−2.9520	−2.6224	−1.7552	−1.5609	−2.5494	−1.6449	−1.5242
	(0.6878)	(0.6925)	(0.6713)	(0.7303)	(0.7994)	(0.8322)	(0.7476)	(0.8115)	(0.8943)
d_D	−30.0531	−10.8175	−1.4542	85.95	58.958	56.1302	85.8984	53.1498	49.1961
	(23.7532)	(20.6163)	(21.1031)	(16.37)	(17.8585)	(18.2188)	(16.5754)	(18.3853)	(18.3187)
Log-likelihood	6873.72	6916.74	6933.99	6846.53	6896.21	6931.18	6846.31	6896.21	6935.81

NOTE: The nine discretized jump-diffusion models are nested by the following specification: $\Delta r_t = \alpha_{-1}/r_{t-1} + (\alpha_0 + \alpha_D) + \alpha_1 r_{t-1} + \alpha_2 r_{t-1}^2 + (\sigma + \sigma_D) r_{t-1}^{(\rho+\rho_D)} \sqrt{h_t} z_t + J(\psi, \gamma^2)\pi(q + q_D)$ **, where** $h_t = \beta_0 + \beta_1 [r_{t-1} - E(r_{t-1} \mid r_{t-2})]^2 + \beta_2 h_{t-1}$, $\{z_t\} \sim iid.N(0,1)$, $J \sim N(\psi, \gamma^2)$ **, and** $\pi(q)$ **is Bernoulli(q) with** $q + q_D = [1 + \exp(-c - c_D - (d + d_D) r_{t-1})]^{-1}$.

Table 6. Difference of Estimation Results in Chinese Spot Rate and USA Spot Rate

China	USA
(a) Single-factor diffusion models	(a) Single-factor diffusion models
1. The contribution of nonlinear drift is evident. 2. The estimate of elasticity is about 0.5	1. the contribution of nonlinear drift is ambiguous：Ait-Sahalia (1996), Stanton (1997), Ahn and Gao (1999) show the existence，Hong, Li & Zhao (2004)， Chapman and Pearson (2000), Pritsker (1998) question；2. elasticity estimate: CKLS (1992)-1.5，Hong, Li & Zhao (2004)-0.25。
(b) GARCH models	(b) GARCH models
Mmean reversion is still significant after the introduction of GARCH	Mean reversion decreases rapidly after introduction of GARCH
(c) Markov regime-switching models	(c) Markov regime-switching models
1.the elasticity for two regimes are 1.5 and 0.5; 2. mean reversion is still significant in two regimes；3. the volatility ratios are unstable in two regimes: for CEV models it is about 8 times, for GARCH models it is about 5 times, for CEV-GARCH models it is unstable; 4. The relationship between volatility and level effect is unstable: for CEV models higher volatility is related to stronger level effect, for CEV-GARCH models it is unstable; 5. CEV models have larger likelihood value than those of GARCH models.	1. the elasticity for two regimes are 0.8 and 0.1; 2. mean reversion is significant in only one regimes；3. the volatility ratios are relatively stable in two regimes: for CEV models it is about 30 times, for GARCH models it is about 4 times, for CEV-GARCH models it is about 3 times；4. The relationship between volatility and level effect is relatively stable: for CEV models higher volatility is related to weaker level effect, for CEV-GARCH models higher volatility is related to stronger level effect it is unstable. 5. GARCH models have larger likelihood value than those of CEV models.
(d)Jump-diffusion models	(d)Jump-diffusion models
1.without GARCH, the elasticity is 1.5; with GARCH, it decreases to about 0.2; 2. The jump size for GARCH models is smaller than that of CEV models; 3. CEV models have larger likelihood value than those of GARCH models.	1. without GARCH, the elasticity is 0.9; with GARCH, it decreases to about 0.1; 2. The jump size for GARCH models is larger than that of CEV models.3. GARCH models have larger likelihood value than those of CEV models.

Table 7 . $\hat{Q}(j)$ Stats of Discrete Spot Rte Models

(a) Discretized Single-Factor Diffusion Models

Lag	RW	Log-normal	Dot	PCEV	Vasicek	CIR	CKLS	Nonlinear drift
1	535.61	5728.87	820.22	460.88	403.85	311.15	337.46	287.81
5	1070.21	11033.40	1576.36	917.45	824.95	635.93	708.81	592.54
10	1423.23	14471.64	2107.14	1222.17	1102.93	841.44	951.43	795.06

(b) GARCH Models

Lag	No drift	linear drift	Nonlinear drift	No drift CEV	linear drift CEV	Nonlinear drift CEV
1	195.39	126.39	136.54	191.04	126.01	143.99
5	373.58	269.71	273.11	368.12	264.83	288.84
10	505.39	374.66	369.54	499.27	367.68	393.79

(c) Markov Regime-Switching Models

Lag	No drift CEV	linear drift CEV	Nonlinear drift CEV	No drift GARCH	linear drift GARCH	Nonlinear drift GARCH	No drift CEV-GARCH	Linear drift CEV-GARCH	nonlinear drift CEV-GARCH
1	33.10	17.31	15.34	30.49	21.94	19.64	28.23	14.34	11.66
5	42.54	17.67	19.02	37.51	24.73	24.22	39.54	11.97	13.50
10	53.10	19.86	23.51	45.62	29.50	28.84	51.01	12.13	15.91

(d) Jump-Diffusion Models

Lag	No drift CEV	linear drift CEV	Nonlinear drift CEV	No drift GARCH	linear drift GARCH	Nonlinear drift GARCH	No drift CEV-GARCH	Linear drift CEV-GARCH	nonlinear drift CEV-GARCH
1	34.20	21.48	20.01	27.64	17.88	20.35	26.97	18.11	18.92
5	39.93	24.84	24.15	40.57	23.85	28.36	39.66	24.25	25.43
10	44.86	27.28	26.52	51.75	30.63	35.52	50.55	31.29	31.36