Specification Analysis of Drift and Volatility Functions of the Interest Rate Processes

by

Licheng Sun

(Under the direction of Stewart Mayhew)

Abstract

Modeling the risk-free interest rate process is very important for the pricing of fixed income securities and interest rate derivatives. Previous studies have identified two features of the short-term risk-free interest rate: nonlinear drift and stochastic volatility. However, little is known about which specification is more fundamental to characterize the dynamics of short term interest rate. In this study, I compare different models of the short rates and test whether these characteristics are of equal importance. Chapter 1 motivates this study by listing several important empirical questions regarding the interest rate process. Chapter 2 proposes a general model that nests both nonlinear drift and stochastic volatility. It is shown that the nonlinear drift is not essential whereas stochastic volatility is indispensable for a parsimonious model of the short rates. Chapter 3 compares the popular stochastic volatility model with its regime switching counterpart. I find strong evidence of regime shifts in the short rate volatilities for four developed countries. Chapter 4 uses a PDE approach to estimate continuous-time short rate models. The results match those found with a discrete-time framework. The last chapter summarizes the empirical findings and gives directions for future research.

Index words: Interest rate process, Nonlinear drift function, Stochastic volatility, Jump diffusion, Partial differential equation approach to estimating continuous-time models

Specification Analysis of Drift and Volatility Functions of the Interest Rate Processes

by

Licheng Sun

B.A., Shanghai Teachers' University, 1994

M.Econ., Shanghai University of Finance and Economics, 1997

A Dissertation Submitted to the Graduate Faculty

of The University of Georgia in Partial Fulfillment

of the

Requirements for the Degree

Doctor of Philosophy

Athens, Georgia

2002

Specification Analysis of Drift and Volatility Functions of the Interest Rate Processes

by

Licheng Sun

Approved:

Major Professor: Stewart Mayhew

Committee: William D. Lastrapes
Marc L. Lipson
Joseph F. Sinkey
Christopher T. Stivers

Electronic Version Approved:

Gordhan L. Patel
Dean of the Graduate School
The University of Georgia
August 2002

Acknowledgments

I would like to thank my major professor, Stewart Mayhew, for his support and encouragement. Helpful comments and suggestions from my dissertation committee members, William D. Lastrapes, Marc L. Lipson, Joseph F. Sinkey, and Christopher T. Stivers, have greatly improved this study. All remaining errors are my own responsibility.

TABLE OF CONTENTS

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Propelled by a rapidly developing market for fixed income securities, modeling interest rate dynamics has become one of the most important and actively researched areas in finance.

Interest rate term structure models often make the simplifying assumption that changes in the entire yield curve are driven by the changes in the short rate. More formally, consider a short-rate process r with $\int_0^T |r_t| dt < \infty$. Under an equivalent martingale measure Q, the price at time t of the zero-coupon bond maturing at s is given by[1]

$$\Lambda_{t,s} \equiv E_t^Q \left[exp \left(\int_t^s -r_u du \right) \right],$$

where Λ is known as the discount function or loosely as the term structure of interest rates. The continuouly compounding yield $y_{t,\tau}$ on a zero-coupon bond maturing at time $t + \tau$ is defined by

$$y_{t,\tau} = -\frac{log(\Lambda_{t,t+\tau})}{\tau}.$$

The specification we choose for the short rate process has a significant impact on the pricing of bonds and interest rate derivatives.

Empirically, as documented by Litterman and Scheinkman (1991), approximately ninety percent of the variation in Treasury yields is associated with changes in the short term interest rates.

[1] For a standard reference on this topic, see Duffie (1996).

Given the importance of the short rate, in this research I mainly focus on the specification of the short rate process.

1.2 Equilibrium versus No-Arbitrage Models

Models of the short rate can be divided into two major categories: equilibrium models and no-arbitrage models. Equilibrium models treat the short rate as an endogenous variable whose dynamics is determined under a general equilibrium framework. Examples include Vasicek (1977) and Cox, Ingersoll and Ross (1985) (CIR). Although economically meaningful, equilibrium models typically do not fit the initial term structure. Taken the initial term structure as an input, no-arbitrage models are very popular among practitioners since by including time-dependent parameters they are capable of giving an exact fit of the yield curve. Examples of no-arbitrage models include Ho and Lee (1986), Hull and White (1990), and Heath, Jarrow, and Morton (1992).[2] However, Dybvig (1997) argues forcefully that essentially any model can give perfect fit to the current yield curve by simply adding more parameters. Backus, Foresi, and Zin (1998) show that no-arbitrage models suffer from the time-inconsistency problem, which could lead to significant biases when applied to the pricing of interest rate derivatives. The crux of the problem is that without any guidance from economic theory model misspecification is a big concern.

Perhaps a more fruitful way of modeling the short rate process is to find a parsimonious model that can fit the data reasonably well without the use of too many free parameters. For this reason, in the current study we choose to focus on identifying a parsimonious model of the short rates.

[2]The orginal Heath, Jarrow, and Morton model is expressed in terms of the forward rates. Carverhill (1995) shows that the model can be equivalently restated using discounted bond prices.

1.3 Empirical Issues Related to Interest Rate Modeling

In a recent review article Chapman and Pearson (2001) list some of the unresolved empirical issues with regard to interest rate modeling:

> What can be learned from this large and growing literature? What model features appear to be essential in describing the fundamental properties of interest rates? First, this new literature does not provide conclusive evidence, based solely on the data, about whether interest rate levels tend to return to a constant long-run level and whether or not this tendency is stronger for extreme levels of interest rates. With respect to interest rate volatility, it is clear that the "absolute" volatility of the short rate, defined as the standard deviation of rate changes scaled by the square root of the time between changes, is increasing in its level.
>
> However, inferences about the relation between the level and volatility of the short rate are sensitive to the treatment of the years between 1979 and 1982, the so-called "Federal Reserve experiment." In particular, the data from this period suggest a very strong relation between volatility and the level of interest rates, while excluding this period or treating it as a distinct (lower probability) "regime" suggests a much weaker relation. Finally, modelling the volatility of interest rates requires more than a simple "level effect", i.e. there appears to be some sort of stochastic volatility. However, the additional volatility component can be described adequately (in a statistical sense) in a variety of competing ways.

The empirical issues addressed by Chapman and Pearson essentially boils down to two related questions.

1. Is the drift function of the short rate process linear or nonlinear?

2. Is modeling regime shifts in the short rate volatility process relevant?

In this study, we provide answers to the above two questions. In chapter 2, we compare models of the short rate with and without nonlinear drifts. It is shown that modeling stochastic volatility is more important than modeling the nonlinear drift. In chapter 3, we show that modeling regime shifts in the short rate volatility is important for the U.S., the U.K., Canada, and Japan short rate data.

In addition to the empirical findings, we also make several methodological contributions. First, we show how to model correlation in a regime-switching stochastic volatility model. Second, we generalize the regime-switching stochastic volatility model by allowing for time-varying transition probabilities. Finally we show how to use the PDE approach to estimate a general jump diffusion model.

Chapter 2

Estimating Discrete-time Interest Rate Models

2.1 Introduction

Recent empirical evidence (Aït-Sahalia (1996), Stanton (1997)) suggests that incorporating nonlinearity in the drift term is needed for the modeling of short-term interest rates. Other researchers (Longstaff and Schwartz (1992), Brenner, Harjes, and Kroner (1996)) (BHK) find that stochastic volatility is a salient feature that should not be ignored in short-rate models. An important empirical question is: which of the two characteristics are more fundamental for a parsimonious model of the short rate? Or are they both indispensable? This question is of theoretical and practical significance since different model specifications could lead to very different results when applied to the pricing of interest rate derivatives. From a modeling perspective, a short rate model with general nonlinear drift may become analytically intractable even in the single-factor case.

Another interesting empirical puzzle surrounding the U.S. short rate was first pointed out in Chan, Karolyi, Longstaff and Sanders (1992) (CKLS). CKLS estimate a general one-factor model for the U.S. risk-free rate. Surprisingly, they find the estimated parameters imply a non-stationary short rate process, which is counter-intuitive.

In this chapter I address the following important empirical questions:

(I) Is the drift term linear or nonlinear for the U.S. short rate?

(II) Is the finding in CKLS robust across different subperiods, data frequencies, and countries?

(III) Is modeling stochastic volatility in the short rate important?

I compare models with nonlinear drift term and/or stochastic volatility with a general model that combines both nonlinear drift and stochastic volatility. It nests various extant interest rate models. I also look at the international evidence on the issue by modeling the interest rate processes for four other industrialized countries in addition to the U.S. case.

The major findings are as follows. I find the CKLS puzzle is present only in a specific dataset for a specific model. I also find that the nonlinear drift specification is not an essential feature of the short rate for any of the five countries under consideration. A linear drift model with stochastic volatility seems to be a good parsimonious model of the short rate at least for countries where the short rates exhibit excessive volatility, such as in the cases of the US and the UK.

Section 2.2 reviews the literature related to the empirical performance of various short term interest rate models. Section 2.3 sets up a general nonlinear drift stochastic volatility model. Section 2.4 describes the data. Section 2.5 tests the ARCH/GARCH effects. Section 2.6 estimates the model using the maximum likelihood estimation. Section 2.7 compares the various model specifications, and performs robustness checks. Section 2.8 discusses the results and concludes this chapter.

2.2 Background and Related Literature

In this section, I give further background regarding the empirical issues involved and review the related literature.

2.2.1 Linear Drift Single-Factor Models

The short rate model of CKLS nests most single-factor models with linear drift, including the Vasicek model and the CIR model. It has the following specification:

$$dr = (\alpha + \beta r) + \sigma r^{\gamma} dW, \tag{2.1}$$

where r is the interest rate level, and W is a Brownian motion. Note that in the CKLS model, volatility is specified as a function of the interest rate level, capturing the so-called "level effect". CKLS use the generalized method of moments (GMM) to estimate a discrete-time version of the above model. Note that if we rewrite $\alpha + \beta r$ as $\beta(r - \theta)$ where $\theta \equiv -\frac{\alpha}{\beta}$, then β can be interpreted as the speed of mean reversion and θ the mean value of the interest rate. The parameter γ has the interpretation of being the elasticity of variance and is usually less than one in theoretical models. For example, $\gamma = 0.5$ in the CIR model and $\gamma = 0$ in the Vasicek model. CKLS find that the unrestricted estimate of γ in their model is approximately 1.5, which contradicts the specifications of most theoretical models. In fact, CKLS are able to reject all models whose γ is less than 1. In addition to the fact that this is incompatible with most theoretical models, a γ greater than 1 has the undesirable implication that the short term interest rate may become non-stationary at high interest rate levels. To solve this puzzle, ongoing research has followed two directions. The first approach is to model the drift term as a nonlinear function such that the mean reversion is faster at high interest rate levels. The other approach argues that the estimated high elasticity of volatility may be misleading because the CKLS model ignores the stochastic nature of the interest rate volatility.

2.2.2 Nonlinear Drift Term

Aït-Sahalia (1996) uses nonparametric techniques to estimate interest rate models. The idea is to compare the density implied by a parametric model and a nonparametric estimator that is valid even if the parametric model is misspecified. In the empirical section of his paper, the drift is parameterized to have the following form: $\mu(r,\theta) = \alpha_0 + \alpha_1 r + \alpha_2 r^2 + \alpha_3/r$. It is claimed that the linearity of the drift is the main source of misspecification of existing models. The paper argues that the nonlinearity of the drift effectively makes the interest rate process stationary.

Stanton (1997) uses Taylor series expansion to approximate the drift and diffusion of the stochastic differential equation. These approximations are then estimated nonparametrically from discretely sampled data. It is found that the drift exhibits evidence of substantial nonlinearity. The estimated volatility elasticity is close to the CKLS result.

Conley, Hansen, Luttmar, and Scheinkman (1997) (CHLS) focus on a class of models in which the local volatility elasticity is constant and the drift has a flexible specification. In their empirical work, they consider the same parameterization of the drift as in Aït-Sahalia (1996). They find evidence for a volatility elasticity between one and two. They claim that the mechanism for inducing stationarity is the increased volatility of the diffusion process. Nonlinearities in the drift are shown to be important for very high-variance elasticities (greater than four) but not for low ones.

Ahn and Gao (1999) derive a closed-form solution for bond prices when the drift term is nonlinear. Specifically, they assume the interest rate process is characterized by the following stochastic differential equation:

$$dr(t) = \kappa(\theta - r(t))r(t)dt + \sigma r(t)^{1.5} dw(t).$$

In their empirical work, they estimate the following model using the GMM.

$$dr(t) = (\alpha_1 + \alpha_2 r(t) + \alpha_3 r(t)^2)dt + \sqrt{\alpha_4 + \alpha_5 r(t) + \alpha_6 r(t)^3}\, dw(t).$$

The specification here is similar to that of Aït-Sahalia (1996) and CHLS. Ahn and Gao show that α_3 is significant and conclude their model outperforms all linear drift models.

The apparent success of nonlinear drift models, however, is inconclusive. In a recent paper, Jones (2000) questions the finding of nonlinear drift as reported in Aït-Sahalia (1996) and CHLS. Using a Bayesian method, he concludes that the large negative drift for high interest rates reported in the above-cited papers represents a non-trivial prior belief about the shape of the drift function, and that this prior belief by itself is strong enough to generate the finding of nonlinearity.

Pritsker (1998) questions the specification test developed in Aït-Sahalia (1996). The argument is that interest rates are known to be highly correlated whereas the nonparametric technique used in Aït-Sahalia's paper is very sensitive to the dependence in the data.

Chapman and Pearson (2000) study the finite-sample properties of the nonparametric estimators used in the Aït-Sahalia (1996) and Stanton (1997) papers by applying them to simulated sample paths of a square-root diffusion. Although the drift is linear, the nonparametric estimators suggest nonlinearities of the type and magnitude reported in Aït-Sahalia (1996) and Stanton (1997). Chapman and Pearson conclude that nonlinearity of the short rate drift is not a robust stylized fact.

2.2.3 Short Rate Volatility and Two-Factor Models

It is well known that interest rate data exhibit volatility clustering or conditional heteroskedasticity. Some of the conditional heteroskedasticity can be accounted for by the level effect as in the CKLS model. However, as evidenced by the high estimate of variance elasticity, some researchers have argued that the level effect may be inadequate to explain the conditional heteroskedasticity and the assumption that variance elasticity is constant may be too restrictive. A natural way to model the conditional heteroskedasticity in asset returns is the ARCH/GARCH models of Engle (1982) and Bollerslev (1986), where the conditional variance is a function of past variance estimates and lagged squared forecast errors. A potential criticism of the (pure) GARCH models of the interest rate volatility is that negative interest rates are permissible. This weakness can be corrected if we combine the features of the CKLS model with the GARCH models, which are accordingly named the Level-GARCH models in the literature.[1]

An alternative modeling technique is the so-called stochastic volatility (SVOL) model. The SVOL model treats the volatility as a latent process which evolves stochastically over time. As an example, Ball and Torous (1999) specify the following discrete-time SVOL model for the short term interest rate:

$$r_t = (a + br_{t-1}) + \sigma_{t-1} r_{t-1}^{\gamma} \epsilon_t$$

$$ln\sigma_t^2 - \mu = \beta(ln\sigma_{t-1}^2 - \mu) + \xi\eta_t,$$

where ϵ_t and η_t are i.i.d. standard Gaussian disturbances. One advantage of the SVOL model is that it has a natural interpretation of being a discrete-time analog to the continuous-time diffusion models. For example, the Ball and Torous model can be viewed as a discrete-time approximation to a two-factor interest rate model, such as Longstaff and Schwartz (1992), where the short rate and its volatility are specified as two factors. The difficulty, however, lies with the estimation of the model. One relatively simple approach put forward by Harvey et al. (1994) is to rewrite the model in state-space form and apply the Kalman filter to build up the likelihood function. We estimate the short rate volatiltiy process with the SVOL model and its regime-switching counterpart in chapter 3.

[1] BHK first study the Level-GARCH model. Since we focus on the Level-GARCH specification, we will use the two terms GARCH and Level-GARCH interchangeably when there is no ambiguity.

In comparison with the SVOL model, GARCH models are relatively easy to estimate. In addition, there exist a variety of extensions to the simple linear GARCH model, which are specialized to capture various stylized empirical facts about financial assets volatilities, such as volatility asymmetry. Furthermore, as shown by Nelson (1990), Nelson and Foster (1994), and Duan (1996), GARCH models converge to the diffusion models in the limit. Thus one can view both the SVOL models and the GARCH models as discrete time approximations to their continuous time counterparts, although it is known that the SVOL model has a faster convergence rate.

Longstaff and Schwartz (1992) develop a two-factor theoretical model of the term structure where the short rate and its volatility are the two factors. A closed-form solution is obtained for the bond price. They estimate the model using GMM. The discrete time model is specified as follows:

$$r_{t+1} - r_t = \alpha_0 + \alpha_1 r_t + \alpha_2 V_t + \epsilon_{t+1},$$

$$V_t = \beta_0 + \beta_1 r_t + \beta_2 V_{t-1} + \beta_3 \epsilon_t^2.$$

Note that the GARCH framework is adopted to parameterize the volatility factor. They find empirical evidence supporting the model.

Fong and Vasicek (1991) argue that from the viewpoint of adopting a rollover strategy on the short rate instrument, the term bond yield will be low when the volatility of the short rate is high, and high if the volatility is low. Thus the volatility of the short rate should be treated as a pricing factor for bond. They suggest a two-factor model using the same factors as in the Longstaff and Schwartz model. The model is specified as follows:

$$dr(t) = \alpha(\bar{r} - r(t))dt + \sqrt{v(t)}dW_1(t)$$

$$dv(t) = \gamma(\bar{v} - v(t))dt + \xi\sqrt{v(t)}dW_2(t)$$

where v denotes the volatility of the short rate. Fong and Vasicek provide a closed-form solution for bond price under this model.

Brenner, Harjes, and Kroner (1996) extend the analysis of CKLS by adding the GARCH effect to the volatility. More formally, they look at the following model:

$$r_t - r_{t-1} = \alpha + \beta r_{t-1} + \epsilon_t,$$

$$E(\epsilon_t | I_{t-1}) = 0, E(\epsilon_t^2 | I_{t-1}) = \sigma_t^2 r_{t-1}^{2\gamma},$$

$$\sigma_t^2 = a_0 + a_1 \epsilon_{t-1}^2 + b\sigma_{t-1}^2.$$

This model improves the CKLS model in that it allows volatility to depend on both interest rate levels and information shocks. They find the sensitivity of volatility to interest rate levels has been overstated in the literature. In particular, the estimated value of the variance elasticity γ turns out to be less than one in their Level-GARCH model, suggesting the unreasonably high estimate of γ obtained in the CKLS model may be a result of model misspecification.

2.3 Model Specifications

Given the various models reviewed in the previous section, one may wonder which model has more explanatory power. Are nonlinear drift and stochastic volatility redundant in the sense that either one is sufficient to explain interest rate movements? Or are they complementary in the sense that both are necessary to describe the stochastic behavior of interest rate?

To address these issues, in this section I introduce a model of the short term riskless interest rate that incorporates both a nonlinear drift function and stochastic volatility. This model can be regarded as a generalization of the BHK specification by including a nonlinear drift. It can also be viewed as an extension to the models suggested by Aït-Sahalia (1996) and CHLS since I allow the volatility to evolve stochastically over time. This general model nests almost all the important models reviewed in the previous section so that we can easily test the various hypotheses related to model specifications.

The proposed general model starts from the following continuous time model of the short rate:

$$dr = (\alpha_0 + \alpha_1 r + \alpha_2 r^2 + \alpha_3 r^{-1})dt + \sigma r^{\gamma} dW, \tag{2.2}$$

where W is a standard Brownian motion. The drift term used here follows Aït-Sahalia (1996) and CHLS. In equation (2.2), when α_3 equals zero and σ is a constant, this model is similar to the Ahn and Gao model.

The discretized version of equation (2.2) takes the following form

$$r_{t+h} - r_t = (\alpha_0 + \alpha_1 r_t + \alpha_2 r_t^2 + \alpha_3 r_t^{-1})h + \sigma_{t+h} r_t^{\gamma} \sqrt{h} \epsilon_{t+h}, \tag{2.3}$$

where h is the length of the time interval and ϵ is a standard normal variable.

Here the Euler approximation scheme is adopted to discretize equation (2.2). Although other higher approximation schemes are available, the Euler scheme is the simplest and most commonly used.

By setting $h = 1$, we have

$$r_{t+1} - r_t = (\alpha_0 + \alpha_1 r_t + \alpha_2 r_t^2 + \alpha_3 r_t^{-1}) + \sigma_{t+1} r_t^{\gamma} \epsilon_{t+1}. \tag{2.4}$$

To allow for stochastic volatility, I augment the model with the GARCH equation:

$$\sigma_{t+1}^2 = \omega + \phi_1 \sigma_t^2 + \phi_2 u_t^2, \tag{2.5}$$

where $u_t \equiv \sigma_t r_{t-1}^{\gamma} \epsilon_t$. This is the familiar linear GARCH(1,1) model, which has been identified as a good parametric representation of the conditional heteroskedasticity prevalent in many financial time series.[2]

[2]The GARCH model in equation (2.5) can be related to certain continuous time models. See, for example, Nelson (1990), Nelson and Foster (1994). One such continuous time limit

The equations (2.4) and (2.5) are the general unrestricted discrete-time model that we estimate in this chapter. Note that it captures both the nonlinear drift and stochastic volatility features reported in the literature. The various nested models are shown Table 2.1.

Table 2.1: Model Specifications

Model	Specification	Restrictions
CKLS	$\Delta r_{t+1} = \alpha_0 + \alpha_1 r_t + \sigma_{t+1} r_t^{\gamma} \epsilon_{t+1}$	$\alpha_2,\ \alpha_3,\ \phi_1,\ \phi_2 = 0$
AG	$\Delta r_{t+1} = \alpha_0 + \alpha_1 r_t + \alpha_2 r_t^2 + \sigma_{t+1} r_t^{\gamma} \epsilon_{t+1}$	$\alpha_3,\ \phi_1,\ \phi_2 = 0$
CHLS	$\Delta r_{t+1} = \alpha_0 + \alpha_1 r_t + \alpha_2 r_t^2 + \alpha_3 r_t^{-1} + \sigma_{t+1} r_t^{\gamma} \epsilon_{t+1}$	$\phi_1,\ \phi_2 = 0$
BHK	$\Delta r_{t+1} = \alpha_0 + \alpha_1 r_t + \sigma_{t+1} r_t^{\gamma} \epsilon_{t+1},$ $\sigma_{t+1}^2 = \omega + \phi_1 \sigma_t^2 + \phi_2 u_t^2$	$\alpha_2,\ \alpha_3 = 0$
Quadratic	$\Delta r_{t+1} = \alpha_0 + \alpha_1 r_t + \alpha_2 r_t^2 + \sigma_{t+1} r_t^{\gamma} \epsilon_{t+1},$ $\sigma_{t+1}^2 = \omega + \phi_1 \sigma_t^2 + \phi_2 u_t^2$	$\alpha_2 = 0$
NLSV	$\Delta r_{t+1} = \alpha_0 + \alpha_1 r_t + \alpha_2 r_t^2 + \alpha_3 r_t^{-1} + \sigma_{t+1} r_t^{\gamma} \epsilon_{t+1},$ $\sigma_{t+1}^2 = \omega + \phi_1 \sigma_t^2 + \phi_2 u_t^2$	

These model specifications include: a quadratic drift stochastic volatility model where $\alpha_3 = 0$; a linear drift stochastic volatility model where both $\alpha_2 = \alpha_3 = 0$ (the BHK model); a model with nonlinear drift but without stochastic volatility where both $\phi_1 = \phi_2 = 0$ (the CHLS model); a model with quadratic drift but without stochastic volatility where α_3, ϕ_1, and $\phi_2 = 0$ (the Ahn and Gao model); and lastly the CKLS model where α_2, α_3, ϕ_1, and $\phi_2 = 0$. Note that the Vasicek, CIR models are nested in the CKLS model.

---

may be specified as follows:

$$d\sigma^2 = (\omega - \theta\sigma^2)dt + \sqrt{2}\alpha\sigma^2 dB,$$

where B is a standard Brownian Motion independent of W. See Bollerslev et al. (1994) for further details.

2.4 Data

The data include 2492 weekly observations of the 3-month Treasury bill rates from the Federal Reserve site, ranging from January 1954 to October 2001. In order to compare my results with other studies in the area, in particular the CKLS study, I also estimate the models with monthly observations of the U.S. T-Bill yields obtained from CRSP risk free rate file ranging from June 1964 to December 1999, a total of 427 observations.

While the majority of studies on interest rate models focus on US data, it is interesting to look at the international evidence with regard to these models. Hence, I also estimate the models using the interest rate data from four other industrialized countries: Canada, Germany, Japan, and the United Kingdom. The data are obtained from Datastream. The interest rate series include: Canada treasury bill 1 month (CN13883) from January 1980 to December 2000, altogether 1095 observations; Germany interbank one month offered rate (FIBOR1M) from November 1990 to December 2000, a total number of 530 observations; Japan interbank 1 month offered rate (JPIBK1M) from December 1985 to December 2000, 782 observations, and UK interbank 1 month middle rate (LDNIB1M) from January 1975 to December 2000, 1356 observations. Because of the relatively short time horizon covered by these interest rate data compared with the US data, I use the weekly frequency to ensure a fairly large number of sample observations.

Figure 2.1 depicts the time series of interest rate levels for the five countries studied in this chapter. A cursory look indicates that these interest rate series do exhibit very distinctive patterns. Both very high interest rate levels and extremely low interest rate levels (less than 1% in the case of Japan) are observed. A common feature is that the interest rate movements typically exhibit very volatile behavior. For example, the US interest rate shows some dramatic swings during the federal reserve experiment period of 1979 to 1982, which coincides with high interest rate levels. Whereas if we look at the cases of Canada and Japan, it seems that excessive volatility can also occur at median to low interest rate levels.

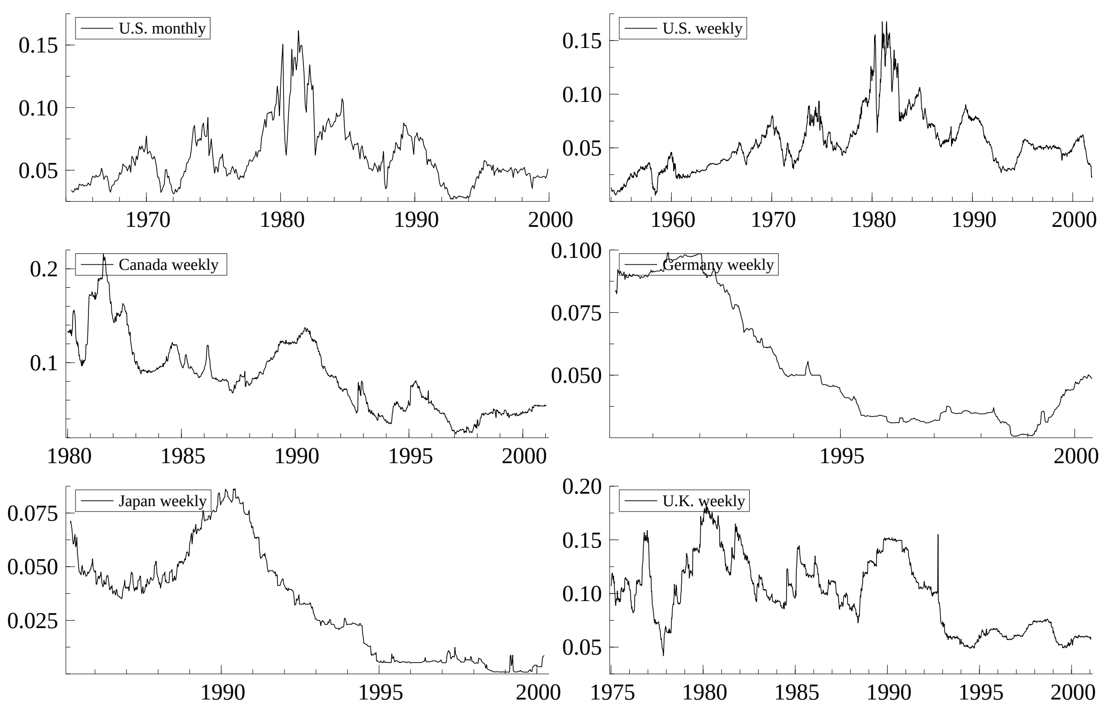

Figure 2.1: Short-term Interest Rates

Table 2.2 reports the descriptive statistics for the interest rate series and their first order differences for the five countries under consideration. The highest interest rate that we can observe is 21.55% from the Canadian interest rate series while the lowest is 0.086% from Japan. All interest rate levels are highly autocorrelated. However, the autocorrelation disappears when we look at their first differences. It is noteworthy that the first differences of the interest rate data exhibits some excess kurtosis, suggesting that normality may not be a good assumption. This motivates the use of t distribution in the empirical analysis of this chapter to check the robustness of the results.

Table 2.2: Descriptive statistics

Panel A: Interest Rate Levels

	US(weekly)	US(monthly)	Canada	Germany	Japan	UK
Mean	0.05485	0.06109	0.08441	0.05368	0.03168	0.09900
Std. Dev.	0.02754	0.02486	0.03928	0.02423	0.02522	0.03382
Minimum	0.0058	0.02670	0.02400	0.02557	0.00086	0.04188
Maximum	0.1676	0.16150	0.21550	0.09901	0.08625	0.18188
Skewness	1.15618	1.41903	0.74012	0.69817	0.42339	0.30458
Kurtosis	4.92721	5.25636	3.16729	1.91209	2.02317	2.08825
ρ_1	0.996	0.95700	0.99700	0.99800	0.99600	0.99300
ρ_2	0.991	0.91400	0.99200	0.99600	0.99200	0.98700
ρ_3	0.985	0.87600	0.98800	0.99200	0.98900	0.98100
ρ_4	0.980	0.85000	0.98300	0.98900	0.98500	0.97400
ρ_5	0.973	0.82800	0.97700	0.98600	0.98200	0.96600
ρ_{36}	0.832	0.24200	0.22800	0.34300	0.28400	0.19900

Panel B: First Order Differences

	US(weekly)	US(monthly)	Canada	Germany	Japan	UK
Mean	0.000004	0.00004	-0.00007	-0.00007	-0.00008	-0.00004
Std Dev	0.0021	0.00716	0.00280	0.00093	0.00148	0.0038
Minimum	-0.0182	-0.05813	-0.01620	-0.00558	-0.00750	-0.0606
Maximum	0.0192	0.03285	0.03050	0.00963	0.00747	0.0537
Skewness	-0.637	-1.41252	1.60243	1.77560	-0.47619	0.2339
Kurtosis	23.377	16.40222	23.00815	30.04524	8.60639	82.0950
ρ_1	0.270	0.00331	0.11200	0.09590	0.01070	-0.0936
ρ_2	0.065	-0.07830	0.11600	0.06340	-0.05340	0.0337
ρ_3	0.050	-0.12200	0.09830	0.07400	0.04460	0.0442
ρ_4	0.087	-0.06450	0.08020	-0.01150	-0.04020	0.0430
ρ_5	0.057	-0.01740	0.06690	-0.02090	-0.06490	0.0350
ρ_{36}	0.058	0.04200	0.04320	0.03900	0.01410	-0.0498

2.5 Diagnostic Tests for the ARCH/GARCH Effect

As a preliminary step, I test for the ARCH/GARCH effect using Engle's LM test and the Ljung-Box Q-statistics. Since OLS estimates are consistent even in the presence of ARCH, I use the squared OLS residuals from the regression equation $dr_t = a_0 + a_1 r_{t-1} + a_2 r_{t-1}^2 + a_3/r_{t-1} + \epsilon_t$ to perform the tests. I choose the number of lags equals to five in the LM test. Similar results are obtained when the number of lags extends 36. For the Ljung-Box Q-statistics, I report the results for lags vary from one to five. I also report results for lags equal to 36.

From Table 2.3, it is obvious that both the Engle's LM test and the Ljung-Box Q-statistics indicate the existence of conditional heteroskedasticity for four of the five countries. The only exception is Germany. However, it is premature to conclude that the stochastic volatility model does not fit the German interest rate since the tests for ARCH/GARCH effect does not take into account the the level effect as described in the CKLS model.

2.6 Model Estimation

I use the maximum likelihood to estimate the models. To set up the likelihood function for the GARCH models, let the $\mathbf{x}_t \equiv (1, r_t, r_t^2, r_t^{-1})'$ be the vector of explanatory variables and $\beta \equiv (\alpha_0, \alpha_1, \alpha_2, \alpha_3)'$ the vector of coefficients. Define $y_t \equiv r_t - r_{t-1}$. In the case of normal distribution, the sample log likelihood is given by

$$L(\theta) = -(T/2)log(2\pi) - (1/2)\sum_{t=1}^{T} log(g_t) - (1/2)\sum_{t=1}^{T}(y_t - \mathbf{x_t'}\beta)^2/g_t,$$

where $g_t \equiv \sigma_t^2 r_{t-1}^{2\gamma}$ and σ_t^2 is defined in equation (2.5).

Table 2.3: Tests for ARCH/GARCH Effects

This table reports the Engle's LM test as well as the Ljung-Box Q statistics to detect the ARCH/GARCH effects.

Panel A: LM ARCH test

	US(weekly)	US(monthly)	Canada	Germany	Japan	UK
LM(1)	148.454	34.960	24.979	0.136	3.342	257.180
p-value	0.000	0.000	0.000	0.713	0.068	0.000
LM(2)	254.526	44.499	25.096	0.137	12.579	316.760
p-value	0.000	0.000	0.000	0.934	0.002	0.000
LM(3)	284.143	51.634	25.126	1.272	18.476	333.898
p-value	0.000	0.000	0.000	0.736	0.000	0.000
LM(4)	304.141	54.592	26.058	4.054	24.085	339.195
p-value	0.000	0.000	0.000	0.399	0.000	0.000
LM(5)	338.999	54.593	27.236	4.678	41.457	354.373
p-value	0.000	0.000	0.000	0.456	0.000	0.000
LM(36)	586.458	102.355	54.792	19.065	66.608	339.022
p-value	0.000	0.000	0.023	0.991	0.001	0.000

Panel B: Ljung-Box

Q(1)	185.74	35.397	25.068	0.1366	3.3581	270.51
p-value	0.000	0.000	0.000	0.712	0.067	0.000
Q(2)	197.54	57.576	26.368	0.1375	13.366	270.65
p-value	0.000	0.000	0.000	0.934	0.001	0.000
Q(3)	204.68	79.258	26.594	0.3084	21.248	270.65
p-value	0.000	0.000	0.000	0.958	0.000	0.000
Q(4)	224.96	96.174	27.793	0.7017	29.998	270.65
p-value	0.000	0.000	0.000	0.951	0.000	0.000
Q(5)	233.92	102.29	29.781	0.8603	54.394	270.66
p-value	0.000	0.000	0.000	0.99	0.000	0.000
Q(36)	442.80	296.74	81.693	3.8689	105.96	271.93
p-value	0.000	0.000	0.000	1	0.000	0.000

I also consider the case when the conditional distribution of y_t is the student-t distribution. The sample log likelihood can be shown to be:

$$\begin{aligned} L(\theta) &= Tlog\left(\frac{\Gamma[(v+1)/2]}{\pi^{1/2}\Gamma(v/2)}(v-2)^{-1/2}\right) - (1/2)\sum_{t=1}^{T} log(g_t) \\ &\quad -[(v+1)/2]\sum_{t=1}^{T} log\left[1+\frac{(y_t-\mathbf{x}_t'\beta)^2}{g_t(v-2)}\right], \end{aligned}$$

where $\Gamma(.)$ is the gamma function, v is the degrees of freedom of the t distribution and is a parameter to be estimated via maximum likelihood. In our actual estimation, we first scale the data by multiplying them by 100. The results are robust to various starting values.

Table 2.4 reports the parameter estimates for the various model specifications listed in Table 2.1. Quad refers to the quadratic model specification. Standard errors of the model parameter estimates are shown in parentheses. Panel A reports the results for the U.S. weekly data, Panel B the results for the U.S. monthly data, Panel C the results for Canada data, Panel D the results for Germany data, Panel E the results for Japan data, and Panel F the results for the U.K. data.

As expected, the estimates of α_1 in the case of linear drift models are all negative for the five countries, with or without stochastic volatility. This is consistent with the interpretation that α_1 is the mean reversion coefficient. In addition, estimates of α_0 are less precise than those for α_1 for linear drift models.

We notice some interesting phenomena related to the parameter estimates of γ, the volatility elasticity. Recall that CKLS find their estimate of γ is significantly higher than one, which is often regarded as an empirical puzzle since it implies the interest rate process is non-stationary.

Table 2.4: Parameter Estimates

Panel A: U.S. Weekly Data

	α_0	α_1	α_2	α_3	γ	ω/σ	ϕ_1	ϕ_2
NLSV	0.014	-0.003	0.0001	0.004	0.152	0.0003	0.740	0.17
	(0.02)	(0.007)	(0.001)	(0.019)	(0.028)	(0.0001)	(0.02)	(0.02)
Quad.	0.018	-0.004	0.0001		0.152	0.0003	0.740	0.166
	(0.01)	(0.004)	(0.000)		(0.028)	(0.0001)	(0.021)	(0.02)
BHK	0.015	-0.0024			0.151	0.0003	0.74	0.166
	(0.01)	(0.001)			(0.028)	(0.0001)	(0.021)	(0.02)
CHLS	-0.027	0.0085	-0.0007	0.0300	0.708	0.003		
	(0.02)	(0.006)	(0.000)	(0.016)	(0.017)	(0.0002)		
AG	0.012	-0.0016	-0.0001		0.708	0.003		
	(0.01)	(0.003)	(0.000)		(0.017)	(0.0002)		
CKLS	0.013	-0.002			0.708	0.003		
	(0.01)	(0.001)			(0.017)	(0.0002)		

Panel B: U.S. Monthly Data

	α_0	α_1	α_2	α_3	γ	ω/σ	ϕ_1	ϕ_2
NLSV	-0.116	0.041	-0.004	0.30	0.7564	0.006	0.40	0.019
	(0.85)	(0.15)	(0.008)	(1.47)	(0.12)	(0.002)	(0.11)	(0.01)
Quad.	0.056	0.011	-0.003		0.757	0.005	0.40	0.019
	(0.13)	(0.047)	(0.004)		(0.119)	(0.002)	(0.11)	(0.01)
BHK	0.14	-0.02			0.757	0.005	0.413	0.019
	(0.06)	(0.012)			(0.12)	(0.002)	(0.10)	(0.01)
CHLS	-0.2	0.06	-0.005	0.44	1.34	0.002		
	(0.99)	(0.18)	(0.01)	(1.67)	(0.08)	(0.001)		
AG	0.06	0.012	-0.003		1.34	0.002		
	(0.13)	(0.051)	(0.004)		(0.08)	(0.001)		
CKLS	0.144	-0.02			1.35	0.002		
	(0.06)	(0.014)			(0.08)	(0.001)		

Table 2.4: (continued)

Panel C: Canada Data

	α_0	α_1	α_2	α_3	γ	ω/σ	ϕ_1	ϕ_2
NLSV	-0.37	0.046	-0.002	0.76	-0.034	0.012	0.59	0.47
	(0.22)	(0.033)	(0.002)	(0.43)	(0.05)	(0.002)	(0.04)	(0.17)
Quad.	0.013	-0.007	0.0004		-0.034	0.011	0.618	0.41
	(0.04)	(0.01)	(0.001)		(0.05)	(0.002)	(0.047)	(0.13)
BHK	-0.01	0.00			-0.029	0.011	0.619	0.39
	(0.02)	(0.003)			(0.05)	(0.002)	(0.048)	(0.13)
CHLS	-0.07	0.006	-0.0003	0.24	0.37	0.016		
	(0.16)	(0.022)	(0.001)	(0.343)	(0.041)	(0.003)		
AG	0.04	-0.008	0.0003		0.37	0.016		
	(0.03)	(0.009)	(0.0005)		(0.04)	(0.003)		
CKLS	0.022	-0.004			0.37	0.016		
	(0.02)	(0.002)			(0.041)	(0.003)		

Panel D: Germany Data

	α_0	α_1	α_2	α_3	γ	ω/σ	ϕ_1	ϕ_2
NLSV	0.43	-0.098	0.006	-0.55	0.26	0.0008	0.57	0.053
	(0.24)	(0.049)	(0.003)	(0.379)	(0.061)	(0.0002)	(0.05)	(0.023)
Quad.	0.073	-0.029	0.0023		0.27	0.0008	0.57	0.05
	(0.03)	(0.011)	(0.001)		(0.062)	(0.0002)	(0.056)	(0.02)
BHK	0.005	-0.002			0.275	0.0008	0.576	0.046
	(0.01)	(0.002)			(0.06)	(0.0002)	(0.056)	(0.02)
CHLS	0.23	-0.06	0.004	-0.26	0.63	0.001		
	(0.25)	(0.05)	(0.003)	(0.390)	(0.056)	(0.0002)		
AG	0.065	-0.026	0.002		0.63	0.001		
	(0.03)	(0.012)	(0.001)		(0.056)	(0.0002)		
CKLS	0.012	-0.004			0.64	0.001		
	(0.01)	(0.002)			(0.056)	(0.0002)		

Table 2.4: (continued)

Panel E: Japan Data

	α_0	α_1	α_2	α_3	γ	ω/σ	ϕ_1	ϕ_2
NLSV	-0.005	-0.004	0.0003	0.003	0.147	0.003	0.74	0.05
	(0.01)	(0.008)	(0.001)	(0.002)	(0.02)	(0.001)	(0.04)	(0.01)
Quad.	0.01	-0.011	0.0011		0.147	0.003	0.74	0.05
	(0.01)	(0.006)	(0.001)		(0.02)	(0.001)	(0.05)	(0.01)
BHK	0.005	-0.0042			0.1468	0.003	0.74	0.047
	(0.01)	(0.002)			(0.02)	(0.001)	(0.048)	(0.01)
CHLS	-0.008	-0.0006	-0.0001	0.003	0.20	0.015		
	(0.01)	(0.008)	(0.001)	(0.002)	(0.017)	(0.001)		
AG	0.008	-0.009	0.0007		0.2	0.02		
	(0.01)	(0.007)	(0.001)		(0.017)	(0.001)		
CKLS	0.005	-0.004			0.198	0.015		
	(0.01)	(0.002)			(0.017)	(0.001)		

Panel F: U.K. Data

	α_0	α_1	α_2	α_3	γ	ω/σ	ϕ_1	ϕ_2
NLSV	-1.09	0.12	-0.0044	3.04	0.97	0.001	0.064	0.002
	(0.61)	(0.07)	(0.002)	(1.72)	(0.04)	(0.0002)	(0.04)	(0.001)
Quad.	-0.011	0.005	-0.0004		0.96	0.001	0.065	0.002
	(0.07)	(0.02)	(0.001)		(0.05)	(0.0002)	(0.046)	(0.00)
BHK	0.023	-0.003			0.96	0.001	0.065	0.002
	(0.02)	(0.003)			(0.045)	(0.0002)	(0.045)	(0.00)
CHLS	-2.04	0.23	-0.008	5.6	1.2	0.0005		
	(0.65)	(0.075)	(0.003)	(1.814)	(0.021)	(0.00)		
AG	-0.022	0.01	-0.0007		1.177	0.0006		
	(0.07)	(0.018)	(0.001)		(0.023)	(0.0001)		
CKLS	0.03	-0.003			1.18	0.0006		
	(0.02)	(0.003)			(0.02)	(0.0001)		

Contrary to the results of CKLS, we find that the estimated γ values are sensitive to the choice of datasets as well as model specifications.

1. We compare the results for the two U.S. interest rate datasets. In the case of weekly data, γ estimates are much lower than the estimates obtained from the monthly data for all models considered. For example, the γ estimate for the unrestricted model (NLSV) is merely 0.152, far below the γ estimate of 0.756 from the monthly data.

2. We find single factor models that ignore the presence of conditional heteroscedasticity in the data give much higher γ estimates than their GARCH counterparts. For the U.S. weekly data, γ estimates from single-factor models are as high as 0.7. Moreover, they exceed the stationary threshold when estimated with the U.S. monthly data, the same dataset used in the CKLS study. Hence our finding is in agreement with the BHK result that CKLS overstate the sensitivity of interest rate volatility to interest rate levels due to model misspecification.

3. We note that variations in the estimates of γ for models with different drift specifications are very small. For example, the γ estimates for the GARCH models using U.S. monthly data are 0.75637, 0.75756, and 0.75672 respecitively. Therefore, we further verify that the BHK result is robust to different specifications in the drift.

For the other four countries under consideration, we obtain similar results. We find that without exception estimates of γ become lower when we include the GARCH equation. For example, in the case of the Germany data, the estimates of γ are as high as 0.63 for models without GARCH. However, if we allow for conditional heteroscedasticity, estimates of γ drop to 0.27 which is significantly less than 1. In addition we note that the standard errors of all γ estimates are relatively small.

In terms of national patterns, I find for single factor models without stochastic volatility γ exceeds the critical value of 1 only in two countries, the US (monthly data) and the UK, which have the most volatile interest rates among the five countries (see Tables 2.2 and 2.3). In addition, in both cases, when we add the GARCH equation, estimates of γ fall below 1, although in the case of the U.K. estimates of γ are still very high (around .96). The case of Canada deserves some attention. The γ estimates turn out to be slightly negative for the three GARCH models. However, these estimates are statistically insignificant and therefore indistinguishable from zero. Hence we conclude that for Canadian interest rates the level effect is ignorable after adjusting for stochastic volatility. This forms a sharp contrast with the UK case where the presence of the level effect is unambiguous. Therefore we are cautioned against trying to extrapolate a successful interest rate model to different countries without first looking for empirical evidence that supports such a model. Otherwise, modeling risks might be a big concern.

Turning to the estimates for the GARCH equation in the case of stochastic volatility models, we find the parameter estimates are all positive, which guarantees that the one-step ahead forecasts of conditional variances are positive. Also note that the sum of estimated coefficients for ϕ_1 and ϕ_2 is less than one, satisfying the stationarity constraint. As it is typically the case, the estimates of ϕ_1 are higher than ϕ_2.

With regard to the shape of the drift functions, we find that estimates of the nonlinear drift parameters α_2 and α_3 are all insignificant. But the linear drift parameters are statistically significant, especially for the linear drift models.

Table 2.5 reports the γ estimates as well as the log-likelihood values under the t distribution for the six models. We elect not to report the estimates for other parameters since in general they look similar to those obtained under the Gaussian distribution. We still observe the same phenomenon that γ estimates are sensitive to the inclusion of stochastic volatility but basically invariant to the different specifications of the drift equation. As is the case under normality, the estimates of γ are lower for GARCH models than for single factor models. For three of the five countries, single factor models imply γ above or very close to 1, whereas all are sufficiently less than one for models with GARCH volatility. Thus, single-factor models exaggerate the sensitivity of interest rate volatility to interest rate levels. With the exception of the U.S. weekly data, we find under the t distribution the estimates for γ are slightly higher than those obtained under the Gaussian distribution.

Table 2.5: γ Estimates under t Distribution

Panel A: Log-likelihood

	US (weekly)	US (monthly)	Canada	Germany	Japan	UK
NLSV	1759.16	-289.899	243.378	809.198	854.755	208.701
Quadratic	1758.64	-290.343	240.457	808.659	854.247	208.390
BHK	1758.39	-290.401	240.456	806.621	853.884	207.813
CHLS	1411.87	-295.832	205.146	758.611	808.405	126.550
AG	1410.08	-295.893	202.187	756.339	807.091	126.529
CKLS	1410.07	-296.032	202.160	753.896	806.991	126.337

Panel B: γ Estimates

	US (weekly)	US (monthly)	Canada	Germany	Japan	UK
NLSV	0.066628	0.87012	0.17345	0.45071	0.35684	0.15259
s.e.	0.012725	0.18366	0.07009	0.12632	0.05867	0.10652
Quadratic	0.065868	0.84981	0.16645	0.44423	0.36565	0.15392
s.e.	0.011973	0.19703	0.06988	0.12586	0.0646	0.10296
BHK	0.064235	0.85263	0.16651	0.43671	0.36989	0.15956
s.e.	0.011589	0.19635	0.06956	0.12731	0.06792	0.10241
CHLS	0.69843	1.4026	0.54135	1.03590	0.75489	1.0001
s.e.	0.0399	0.11353	0.06770	0.12281	0.03727	0.08822
AG	0.69753	1.4063	0.53410	1.0226	0.74984	0.99973
s.e.	0.03976	0.11296	0.06778	0.12343	0.03713	0.08820
CKLS	0.69781	1.4097	0.53259	1.0603	0.74936	1.00020
s.e.	0.03968	0.11237	0.06746	0.12392	0.03713	0.08825

2.7 Model Comparision

In this section, I compare the various models using three metrics: AIC, SBC, and the likelihood ratio test statistics.

2.7.1 LR Tests and Information Criteria

I compare the various model specifications based on two benchmarks. The first one is the likelihood ratio (LR) test, which is readily available since we have estimated both the unrestricted and the restricted models. The second benchmark is the information criteria for model selection. We use both the Akaike Information Criterion (AIC) and the Schwartz Bayesian Criterion (SBC).[3]

Tables 2.6 and 2.7 report the LR test statistics where the error terms for various models are assumed to follow the normal and t distributions respectively. D.f. refers to the degree of freedom for the test.

Under the normal distribution, one can reject all single-factor models without stochastic volatility at the 1% significance level. Similar conclusions can be reached under the t distribution, where only the CKLS model in the case of the US monthly data shows some marginal significance at the 1% level. At the five percent level all of the single-factor models are squarely rejected regardless of the specification of the drift term. On the other hand, under normality the LR tests fail to reject the linear drift stochastic volatility (BHK) model at both five and ten percent levels for all countries except Germany , which has a p-value of 1.8%. It is noteworthy that for the US data, the linear drift GARCH model is almost indistinguishable from the nonlinear drift GARCH model. Since most studies focus on the U.S. data, it seems that adding nonlinear drift to the short rate model is not justified by the empirical evidence presented here. Also for three of the five countries, Canada, Japan, and the

[3] AIC $= -2(L/T) + 2(k/T)$ and SBC $= -2(L/T) + klog(T)/T$, where L refers to the log-likelihood value, T the number of observations, and k the number of parameters.

U.K., the tests seem to favor the linear drift GARCH model more than its quadratic counterpart. Under the t distribution, the linear drift GARCH model does better than the quadratic model in three of the five countries, the U.S., Canada, and Japan. Once again we find strong evidence that the nonlinear drift specification does not seem to fit the U.S. data. And at a five percent level the BHK model is not rejected for all countries.

Table 2.6: Likelihood Ratio Tests under Normal Distribution

	US(monthly)			US(weekly)		
Model	LR	d.f.	p-value	LR	d.f.	p-value
Quadratic	0.042	1	0.8376	0.04	1	0.8415
BHK	0.52	2	0.7711	0.16	2	0.9231
CHLS	32.18	2	0.0000	1291.514	2	0.0000
AG	32.25	3	0.0000	1295.07	3	0.0000
CKLS	32.71	4	0.0000	1295.126	4	0.0000
	Germany			Canada		
Model	LR	d.f.	p-value	LR	d.f.	p-value
Quadratic	2.028	1	0.1544	3.2794	1	0.0702
BHK	7.972	2	0.0186	3.7644	2	0.1523
CHLS	72.508	2	0.0000	166.9172	2	0.0000
AG	72.94	3	0.0000	167.4272	3	0.0000
CKLS	76.478	4	0.0000	167.7252	4	0.0000
	UK			Japan		
Model	LR	d.f.	p-value	LR	d.f.	p-value
Quadratic	3.162	1	0.0754	2.534	1	0.1114
BHK	3.412	2	0.1816	3.962	2	0.1379
CHLS	102.468	2	0.0000	35	2	0.0000
AG	111.754	3	0.0000	37.494	3	0.0000
CKLS	112.312	4	0.0000	38.074	4	0.0000

Table 2.7: Likelihood Ratio Tests under t Distribution

This table reports the likelihood ratio test for various models under the assumption of t distribution. LR refers to the likelihood ratio test statistics. D.f. refers to the degree of freedom for the test.

	US(monthly)			US(weekly)		
Model	LR	d.f.	p-value	LR	d.f.	p-value
Quadratic	0.888	1	0.3460	1.04	1	0.3078
BHK	1.004	2	0.6053	1.54	2	0.4630
CHLS	11.866	2	0.0027	694.58	2	0.0000
AG	11.988	3	0.0074	698.16	3	0.0000
CKLS	12.266	4	0.0155	698.18	4	0.0000
	Germany			Canada		
Model	LR	d.f.	p-value	LR	d.f.	p-value
Quadratic	1.078	1	0.2991	5.842	1.0000	0.01565
BHK	5.154	2	0.0760	5.844	2.0000	0.0538
CHLS	101.174	2	0.0000	76.464	2.0000	0.0000
AG	105.718	3	0.0000	82.382	3.0000	0.0000
CKLS	110.604	4	0.0000	82.436	4.0000	0.0000
	UK			Japan		
Model	LR	d.f.	p-value	LR	d.f.	p-value
Quadratic	0.622	1	0.4303	1.016	1.0000	0.3135
BHK	1.776	2	0.4115	1.742	2.0000	0.4185
CHLS	164.302	2	0.0000	92.7	2.0000	0.0000
AG	164.344	3	0.0000	95.328	3.0000	0.0000
CKLS	164.728	4	0.0000	95.528	4.0000	0.0000

Table 2.8 and Table 2.9 report the AIC and SBC for all the countries and all the models (including the general unrestricted model) under the normal and t distributions respectively.

Under the Gaussian distribution, the SBC indicates that the linear drift stochastic volatility model is the best, and from the AIC one can draw a similar conclusion except for the case of Germany, where the NLSV and the Quadratic models slightly outperform the BHK model. This is natural since the AIC puts less penalty on models with more parameters than the SBC does. In all situations, models without stochastic volatility performs badly.

The situation is a little more complicated when we look at models under the t distribution. The AIC suggests the BHK model is best for the U.S., the U.K., and Japan, while in the second and third places for Canada and Germany. Using the SBC, the BHK model is best for US (weekly), Canada, U.K., and Japan.

To sum up, the evidence appears to be in favor of the linear drift GARCH (BHK) model, particularly for the U.S., the U.K., and Japan where short term interest rates are volatile.

Table 2.8: Information Criteria under Normal Distribution

This table reports the information criteria for the various models under normal distribution. AIC refers to the Akaike information criterion, and SBC refers to the Schwartz Bayesian criterion. $AIC = -2(L/T) + 2(k/T)$ and $SBC = -2(L/T) + klog(T)/T$, where L refers to the log-likelihood value, T the number of observations, and k the number of parameters.

	US(monthly)		US(weekly)		Canada	
Model	AIC	SBC	AIC	SBC	AIC	SBC
NLSV	1.451977	1.527982	-1.30485	-1.30036	0.076415	0.112934
Quadratic	1.447391	1.513896	-1.30563	-1.30171	0.077584	0.109538
BHK	1.443827	1.500831	-1.30639	-1.30303	0.0762	0.103589
CHLS	1.517972	1.574976	-0.78819	-0.78483	0.225198	0.252587
AG	1.513452	1.560955	-0.78756	-0.78476	0.223837	0.246662
CKLS	1.509845	1.547848	-0.78834	-0.7861	0.222283	0.240543
	Germany		Japan		UK	
Model	AIC	SBC	AIC	SBC	AIC	SBC
NLSV	-2.19447	-2.12997	-1.14057	-1.09287	0.58381	0.614561
Quadratic	-2.19442	-2.13798	-1.13988	-1.09815	0.584667	0.611574
BHK	-2.18697	-2.1386	-1.14061	-1.10485	0.583376	0.606439
CHLS	-2.06521	-2.01684	-1.10092	-1.06515	0.656426	0.67949
AG	-2.06817	-2.02786	-1.10029	-1.07048	0.661799	0.681019
CKLS	-2.06526	-2.03302	-1.10211	-1.07826	0.660736	0.676111

Table 2.9: Information Criteria under t Distribution

This table reports the information criteria for the various models under t distribution. AIC refers to the Akaike information criterion, and SBC refers to the Schwartz Bayesian criterion. $AIC = -2(L/T) + 2(k/T)$ and $SBC = -2(L/T) + klog(T)/T$, where L refers to the log-likelihood value, T the number of observations, and k the number of parameters.

	US(monthly)		US(weekly)		Canada	
Model	AIC	SBC	AIC	SBC	AIC	SBC
NLSV	1.395311	1.471317	-1.40543	-1.40094	-0.42991	-0.3934
Quadratic	1.392707	1.459212	-1.40581	-1.40189	-0.42641	-0.39445
BHK	1.388295	1.445299	-1.40641	-1.40305	-0.42823	-0.40084
CHLS	1.413733	1.470737	-1.12831	-1.12494	-0.36374	-0.33635
AG	1.409335	1.456838	-1.12767	-1.12487	-0.36016	-0.33733
CKLS	1.405302	1.443305	-1.12847	-1.12623	-0.36194	-0.34368
	Germany		Japan		UK	
Model	AIC	SBC	AIC	SBC	AIC	SBC
NLSV	-3.02339	-2.95889	-2.16561	-2.11792	-0.29602	-0.26527
Quadratic	-3.02513	-2.96869	-2.16687	-2.12514	-0.29704	-0.27013
BHK	-3.02121	-2.97284	-2.1685	-2.13273	-0.29766	-0.2746
CHLS	-2.84004	-2.79167	-2.05219	-2.01642	-0.1778	-0.15474
AG	-2.83524	-2.79493	-2.05138	-2.02158	-0.17925	-0.16003
CKLS	-2.8298	-2.79755	-2.05369	-2.02984	-0.18044	-0.16506

2.7.2 Regime Shifts and the Fed Experiment

Lamoureux and Lastrapes (1990) point out that one has to be careful when interpreting the finding of persistence in variance as in the GARCH model. Their study on stock-return data suggests that the pronounced phenomenon of variance persistence may be overstated because of the failure to account for structural shifts in the data. Since in this research I find the GARCH specification is important for the short rate models, whereas nonlinearity in drift is less significant, it is possible that the result might be affected due to the structural breaks. In the case of the US data, the Fed experiment from 1979 to 1982 stands out as a typical example of a structural break. In fact, during the Fed experiment we observe very high short rate volatility coincided with high interest rate levels. Duffee (1993) notes that the unreasonably high volatility elasticity estimates obtained by CKLS is attributable to their failure to account for structure breaks. I choose to reexamine the US data by introducing a dummy variable into the GARCH equation to take into account the structural break during the Fed experiment. Basically, this allows for changes in the unconditional variance during the Fed experiment for the short rate models. The results are obtained under the Gaussian distribution. Table 2.10 and 2.11 report the results for the U.S. weekly and monthly data respectively.

From Table 2.10, I find that estimates of β, the coefficient of the dummy variale, are not statistically significant. The sum of ϕ_1 and ϕ_2 remains unchanged. On the other hand, the Federal Reserve experiment dummy does seem to have a bigger impact on the parameters of interest for the one-factor models than the two-factor models. For example, in the case of the U.S. weekly data the estimates of γ decrease slightly from 0.15 to 0.14 for the stochastic volatility models. In contrast, they drop sharply from 0.7 to 0.43 for those single-factor models. The same conclusion applies to the case of monthly data. The mean reversion coefficient α_1 still has the correct sign. And the estimated coefficients of the GARCH equation all satisfy the usual constraints. Interestingly, we still find the familiar result that the inclusion of stochastic volatility seems to have a greater impact of the estimates of coefficients of interest than the specification of the drift term does. The LR tests solidly reject all single factor models regardless of their drift function specification, but fail to reject the BHK model. In addition both the AIC and the SBC pick the linear drift stochastic volatility model as the best one. Therefore we conclude that stochastic volatility is still the most important feature of the short rate after adjusting for the Fed experiment, whereas nonlinearity in the drift seems relatively unimportant for a parsimonious model of the short rate at least in the case of the US data.

Table 2.10: Parameter Estimates Subject to Changes in Unconditional Variance

This table reports the parameter estimates when including a dummy variable for the Federal Reserve Experiment during 1979 to 1982. β is the coefficient estimate for the dummy variable of the Federal Reserve experiment. AIC refers to the Akaike information criterion, and SBC refers to the Schwartz Bayesian criterion.

Panel A: U.S. Monthly Data

Model	NLSV	Quadratic	BHK	CHLS	AG	CKLS
α_0	-0.22119	0.050939	0.14324	-0.02211	0.066368	0.17555
s.e.	0.88514	0.13609	0.06149	1.05286	0.15136	0.06574
α_1	0.060852	0.013841	-0.02122	0.029677	0.01399	-0.02933
s.e.	0.15871	0.0478	0.01214	0.19288	0.05577	0.0136
α_2	-0.00539	-0.00298		-0.00469	-0.00386	
s.e.	0.00873	0.00394		0.01088	0.00482	
α_3	0.47185			0.15022		
s.e.	1.51587			1.76921		
γ	0.70499	0.70875	0.71286	1.0175	1.0181	1.0239
s.e.	0.12689	0.12618	0.12569	0.11282	0.11255	0.11213
ω	0.006874	0.006777	0.00658	0.006411	0.006396	0.006284
s.e.	0.00289	0.00283	0.00272	0.00244	0.00243	0.00238
ϕ_1	0.3805	0.38167	0.39117			
s.e.	0.11063	0.11067	0.10899			
ϕ_2	0.022667	0.022297	0.021582			
s.e.	0.01326	0.01298	0.01254			
β	0.008278	0.007934	0.00735	0.013106	0.013082	0.01271
s.e.	0.01172	0.01141	0.01093	0.00891	0.00888	0.00862
log-likelihood	-301.421	-301.47	-301.758	-308.76	-308.764	-309.085
LR		0.098	0.674	14.678	14.686	15.328
p-value		0.754243	0.713909	0.00065	0.002106	0.004067
AIC	1.453963	1.449508	1.446173	1.47897	1.474304	1.471124
SBC	1.539469	1.525514	1.512678	1.545474	1.531308	1.518627

Table 2.10: (continued)

Panel B: U.S. Weekly Data

	NLSV	Quadratic	BHK	CHLS	AG	CKLS
α_0	0.015199	0.018255	0.015071	-0.02168	0.01011	0.012173
s.e.	0.02188	0.00923	0.00484	0.0256	0.00883	0.00568
α_1	-0.00302	-0.00383	-0.00231	0.007443	-0.0011	-0.00224
s.e.	0.00654	0.00391	0.00109	0.00756	0.00393	0.00125
α_2	9.67E-05	0.000155		-0.00072	-0.00012	
s.e.	0.00054	0.00038		0.0006	0.0004	
α_3	0.002959			0.02619		
s.e.	0.0192			0.01979		
γ	0.14067	0.14052	0.13995	0.43966	0.4388	0.43914
s.e.	0.02938	0.02937	0.02934	0.02397	0.02394	0.02392
ω	0.000358	0.000359	0.000359	0.004797	0.004812	0.004807
s.e.	0.00005	0.00005	0.00005	0.00037	0.00037	0.00037
ϕ_1	0.74091	0.7408	0.74034			
s.e.	0.02137	0.02138	0.02142			
ϕ_2	0.16881	0.16896	0.16974			
s.e.	0.02219	0.02219	0.02224			
β	0.001953	0.001961	0.001956	0.029295	0.029477	0.029446
s.e.	0.0016	0.0016	0.00161	0.00499	0.00501	0.00501
log-likelihood	1635.48	1635.47	1635.39	1189.86	1188.99	1188.94
LR		0.02	0.18	891.24	892.98	893.08
p-value		0.887537	0.913931	2.9E-194	2.9E-193	5.3E-192
AIC	-1.30536	-1.30616	-1.30689	-0.94933	-0.94943	-0.95019
SBC	-1.28434	-1.28747	-1.29054	-0.93298	-0.93542	-0.93851

2.7.3 Sub-period Analysis

As an additional robustness check, we divide the US weekly data into two sub-periods: January 1954 to December 1982 and January 1983 to November 2001. The first sub-period has 1514 observations and the second has 978 observations. Note that Federal Reserve experiment is included in the first sub-period. In addition, the second sub-period covers approximately the same period as for the other countries that we examine, making the results directly comparable to those of other countries.

The results are reported in Table 2.11. Overall, the results for the two sub-periods are broadly consistent with those for the whole sample. Once again, we find the inclusion of GARCH effect lowers the γ estimates, and all the statistical tests are supportive of the linear drift stochastic volatility model specification. This suggests that our results are robust to different sample periods.

Table 2.11: Sub-period Analysis

This table reports the parameter estimates when we divide the U.S. weekly data into two sub-periods. The first sub-period is from 1954 to 1982, and the second sub-period is from 1983 to 2001.

Panel A: 1954 - 1982

	NLSV	Quadratic	BHK	CHLS	AG	CKLS
α_0	-0.0322	0.006722	0.009875	-0.03653	0.010828	0.013557
	0.02701	0.0108	0.00634	0.02739	0.00814	0.0059
α_1	0.012119	0.001276	-0.00033	0.013112	0.000139	-0.00169
	0.00839	0.00477	0.00173	0.00827	0.00413	0.0017
α_2	-9.32E-04	-0.00017		-0.001	-0.00019	
	0.00068	0.00048		0.0006	0.0004	
α_3	0.034676			0.034918		
	0.0219			0.01928		
γ	0.14721	0.1466	0.14751	0.75578	0.75448	0.75501
	0.03152	0.03155	0.03148	0.02036	0.02032	0.02028
ω	0.000224	0.000229	0.000227	0.003423	0.003444	0.003439
	0.00003	0.00003	0.00003	0.00024	0.00024	0.00024
ϕ_1	0.76784	0.76726	0.76778			
	0.01909	0.01969	0.01955			
ϕ_2	0.16546	0.16602	0.16513			
	0.02369	0.02401	0.02375			
loglik	807.903	806.675	806.61	456.894	455.258	455.139
LR		2.456	2.586	702.018	705.29	705.528
p-value		0.117077	0.274446	0.0000	0.0000	0.0000
AIC	-1.05668	-1.05637	-1.05761	-0.59563	-0.59479	-0.59596
SBC	-1.02855	-1.03177	-1.03652	-0.57454	-0.57721	-0.58189

Table 2.11: (continued)

Panel B: 1983 - 2001

	NLSV	Quadratic	BHK	CHLS	AG	CKLS
α_0	0.39313	0.027448	0.004885	0.091058	-0.0236	-0.00451
	0.27561	0.02909	0.01101	0.27693	0.03155	0.01147
α_1	-0.07799	-0.00947	-0.00097	-0.01437	0.007054	-0.00022
	0.05237	0.01033	0.00199	0.05264	0.0114	0.00208
α_2	4.69E-03	0.000738		0.000602	-0.00063	
	0.00309	0.00088		0.00311	0.00097	
α_3	-0.59585			-0.18769		
	0.44678			0.45039		
γ	0.22455	0.23724	0.24735	0.545	0.54479	0.54413
	0.08041	0.07964	0.07859	0.05914	0.05916	0.0592
ω	0.001044	0.001071	0.001078	0.002003	0.002005	0.00201
	0.00024	0.00024	0.00024	0.00041	0.00041	0.00041
ϕ_1	0.49606	0.47874	0.46714			
	0.07337	0.07429	0.0748			
ϕ_2	0.16251	0.15731	0.15319			
	0.05009	0.04831	0.04674			
loglik	841.383	840.489	840.143	736.227	736.14	735.93
LR		1.788	2.48	210.312	210.486	210.906
p-value		0.18117	0.289384	0.0000	0.0000	0.0000
AIC	-1.70426	-1.70448	-1.70581	-1.49331	-1.49517	-1.49679
SBC	-1.6643	-1.66951	-1.67584	-1.46333	-1.4702	-1.47681

2.8 Conclusions

Our findings shed some new light on the empirical evidence with regard to interest rate modeling.

First, we find evidence that the nonstationarity puzzle found in CKLS may not be a robust stylized fact of US short-term interest rates. We find the parameter estimates for γ belong to the nonstationary region only if we use the monthly risk-free dataset to estimate a one-factor model. Either the extension to a two-factor model or using an weekly dataset leads to strong rejection of the nonstationarity in the U.S. short rate. Hence our result supports BHK's conclusion that the level effect is exaggerated in CKLS. We also find that BHK's conclusion is robust to different specifications of the drift functions.

Second, our international evidence seems to supports the view that in order to avoid model risks it is of great importance for financial institutions to carefully study the related empirical evidence on short rates in different countries when pricing interest rate derivatives or hedging interest rate related risks. As an example we show that the level effect is evident in the U.S. and U.K. data but not in the Canadian data. Thus, the success of a specific model in a particular country does not necessarily guarantee its success in another.

Third, we provide new evidence with regard to the shape of the drift function. Aït-Sahalia, CHLS, and Stanton argue that the drift function of the US short rate is nonlinear. Chapman and Pearson (2000), Jones (2000), and Pritsker (1998) question the econometric techniques used in finding the nonlinear drift. We show that the empirical evidence in support of nonlinear drift function is very weak. We find that after taking into account the stochastic volatility in the data, the linear drift stochast volatility model seems to fit the data best. In fact, our model selction criteria indicate that even among the single-factor models, the nonlinear drift models do not outperform their linear drift counterparts. Figure 2.2 plots the drift functions for the six models estimated with U.S. weekly data. Interestingly, we find the only model that exhibit significant nonlinear drift is the CHLS model. The other five models, including models with nonlinear drift and stochastic volatility, have linear or very close to linear drift terms. However, if we plot the drift function for the CHLS model along with its confidence band (one standard deviation). The very wide confidence band suggests that the support for nonlinear drift is indeed very weak even in the case of the CHLS model.

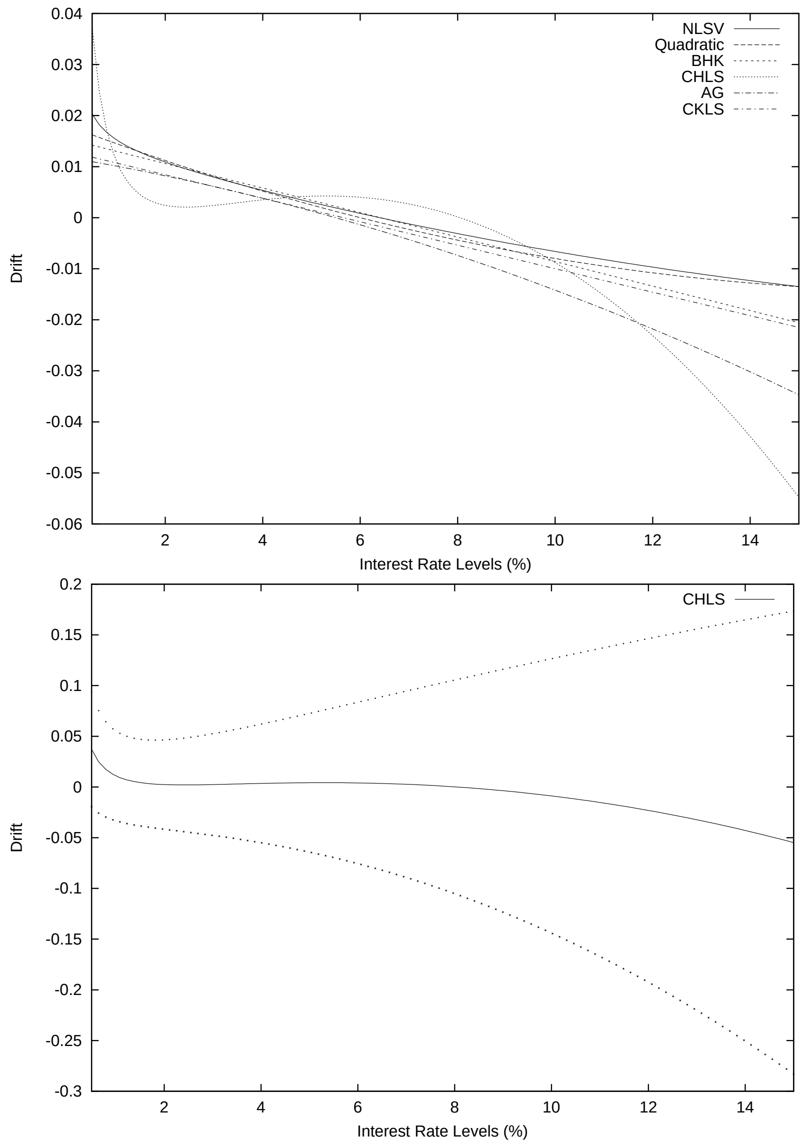

Figure 2.2: Drift Functions for US Weekly Data

Figures 2.3 to 2.6 plot the drift functions for the other four countries. Interestingly, we find sometimes imposing the nonlinear drift model on the data may imply a nonstationary interest rate process. For instance, for the German data, the nonlinear drift models imply negative drift functions when interest rates are low and positive drift functions when rates are higher. Similar situation is also found in the case of Japan for high interest rate levels.

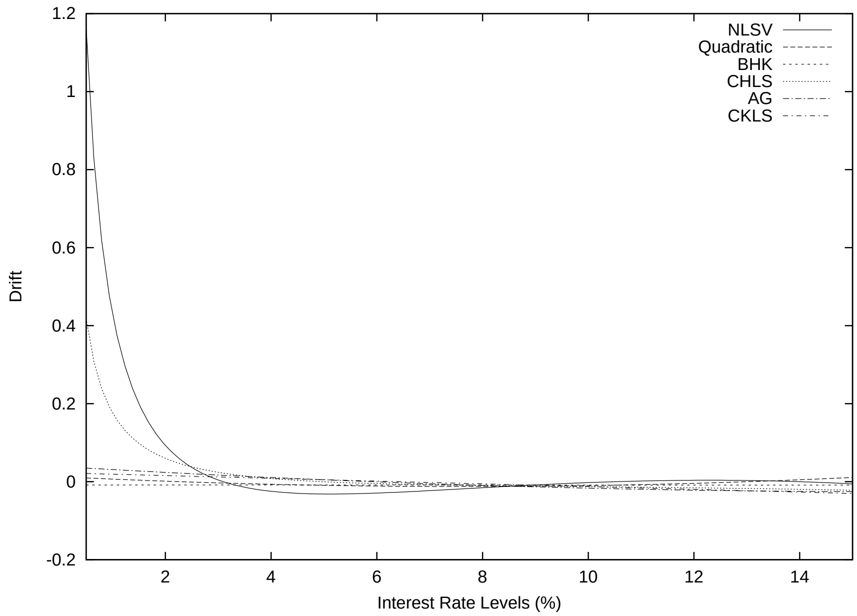

Figure 2.3: Drift Functions for Canadian Weekly Data

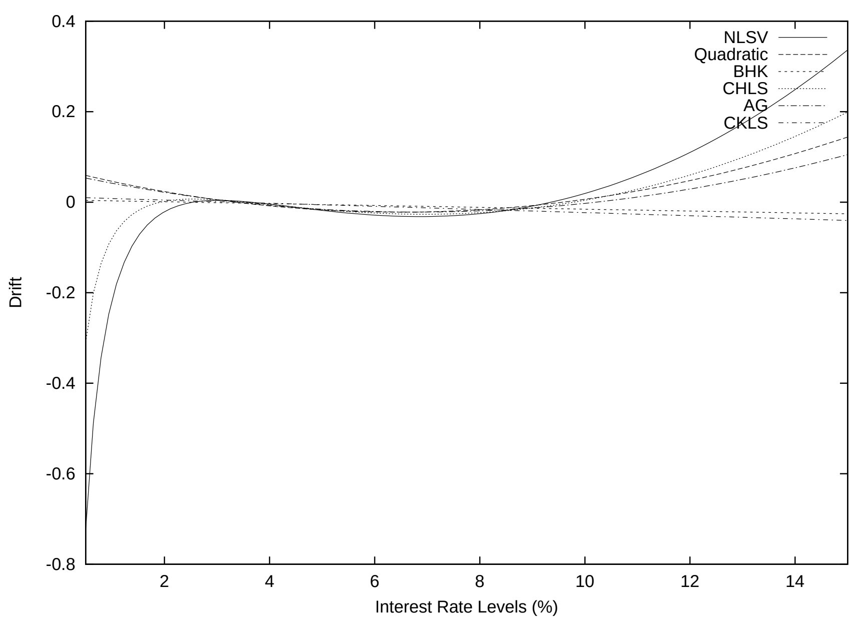

Figure 2.4: Drift Functions for German Weekly Data

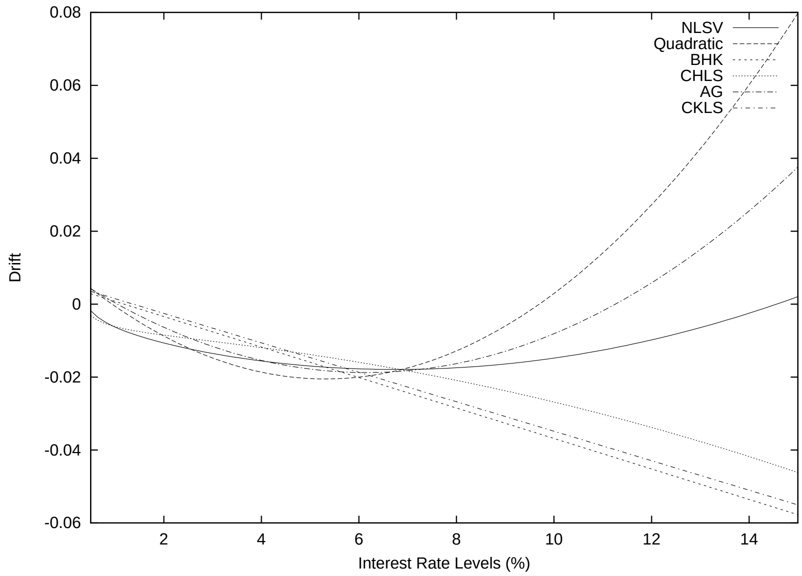

Figure 2.5: Drift Functions for Japanese Weekly Data

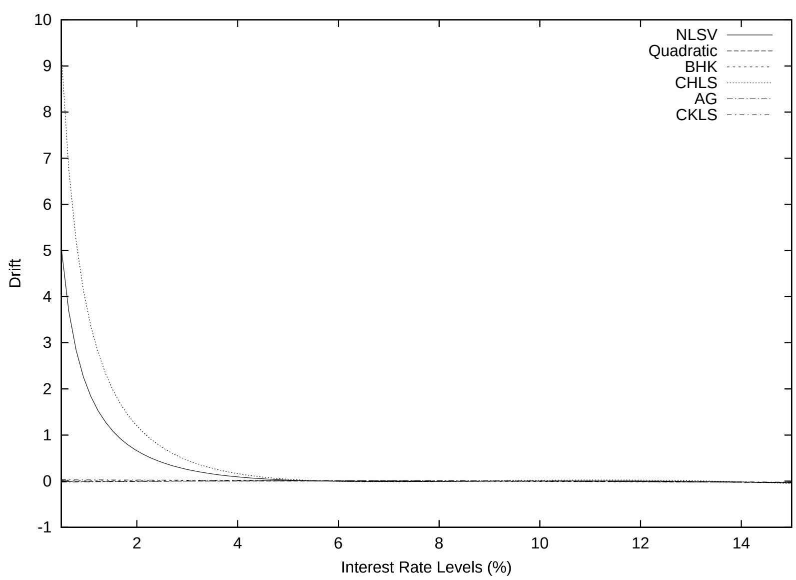

Figure 2.6: Drift Functions for UK Weekly Data

Chapter 3

Regime Shifts in Interest Rate Volatility

3.1 Introduction

It has been a well-established empirical fact that the volatility of the U.S. short term interest rate is itself volatile (e.g. Brenner, Harjes, and Kroner (1996) , Andersen and Lund (1997)). Moreover, Ball and Torous (1999) provide evidence that stochastic volatility is also a salient feature characterizing the short rate dynamics of some other developed countries, such as the U.K., Japan, and so on. In the previous chapter, we model the stochastic volatility in short rates with a GARCH model. A more attractive alternative probably is the stochastic volatility (SVOL) model of Taylor (1986). In this chapter we take a look at the SVOL model and compare it with its regime-switching counterpart.

As a competitor of the ARCH/GARCH models, the SVOL model has a natural interpretation of being a discrete-time analog of continuous-time stochastic volatility diffusion models, which are widely used in derivative pricing. In the SVOL approach, the underlying volatility is modeled as an unobserved state variable. Ball and Torous (1999) adopt this approach to model the volatility process of the short term interest rates. A potential problem, however, is that the persistence in volatility may be overstated due to structural breaks in the volatility process. Lamoureux and Lastrapes (1990) first point out this misspecification problem in the context of the GARCH models. But the same criticism may also be applied to the SVOL model.

One way to account for structural breaks is to use the regime-switching model of Hamilton (1989). Several authors have attempted to model the conditional heterescedasticity in financial assets in the regime-switching framework. Gray (1996) argues that structural breaks, such as the Federal Reserve experiment of 1979 to 1982, justify the use of a regime-switching model to model the short rate volatility. Duffee (1993) also presents similar evidence on this issue. He indicates that the unreasonably high volatility elasticity estimates obtained by Chan, Karolyi, Longstaff and Sanders (1992) may be attributable to their failure to account for structure breaks. Regime shifts in the ARCH model has been studied in Hamilton and Susmel (1994) as well as in Cai (1994). Cai applies the regime-switching ARCH model to the monthly returns of the three-month U.S. Treasury bills, and identifies two periods of shifting regimes, associated with the oil crisis and the Federal Reserve monetary policy experiment from 1979 to 1982 respectively. Gray (1996) develops a generalized regime-switching model where regime shifts are incorporated into the GARCH model. He concludes that the regime-switching GARCH model outperform the its single-regime counterparts for the U.S. short rates.

Modeling regime shifts in the SVOL model has been considered in So, Lam, and Li (1998). However, their approach is based on Bayesian methods and, therefore, computationally intensive. More importantly, compared with the method used in this chapter, their approach is restrictive in the sense that their model do not allow for non-zero correlation between the disturbance terms. Nor do they consider the case of time-varying transition probabilities.

In this chapter, we compare the SVOL approach with the regime-switching stochastic volatility (RSSV) approach in modeling short rate volatilities. Our contributions to the literature are as follows:

First, we find strong evidence that there exist regime shifts in short rate volatilities for the four countries under consideration. These volatility regimes are sharply defined and seem to last for a long period of time.

Second, by comparing the results from the SVOL and RSSV models, we find strong evidence that using the SVOL model may lead to spuriously high volatility persistence if regime shifts in the volatility process are unaccounted for. In fact, for three out of the four countries, the apparent volatility persistence drop significantly after we allow for regime shifts.

Third, we extend the RSSV model by relaxing the independence assumption commonly used in the SVOL/RSSV models. We find evidence that modeling the correlation is important at least for the U.S short rate volatilities.

Fourth, apart from a constant transition probability RSSV model, we also put forward a time-varying transition probability RSSV model where the transition probabilities are allowed to vary with other exogenous variables. We find evidence that a constant transition probability RSSV model is the most parsimonious specification.

This chapter is organized as follows. Section 3.2 gives further background and sets up the models. Section 3.3 compares SVOL model with the RSSV model using the short rate data for the U.S., the U.K., Canada, and Japan. Section 3.4 extends the RSSV model by allowing for non-zero correlations. Section 3.5 estimates two alternative specifications of the basic RSSV model. Section 3.6 summarizes this chapter. Section 3.7 describes some of the technical details.

3.2 Background and Model Specifications

This section gives further background regarding the stochastic volatility model and its regime-switching counterpart.

3.2.1 The Stochastic Volatility Model

We are interested in modeling the volatility of a financial time series. The simplest framework is to start with a (Gaussian) white noise process

$$y_t = \sigma_t \epsilon_t, \tag{3.1}$$

where ϵ_t is usually assumed to be a unit variance (Gaussian) white noise. σ_t is known as the volatility. In the ARCH/GARCH class of models, σ_t^2 is modeled as a deterministic function of lagged squared errors as well as the lagged values of σ_t^2.

In the SVOL model, however, σ_t is treated as an unobserved stochastic process. This feature of the SVOL model makes it fit naturally to the theoretical work in finance, especially those stochastic volatility models in option pricing theory, such as the Hull and White (1987) model. To make sure that σ_t^2 is always positive, we define $h_t \equiv ln(\sigma^2)$. The SVOL model can be specified as follows:

$$y_t = e^{h_t/2} \epsilon_t, \tag{3.2}$$

$$h_{t+1} = \mu + \phi h_t + \eta_t, \tag{3.3}$$

where ϵ_t and η_t are typically assumed to be uncorrelated Gaussian innovations. Harvey and Shephard (1996) shows how to allow for non-zero correlations between ϵ_t and η_t, which is important, for example, to model the pronounced asymmetric volatility phenomenon for stock returns.

3.2.2 Quasi-maximum Likelihood Estimation of the SVOL Model

To estimate the SVOL model specified in equations (3.2) and (3.3) , we first linearize equation (3.2) by squaring and taking the natural logarithm of y_t. We obtain

$$ln(y_t^2) = \omega + h_t + \xi_t, \tag{3.4}$$

where $\omega \equiv E(ln(\epsilon_t^2))$ and $\xi_t \equiv ln(\epsilon_t^2) - E(ln(\epsilon_t^2))$.

If $\epsilon_t \sim NID(0,1)$, $ln(\epsilon_t^2)$ is known to have a log χ^2 distribution with mean -1.2704 and variance $\pi^2/2$.[1] Estimation of the system of equations (3.4) and (3.3) is not an easy task because of the fact that h_t is an unobservable latent process.

One estimation strategy is based on the Kalman filter by noticing that equations (3.4) and (3.3) conform to a linear state-space form. By assuming $\xi_t \sim NID(0, \pi^2/2)$, estimation of the SVOL model can be carried out by quasi-maximum likelihood (QML). The Kalman filter is used to obtain the prediction error decomposition of the Gaussian likelihood function, which is then numerically maximized.[2] The QML approach to the estimation of the SVOL model has been discussed in Ruiz (1994) and extended to the multivariate case by Harvey, Ruiz, and Shephard (1994). We discuss technical details in the section 3.7.

To apply the SVOL model to the short rate volatility, we start with the Chan, Karolyi, Longstaff and Sanders (1992) (CKLS) model of the short rate:

$$dr = (\alpha + \beta r) + \sigma r^{\gamma} dW.$$

A more detailed discussion of this model can be found in the previous chapter. Here we only mention that CKLS find the unrestricted estimate of γ, the elasticity of volatility, is approximately 1.5, which implies nonstationarity in short rate process. In the previous chapter we have shown that this result is probably due to CKLS's failure to account for the presence of stochastic volatility in short rate process. As in Ball and Torous (1999), we augment the discrete-time CKLS model by incorporating the stochastic volatility into the short rate process:

$$\Delta r_t = a + b r_{t-1} + \sigma_{t-1} r_{t-1}^{\gamma} \epsilon_t \tag{3.5}$$

[1] See Abramovitz and Stegun (1970).

[2] See Hamilton (1994). Kitagawa (1987) discusses the non-Gaussian state-space approach. An application of Kitagawa's method to the SVOL model can be found in Fridman and Harris (1998). Typically this method is computationally more intensive than the QML method discussed here.

$$h_t = \mu + \phi h_{t-1} + \sigma_\eta \eta_t, \tag{3.6}$$

where $\Delta r_t \equiv r_t - r_{t-1}$, $h_t \equiv ln(\sigma_t^2)$, ϵ_t and η_t are i.i.d. standard Gaussian innovations. We will relax the independence assumption later. This SVOL model can be viewed as a discrete-time approximation to a two-factor interest rate model, such as the Longstaff and Schwartz (1992) model, where the short rate and its volatility are specified as two factors.

To estimate the model given by equations (3.5) and (3.6), we first estimate the drift parameters a and b by OLS. Let $y_t \equiv \Delta r_t - (a + br_{t-1})$, and $x_t \equiv ln(y_t^2)$. We obtain the following state-space model:

$$x_t = h_{t-1} + 2\gamma ln(r_{t-1}) + ln(\epsilon_t^2) \tag{3.7}$$

$$h_t = \mu + \phi h_{t-1} + \sigma_\eta \eta_t. \tag{3.8}$$

Rewrite equation (3.7) as follow:

$$x_t = h_{t-1} + 2\gamma ln(r_{t-1}) - 1.2704 + \xi_t, \tag{3.9}$$

where $\xi_t = ln(\epsilon_t^2) + 1.2704$. We note the similarity between equations (3.9) and (3.4). Hence we can estimate the SVOL model for short rates with QML.

3.2.3 The Regime-Switching Stochastic Volatility Model

Although the ARCH/GARCH and the SVOL models provide a nice way to account for volatility persistence that is commonly observed in financial data, there is a concern that the apparent volatility persistence may be overestimated because of the failure to account for structural shifts in volatility. Lamoureux and Lastrapes (1990) investigate this possibility in the case of the GARCH models subject to deterministic structural breaks. But obviously the same problem may also plague the SVOL model with random structural breaks.

In the absence of a perfect knowledge when such structural shifts might occur, the regime-switching model of Hamilton (1989) may be a useful tool to account for random structural shifts. Attempts to incorporate regime shifts into the ARCH/GARCH models have been made by Hamilton and Susmel (1994), Cai (1994) Gray (1996), among others.

So, Lam, and Li (1998) generalize the SVOL model by adding the regime-switching properties, which we refer to as the regime-switching stochastic volatility (RSSV) model. The switching dynamics is governed by a first-order Markov process. They estimate the RSSV model with a Bayesian approach (Markov-chain Monte Carlo), which is computationally intensive.

A more convenient way to estimate the RSSV model is to use Kim's filter. Kim (1994) extends Hamilton's regime-switching model to a general state-space form. Since the SVOL model is typically written in the linear state-space form, we can apply Kim's filter to the estimation of the RSSV model. Technical details are discussed later.

In our model specification, we focus on the case where the parameter μ in equation (3.8) is allowed to be regime-dependent. Later we also estimate an alternative specification where both μ and ϕ are regime dependent. Let s be an unobserved regime variable, where $s = 0$ and $s = 1$ denote two distinct volatility regimes. We rewrite equation (3.8) as follows:

$$h_t = \mu_s + \phi h_{t-1} + \sigma_\eta \eta_t, \tag{3.10}$$

where μ_s and hence the unconditional log variance $(\frac{\mu_s}{1-\phi})$ is regime-dependent.

Equations (3.9) and (3.10) form the RSSV model for short rates that will be estimated in this chapter.

In the following discussion, we concentrate on a simple two-state RSSV model, i.e. the regime variable s is assumed to follow a two-state, first-order Markov process. The transition probability matrix $\mathbf{X}$ of the Markov process can be written as follows:

$$\mathbf{X} = \begin{pmatrix} p & 1-p \\ 1-q & q \end{pmatrix} \tag{3.11}$$

where $p = Pr(s_t = 0|s_{t-1} = 0, I_{t-1})$, $q = Pr(s_t = 1|s_{t-1} = 1, I_{t-1})$, and I_{t-1} is the information set.

In So, Lam, and Li (1998), the transition probabilities are assumed to be constants. Later we show how to relax this assumption by allowing for time-varying transition probabilities.

3.3 A Comparison of the Stochastic Volatility and the Regime-Switching Stochastic Volatility Models

In this section, we compare the SVOL model with its regime-switching counterpart the RSSV model using short term risk-free interest rate data from four developed countries, the United States, Canada, Japan, and the United Kingdom.

3.3.1 Data

The interest rate data used in this chapter is basically the same as those used in the previous chapter.[3] The U.S. dataset includes 2492 weekly observations of the 3-month Treasury bill rates obtained from the Federal Reserve site, ranging from January 1954 to October 2001. I also exam the interest rate data from three other industrialized countries: Canada, Japan, and the United Kingdom. The data are obtained from Datastream. The interest rate series include: Canada Treasury

[3] We exclude the dataset for Germany due to the convergence problem when estimating the RSSV model, probably because of the relatively small sample size.

bill 1 month (CN13883) from January 1980 to December 2000, altogether 1095 observations; Japan Interbank 1 month offered rate (JPIBK1M) from December 1985 to December 2000, 782 observations, and UK interbank 1 month middle rate (LDNIB1M) from January 1975 to December 2000, 1356 observations. All data are weekly observations.

The plots of these interest rate data as well as their first order difference have been shown in the previous chapter. A cursory look indicates that these interest rate series do exhibit very distinctive patterns. Both very high interest rate levels and extremely low interest rate levels (less than 1% in the case of Japan) are observed. A common feature is that the interest rate movements typically exhibit very volatile behavior, especially in the case of the U.S. and the U.K.. For example, the U.S. short rates show some dramatic swings during the Federal Reserve experiment period of 1979 to 1982. However, starting from the 90s the US short rates look relatively tamed. This suggests that the RSSV model might be a useful tool to characterize the interest rate volatilities.

In the previous chapter,[4] we have reported the summary statistics for these interest rate series and their first order differences. So we choose not to repeat here.

3.3.2 Model Estimation

We first estimate the drift function with OLS. As we have shown in the previous chapter, non-linear drift does not seem to be a robust fact for the interest rate data under consideration. Hence we focus on the linear drift specification. We then subtract the estimated drift terms to obtain the residuals. Figure 3.1 plots the residuals from the OLS model.

[4]See table 2.1.

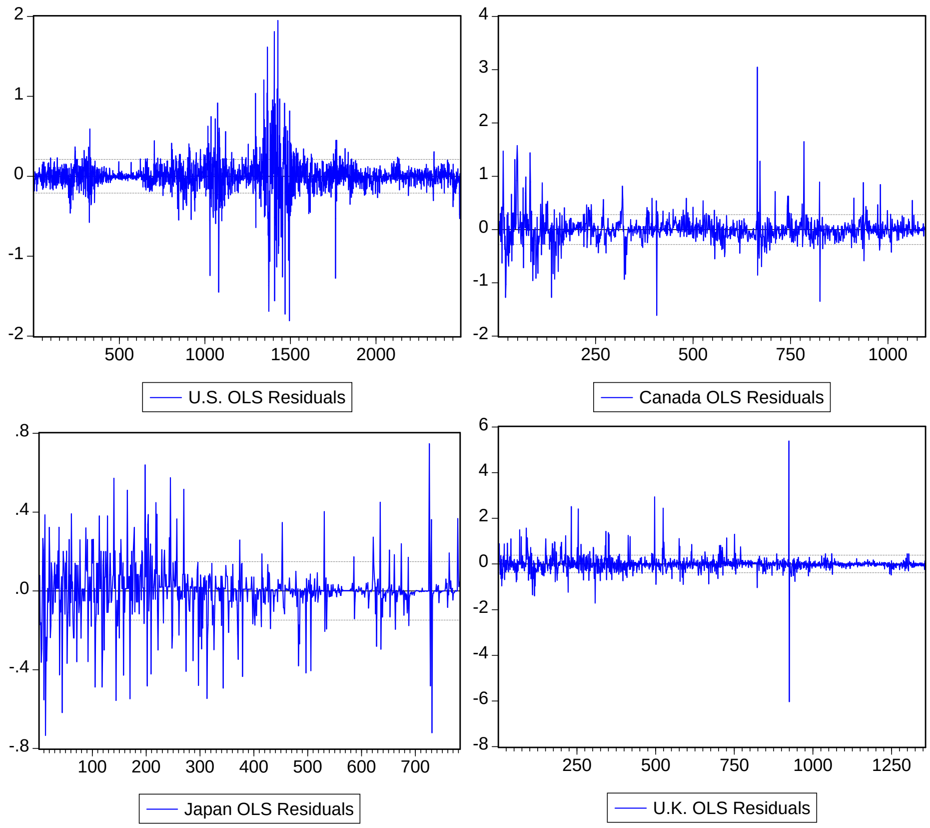

Figure 3.1: Residuals from OLS

Eyeball statistics seem to suggest that constant volatility is not a good assumption. For all the four nations, volatility seems to evolve stochastically over time. In addition we can also visually identify periods of higher volatility versus periods of lower volatility.

The OLS estimates for a and b are reported in panel A of table 3.1. We notice that the estimates for b are negative in all four countries, which is consistent with the interpretation that b is the mean reversion parameter for linear drift models.

Table 3.1: Parameter Estimates of the SVOL Model

This table reports the parameter estimates of the stochastic volatility (SVOL) model for short term interest rates from four countries. Panel A reports the drift parameter estimate for the following model with OLS:

$$\Delta r_t = a + b r_{t-1} + \epsilon_t.$$

Panel B reports the parameter estimate using the following SVOL model:

$$x_t = h_{t-1} + 2\gamma ln(r_{t-1}) - 1.2704 + \xi_t,$$

$$h_t = \mu + \phi h_{t-1} + \sigma_\eta \eta_t,$$

where x_t is the log of squared OLS residuals, and h_t denotes the log variance. Standard errors are in parentheses.

Panel A: OLS drift parameter estimates

Parameter	US	Canada	Japan	UK
a	0.0174	0.0179	0.0007	0.0558
	(0.0094)	(0.0201)	(0.0085)	(0.0324)
b	-0.0031	-0.0030	-0.0027	-0.0060
	(0.0015)	(0.0022)	(0.0021)	(0.0031)

Panel B: SVOL model parameter estimates

Parameter	US	Canada	Japan	UK
ϕ	0.9662	0.9803	0.9905	0.9580
	(0.0386)	(0.0201)	(0.0065)	(0.0709)
μ	-0.2272	-0.1165	-0.0649	-0.3237
	(0.2597)	(0.1267)	(0.0445)	(0.5474)
σ_η	0.3458	0.2231	0.2787	0.3916
	(0.1949)	(0.1219)	(0.0961)	(0.3247)
γ	0.7093	0.5486	0.7542	0.9211
	(0.0483)	(0.0780)	(0.0417)	(0.1000)
Log-likelihood	-5907.69	-2548.67	-1966.10	-3219.08

Panel B of Table 3.1 reports the parameter estimates for the SVOL model consisting of equations (3.8) and (3.9). We find that, compared with the result from the GARCH model reported in the previous chapter, the γ estimates reported here seems to be higher except in the case of the U.K. data. However, all the estimates are less than one, which is consistent with the results from the previous chapter and implies stationary short rate process.

We also notice that the estimates for ϕ are very high for all countries. The highest number is 0.9905 for Japan and the lowest is 0.9580 for the U.K. data with the estimates for the U.S. and Canada in the middle. Such high estimates of ϕ imply very persistent volatility processes. Moreover, the standard errors for these ϕ estimates are relatively small.

As pointed out by Lamoureux and Lastrapes (1990), the apparent highly persistent volatilities found here could be misleading if we do not account for possible structural breaks in the volatility process. Therefore we proceed by estimating the RSSV model to account for possible regime shifts in short rate volatilities.

Table 3.2 reports the parameter estimates of the RSSV model (equations (3.9) and (3.10)) using the interest rate data from the four countries under consideration. We find strong evidence of regime shifts in short rate volatilities. We notice that the estimated transition probabilities are very close to 1. In fact for all four countries their transition probabilities p and q are above 0.99, which implies very distinct volatility regimes. In addition, the standard errors for these transition probability estimates are very small, suggesting the transition probabilities are precisely estimated. High transition probabilities imply the regimes are indeed persistent.

Table 3.2: Parameter Estimates of the RSSV Model

This table reports the parameter estimates of the following regime-switching stochastic volatility (RSSV) model for short term interest rates from four countries:

$$x_t = h_{t-1} + 2\gamma ln(r_{t-1}) - 1.2704 + \xi_t,$$

$$h_t = \mu_s + \phi h_{t-1} + \sigma_\eta \eta_t,$$

where we allow the intercept of the volatility equation to be regime dependent. The lower case s is an unobserved regime variable that follows a first-order, two-state Markov process, where we assume constant transition probabilities. h_t denotes the log variance. x_t is the log of squared residuals from the following OLS regression

$$\Delta r_t = a + br_{t-1} + \epsilon_t.$$

Standard errors are in parentheses.

	US	Canada	Japan	UK
p	0.9985	0.9951	0.9975	0.9947
	(0.0012)	(0.0051)	(0.0029)	(0.0034)
q	0.9975	0.9960	0.9957	0.9916
	(0.0020)	(0.0036)	(0.0039)	(0.0060)
ϕ	0.9592	0.7585	0.6071	0.6205
	(0.0185)	(0.0850)	(0.0606)	(0.1260)
μ_0	-0.2156	-1.4719	-2.3735	-2.2464
	(0.1069)	(0.5578)	(0.3767)	(0.8053)
μ_1	-0.2902	-1.7468	-3.0907	-2.9305
	(0.1395)	(0.6718)	(0.4953)	(1.0365)
σ_η	0.2758	0.6462	1.4675	0.8235
	(0.0631)	(0.1425)	(0.1266)	(0.1568)
γ	0.4851	0.7533	0.8447	0.6682
	(0.1449)	(0.2176)	(0.0684)	(0.1539)
Log-likelihood	-5682.71	-2516.74	-1907.53	-3153.45

In fact, we can calculate the expected durations of these different regimes using the transition probability estimates. The expected duration of a particular regime is given by $\frac{1}{1-pr}$ where pr is the estimated transition probability for that regime. We found the expected duration in our datasets last for years instead of months, which implies that these regimes might be related to some fundamental state variables with long-lasting effects rather than short term transient market movements. Formal hypothesis testing regarding the number of regime states requires the use of numerical methods that are very costly to compute. See Hansen (1992) and the related erratum (1996). Here we simply note that the psudo-likelihood ratio test statistics are all extremely significant, with the highest p-value being 8.8E-14 in the case of Canada and lowest p-value 3.3E-97 in the case of the U.S. data.

To further convince ourselves that regime-shifts are indeed a robust fact. We plot the high-volatility regime probabilities along with the filtered conditional volatilities for the four countries in Figures 3.2 to 3.5. First, we notice that the probability plots are consistent with the conditional volatility plots. We find periods of high conditional volatilities matching with high regime probabilities. Second, the evidence from the regime probability plot for the U.S. data seems to suggest a correlation between high volatility regimes and macroeconomic shocks. For example, the first high volatility regime period (from approximately 1954 to 1961) coincides with two NBER dated recessions during this period. The second high volatility regime period (from about 1970 to 1976) corresponds to two recessions as well as the oil crisis. The third period (from 1979 to the mid 80's) seems related to the Federal Reserve experiment and two economic recessions. The low volatility regime dominates the 90s when we have a sustained economic expansion.

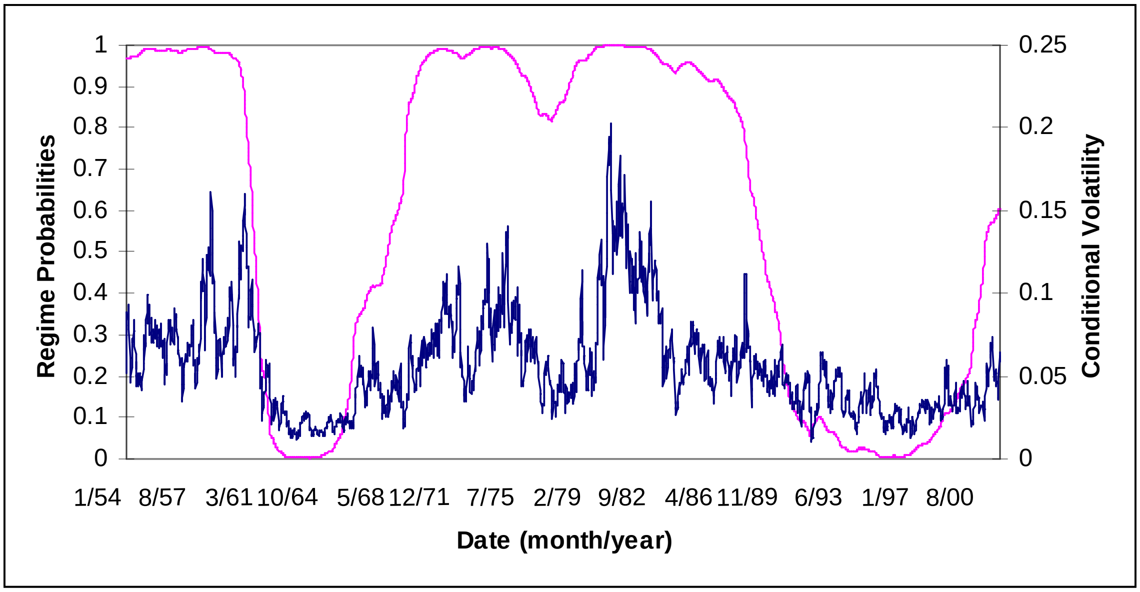

Figure 3.2: Regime Probabilities and Conditional Volatility for U.S. Short Rates

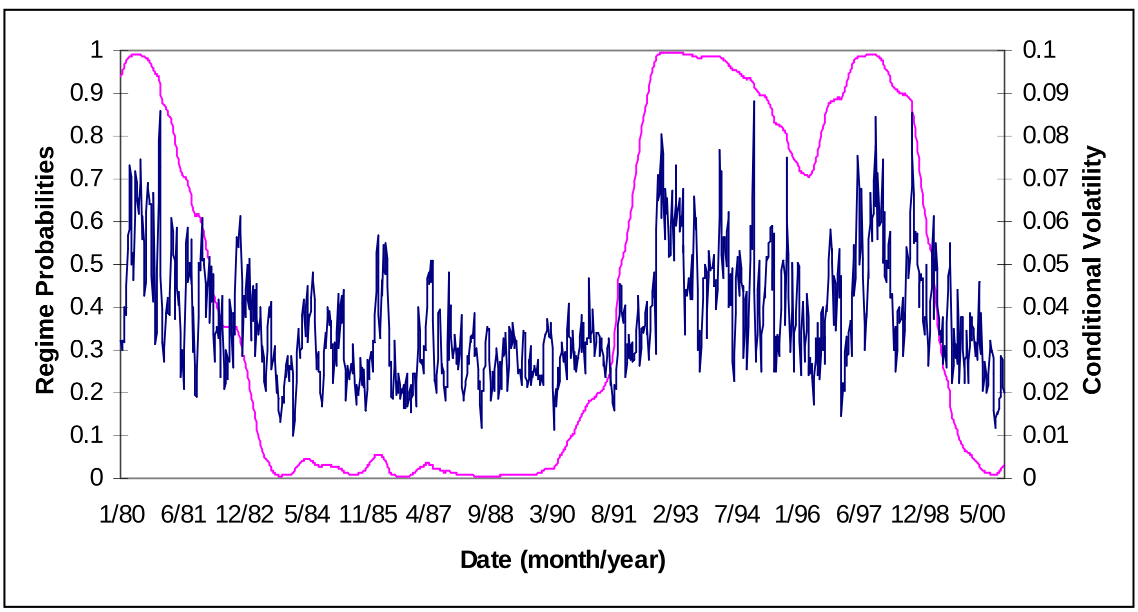

Figure 3.3: Regime Probabilities and Conditional Volatility for Canada Short Rates

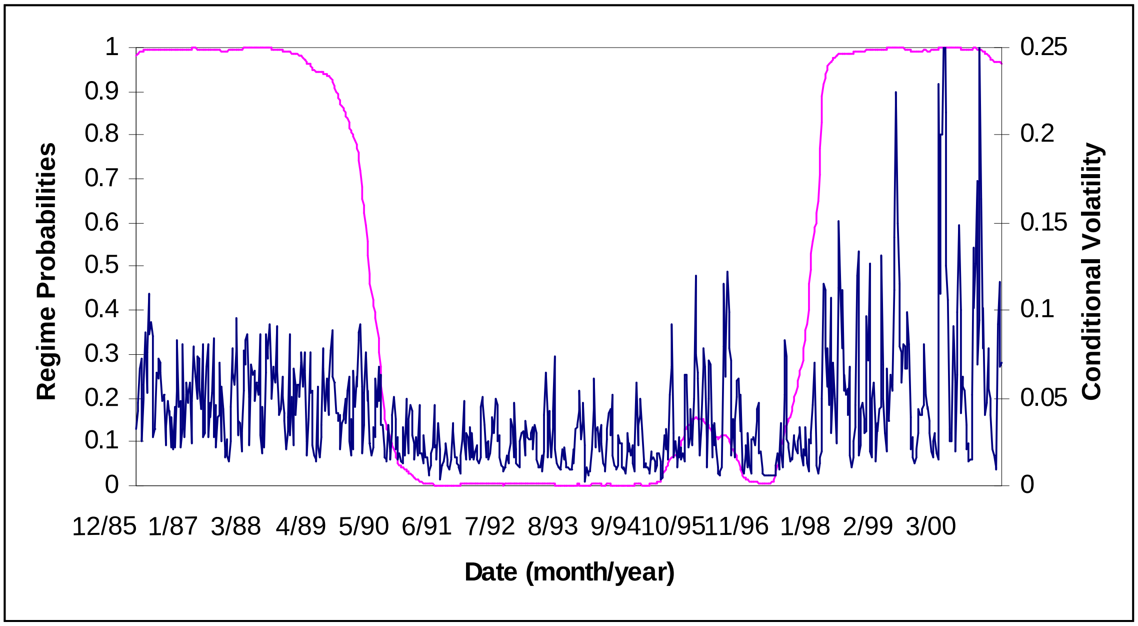

Figure 3.4: Regime Probabilities and Conditional Volatility for Japan Short Rates

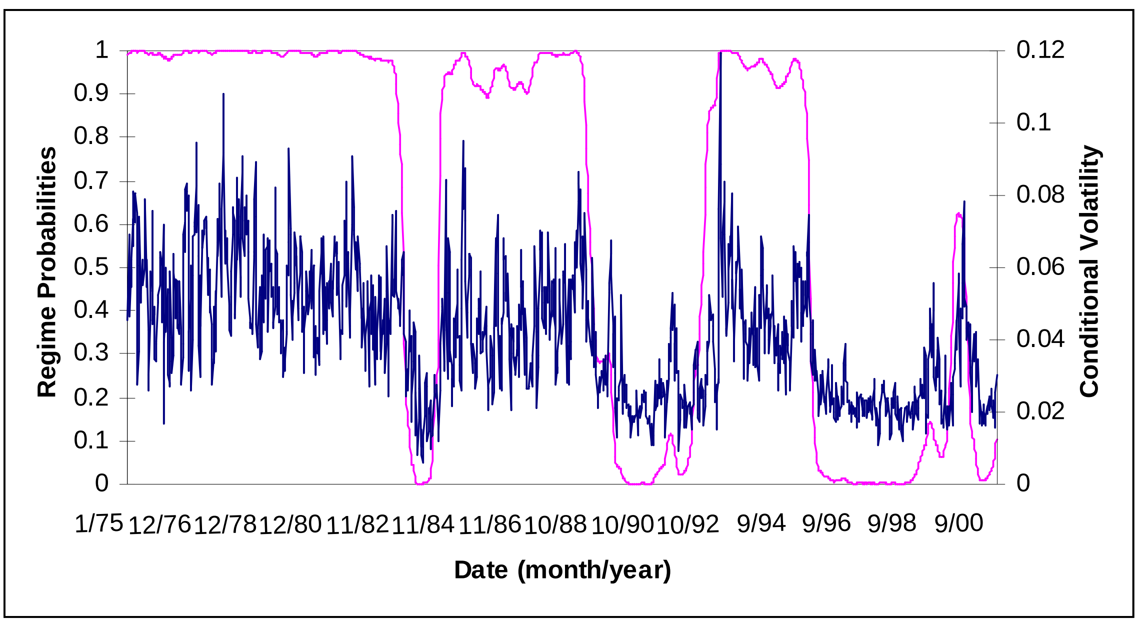

Figure 3.5: Regime Probabilities and Conditional Volatility for U.K. Short Rates

Another interesting phenomenon is related to the parameter estimates for ϕ. If we compare the ϕ estimates obtained from the SVOL model versus those from the RSSV model, we find the parameter estimates drop dramatically for three out of the four countries. For the Canada short rate volatilty, ϕ decreases from 0.98 to 0.76, for Japan down from 0.99 to 0.61, and for the United Kingdom from 0.96 to 0.62. These represent decreases of 22% to 38%. In addition the size of the standard errors are relatively small compared with the magnitude of the decrease. The same parameter estimates decrease slightly in the U.S. case, and are not statistically different from each other. This seems to confirm the our concern that the SVOL model could lead to overstated volatility persistence due to its failure to account for regime shifts in the volatilities.

The γ estimates are broadly similar for the two models under consideration. We note that all the estimates of γ are less than one. In other words, they all imply stationary interest rate processes. This reaffirms the results from the previous chapter. Namely, the CKLS puzzle is possibly due to their failure to adjust for stochastic volatilities in the interest rates.

3.4 Modeling Correlation in the Regime-Switching Stochastic Volatility Model

The SVOL model chooses to model the log variance instead of volatility itself to ensure positive variances. However, by squaring the data, we may lose useful information unless the true correlation between ϵ_t and η_t is zero. Most SVOL models impose the zero-correlation assumption. Nevertheless, *a priori*, we have no reason to believe that this assumption is valid. As a matter of fact, empirical evidence from stock returns indicates that this zero-correlation assumption may be false because of the well documented phenomenon of asymmetric volatilities for stock returns. Namely stock market volatilities increase (decrease) as stock prices drop (go up).[5]

In the case of interest rate data, it is unclear whether or not the correlation between ϵ_t and η_t is zero. For example, Ball and Torous (1999) assume zero correlation when estimating the SVOL model. They argue that in their sample the correlations are low. However, their reported correlations (standard errors in parentheses) for the Euro-mark and the Euro-yen series are 0.163 (0.082) and 0.224 (0.091). It looks like at least for these two interest rate series, zero-correlation is not a very good assumption. Ball and Torous do not report the correlation for U.S. T-bill yields due to convergency problem.

Harvey and Shephard (1996) propose a method to handle the correlation between the two disturbances for the SVOL model. They show that the loss of information due to squaring may be recovered if we carry out inference conditional on the signs of the observations.

Consider the SVOL model as expressed in equations (3.3) and (3.4). Harvey and Shephard show that if we condition on the sign of the residuals to recover the information about correlation, then we only need to modify the SVOL model as follows:

$$ln(y_t^2) = \omega + h_t + \xi_t, \tag{3.12}$$

$$h_t = \phi h_{t-1} + g_t u^* + \eta_t^*, \tag{3.13}$$

$$\begin{pmatrix} \xi_t \\ \eta_t^* \end{pmatrix} \Bigg| g_t \sim ID\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\xi^2 & \gamma^* g_t \\ \gamma^* g_t & \sigma_\eta^2 - u^{*2} \end{pmatrix} \right), \tag{3.14}$$

where g_t is a variable that takes 1 (-1) if y_t is positive (negative). $u^* = E_+(\eta_t)$ and $\gamma^* = cov_+(\eta_t, \xi_t)$. E_+ and cov_+ denote the expectation and covariance conditional on ϵ_t being positive. Note that in our model σ_ξ^2 equals $\frac{\pi^2}{2}$ and is not a parameter to be estimated.

[5]See, for example, Black (1976) and Schwert (1989).

When ϵ_t and η_t are bivariate normal with $corr(\epsilon_t, \eta_t) = \rho$, Harvey and Shephard show that

$$u^* = 0.7979\rho\sigma_\eta, \tag{3.15}$$

$$\gamma^* = 1.1061\rho\sigma_\eta. \tag{3.16}$$

Because of the fact that equations (3.12) and (3.13) still form a state-space form, QML estimation of the SVOL model with correlation can be carried out as usual, using the results in equations (3.15) and (3.16).

Extending Harvey and Shephard's approach to the case of the RSSV model is fairly straightforward. We rewrite equation (3.10) as follows:

$$h_t = \mu_s + \phi h_{t-1} + g_t u^* + \sigma_\eta \eta_t^*. \tag{3.17}$$

Obviously the above equation and equation (3.9) remain a linear state-space model. Therefore we can still apply Kim's filter to estimate the RSSV model with correlation.

The results for this new model are reported in Table 3.3. We notice that the parameter estimates looks very similar to those reported in Table 3.2. If we take a look at the estimated correlation coefficients, only the correlation for the U.S. short rate data is statistically different from zero. For the other three countries, zero-correlation does not seem to be a bad assumption. In fact the regime probability plots for this model also look similar to those plots shown in Figures 3.2 to 3.5. Hence we do not plot the regime probability graphs for this model. Overall, in our sample, we find no evidence of strong correlation in the data. However, in other cases, such as stock returns, modeling correlation might be important, and the approach taken in this section could prove to be useful under those circumstances.

Table 3.3: Parameter Estimates of the RSSV Model with Correlation

This table reports the parameter estimates of the following regime-switching stochastic volatility (RSSV) model with correlation for short term interest rates from four countries:

$$x_t = h_{t-1} + 2\gamma ln(r_{t-1}) - 1.2704 + \xi_t,$$

$$h_t = \mu_s + \phi h_{t-1} + g_t u^* + \sigma_\eta \eta_t^*,$$

where we allow the intercept of the volatility equation to be regime dependent. The lower case s is an unobserved regime variable that follows a first-order, two-state Markov process, where we assume constant transition probabilities. h_t denotes the log variance. x_t is the log of squared residuals. g_t is a variable that takes 1 (-1) if y_t is positive (negative). $u^* = E_+(\eta_t)$, and E_+ denotes the expectation conditional on ϵ_t being positive. η_t^* is defined in equation (3.14). Standard errors are in parentheses.

	US	Canada	Japan	UK
p	0.9985	0.9953	0.9974	0.9947
	(0.0012)	(0.0048)	(0.0031)	(0.0035)
q	0.9977	0.9960	0.9958	0.9917
	(0.0019)	(0.0035)	(0.0038)	(0.0061)
ϕ	0.9524	0.7595	0.6426	0.6333
	(0.0158)	(0.0847)	(0.0630)	(0.1254)
μ_0	-0.2725	-1.4584	-2.1519	-2.1388
	(0.0945)	(0.5536)	(0.3929)	(0.8164)
μ_1	-0.36145	-1.7304	-2.8024	-2.7975
	(0.1222)	(0.6642)	(0.5124)	(1.0467)
σ_η	0.2944	0.6466	1.4290	0.8079
	(0.0541)	(0.1425)	(0.1302)	(0.1588)
γ	0.6287	0.7432	0.8397	0.6508
	(0.1312)	(0.2118)	(0.0721)	(0.1608)
ρ	-0.1683	0.0172	0.1171	0.0452
	(0.0750)	(0.0916)	(0.0776)	(0.0791)
Log-likelihood	-5680.28	-2516.72	-1906.36	-3153.28

3.5 Alternative Model Specifications

As robustness checks of the results reported in this chapter, we consider two alternative model specifications to the RSSV model in this section. The first is a time-varying transition probability model. The second is a model where the autoregressive parameter of the volatility equation is also allowed to be regime dependent.

3.5.1 Time-varying RSSV model

In equation (3.11) the transition probabilities p and q of the RSSV model are constants, which is not very flexible. It is possible to let these transition probabilities be time-varying. Following Diebold et al. (1994), we specify the time-varying transition probabilities as follows:

$$p(s_t = j | s_{t-1} = j; I_{t-1}) = \frac{e^{a_j + b_j r_{t-1}}}{1 + e^{a_j + b_j r_{t-1}}}, \ j = 0, 1. \tag{3.18}$$

Thus, the transition probabilities in this model specification are allowed to vary with the lagged interest rate levels. In fact we can let the transition probabilities be a function of any other exogenous variables as well. Hence this model specification is flexible and encompasses a constant transition probability model.

When we estimate this time-varying transition probability model, it turns out that the results are very similar to the constant transition probability model. For example, in the case of the U.S. short rate data, the parameter estimates for a_0, a_1, b_0, and b_1 are 6.09, 5.74, 0.07 and 0.05 respectively. For a_j's close to 6, the implied transition probability is about 0.9975, which is very close to what we get from the constant transition probability model. In fact a likelihood ratio test can not reject the null of a constant transition probability model and the p-value is as high as 0.96. In addition the regime probabilities are almost identical to those in the constant transition probability model. Hence modeling time-varying probability does not seem to have an additional advantage at least for our sample.

This time-varying transition probability model may be a useful framework to investigate the relation between the volatility regimes and the fundamental economic state variables. We have shown that in the U.S. case the volatility regimes seem to be related to large shocks to the macroeconomy. If this conjecture contains some truth, we could allow the transition probabilities to vary with those underlying state variables of interest. Hence this time-varying model might be useful towards identifying the determinants of volatility regimes.

3.5.2 Regime-Switching ϕ

Next we consider the following specification of the RSSV model where the both the μ and the ϕ parameters are allowed to be regime dependent:

$$h_t = \mu_s + \phi_s h_{t-1} + \sigma_\eta \eta_t. \tag{3.19}$$

Replacing equation (3.10) with the above equation, we can proceed as usual. The results for this model are reported in Table 3.4. There seems to exist little evidence that this model actually outperforms the more parsimonious specification that we estimate earlier. For all four countries under consideration, the two ϕ estimates are not statistically different from each other. In addition all the other parameter estimates look very much the same as in the old specification. The regime probabilities (not reported) are almost identical to the old model. Hence we conclude that the results reported for the basic RSSV model are robust to this new specification.

Table 3.4: Parameter Estimates of the RSSV Model: Robustness Check

This table reports the parameter estimates of the following regime-switching stochastic volatility (RSSV) model for short term interest rates from four countries:

$$x_t = h_{t-1} + 2\gamma ln(r_{t-1}) - 1.2704 + \xi_t,$$

$$h_t = \mu_s + \phi_s h_{t-1} + \sigma_\eta \eta_t,$$

where we allow both the intercept and the autoregressive parameter in the volatility equation to be regime dependent. The lower case s is an unoberved regime variable that follows a first-order, two-state Markov process, where we assume constant transition probabilities. h_t denotes the log variance. x_t is the log of squared residuals from the following OLS regression

$$\Delta r_t = a + b r_{t-1} + \epsilon_t.$$

Standard errors are in parentheses.

	US	Canada	Japan	UK
p	0.9985	0.9953	0.9976	0.9940
	(0.0012)	(0.0048)	(0.0028)	(0.0037)
q	0.9975	0.9960	0.9959	0.9897
	(0.0021)	(0.0035)	(0.0036)	(0.0076)
ϕ_0	0.9581	0.74302	0.5797	0.6253
	(0.0204)	(0.0969)	(0.0815)	(0.1143)
ϕ_1	0.9601	0.7752	0.6313	0.5092
	(0.0195)	(0.0928)	(0.0719)	(0.2764)
μ_0	-0.2106	-1.3866	-2.5369	-2.2157
	(0.1119)	(0.5867)	(0.4970)	(0.7204)
μ_1	-0.2980	-1.882	-2.8945	-3.8143
	(0.1534)	(0.7952)	(0.5805)	(2.1577)
σ_η	0.2752	0.64559	1.4680	0.8426
	(0.0633)	(0.1424)	(0.1266)	(0.1491)
γ	0.4833	0.77493	0.8412	0.6646
	(0.1475)	(0.2419)	(0.0661)	(0.1577)
Log-likelihood	-5682.70	-2516.67	-1907.38	-3153.31

3.6 Concluding Remarks

In this chapter we investigate the possibility of regime shifts in short rate volatility. In particular, we are interested in the empirical question whether or not the apparent volatility persistence in interest rate volatility, as often suggested by the various GARCH and SVOL models, are due to the failure to control for regime shifts in the volatility processes. We look at the empirical evidence using interest rate data from four countries. Parameter estimates from the SVOL model imply highly persistent short rate volatilities for all the four countries under consideration. By contrast, we estimate the RSSV model using the same dataset and find the previously found persistence in volatilities falls dramatically for the U.K., Canada and Japan data. In the case of the U.S. short rate volatility, the volatility is still highly persistent. The evidence presented here highlights the importance of accounting for possible structural breaks in the volatility process.

We also contribute to the literature by showing how to account for correlation in the RSSV model. This may be important for modeling stock market volatilities, where there is well documented evidence of asymmetric volatility. In our dataset, we find the evidence of negative correlation for the U.S. short rate volatility, but essentially zero correlation for other countries.

The volatility regimes identified in this chapter are very distinct and seem to last for a fairly long period of time. We show that, in the case of the United States, these regimes may be influenced by some persistent fundamental economic state variables. Intuitively the volatility regimes may be influenced by the central bank's monetary policy, variations in inflation rates, and various macroeconomic shocks. Future research may be directed towards identifying the relation between these volatility regimes and the underlying state variables.

3.7 Estimation Procedures

In this section, I provide the technical details regarding the estimation of the various models.

3.7.1 State-Space Representations and the Kalman Filter

State-space models are useful tool to deal with dynamic time series models with unobserved variables. Hamilton (1994) gives a detailed discussion on this topic as well as its applications in economics and finance. The Kalman filter is a recursive procedure that can handle the state-space models.

Let's consider the following simple state-space model

$$y_t = \beta h_t + \alpha z_t + \epsilon_t, \tag{3.20}$$

$$h_t = \mu + \phi h_{t-1} + \eta_t, \tag{3.21}$$

where y_t is the observed variable (data), h_t is the unobserved state variable, z_t is an exogenous variable. Equation (3.20) is the observation equation that describes the relation between the observable and unobservable variables. Equation (3.21) is the state equation that describes the dynamics of the state variables. For the SVOL model considered in this chapter, equation (3.9) is the observation equation and equation (3.8) is the state equation.

It is often assumed that

$$\epsilon_t \sim i.i.d.N(0, R),$$

$$\eta_t \sim i.i.d.N(0, Q),$$

$$E(\epsilon_t, \eta_t) = 0.$$

The Kalman filter consists of two basic steps: prediction and updating.

1. Prediction: In this step we form an optimal predictor[6] of y_t based on information up to time $t-1$.

$$h_{t|t-1} = \mu + \phi h_{t-1|t-1}, \tag{3.22}$$

$$P_{t|t-1} = \phi P_{t|t-1} \phi' + Q, \tag{3.23}$$

$$e_{t|t-1} = y_t - y_{t|t-1} = y_t - \beta h_{t|t-1} - \alpha z_t, \tag{3.24}$$

$$f_{t|t-1} = \beta P_{t|t-1} \beta' + R, \tag{3.25}$$

where $h_{t|t-1}$ is the conditional expectation of h_t based on information up to $t-1$, $P_{t|t-1}$ is the conditional covariance of h_t, $e_{t|t-1}$ is the prediction error, and $f_{t|t-1}$ is the conditional variance of the prediction error.

2. Updating: After observing y_t, we make a more accurate inference of h_t.

$$h_{t|t} = h_{t|t-1} + K_t e_{t|t-1}, \tag{3.26}$$

$$P_{t|t} = P_{t|t-1} - K_t \beta P_{t|t-1}, \tag{3.27}$$

where $K_t = P_{t|t-1} \beta' f_{t|t-1}^{-1}$ is the Kalman gain that determines the weight assigned to the new information.

As a by-product of the Kalman filter, we can set up the following log-likelihood function

$$lnL = -\frac{1}{2} \sum_{t-1}^{T} ln(2\pi f_{t|t-1}) - \frac{1}{2} \sum_{t-1}^{T} e'_{t|t-1} f_{t|t-1}^{-1} e_{t|t-1},$$

which is then maximized using some hill-climbing numerical procedures. Although for the SVOL model one of the disturbance is non-Gaussian, the procedure discussed here are still valid and can be used to generate consistent and asymptotically normal estimates.[7]

[6]The Kalman filter is based on linear projection, which is optimal under normality.

[7]This is known as the quasi-maximum likelihood (QML) estimates.

3.7.2 RSSV Model and Kim's Filter

Kim (1994) extends the regime-switching model of Hamilton (1989) to the general case of a state-space model. See also Kim and Nelson (1999).

The RSSV model consists of equations (3.9) and (3.10). Compared with the SVOL model, the only difference is that we now have an unobserved regime variable s_t that is assumed to follow a Markov process as specified in equation (3.11). In the following discussion, we assume only μ is regime dependent. However, the framework can be easily extended to more general specifications. For a state-space model with regime-switching, the goal is to forecast h_t conditional both on the information set at $t-1$ and on the value taken by the regime variable s_{t-1}. Let $h_{t|t-1}^{(i,j)}$ denote $E[h_t|I_{t-1}, s_t = j, s_{t-1} = i]$, and similarly for $P_{t|t-1}^{(i,j)}$, and so on. The Kalman filter now looks as follows:

$$h_{t|t-1}^{(i,j)} = \mu_j + \phi h_{t-1|t-1}^{i}, \tag{3.28}$$

$$P_{t|t-1}^{(i,j)} = \phi P_{t|t-1}^{i} \phi' + Q, \tag{3.29}$$

$$e_{t|t-1}^{(i,j)} = y_t - \beta h_{t|t-1}^{(i,j)} - \alpha z_t, \tag{3.30}$$

$$f_{t|t-1}^{(i,j)} = \beta P_{t|t-1}^{(i,j)} \beta' + R, \tag{3.31}$$

$$h_{t|t}^{(i,j)} = h_{t|t-1}^{(i,j)} + K_t^{(i,j)} e_{t|t-1}^{(i,j)}, \tag{3.32}$$

$$P_{t|t}^{(i,j)} = P_{t|t-1}^{(i,j)} - K_t^{(i,j)} \beta P_{t|t-1}^{(i,j)}, \tag{3.33}$$

The path-dependent nature of the above procedure makes the filtering problem numerically difficult to handle. Kim (1994) suggests that we first approximate $h_{t|t}^{(i,j)}$ in equation (3.32) with $E[h_t|I_t, s_t = j, s_{t-1} = i]$, and then integrate out $s_{t-1} = i$ to get $h_{t|t}^{j}$. Similarly we can replace $P_{t|t}^{(i,j)}$ in equation (3.33) with $E[(h_t - h_{t|t}^{(i,j)})(h_t - h_{t|t}^{(i,j)})'|I_t, s_t = j, s_{t-1} = i]$, and integrate out $s_{t-1} = i$ to get $P_{t|t}^{j}$.

The above approximation is then combined with Hamilton's filter for the regime-switching model to estimate the RSSV model.

Chapter 4

Estimating Continuous-time Interest Rate Models

4.1 Introduction

Theoretical models in finance often model the movements in financial asset prices as a continuous time diffusion process. The celebrated Black-Scholes model, for example, characterizes the stock price movements with the following stochastic differential equation (SDE):

$$dS = \mu S dt + \sigma S dW, \tag{4.1}$$

where W is the standard Brownian motion. Albeit its success, a potential weakness of this modeling approach is that discontinuities in the underlying asset price movements is not captured. Merton (1978) first puts forward a model that allows for jumps. A general jump diffusion model may be expressed as the following stochastic integral equation:

$$X(t) = X(t_0) + \int_{t_0}^{t} \mu(X, \tau; \theta) d\tau + \int_{t_0}^{t} \sigma(X, \tau; \theta) dW(\tau) + \int_{t_0}^{t} J(X, \tau; \theta) dN_\lambda(\tau), \tag{4.2}$$

where $W(t)$ is a Wiener process and $N_\lambda(t)$ a Poisson counter. μ, σ, and J are known functions of X and t. θ is a parameter vector to be estimated.

The above equation can be written in the more familiar form of a SDE:

$$dX(t) = \mu(X, t; \theta)dt + \sigma(X, t; \theta)dW(t) + J(X, t; \theta)dN_\lambda(t). \tag{4.3}$$

Continuous-time models such as equations (1) and (3) are convenient tools for developing theoretical models and deriving closed-form solutions when the purpose is to price derivative securities. However, from an econometrician's point of view, continuous-time models turn out to be extremely difficult to estimate using discretely observed data since the likelihood functions for the continuous-time models in most cases are not explicitly computable. To see this, recall that the maximum likelihood estimation of the true parameter vector θ is based on the likelihood function $logL = \sum_{t=1}^{T} logf(y_{t+1}|y_t, \theta)$, where $f(X_{t+1}|X_t, \theta)$ are the Markovian transition density functions. If analytical solutions of the SDE's are available, then $f(Y_{t+1}|Y_t, \theta)$ can be solved in closed-forms. However, the challenge is that analytical solutions of the general SDEs are rarely available.

One popular estimation technique is to discretize the SDE's, typically using the Euler scheme. However, as shown in Lo (1988), this method only works if the time interval converges to zero. In practice, the time interval is usually kept fixed. Hence the Euler as well as other discretization methods may lead to inconsistent parameter estimates. The extant estimation techniques are all numerically intensive. To reduce the computational costs, we will focus on estimating the one-factor continuous-time interest rate models in this chapter.

The approach adopted in this chapter is based on a theoretical result given by Lo (1988), who shows that we can construct likelihood functions of the jump-diffusion models by solving the Kolmogrov forward equations corresponding to the conditional densities. Henceforth we refer to this method as the partial differential equation (PDE) approach. Lo does not further pursue the idea, only suggesting that the forward equations may be solved using standard methods such as Fourier Transform. To my knowledge, so far there are only two attempts to numerically implement Lo's method in the special case of a pure diffusion model. Mella-Barral and Perraudin (1994) suggest that we solve the Kolmogrov equations using explicit finite difference schemes. Poulsen (1999) refines the PDE approach by using the numerically more stable Crank-Nicolson method. Both articles estimate the continuous-time version of the Chan, Karolyi, Longstaff and Sanders (1992) (CKLS) interest rate model.

In this chapter we make two contributions to the literature. First, we show how to extend the PDE approach to the general case of a jump diffusion model. Second, we compare the results obtained by the Euler-discretization method versus the more numerically intensive PDE approach for various specifications of the single-factor interest rate models. Interestingly, We find the results are not strikingly different.

This chapter is organized as follows. Section 4.2 briefly reviews several methods in the existing literature to estimate SDE's. Section 4.3 shows how to implement Lo's method to estimate jump-diffusion models. Section 4.4 applies the PDE approach to estimate various interest rate models and compares the result with the widely used discretization-based method. Section 4.5 concludes the chapter.

4.2 Existing Methods

In this section, I briefly summarize the various econometric methods used in the literature to estimate SDE's.

4.2.1 Maximum Likelihood Estimation with Discretization

The Euler-Maruyama discretization scheme is probably the most popular method to estimate SDE's, among practitioners as well as academic researchers. For example, if $J(X,\tau;\theta) = 0$ in equation (4.2), i.e. a pure diffusion SDE, the Euler approximation of the SDE is:

$$X_{t-1} = X_t + \mu(X,\tau;\theta)\Delta_t + \sigma(X,\tau;\theta)(W_{t+1} - W_t),$$

where Δ_t is the time increment. Since $W_{t+1} - W_t$ follows the normal distribution, the maximum likelihood approach can be applied to estimate the model. The Euler-Maruyama discretization scheme weakly converges to the corresponding continuous-time process with order 1.[1] Although higher-order approximation schemes may be considered, in practice these methods are rarely adopted.

Lo (1988) demonstrates that, if what is of interest is the true continuous-time model parameters, then for discretely sampled data, maximum likelihood estimation based on fixed time interval discretization could lead to inconsistent estimates of the true parameters.

4.2.2 Generalized Method of Moment

Chan, Karolyi, Longstaff and Sanders (1992) provide an application of the Generalized Method of Moment (GMM) of Hansen (1982) to the estimation of the following continous-time model of the interest rate:

$$dr = (\alpha + \beta r)dt + \sigma r^{\gamma} dW, \tag{4.4}$$

where r is the risk-free interest rate, and W is a Brownian motion. CKLS discretize equation (4.3) using the Euler approximation scheme. Then they use the conditional moment restriction $E(f_t(\theta)|\Omega_t) = 0$, where $f_t(\theta) = [\epsilon_{t+1},\ \epsilon_{t+1} r_t,\ \epsilon_{t+1}^2 - \sigma^2 r_t^{2\gamma},\ (\epsilon_{t+1}^2 - \sigma^2 r_t^{2\gamma}) r_t]'$, to set up the GMM estimation.

Conditional moments based on discretization are crude unless the discretization error disappears. Hence the same criticism also applies here. To address this problem, Conley et al. (1997) suggest using the infinitesimal generator for a candidate Markov process to build moment conditions.

[1] A standard reference on this topic is Kloeden and Platen (1992).

4.2.3 Nonparametric Methods

Aït-Sahalia (1996) uses nonparametric techniques to estimate one-factor interest rate diffusion models. The idea is to compare the density implied by a parametric model and a nonparametric estimator valid even if the parametric model is misspecified.[2] Hence the test statistic has a minimum distance flavor.

Pritsker (1998) questions the specification test developed in Aït-Sahalia (1996). The argument is that interest rates are known to be highly correlated whereas the nonparametric technique used in Aït-Sahalia's paper is very sensitive to the dependence in the data. Chapman and Pearson (2000) conduct a Monte Carlo study. Their result cast further doubts on the robustness of Aït-Sahalia's approach.

4.2.4 Efficient Method of Moments

The efficient method of moments (EMM) of Gallant and Tauchen (1996)) is another popular way of estimating SDE's. EMM uses the expectation of the score vector from the auxiliary model as the moment conditions for the GMM estimator. As long as the auxiliary model closely approximates the actual distribution of the data, even if it does not nest the structural model, the EMM estimator is nearly fully efficient. See Gallant and Long (1997). If the auxiliary model does encompass the structural model, then the EMM estimator is as efficient as the maximum likelihood method.

In the context of estimating the SDE's, the structural model is the continuous time models of interest, and the auxiliary model is usually the semi-nonparametric (SNP) model of Gallant and Tauchen (1992). However, Duffee and Stanton (2000) find out that using the Kalman Filter as the auxiliary model may be a better choice than the SNP for interest rate data.

[2] See also the articles by Stanton (1997) and Ahn and Gao (1999)

4.2.5 Closed-Form Density Approximation

Aït-Sahalia (2002) proposes to construct a sequence of Hermite approximation to the true transition function for continous-time diffusions. The difficulty with regard to applying this method is that two transformations of the original SDE's are required, and the resulting likelihood function looks complicated even for low order polynomial approximation and relatively simple models. In addition, it is unclear how to apply the method to jump-diffusion models.

4.3 The PDE Approach to Estimation of Continuous-time Models

This section describes the PDE approach to estimating the SDE's.

4.3.1 Kolmogrov Forward Equation

Given $n+1$ discrete observations of the process $X(t)$ sampled at time t_0, $t1$, $\ldots$, t_n and the stochastic specification of $X(t)$ in equation (4.2), the joint density p is the likelihood function of X when viewed as a function of θ. Since $X(t)$ is a Markov process, we can write p as a product of conditional densities:

$$p(X) = p_0(X_0) \prod_{k=1}^{n} p_k(X_k, t_k | X_{k-1}, t_{k-1}). \tag{4.5}$$

Hence all likelihood approaches to estimating SDE essentially boil down to how to approximate the conditional densities.

It is known in the literature that under some regularity conditions, the conditional density p_k can be characterized by the corresponding Kolmogrov (or Fokker-Planck) forward equation:

$$\frac{\partial}{\partial t}(p_k) = -\frac{\partial}{\partial X}(\mu p_k) + \frac{1}{2}\frac{\partial^2}{\partial X^2}(\sigma^2 p_k) - \lambda p_k + \lambda \tilde{p_k} \left| \frac{\partial}{\partial X}(\tilde{J^{-1}}) \right| \tag{4.6}$$

subject to

$$p_k(X, t_{k-1}) = \delta(X - X_{k-1}), \tag{4.7}$$

and any other relevant boundary coditions. $\tilde{J}$ is defined as $X + J$, $\tilde{p_k} \equiv p_k(\tilde{J}^{-1}, t)$, and $\tilde{J}^{-1}$ is the inverse function of J satisfying $X = \tilde{J^{-1}}(\tilde{J})$. $\delta(X - X_{k-1})$ is the Dirac-δ generalized function centered at X_{k-1}. A formal proof of this result can be found in Lo (1988).

Lo argues that one can solve the forward equation via standard methods such as Fourier transform. In the special case of a pure diffusion model (when $J = 0$ in equation (4.2)), both Mella-Barral and Perraudin (1994) and Poulsen (1999) suggest that we numerically solve the PDE. Poulsen's method is better since he uses the Crank-Nicolson method in stead of the explicit finite difference method, which is numerically less stable. Later we show how to extend the framework to the general case of jump diffusions.

4.3.2 Estimating Jump Diffusion Models with the PDE Approach

In what follows, we discuss the steps involved in constructing the likelihood functions. Using subscripts to stand for partial derivatives, we can rewrite the foward equation (4.6) as follows:

$$p_t(t, y) = a(y)p + b(y)p_y + c(y)p_{yy} + d(y)\tilde{p} \tag{4.8}$$

where

$$a(y) = \frac{1}{2}(\sigma^2)_{yy} - \mu_y - \lambda,$$

$$b(y) = (\sigma^2)_y - \mu,$$

$$c(y) = \frac{1}{2}\sigma^2,$$

$$d(y) = \lambda|\tilde{J}_y^{-1}|.$$

Note that if we take away λ and $d(y)$, the above equation reduces to the case of a pure diffusion model. Since our purpose is to numerically solve the forward equation, the next step is to use difference operators to replace differential operators.

Finite difference methods are probably the most popular approch to numerically solve the partial differential equations that frequently arise in Finance. Following Poulsen (1999) we choose the Crank-Nicolson method, which is superior to the alternative explicit and implicit schemes. It can be shown that Crank-Nicolson is both numerically stable and converges faster than the explicit and implicit methods.[3]

Let's consider a grid in time and space dimensions with time step k and space step h. We can rewrite the differential equation as a difference equation:

$$\frac{u_m^{n+1} - u_m^n}{k} = a\Delta_0(h) + b\Delta_1(h) + c\Delta_2(h) + d\Delta_0(h+j) \tag{4.9}$$

where $u_m^n \approx p(nk, X_0 + mh)$ is the grid point that approximates the true conditional density value at time step n and space step m, and a, b, c, and d are given by in equation (4.8).

For the Crank-Nicolson method, Δ_i, the ith order difference operators are defined as follows:

$$\Delta_0(h) = \frac{1}{2}u_m^{n+1} + \frac{1}{2}u_m^n, \tag{4.10}$$

$$\Delta_1(h) = \frac{1}{2}\frac{u_{m+1}^{n+1} - u_{m-1}^{n+1}}{2h} + \frac{1}{2}\frac{u_{m+1}^n - u_{m-1}^n}{2h}, \tag{4.11}$$

$$\Delta_2(h) = \frac{1}{2}\frac{u_{m+1}^{n+1} - 2u_m^{n+1} + u_{m-1}^{n+1}}{h^2} + \frac{1}{2}\frac{u_{m+1}^n - 2u_m^n + u_{m-1}^n}{h^2}, \tag{4.12}$$

$$\Delta_0(h+j) = \frac{1}{2}u_{m+j}^{n+1} + \frac{1}{2}u_{m+j}^n. \tag{4.13}$$

Plugging these terms into equation (4.9) and rearrange, we have

$$\alpha_1 u_{m-1}^{n+1} + \alpha_2 u_m^{n+1} + \alpha_3 u_{m+1}^{n+1} + \alpha_4 u_{m+j}^{n+1} = -\alpha_1 u_{m-1}^n + (2-\alpha_2)u_m^n - \alpha_3 u_{m+1}^n - \alpha_4 u_{m+j}^n. \tag{4.14}$$

The coefficients in equation (4.14) are defined as follows:

$$\alpha_1 \equiv \frac{bhq}{4} - \frac{cq}{2},$$

[3] For a standard reference on numerical solutions for partial differential equations, see Ames (1977).

$$\alpha_2 \equiv 1 - \frac{ak}{2} + cq,$$
$$\alpha_3 \equiv -\frac{bhq}{4} - \frac{cq}{2},$$
$$\alpha_4 \equiv -\frac{1}{2}dk,$$

where $q = k/h^2$.

We should pay some attention to the presence of the grid points u_{m+j}^{n+1} and u_{m+j}^{n} in equation (4.14). When there are no jumps, solving this difference equation with the Crank-Nicolson method involves only six neighboring grid points. However, when jumps are allowed, we need to evaluate two additional grid points. Depending on the size of the jump, the two additional grid points may not be adjacent to original six points.

To evaluate the grid points, we note that it can be written as a system of linear equations:

$$\mathbf{A}\mathbf{u}^{k+1} = \mathbf{B}\mathbf{u}^{k}, \tag{4.15}$$

where $\mathbf{u}^k$ is the vector of conditional density values at time step k. The coefficient matrices $\mathbf{A}$ and $\mathbf{B}$ have some interesting properties. In the pure diffusion case, $\mathbf{A}$ and $\mathbf{B}$ are two tridiagnal matrices:

$$\mathbf{A} = \begin{pmatrix} \alpha_2 & \alpha_1 & 0 & \dots & 0 \\ \alpha_3 & \alpha_2 & \alpha_1 & & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & \alpha_1 \\ 0 & 0 & & \alpha_3 & \alpha_2 \end{pmatrix}$$

$$
\mathbf{B} = \begin{pmatrix}
2-\alpha_2 & -\alpha_1 & 0 & \dots & 0 \\
-\alpha_3 & 2-\alpha_2 & -\alpha_1 & & \vdots \\
0 & \ddots & \ddots & \ddots & 0 \\
\vdots & & \ddots & \ddots & -\alpha_1 \\
0 & 0 & & -\alpha_3 & 2-\alpha_2
\end{pmatrix}
$$

Inverse of the tridiagnal matrix **A** can be found efficiently using various numerical techniques.[4]

When there are jumps, however, the matrices are no longer tridiagnal, not even symmetric.

$$
\mathbf{A} = \begin{pmatrix}
\alpha_2 & \alpha_1 & 0 & \dots & \alpha_4 & 0 & \dots & 0 \\
\alpha_3 & \alpha_2 & \alpha_1 & \ddots & 0 & \ddots & & \vdots \\
0 & \ddots & \ddots & \ddots & & & \ddots & \vdots \\
0 & & \ddots & \ddots & \ddots & & & \alpha_4 \\
\vdots & & & \ddots & \ddots & \ddots & & 0 \\
\vdots & & & & \ddots & \ddots & \ddots & \vdots \\
\vdots & & & & & \ddots & \ddots & \alpha_1 \\
0 & \dots & & & & & \alpha_3 & \alpha_2
\end{pmatrix}
$$

$$
\mathbf{B} = \begin{pmatrix}
2-\alpha_2 & -\alpha_1 & 0 & \dots & -\alpha_4 & 0 & \dots & 0 \\
-\alpha_3 & 2-\alpha_2 & -\alpha_1 & \ddots & 0 & \ddots & & \vdots \\
0 & \ddots & \ddots & \ddots & & & \ddots & \vdots \\
\vdots & & \ddots & \ddots & \ddots & & & -\alpha_4 \\
\vdots & & & \ddots & \ddots & \ddots & & 0 \\
\vdots & & & & \ddots & \ddots & \ddots & \vdots \\
\vdots & & & & & \ddots & \ddots & -\alpha_1 \\
0 & \dots & & & & & -\alpha_3 & 2-\alpha_2
\end{pmatrix}
$$

[4]See, for example, the numerical routines in Press et al. (1992).

When solving for the difference equation, we assume the boundaries are inaccessible, i.e. the conditional densities at both the upper bound and the lower bound are set to zero. For the initial condition we use the following:

$$u_m^1 = N(X_0 + \mu k, \sigma^2 k), \tag{4.16}$$

where $N(.)$ refers to the Normal density function with mean $x + \mu k$ and variance $\sigma^2 k$. Note that this converges to the Dirac-δ function when k approaches zero, which is the true initial condition.

Thus given two observations X_k and X_{k+1}, we numerically solve the forward equation with the above-mentioned procedure to get an approximation of the true conditional density at time $k+1$. If we take a log transformation of the density values and add them up for all observations except the first one, we obtain the desired log likelihood function, which is then numerically maximized. Hence maximum likelihood estimates of the true model parameters can be found as usual. Poulsen (1999) shows that the parameter estimates obtained with this method retains all the desirable properties of the maximum likelihood estimates. Namely we have consistency, asymptotically normality, and efficiency (if the approximation error is ignorably small). More importantly, with this method the econometrician can control the precision parameters such as k and h (at the cost of CPU time).

4.3.3 Approximating the Conditional Density of the Vasicek Model

The sucess of the PDE method depends on how big the approximation errors are when we use the difference operators to replace the differential operators. In order to evaluate the precision at which we can practically approximate the true conditional densities. We do the following numerical experiment.

Although the exact solution to a general SDE is unknown, there exist several special cases in which we do know the exact solution. One of these exceptions is the Ornstein-Uhlenbeck process. In term structure modeling, this is known as the Vasicek (1977) model

$$dr = \kappa(\theta - r)dt + \sigma dW,$$

where θ is the mean of interest rate levels, σ is the interest rate volatility, and θ is the so-called mean-reversion speed parameter. The associated conditional density for the Vasicek model is given by the following normal density[5]

$$p(\tau, x, y) = N(y; e^{-\kappa\tau}x + \theta(1 - e^{-\kappa\tau}), \frac{\sigma^2(1 - e^{-2\kappa\tau})}{2\kappa}).$$

In Figure 4.1, we plot both the exact conditional density and the approximated condition density for the Vasicek model using the following parameters: $\kappa = 0.15$, $\theta = 0.08$, $\sigma = 0.03$, and $x = 0.08$. We choose $\tau = 1/12$, which corresponds to the monthly data frequency. The conditional densities are approximated with $h = 1/200$ and $k = 1/50$. On a 733 MHZ PC, the computation takes 2.14 seconds. The density plot suggests that the approximated density matches the true density pretty well. In terms of log density the difference is even smaller.

To stress-test our method, we also look at the case where $x = 0.0267$, which is the minimum interest rate level for the U.S. monthly risk-free rate data described in chapter 2. We use the same model parameter and the density is plotted in Figure 4.2. It looks like the approximation gives slight higher density value for interest rate levels below the mean than for those levels above the mean. This is probably due to fact that we only model positive interest rates and hence force conditional probabilities below the lower bound (0.0001 in this case) to zero, whereas in the Vasicek model negative interest rates may still have positive conditional probabilities. Overall the numerical experiments shown here indicate that the PDE approach does give reasonably good estimates of the conditional densities.

[5] For example, see Karlin and Taylor (1981).

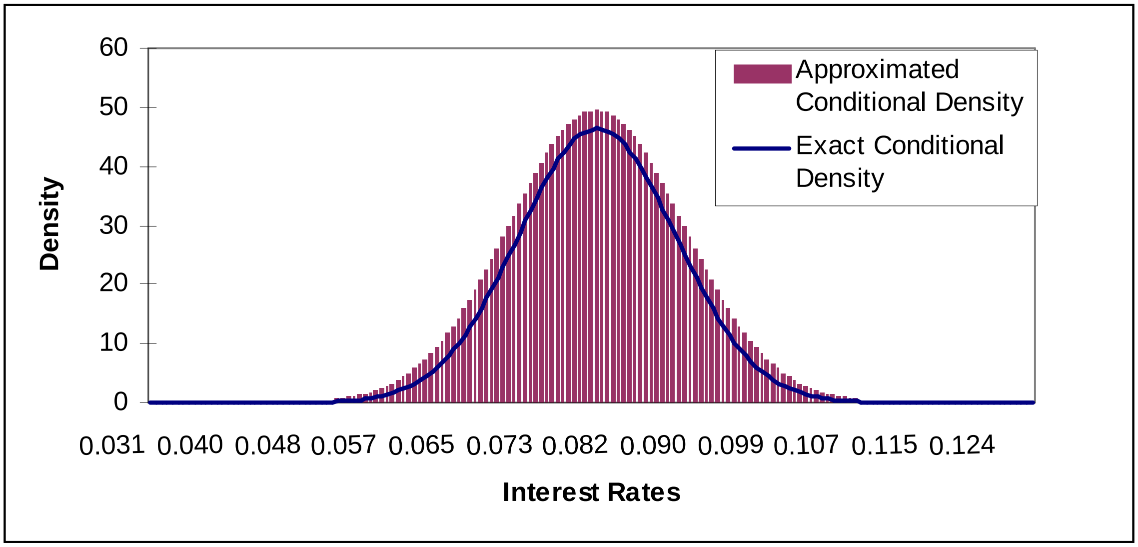

Figure 4.1: Approximated and Exact Conditional Densities for the Vasicek Model when $x = 0.08$

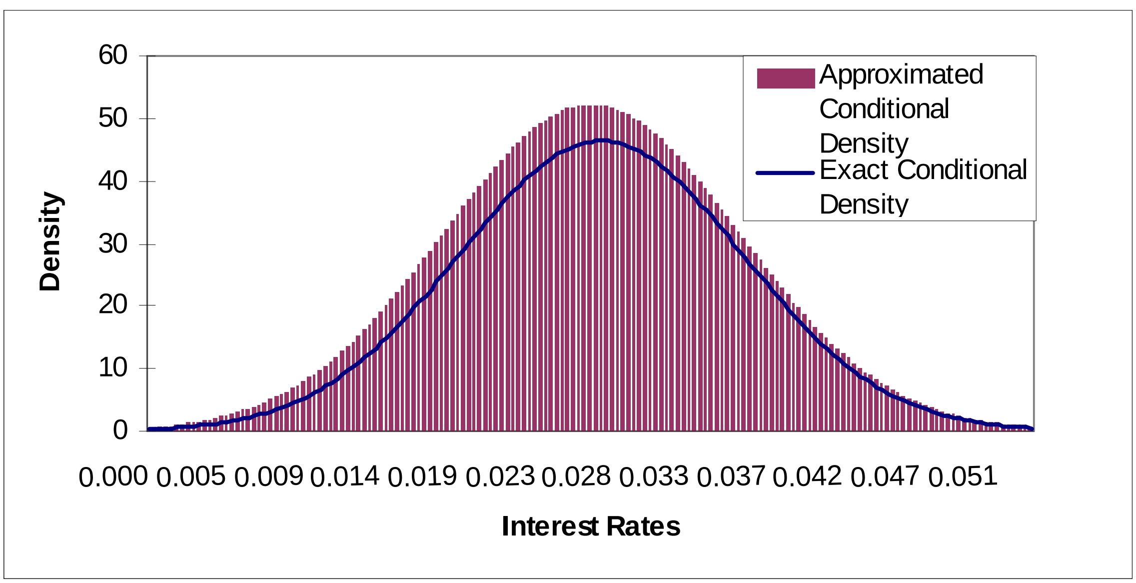

Figure 4.2: Approximated and Exact Conditional Densities for the Vasicek Model when $x = 0.0267$

4.4 Estimating Continuous-time One-Factor Interest Rate Models

In this section we use the PDE approach developed in previous sections to estimate a general one-factor interest rate model.

$$dr = (\alpha_0 + \alpha_1 r + \alpha_2 r^2 + \alpha_3/r)dt + \sigma r^\gamma dW. \tag{4.17}$$

This is the CKLS model plus a nonlinear drift function. In chapter 2 we have estimated the same model using discretized maximum likelihood method. This general model also includes several sub-models. When $\alpha_2 = \alpha_3 = \gamma = 0$, it becomes the Vasicek model. When $\alpha_2 = \alpha_3 = 0$ and $\gamma = 0.5$, it is the CIR model. CKLS model is the case when $\alpha_2 = \alpha_3 = 0$.

We use the weekly U.S. observations to estimate the four models. The summary statistics and the plot of the data are in chapter 2. The results are reported in Table 4.1. Interestingly, we find the parameter estimates using the more computationally intensive PDE approach look very similar to the old results (Table 2.3). The likelihood ratio tests strongly reject the Vasicek and the CIR models. However, the nonlinear drift CKLS model does not seem to outperform the linear drift CKLS model.

Table 4.1: Parameter Estimates of One-Factor Interest Rate Models

This table reports the parameter estimates of the one-factor continuous time interest rate models using the PDE approach. The nonlinear-CKLS model is specified as follows:

$$dr = (\alpha_0 + \alpha_1 r + \alpha_2 r^2 + \alpha_3/r)dt + \sigma r^{\gamma} dW.$$

When $\alpha_2 = \alpha_3 = \gamma = 0$, it becomes the Vasicek model. When $\alpha_2 = \alpha_3 = 0$ and $\gamma = 0.5$, it is the CIR model. CKLS model is the case when $\alpha_2 = \alpha_3 = 0$. LRT refers to the likelihood ratio test. Standard errors are in parentheses.

	Vasicek	CIR	CKLS	Nonlinear-CKLS
α_0	0.0182	0.0174	0.0147	-0.0317
	(0.0095)	(0.0082)	(0.0051)	(0.0254)
α_1	-0.0028	-0.0023	-0.0017	0.0089
	(0.0012)	(0.0009)	(0.0008)	(0.0071)
α_2				-0.0007
				(0.0006)
α_3				0.0419
				(0.0192)
γ	0	0.5	0.7506	0.7491
			(0.0216)	(0.0211)
σ	0.0021	0.0013	0.0024	0.0025
	(0.0001)	(0.0001)	(0.0002)	(0.0002)
LRT (p-value)	0.000	0.000	0.719	

4.5 Conclusion

In this chapter, we propose to estimate the continuous-time interest rate model using the PDE approach.

Although the PDE approach has been considered in the literature before, all previous research focuses on the special case of pure diffusions. We show how to extend the framework to incorporate jumps.

Our empirical results confirm our early finding that modeling nonlinear drift function for interest rate are relatively unimportant. Surprisingly, the parameter estimates using the discretized method and the PDE approach generate very similar results.

There are two possible explanations. First, continuous-time models are theoretical abstraction of the real world situation. They have proven to be useful tools for theoretical construction, but awkward to implement empirically. We could argue that the true data generating process (DGP) may in fact be a discrete-time one. Alternatively, the true DGP might be a continuous-time process, but if we use high-frequency data, the discretization errors may be too small to have any real impact on the results.

Overall the results obtained here suggest that at least for real world applications, where estimation time is important, using discretization-based methods may be the appropriate choice.

Chapter 5

Concluding Remarks

The results obtained in this study are interesting and may have important implications for future academic research as well as real world applications. I briefly summarize the major empirical findings as follows.

First, we compare various model specifications for the short-term riskless interest rates. It is found that modeling stochastic volatility is far more important than modeling nonlinear drift. In fact we can reject all models with constant volatility regardless of their drift specification. For models within the same class (with or without stochastic volatility) linear drift models perform as well as non-linear drift models.

Second, the empirical evidence found here suggests that the CKLS puzzle not be considered a robust empirical fact. In fact, the CKLS puzzle is sensitive to the choices of both model specifications and datasets.

Third, we show that interest rate processes exhibit very distinct national patterns. As an example we find the level effect is evident in the U.S. data but not in the Canadian data. Hence the success of a specific model in a particular country does not necessarily guarantee its success in another. To minimize modeling risk, financial institutions need to use different models to characterize the unique features of the short rate processes for different countries.

Fourth, we find strong evidence of regime shifts in short rate volatility for four countries under consideration. The volatility regimes are very persistent and do not seem to be related to the levels of short rate.

Last, but not least, the empirical results obtained from the discrete-time models look very similar to those from the continuous-time diffusion models. This finding is consistent with the results obtained by Engle and Lee (1996), Duffee and Stanton (2001), among others. For example, Engle and Lee (1996) conclude that estimates of discrete-time stochastic volatility models and continuous time diffusions are typically not very different. In our view, while discretization based estimation methods are unsatisfactory from a theoretical perspective, the discretization errors may be small for the data frequency commonly used in financial studies. In contrast, the existing estimation methods for continuous-time model parameters are undesirable both because they are numerically very demanding and because the small sample properties of these estimators may be quite poor. As an example, Duffee and Stanton note that the approximate Kalman filter method outperforms the EMM approach despite the latter's good asymptotitic properties. From a practical viewpoint, for market participants who need model parameter estimates in a timely fashion, discretization-based model combined with some high frequency data may be the best choice.

A partial list of things-to-do is as follows:

1. Generalize and implement of the PDE approach to estimate jump diffusion models using equity and exchange rate data.

2. Identify the underlying state variables that influence the volatility regimes.

3. Study how regime shifts in volatility affect the pricing of interest rate derivatives.

4. Apply the regime-switching stochastic volatility model with correlation to equity and exchange rate data.

Bibliography

[1] M. Abramovitz and N. Stegun. *Handbook of Mathematical Functions*. Dover Publications, New York, NY, 1970.

[2] D.-H. Ahn and B. Gao. A parametric nonlinear model of term structure dynamics. *Review of Financial Studies*, 12(4):721–762, 1999.

[3] Y. Aït-Sahalia. Testing continuous-time models of the spot interest rate. *Review of Financial Studies*, 9(2):385–426, 1996.

[4] Y. Aït-Sahalia. Maximum-likelihood estimation of discretely-sampled diffusions: A closed-form approximation approach. *Econometrica*, 70(1):223–262, 2002.

[5] W. F. Ames. *Numerical Methods for Partial Differential Equations*. Academic Press, New York, NY, second edition, 1977.

[6] T. G. Andersen and J. Lund. Estimating continuous time stochastic volatility models of the short-term interest rate. *Journal of Econometrics*, 77:343–377, 1997.

[7] D. Backus, S. Foresi, and S. Zin. Arbitrage opportunities in arbitrage-free models of bond pricing. *Journal of Business and Economic Statistics*, 16:13–26, January 1998.

[8] Ball and Torous. The stochastic volatility of short-term interest rates: Some international evidence. *Journal of Finance*, 54:2339–2359, 1999.

[9] F. Black. Studies of stock price volatility changes. In *Proceedings of the Business and Economic Statistics*, pages 177–181. American Statistical Association, 1976.

[10] T. Bollerslev, R. F. Engle, and D. B. Nelson. *ARCH Models*, volume 4 of *Handbook of Econometrics*, chapter 49. Elsevier, 1994.

[11] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31:307–327, 1986.

[12] R. Brenner, R. Harjes, and K. Kroner. Another look at models of the short-term interest rate. *Journal of Financial and Quantitative Analysis*, 31:85–107, 1996.

[13] J. Cai. A markov model of switching-regime arch. *Journal of Business and Economic Statistics*, 12(3):309–316, July 1994.

[14] A. P. Carverhill. A simplified exposition of the heath-jarrow-morton model. *Stochastics and Stochastic Reports*, 53:227–240, 1995.

[15] K. Chan, G. Karolyi, F. Longstaff, and A. Sanders. An empirical comparison of alternative models of the short-term interest rate. *Journal of Finance*, 47:1209–1227, 1992.

[16] D. A. Chapman and N. D. Pearson. Is the short rate drift actually nonlinear. *Journal of Finance*, 55(1):355–388, 2000.

[17] D. A. Chapman and N. D. Pearson. Recent advances in estimating term-structure models. *Financial Analysts Journal*, 57(4):77–95, 2001.

[18] T. Conley, L. Hansen, E. Luttmer, and J. Scheinkman. Short-term interest rate as subordinated diffusions. *Review of Financial Studies*, 10:525–577, 1997.

[19] J. C. Cox, J. Jonathan E. Ingersoll, and S. A. Ross. A theory of the term structure of interest rates. *Econometrica*, 53:385–406, 1985.

[20] F. X. Diebold, J. H. Lee, and G. C. Weinbach. Regime switching with time-varying transition probabilities. In C. Hargreaves, editor, *Time Series Analysis and Cointegration*. Oxford University Press, 1994.

[21] J.-C. Duan. A unified theory of option pricing under stochastic volatility - from GARCH to diffusion. Working Paper, Hong Kong University of Science and Technology, October 1997.

[22] G. Duffee and R. Stanton. Estimation of dynamic term structure models. Working Paper, University of California - Berkeley, May 2001.

[23] G. Duffee. On the relation between the level and volatility of short-term interest rates: A comment on chan, karolyi, longstaff and sanders. Working Paper, Federal Reserve Board Washington D.C., 1993.

[24] D. Duffie. *Dynamic Asset Pricing Theory*. Princeton University Press, Princeton, New Jersey, second edition, 1996.

[25] P. H. Dybvig. Bond and bond option pricing based on the current term. In M. A. H. Dempster and S. R. Pliska, editors, *Mathematics of Derivative Securities*. Cambridge University Press, U.K., 1997.

[26] R. Engle and G. Lee. Estimating diffusion models of stochastic volatility. In P. Rossi, editor, *Modeling Stock Market Volatility: Bridging the Gap in Continuous Time*. Academic Press, San Diego, 1996.

[27] R. Engle. Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica*, 50:987–1008, 1982.

[28] H. C. Fong and O. A. Vasicek. Interest rate volatility as a stochastic factor. Working Paper, Gifford Fong Associates, 1991.

[29] M. Fridman and L. Harris. A maximum likelihood approach for non-gaussian stochastic volatility models. *Journal of Business and Economic Statistics*, 16(3):284–291, July 1998.

[30] A. R. Gallant and J. R. Long. Estimation stochastic differential equations efficiently by minimum chi-square. *Biometrika*, 84:125–141, 1997.

[31] A. R. Gallant and G. Tauchen. A nonparametric approach to nonlinear time series analysis: Estimation and simulation. In D. Brillinger, P. Caines, J. Geweke, E. Parzen, M. rosenblatt, and M. Taqqu, editors, *New Directions in Time Series Analysis*, pages 71–92, New York, NY, 1992. Springer-Verlag.

[32] A. R. Gallant and G. Tauchen. Which moment to match? *Econometric Theory*, 12:657–681, 1996.

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

[34] J. D. Hamilton and R. Susmel. Autoregressive conditional heteroscedasticity and changes in regime. *Journal of Econometrics*, 64:307–333, 1994.

[35] J. D. Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57:357–384, 1989.

[36] J. D. Hamilton. *State-Space Models*, volume IV of *Handbook of Econometrics*, chapter 50, pages 3039–3079. Elsevier Science B. V., 1994.

[37] B. E. Hansen. The likelihood ration test under nonstandard conditions: Testing the markov switching model of GNP. *Journal of Applied Econometrics*, 7:S61–S82, 1992.

[38] B. E. Hansen. Erratum: The likelihood ration test under nonstandard conditions: Testing the markov switching model of GNP. *Journal of Applied Econometrics*, 11:195–198, 1996.

[39] L. P. Hansen. Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4):1029–1054, July 1982.

[40] A. Harvey, E. Ruiz, and N. Shephard. Multivariate stochastic variance models. *Review of Economic Studies*, 61:247–264, 1994.

[41] A. Harvey and N. Shephard. Estimation of an asymmetric stochastic volatility model for asset returns. *Journal of Businesss and Economic Statistics*, 14:429–434, October 1996.

[42] D. Heath, R. Jarrow, and A. Morton. Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation. *Econometrica*, 60:77–105, 1992.

[43] T. Ho and S. Lee. Term structure movements and pricing interest rate contingent claims. *Journal of Finance*, 41:1011–1029, 1986.

[44] J. Hull and A. White. The pricing of options on assets with stochastic volatilities. *Journal of Finance*, 42:281–300, 1987.

[45] J. Hull and A. White. Pricing interest rates derivative securities. *Review of Financial Studies*, 3:573–592, 1990.

[46] C. S. Jones. Nonlinear mean reversion in the short-term interest rate. Working Paper, University of Rochester, August 2000.

[47] S. Karlin and H. M. Taylor. *A Second Course in Stochastic Processes*. Academic Press, New York, NY, 1981.

[48] C. J. Kim and C. R. Nelson. *State-Space Models with Regime Switching*. The MIT Press, Cambridge, Massachusetts, 1999.

[49] C. J. Kim. Dynamic linear models with markov-switching. *Journal of Econometrics*, 60:1–22, 1994.

[50] G. Kitagawa. Non-gaussian state-space modeling of nonstationary time series. *Journal of the American Statistical Association*, 82(400):1032–1063, December 1987.

[51] P. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Springer-Verlag, 1992.

[52] C. G. Lamoureux and W. D. Lastrapes. Persistence in variance, structural change and the GARCH model. *Journal of Business and Economic Statistics*, 8:225–234, 1990.

[53] R. Litterman and J. Scheinkman. Common factors affecting bond returns. *Journal of Fixed Income*, 3(3):54–61, December 1991.

[54] F. Longstaff and E. Schwartz. Interest rate volatility and the term structure: A two-factor general equilibrium model. *Journal of Finance*, 47:1259–1282, 1992.

[55] A. W. Lo. Maximum likelihood estimation of generalized Itô processes with discretely sampled data. *Econometric Theory*, 4:231–247, 1988.

[56] P. Mella-Barral and W. Perraudin. New methods for estimating nonlinear continuous time interest rate processes. Working Paper, University of Cambridge, 1994.

[57] D. B. Nelson and D. P. Foster. Asymptotic filtering theory for univariate ARCH models. *Econometrica*, 62(1):1–41, 1994.

[58] D. Nelson. ARCH models as diffusion approximations. *Journal of Econometrics*, 45:7–38, 1990.

[59] R. Poulsen. Approximate maximum likelihood estimation of discretely observed diffusion processes. Working Paper, University of Aarhus, 1999.

[60] W. H. Press, S. A. Teukolsky, W. T. Vettering, and B. P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, Cambridge, U.K., second edition, 1992.

[61] M. Pritsker. Nonparametric density estimation and tests of continuous time interest rate models. *Review of Financial Studies*, 11:449–487, 1998.

[62] E. Ruiz. Quasi-maximum likelihood estimation of stochastic volatility models. *Journal of Econometrics*, 63:289–306, 1994.

[63] W. G. Schwert. Why does stock market volatility change over time? *Journal of Finance*, 44:1115–1153, 1989.

[64] M. So, K. Lam, and W. Li. A stochastic volatility model with markov switching. *Journal of Business and Economic Statistics*, 16(2):244–253, April 19998.

[65] R. Stanton. A nonparametric model of term structure dynamics and the market price of interest rate risk. *Journal of Finance*, 52:1973–2000, 1997.

[66] S. J. Taylor. *Modelling Financial Time Series*. John Wiley, Chichester, U.K., 1986.

[67] O. Vasicek. An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5:177–188, 1977.