

GREQAM

Groupement de Recherche en Economie
Quantitative d'Aix-Marseille - UMR-CNRS 6579
Ecole des Hautes Etudes en Sciences Sociales
Universités d'Aix-Marseille II et III

**Document de Travail
n°2009-13**

Modeling Multivariate Interest Rates using Time-Varying Copulas and Reducible Stochastic Differential Equations

Ruijun BU

Ludovic GIET

Kaddour HADRI

Michel LUBRANO

May 2009

DT-GREQAM

Modeling Multivariate Interest Rates using Time-Varying Copulas and Reducible Stochastic Differential Equations

Ruijun Bu* Ludovic Giet[†] Kaddour Hadri[‡]
Michel Lubrano[§]

May 2009

Abstract

We propose a new approach for modeling non-linear multivariate interest rate processes based on time-varying copulas and reducible stochastic differential equations (SDEs). In the modeling of the marginal processes, we consider a class of non-linear SDEs that are reducible to Ornstein-Uhlenbeck (OU) process or Cox, Ingersoll, and Ross (1985) (CIR) process. The reducibility is achieved via a non-linear transformation function. The main advantage of this approach is that these SDEs can account for non-linear features, observed in short-term interest rate series, while at the same time leading to *exact discretisation* and *closed form likelihood functions*. Although a rich set of specifications may be entertained, our exposition focuses on a couple of non-linear constant elasticity volatility (CEV) processes, denoted OU-CEV and CIR-CEV, respectively. These two processes encompass a number of existing models that have closed form likelihood functions. The statistical properties of the two processes are investigated. In order to obtain more flexible functional form over time, we allow the transformation function to be time-varying. Results from our study of US and UK short term interest rates suggest that the new models outperform existing parametric models with closed form likelihood functions. We also find the time-varying effects in the transformation functions statistically significant. We study the conditional dependence structure of the two rates using Patton (2006a) time-varying Symmetrised Joe-Clayton copula. We find evidence of asymmetric dependence between the two rates, and that the level of dependence is positively related to the level of the two rates.

*University of Liverpool. Corresponding author: Management School, Chatham Street, Liverpool, L69 7ZH, UK, Tel: +44-151-795-3122, Fax: +44-151-795-3005, Email: RuijunBu@liv.ac.uk (Ruijun Bu).

[†]GREQAM, Marseille

[‡]Queen's University, Belfast

[§]GREQAM, Marseille and CNRS

Abstract

Nous proposons une nouvelle approche pour modéliser de façon non-linéaire et multivariée les taux d'intérêt de court terme en reliant au moyen d'un copule à paramètre variable des équations différentielles stochastiques réductibles (SDEs). Dans la modélisation des processus marginaux, nous considérons une classe de SDES non-linéaires réductible, soit à un Ornstein-Uhlenbeck (OU), soit à un processus de Cox, Ingersoll et Ross (1985) (CIR). La réduction est réalisée via une fonction de transformation non-linéaire. L'avantage principal de cette approche consiste en ce que ces SDES peuvent représenter des caractéristiques non-linéaires, observées dans les séries de taux d'intérêt à court terme, tout en conduisant en même temps à *une discretisation exacte* et à *fonctions de vraisemblance analytique*. Bien qu'une riche palette de spécifications puisse être considérée, nous nous concentrons sur deux fonctions de volatilité à élasticité constante non-linéaire (CEV), dénotées respectivement OU-CEV et CIR-CEV. Ces deux processus enveloppent un nombre important de modèles existants qui possèdent des fonctions de vraisemblance analytiques. Nous examinons les propriétés statistiques de ces deux processus. Afin d'obtenir des formes fonctionnelles flexibles, nous autorisons la fonction de transformation à varier dans le temps. Nos résultats empiriques portant sur les taux courts américains et britanniques suggèrent que nos nouveaux modèles dominent les modèles existants ayant des fonctions de vraisemblance analytiques. Nous trouvons aussi que les effets variables dans le temps de la fonction de transformation sont statistiquement significatifs. Nous étudions la structure de dépendance conditionnelle des deux taux au moyen de la copule de Joe-Clayton symétrisée à paramètres variables proposée par Patton (2006a). Nous montrons que la dépendance est asymétrique et que son niveau est relié positivement au niveau des deux taux.

JEL Classification: C13, C32, G12

Keywords: Interest Rate Models; Reducible Stochastic Differential Equations; Maximum Likelihood Estimation; Constant Elasticity Models; Copula;

1 Introduction

Continuous-time models have proved to be an enormously useful tool in modeling financial and more generally economic variables. They are widely used to study issues that include the decision to optimally consume, save, and invest, portfolio choice under a variety of constraints, contingent claim pricing, capital accumulation, resource extraction, game theory, and recently contract theory. The short term risk-free interest rate is one of the key variable in economics and finance. More models have been put forward to explain its behaviour than for any other issue in finance. Although many refinements and extensions are possible, the basic dynamic model for an interest rate process $\{r_t, t \geq 0\}$ is described by a stochastic differential equation

$$dr_t = \mu(r_t, \phi) dt + \sigma(r_t, \phi) dW_t, \quad (1)$$

where $\{W_t, t \geq 0\}$ is a standard Brownian motion. Both parametric and non-parametric methods of estimation have been developed in the literature. Parametric approaches, which form the majority, assume that the drift $\mu(r_t, \phi)$ and diffusion $\sigma^2(r_t, \phi)$ are known functions except for an unknown parameter vector ϕ in a bounded set $\phi \subset \mathbf{R}^K$. Examples include Merton (1973), Brennan and Schwartz (1979), Vasicek (1977), Cox (1975), Dothan (1978), Cox, Ingersoll, and Ross (1980, 1985), Courtadon (1982), Constantinides and Ingersoll (1984), Constantinides (1992), Duffie and Kan (1996) and Ait-Sahalia (1996b). Non-parametric and semi-parametric approaches, which do not constrain the functional form of the drift and diffusion functions to be within a parametric class, have been developed by for example Ait-Sahalia (1996a,b) and Stanton (1997). In the paper, we also allow for ϕ to be time-varying obtaining hence more flexible functional forms.

Parametric SDEs often provide a convenient way to describe the dynamics of economic and financial data, and a great deal of effort has been expended searching for efficient ways to estimate models based on them. Maximum likelihood is typically the estimator of choice. Although the process specified by a SDE is defined in continuous time, the data which are available are typically sampled at discrete time intervals. Little can be said about the implications of the dynamics in equation (1) for longer time intervals except for only a few rare cases. In finance, the well-known exceptions include Black and Scholes (1973), Vasicek (1977), and Cox, Ingersoll, and Ross (1985) which rely on existing closed form expressions. Since the majority of the SDEs do not lead to closed form likelihood functions, the resulting estimation problem turns out to be non-trivial. Considerable energy has been employed in developing computationally and statistically efficient approximation schemes. Examples include Lo (1988), Pedersen (1995), Brandt and Santa-Clara (2002), Shoji and Ozaki (1998), Kessler (1997), Elerian et al (2001). Durham and Gallant (2002) provides a survey on existing numerical techniques. More recently, Ait-Sahalia in several of his papers, Ait-Sahalia

(1999, 2002, 2008) developed methods for generating closed form approximation of likelihood functions for univariate and multivariate diffusions.

Following the recent trend in continuous-time models in finance, we develop a multivariate model. Multivariate models can deal, for instance, with multiple explanatory factors in asset pricing or with multiple yields or factors in term structure models. Their use permit the analysis of the movement and co-movement of these state variables over-time, across states and markets. To account for the possibility of changing financial and economic regime during the period of observation we permit the parameters to be time-varying. There is a substantial evidence of non-linearity in the drift and the diffusion components in diffusion processes representing the stochastic dynamics of short-term interest-rates. Therefore, in an attempt to broaden the scope of the class of non-linear diffusion processes that could lead to closed form likelihood functions and therefore avoid to resort to generally computationally intensive approximation methods. A new approach for modeling non-linear multivariate time-varying processes is considered. We first construct univariate non-linear processes of interest rates leading to closed form likelihood functions based on reducible SDEs. As far as we are aware this is the first time this technique is used. Although, Ait-Sahalia (1999, 2008) already employs the concept of reducibility but in a very different way than here. For Ait-Sahalia a diffusion r is said to be reducible to *unit* diffusion if there exists a one to one transformation of the diffusion r into a diffusion x , where $x = U(r)$ and $U(r)$ is an invertible function (see Ait-Sahalia (1999, 2008) for more details). This concept is then used as a condition for finding closed-form expansion for the likelihood function. In this paper, the concept of reducibility is defined in a slightly different way and is employed directly to find a non-linear SDE with a closed form likelihood function as explained further down.

The derived univariate marginal distributions are then welded together by an appropriate copula function to obtain a non-linear joint distribution of interest rates which accounts for the likely possibility of correlation (co-movement) across short-term interest rates. Evidently, our method can be used to estimate other multivariate financial models like asset pricing models, term structure models ect.

Reducible SDEs are typically non-linear SDEs which by definition can be reduced via a non-linear transformation to SDEs that are solvable analytically. These SDEs usually have closed form likelihood functions. Therefore, the likelihood function for the original non-linear SDEs can be expressed also in closed form via a distribution transformation, provided that the transformation function satisfies weak regularity conditions. Kloeden and Platen (1992) provide a detailed discussion of the techniques. In the context of financial modeling, we consider the type of non-linear SDEs that are reducible either to Ornstein-Uhlenbeck (OU) process or to Cox, Ingersoll, and Ross (1985) (CIR) process. A process that is reducible to OU is hence-

forth named "OU-reducible" in this paper, and similarly a process that is reducible to CIR is termed "CIR-reducible". The OU and CIR processes are therefore called the "*basic processes*". We derive the specifications for general OU-reducible and CIR-reducible processes, as well as the transformation functions that lead to the reduction.

Although a rich set of specifications for the desired non-linear SDEs (based on alternative transformation functions) may be entertained under the general framework, our exposition of this approach focuses on models that have constant elasticity volatility (CEV), also known as the CEV models. Depending on whether the specifications are OU-reducible or CIR-reducible, they are denoted as OU-CEV and CIR-CEV, respectively. It should be noted that the two CEV models considered here (relatively simple choice of the diffusion function) already nest, as special cases, a number of well-documented continuous-time models that are known to have closed form likelihood functions. These models include not only the OU and CIR, the two basic processes, but also the process of Ahn and Gao (1999), henceforth denoted as AG process. The sufficient conditions for the stationarity of the CIR-CEV process are provided and the same issue for the OU-CEV process is discussed.

In an attempt to generate more flexible dynamics of the process which allows for time-variation in the transformation applied to the basic processes, we extended our theory to allow for conditioning variables in the transformation function. This has been done in the spirit of an *Autoregressive Conditional specification*. This modification of the original OU-CEV and CIR-CEV processes lead to time-varying (TV) version of the processes, denoted as TV-OU-CEV and TV-CIR-CEV, respectively.

As an illustration of the approach developed in this paper, we apply our framework to analyse the dynamic co-movements between US and UK short term interest rates data. We measure the US and UK short rates by 1-Month Eurodollar Rate (EDR) and 1-Month London Interbank Offered Rate (LIBOR) in British Sterling, respectively. Two different frequencies (monthly and weekly) of the two rates are considered in this study. We found that simple models like OU and CIR are strongly rejected by the data against the more general CEV model. The AG model is also rejected by most data series with the only exception happening for monthly LIBOR series. When we extend our model to allow for time-varying transformation, we found that for all four data series the time-varying effect of the transformation parameter is significant.

The dependence of the US and UK short rates are studied on the basis of a fitted conditional copula. We found that the best fitting copula model in terms of information criteria is the Patton (2006a) conditional Symmetrised Joe-Clayton copula (SJC). We found that the time-varying effects in the conditional SJC copula are significant. Also significant is the asymmetry in the tail dependence implied by the copula. From the fitted tail dependence

coefficients, we found that the tail dependencies tend to be higher when the level of the two rates are relatively high. When the level of the two rates are relatively low, the tail dependencies tend to be lower. Similar relationship is also found in the conditional linear correlation coefficients implied by the conditional SJC copula.

The remainder of the paper is organised as follows. In Section 2, we propose a copula-based approach for modeling non-linear multivariate interest rate processes. For modeling the marginal processes, we derive a general specification of a class of univariate non-linear SDEs that are reducible to either an OU or a CIR process. We also discuss how copula functions can be used to construct flexible non-linear multivariate models. Section 3 gives a detailed exposition of the CEV specification for the marginal process and analyses the characteristics of the implied interest rate distributions. A time-varying transformations, which can be very useful in empirical dynamic applications, is also proposed. In Section 4, we present an application of our approach dealing with the study of the dynamic co-movements between US and UK short term interest rates via a time-varying copula. Some concluding remarks are given in Section 5. All theoretical proofs are provided in the Appendices.

2 Non-linear Multivariate Modeling

In this section we propose a copula-based approach for modeling multivariate interest rates. We start first by deriving the marginal processes employing the reduction technique then we weld the univariate processes obtained in the first step via copulas to obtain a multivariate distribution. In both cases we allow for time-varying parameters to make our models more flexible and dynamic.

2.1 Modeling Marginal Processes using Reducible SDEs

We assume that the dynamics of a stochastic process of interest $\{r_t, t \geq 0\}$, can be described by the SDE in (1). The functions $\mu(r_t, \phi)$ and $\sigma^2(r_t, \phi)$, typically non-linear in r_t , are respectively the drift and the diffusion functions of the process, and $\phi \in R^K$ is a vector of unknown parameters. The only assumption we impose at this stage of the analysis is that (1) belongs to the class of the so called *reducible* SDEs as defined in Kloeden and Platen (1992). These reducible SDEs can be changed via appropriate transformation functions, into SDEs that have closed form likelihood functions. Provided that the transformation function satisfies weak regularity conditions, which are explained below, the likelihood function for non-linear SDE in (1) may also be expressed in closed form by employing this technique.

The class of SDEs that have closed form transition probability density

functions can be represented by

$$dx_t = (a_1 x_t + a_2) dt + b x_t^\alpha dW_t. \quad (2)$$

Special cases of interest that arise in the finance literature include: (i) the Ornstein-Uhlenbeck (OU) process ($\alpha = 0$) which has both the transition probability density and the marginal density normally distributed; (ii) the Cox, Ingersoll and Ross (1985) (CIR) process ($\alpha = 1/2$) which has a non-central χ^2 transition density with fractional degrees of freedom and its marginal density follows a Gamma distribution; and (iii) the *Geometric Brownian motion* ($\alpha = 1$) and ($a_2 = 0$) which leads to a *log-normal* transition density function.

If there exists an appropriate transformation function $U(\cdot, \phi)$ such that the process $\{x_t = U(r_t, \phi), t \geq 0\}$ follows a SDE that is solvable analytically, then the process $\{r_t, t \geq 0\}$ governed by (1) is said to be reducible. A process that is reducible to OU is called "OU-reducible" and a process that can be reduced to CIR is called "CIR-reducible"¹. It can be shown that under minor conditions such processes would possess an explicit analytic likelihood function via a trivial transformation of the distribution. If $\partial U(r_t, \phi) / \partial r_t \neq 0$, the *Inverse Function Theorem* ensures the existence of a local inverse $r_t = U^{-1}(x_t, \phi)$.

2.1.1 The Transformation Function

For ease of exposition, we re-write (1), with no loss of generality, in the following way:

$$dr_t = \mu(r_t, \phi) dt + \sigma_0 \sigma(r_t, \theta) dW_t, \quad (3)$$

where $\phi = (\theta, \sigma_0)'$ and σ_0 is a normalizing scalar.

Define an analytic *transformation function* $U(\cdot, \phi)$, where typically $U(\cdot, \phi)$ only depends on a subset of ϕ , and let $x_t = U(r_t, \phi)$. Then according to Itô's lemma, we obtain the following dynamics for $\{x_t, t \geq 0\}$:

$$\begin{aligned} dx_t = & \left[\mu(r_t, \phi) \frac{\partial U(r_t, \phi)}{\partial r_t} + \frac{\sigma_0^2 \sigma^2(r_t, \theta)}{2} \frac{\partial^2 U(r_t, \phi)}{\partial r_t^2} \right] dt \\ & + \sigma_0 \sigma(r_t, \theta) \frac{\partial U(r_t, \phi)}{\partial r_t} dW_t. \end{aligned} \quad (4)$$

It follows that the non-linear SDEs in (3) that are reducible to (2) via transformation function $x_t = U(r_t, \phi)$ must satisfy the following two equa-

¹The Geometric Brownian Motion is not mean reverting, neither is the SDE that is reducible to GBM in general. This lack of mean-reversion generally makes the GBM-reducible SDEs not attractive for modeling financial variables that may be interpreted as a return. Therefore, it is not pursued further in this study. However, non-linear SDEs that are reducible to non-stationary processes is a topic for future research.

tions:

$$\sigma_0 \sigma(r_t, \boldsymbol{\theta}) \frac{\partial U(r_t, \phi)}{\partial r_t} = b U^\alpha(r_t, \phi) \quad (5)$$

$$\mu(r_t, \phi) \frac{\partial U(r_t, \phi)}{\partial r_t} + \frac{1}{2} \sigma_0^2 \sigma^2(r_t, \boldsymbol{\theta}) \frac{\partial^2 U(r_t, \phi)}{\partial r_t^2} = a_1 U(r_t, \phi) + a_2. \quad (6)$$

It should be noted that the three unknown functional forms $\mu(r_t, \phi)$, $\sigma(r_t, \boldsymbol{\theta})$ and $U(r_t, \phi)$ cannot be uniquely identified from only two equations (5) and (6) unless an additional assumption is imposed on them. There are three approaches for dealing with this identification problem. The first approach is to start with a desired drift function $\mu(r_t, \phi)$. This is the most difficult route as it involves solving a higher-order differential equation for either $\sigma(r_t, \boldsymbol{\theta})$ or $U(r_t, \phi)$. An analytic solution is hardly obtained except in very rare cases. The most general approach is to make assumptions directly on $U(r_t, \phi)$. Then the specification of (3) or equivalently (1) can be uniquely determined under minor identification conditions. However, it is not always straightforward to formulate such a specification without prior knowledge on the desired features that the resulting SDEs should possess. A slightly less ambitious but substantially simplified approach is to start with a desired specification of the volatility function $\sigma(r_t, \boldsymbol{\theta})$. Then, finding $U(r_t, \phi)$ will only involve solving the first-order differential equation in (5). The drift function $\mu(r_t, \phi)$ can then be trivially inferred from equation (6). Given the significance of the volatility properties in financial applications, this approach appears to be fairly reasonable. Moreover, knowledge about the $U(r_t, \phi)$ for some particular $\sigma(r_t, \boldsymbol{\theta})$ provides useful insights into how to construct a transformation function that gives a non-linear SDEs with wanted properties. It also helps to understand the restrictions required on a general transformation function specification so that the resulting SDEs satisfy regularity conditions regarding, for instance stationarity and boundary behaviours.

For a chosen standardised volatility function $\sigma(r_t, \boldsymbol{\theta})$ and letting $b = \sigma_0$, the transformation function $U(\cdot, \phi)$ can be found by solving the following ordinary differential equation:

$$\frac{\partial U(r_t, \phi)}{\partial r_t} = U^\alpha(r_t, \phi) \frac{1}{\sigma(r_t, \boldsymbol{\theta})},$$

which yields for $\alpha \neq 1$

$$U(r_t, \phi) = U(r_t, \boldsymbol{\theta}) = \left\{ (1 - \alpha) \left[\int \frac{1}{\sigma(r_t, \boldsymbol{\theta})} dr_t + c \right] \right\}^{\frac{1}{1-\alpha}},$$

where c is the *constant of integration*. Note here that the transformation function only depends on $\boldsymbol{\theta}$. When $\alpha = 0$, the original process is reducible

to the OU process and the required transformation is given by

$$U(r_t, \boldsymbol{\theta}) = \int \frac{1}{\sigma(r_t, \boldsymbol{\theta})} dr_t + c. \quad (7)$$

When $\alpha = 1/2$, the original process is reducible to CIR and the corresponding transformation is²

$$U(r_t, \boldsymbol{\theta}) = \left\{ \frac{1}{2} \left[\int \frac{1}{\sigma(r_t, \boldsymbol{\theta})} dr_t + c \right] \right\}^2. \quad (8)$$

Replacing the transformation function and its first and second derivatives reveals the non-linear drift function $\mu(r_t, \boldsymbol{\phi})$. The complete specification of the process can then be written as

$$dr_t = \frac{1}{\frac{\partial U(r_t, \boldsymbol{\theta})}{\partial r_t}} \left[a_1 U(r_t, \boldsymbol{\theta}) + a_2 - \frac{1}{2} b^2 \sigma^2(r_t, \boldsymbol{\theta}) \frac{\partial^2 U(r_t, \boldsymbol{\theta})}{\partial r_t^2} \right] dt + b \sigma(r_t, \boldsymbol{\theta}) dW_t, \quad (9)$$

where $U(r_t, \boldsymbol{\theta})$ is given by either (7) or (8). Note that the unknown parameter vector $\boldsymbol{\phi}$ is in fact identified as $\boldsymbol{\phi} = (\boldsymbol{\theta}', a_1, a_2, b)'$.

2.1.2 Exact Discretisation and Closed Form Transition Densities

Using reducible SDEs as a modeling tool has the following advantages. Firstly, since the non-linear diffusion process in (9) is a transformed process of a basic process, either OU or CIR, via a transformation function (7) or (8), many useful mathematical and statistical properties of the basic processes are preserved after the transformation. For instance, since both OU and CIR processes have exact discretisation, the process in (9) also has exact discretisation as a result of straight forward mapping by function $r_t = U^{-1}(x_t, \boldsymbol{\phi})$.

For OU-reducible and CIR-reducible processes, the Jacobians of the transformations are given by

$$\mathbf{J}_{OU} = |\partial U(r_t) / \partial r_t| = 1 / \sigma(r_t, \theta),$$

and

$$\mathbf{J}_{CIR} = |\partial U(r_t) / \partial r_t| = [1 / 2 \sigma(r_t, \theta)] \left| \int 1 / \sigma(r_t, \theta) dr_t \right|,$$

²The transformation function in (7) suggests that the integration constant c only causes a shift in the level (unconditional mean) of the basic process x_t , which is already accounted for by the value of a_2 . It is therefore irrelevant and can be set equal zero for OU-reducible processes. However, a quick inspection of (8) reveals a clear non-linear effect of c in the transformation. The effect on the transformed process x_t is clearly more complicated than just a level shift. Some preliminary experiments suggest that to include c as a free parameter does not improve the goodness of fit significantly and it is likely to cause some identification problems. We therefore chose to set $c = 0$ in the subsequent analysis.

respectively. The transition density for the proposed model can be easily obtained by the standard transformation method of the distribution. Monotonicity in $U(r_t, \boldsymbol{\theta})$ ensures that the transformation is unique. The corresponding marginal density function can be obtained by taking the step length Δ to the limit providing that the process is stationary and therefore the limit exists.

2.2 Copulas

Flexible non-linear multivariate models can be constructed by joining univariate non-linear processes via appropriate copulas. The concept of copula was made operational by Sklar (1959) who showed that if an n -dimensional joint distribution exists then it can be recovered by joining the n marginal distributions by a suitable copula which embodies the dependence between the n variables. Recent methodological reviews are given for example by Joe (1997) and Nelson (1999). A review of applications of copulas to finance can be found in Embrechts et al. (2003) and Cherubini et al. (2004).

A more useful implication of Sklar's theorem is that we may link together any two univariate distributions, of any type (not necessarily from the same family), using any copula and we will have defined a valid bivariate distribution. This flexibility of copulas makes it an extremely useful tool for modeling jointly distributed interest rates. The usefulness of this result stems from the fact that although in economics and statistics literatures we have a vast selection of flexible parametric univariate distributions, the set of parametric multivariate distributions available is much smaller. With Sklar's theorem, the set of possible parametric multivariate distributions is increased substantially, some of which may be empirically attractive.

Focusing on the bivariate case, we denote $F_X(x)$ and $F_Y(y)$ as the continuous marginal distribution functions of X and Y , and $F_{XY}(x, y)$ the joint distribution function. Also let $f_x(x)$ and $f_y(y)$ be the marginal density functions, and $f_{xy}(x, y)$ the joint probability density function. The Sklar's Theorem states:

$$\begin{aligned} F_{XY}(x, y) &= C(F_X(x), F_Y(y)) \\ f_{xy}(x, y) &= f_x(x) \cdot f_y(y) \cdot c(F_X(x), F_Y(y)), \end{aligned}$$

where $C : [0, 1]^2 \rightarrow [0, 1]$ is the copula function for the bivariate random vector (X, Y) , and c is the corresponding copula density.

The procedure employed to construct the joint distribution is a two-step method of estimation. In the first stage we estimate the two marginal distribution models separately, and in the second stage we estimate the copula model. Although estimating all of the coefficients simultaneously yields the most efficient estimates, a larger number of parameters can make numerical optimisation of the loglikelihood function difficult. Under standard condi-

tions, the estimates obtained in two-steps are consistent and asymptotically normal (see Patton (2006b) for more details).

3 The CEV Specification for the Marginals

The constant elasticity volatility model was introduced by Chan et al. (1992). They claim that it was their best fitting model in the study carried out in their paper. This model was further studied by Aït-Sahalia (1996b) who promoted the use of a non-linear drift function to provide a better mean-reversion effect. The same type of specification was also estimated by Conley et al. (1997) and Gallant and Tauchen (1998).

3.1 Transformation Functions and SDEs

The CEV specification of the diffusion is given by $\sigma(r_t, \gamma) = r_t^\gamma$, where $\gamma \in (0, 1) \cup (1, \infty)$. It follows from (7) that for a non-linear CEV process that is reducible to OU, henceforth denoted as OU-CEV, the transformation is given by

$$x_t = U(r_t, \gamma) = r_t^{1-\gamma} / (1 - \gamma). \quad (10)$$

Aït-Sahalia (1999) suggests to define

$$x_t = U(r_t, \gamma) = r_t^{1-\gamma} / (\gamma - 1), \quad (11)$$

for $\gamma > 1$. Since $r_t \in \mathbb{R}^+$, such treatment ensures that x_t and r_t covers the same domain. Nevertheless, to cover the same domain is not essential because it only affects the sign of the parameter a_2 in the estimated model. Such a treatment is a matter of convenience. We therefore stick to (10) and (11) according to the range of γ for the rest of the paper. It is easily verified that $\partial U(r_t, \gamma) / \partial r_t = r_t^{-\gamma}$. Since $r_t \in \mathbb{R}^+$, the above transformation is always strictly monotonic, which ensures identification of all parameters. It follows from (9) that the dynamics of OU-CEV process is governed by the following diffusion.

$$dr_t = \left(\frac{1}{2} b^2 \gamma r_t^{2\gamma-1} - a_2 r_t^\gamma \text{sgn}(\gamma - 1) + \frac{a_1}{1 - \gamma} r_t \right) dt + b r_t^\gamma dW_t, \quad (12)$$

where $\text{sgn}(\cdot)$ is the sign function.

For a non-linear CEV model that is reducible to CIR, henceforth denoted by CIR-CEV, we have

$$x_t = U(r_t) = (1/4) \left[r_t^{1-\gamma} / (1 - \gamma) \right]^2. \quad (13)$$

It is easily verified that $\partial U(r_t) / \partial r_t = r_t^{1-2\gamma} / (2 - 2\gamma)$. For $r_t \in \mathbb{R}^+$, the above transformation is also strictly monotonic. The dynamics of the CIR-CEV process is therefore given by

$$dr_t = \left\{ \left[2a_2 (1 - \gamma) + \frac{1}{2} b^2 (2\gamma - 1) \right] r_t^{2\gamma-1} + \frac{a_1 r_t}{(2 - 2\gamma)} \right\} dt + b r_t^\gamma dW. \quad (14)$$

The SDEs defined in (12) and (14) encompasses a number of interest rate processes that are known to have closed form likelihood functions. These models can be obtained from (12) and (14) by simply placing the appropriate restrictions on the four parameters, a_1 , a_2 , b , and γ . *Table 1* provides the specifications of nested models and the corresponding restrictions.

Table 1: Models Nested in OU-CEV and CIR-CEV Specifications

Models Nested in OU-CEV		
Model	Specification	Restriction(s)
1. Merton	$dr_t = a_2 dt + b dW_t$	$\gamma = 0, a_1 = 0$
2. OU (Vasicek)	$dr_t = (a_1 r_t + a_2) dt + b dW_t$	$\gamma = 0$
Models Nested in CIR-CEV		
Model	Specification	Restriction(s)
3. CIR (SR)	$dr_t = (a_1 r_t + a_2) dt + b \sqrt{r_t} dW_t$	$\gamma = 1/2$
4. CIR (VR)	$dr_t = br_t^{3/2} dW_t$	$\gamma = 3/2, a_1 = 0, a_2 = b^2$
5. CEV	$dr_t = a_1 r_t dt + br_t^\gamma dW_t$	$a_2 = b^2 (2\gamma - 1)/4 (\gamma - 1)$
6. AG	$dr_t = [(b^2 - a_2) r_t^2 - a_1 r_t] dt + br_t^{3/2} dW$	$\gamma = 3/2$

Model 1 and 2 are nested in the OU-CEV specification. Model 1 is used in Merton (1973) to derive a model of discount bond prices. This stochastic process for the risk free rate is simply a Brownian motion with drift. Model 2 is the Ornstein-Uhlenbeck process used by Vasicek (1977) in deriving an equilibrium model of discount bond prices. This Gaussian process has been used extensively by others in valuing bond options, futures, futures options, and other types of contingent claims.

Model 3 to 6 are nested in the CIR-CEV specification. Model 3 is the square root (SR) process which appears in the Cox, Ingersoll, and Ross (CIR) (1985) single-factor general-equilibrium term structure model. This model has also been used extensively in developing valuation models for interest-rate-sensitive contingent claims. Model 4 is introduced by CIR (1980) in their study of variable-rate (VR) securities. A similar model is also used by Constantinides and Ingersoll (1984) to value bonds in the presence of taxes. Model 5 is the constant elasticity of variance (CEV) process introduced by Cox (1975) and by Cox and Ross (1976). The application of this process to interest rates is discussed in Marsh and Rosenfeld (1983). Model 6 is relatively recent compared to Model 1-5. It is introduced by Ahn and Gao (1999) and is of particular interest due to its quadratic drift function. In their study, they considered a non-linear term structure model which was documented to be able to generate more realistic dynamics of the interest rate. The drift and the diffusion functions were both non-linear, and the market price of risk generated by the model were close to the empirical findings reported in Chan et al. (1992), Ait-Sahalia (1996a,b), Conley et al. (1997), and Stanton (1997). Their empirical analysis indicates that

their model can dominate the affine-class term structure models in both the time series and cross-sectional dimension. It is easily seen that the CIR-CEV model proposed in the current work encompasses the Ahn and Gao (1999) model as a special case where $\gamma = 1.5$. Clearly, the CIR-CEV model is a more general setup which not only preserves the non-linear features of the drift function discussed in Ahn and Gao (1999) but also allows for an extra degree of freedom in the (data-driven) choice of the value of γ . In our application, we found that the empirical estimate of γ can be very different from the suggested value.

3.2 Analysis of the Distributions

This section provides an analysis of the distributions implied by the two stochastic processes proposed in the previous section³.

3.2.1 CIR-CEV

We apply the methodology of Ait-Sahalia (1996b) concerning the constraints on the drift and the diffusion to the CIR-CEV model in (14) to derive the sufficient conditions for stationarity and unattainability of 0 and ∞ in finite expected time. The results are given in the following proposition.

Proposition 3.1 Let $\{r_t, t \geq 0\}$ be a CIR-CEV process defined in (14). The necessary and sufficient conditions for stationarity and unattainability of 0 and ∞ in finite expected time are: (i) $a_1 < 0$ and $4a_2/b^2 > (2\gamma - 1)/(\gamma - 1)$ if $\gamma > 1$; (ii) $a_1 < 0$ and $4a_2/b^2 > 1/(1 - \gamma)$ if $\gamma < 1$.

Proof See Appendix ■

Since the original CIR process is a special case of (14) when $\gamma = 0.5$, the relevant conditions deduced from the proposition is $2a_2/b^2 > 1$ as stated in Cox et al. (1985). The process in (14) reduces to the Ahn and Gao (1995) process if $\gamma = 1.5$, the corresponding conditions become $a_2/b^2 > 1$ as stated in their paper. Clearly, our results provide conditions for arbitrary value of $\gamma \in (0, 1) \cup (1, \infty)$.

For the CIR-CEV process, according to the transformation in (13) the transition density of the process is given by

$$f(r_t | r_{t-\Delta}) = \frac{1}{2} \frac{r_t^{1-2\gamma}}{|1-\gamma|} c e^{-u-v} \left(\frac{v}{u}\right)^{q/2} I_q \left[2(uv)^{1/2}\right],$$

³The analysis carried out in this section assumes that the parameter a_1 in the two processes are non-zero. Therefore, the results derived here may not be valid for Model 1 and 4 which assume $a_1 = 0$. Meanwhile, for the same reason discussed in footnote 1, Model 1, 4, and 5 are not considered in the rest of the paper.

where

$$\begin{aligned} c &= 2a_1 / [b^2 (e^{a_1 \Delta} - 1)], u = (ce^{a_1 \Delta} / 4) \left[r_{t-\Delta}^{1-\gamma} / (1-\gamma) \right]^2 \\ v &= (c/4) \left[r_t^{1-\gamma} / (1-\gamma) \right]^2, q = 2a_2 / b^2 - 1. \end{aligned}$$

and $I_q(\cdot)$ is the modified Bessel function of the first kind of order q .

Similar to the OU-CEV process, the CIR-CEV process also permits a closed form expression for its conditional distribution function which can be written as

$$F(r_t | r_{t-\Delta}) = \begin{cases} D(2cx_t; 2q+2, 2u) & \text{for } \gamma < 1 \\ 1 - D(2cx_t; 2q+2, 2u) & \text{for } \gamma > 1 \end{cases},$$

where x_t is defined by (13) and $D(\cdot; 2q+2, 2u)$ is the non-central χ^2 distribution function with $2q+2$ degrees of freedom and non-centrality parameter $2u$.

Straightforward calculation yields the m th conditional moments for r_t following the CIR-CEV process:

$$\begin{aligned} &E(r_t^m | r_{t-\Delta}) \\ &= [2|1-\gamma|]^{\frac{m}{1-\gamma}} c^{-\frac{m}{2(1-\gamma)}} e^{-u} \frac{\Gamma\left(q + \frac{m}{2(1-\gamma)} + 1\right)}{\Gamma(1+q)} {}_1F_1\left(q + \frac{m}{2(1-\gamma)} + 1, 1+q, u\right), \end{aligned}$$

where ${}_1F_1(\cdot, \cdot, \cdot)$ is the *confluent hypergeometric function*⁴ (or *Kummer's function*) which can be represented in the *Pochhammer symbols* as follows,

$${}_1F_1(a, b, y) = \sum_{j=0}^{\infty} \frac{(a)_j y^j}{(b)_j j!} = \frac{\Gamma(b)}{\Gamma(b-a)\Gamma(a)} \int_0^1 e^{yz} z^{a-1} (1-z)^{b-a-1} dz,$$

and $\Gamma(\cdot)$ is the gamma function. Since the CIR-CEV process displays mean reversion, then as $\Delta \rightarrow \infty$, its distribution is well defined. It can be shown that the steady-state density function is given by

$$\pi(r_t) = \frac{1}{2|1-\gamma|} \frac{r_t^{1-2\gamma} \left(-\frac{2a_1}{b^2}\right)^{\frac{2a_2}{b^2}}}{\Gamma\left(\frac{2a_2}{b^2}\right)} x_t^{\frac{2a_2}{b^2}-1} \exp\left(\frac{2a_1}{b^2} x_t\right),$$

and the m th unconditional moments are

$$E[r_t^m] = [2|1-\gamma|]^{\frac{m}{1-\gamma}} \left(-\frac{2a_1}{b^2}\right)^{-\frac{m}{2(1-\gamma)}} \frac{\Gamma\left(\frac{2a_2}{b^2} + \frac{m}{2(1-\gamma)}\right)}{\Gamma\left(\frac{2a_2}{b^2}\right)}.$$

⁴Abadir (1999) provided a detailed account of the hypergeometric functions and their applications in economics. Also see Abadir and Rockinger (2003) and Bu and Hadri (2007) for applications in finance.

3.2.2 OU-CEV

The analysis of the OU-CEV process is less straightforward than that of the CIR-CEV process. An inspection of the two transformation functions reveals that the process is only defined when the driving process x_t is on the domain $(0, \infty)$. Since zero and negative values are attainable for the OU process in finite expected time, in principle we can only concentrate on the realizations of the process that do not lead to absorption at zero. That is, we need at least in theory define a conditional process $\{r_t = U(x_t) : t > 0, x_t > 0\}$. Accordingly, $\{r_t = U(x_t) : t > 0, x_t > 0\}$ is prescribed to be the process confined to the sample paths not involving ultimate absorption at zero. This design is similar at least in spirit to the *Conditioned Diffusion Processes* discussed in Karlin and Taylor (1981). The separate treatment in (10) and (11) guarantees that $-a_2/a_1 > 0$ and thus the conditional process is mean-reverting on $(0, \infty)$.

The techniques required to formally express the desired conditional diffusion process are beyond the scope of this study. However, it is not difficult to realise that if the probability that $x_t > 0$ becomes negligible, i.e. under the assumption

$$P(x_t \leq 0) = \int_{-\infty}^0 \pi(x_t) dx_t = 0, \quad (15)$$

then the difference between the unconditional and conditional processes becomes insignificant in almost all real world applications.

For this reason, implementation of the OU-CEV model as well as all subsequent analysis of the distribution of the process rely on the assumption that the parameters a_1, a_2 and b are such that condition (15) holds approximately, where $\pi(x_t)$ is the marginal density of x_t . This essentially requires that the unconditional mean of the process $-a_2/a_1$ is sufficiently large in absolute value compared to the unconditional standard deviation $\sqrt{-b^2/2a_1}$. It turns out that in the practice of modeling interest rate such restriction is hardly ever binding, since the above probability implied by the estimated parameters for the embedded OU process is typically negligibly small. For instance, in our application to the US and UK short term interest rates the corresponding probabilities calculated from the fitted models are found to be practically zero (of order lower than 10^{-8}). The implication is that in almost all practical work concerning only finite time intervals, absorption of the process should not be a serious concern.

For the OU-CEV process, according to the transformations defined in (10) and (11) the probability density of the interest rate r_t conditional on $r_{t-\Delta}$, where Δ is the step length, is given by the following

$$f(r_t|r_{t-\Delta}) = r_t^{-\gamma} \frac{1}{\sqrt{2\pi\sigma_{ou}^2}} \exp\left[-\frac{1}{2}\left(\frac{x_t - \mu_{ou}}{\sigma_{ou}}\right)^2\right],$$

where

$$\mu_{ou} = e^{a_1\Delta}x_{t-\Delta} - \frac{a_2}{a_1}(1 - e^{a_1\Delta}) \quad \text{and} \quad (16)$$

$$\sigma_{ou}^2 = \frac{b^2(e^{2a_1\Delta} - 1)}{2a_1}. \quad (17)$$

Since the OU-CEV process is a continuous and monotonic transformation of the OU process, it has a closed form expression for its conditional distribution function which is given by

$$F(r_t|r_{t-\Delta}) = \begin{cases} \Phi(x_t; \mu_{ou}, \sigma_{ou}) - \Phi(0; \mu_{ou}, \sigma_{ou}) & \text{for } \gamma < 1 \\ 1 - \Phi(x_t; \mu_{ou}, \sigma_{ou}) & \text{for } \gamma > 1 \end{cases},$$

where $\Phi(\cdot; \mu_{ou}, \sigma_{ou})$ is the distribution function for a normally distributed random variable with mean μ_{ou} and standard deviation σ_{ou} . Here x_t is defined by (10) for $\gamma < 1$ and (11) for $\gamma > 1$, respectively.

Straightforward calculation also yields the m th conditional moments for r_t following the OU-CEV process⁵:

$$\begin{aligned} E(r_t^m|r_{t-\Delta}) &= \frac{[\sqrt{2}\sigma_{ou}|1-\gamma|]^v}{\sqrt{\pi}} \frac{1}{2} e^{-w^2} \left[\Gamma\left(\frac{v+1}{2}\right) {}_1F_1\left(\frac{v+1}{2}, \frac{1}{2}, w^2\right) \right. \\ &\quad \left. + wv\Gamma\left(\frac{v}{2}\right) {}_1F_1\left(1 + \frac{v}{2}, \frac{3}{2}, w^2\right) \right], \end{aligned}$$

where

$$v = \frac{m}{1-\gamma}, \quad w = \frac{\mu_{ou}}{\sqrt{2}\sigma_{ou}}.$$

Since the unconditional distribution of the OU process is also normal, the marginal density $\pi(r_t)$ and the unconditional moments $E(r_t^m)$ have similar expressions to their conditional counterparts. The only difference is that we will have to replace the conditional mean and variance in (16) and (17) by their corresponding limits as $\Delta \rightarrow \infty$.

4 The Conditional Transformation

The discussion up to this point assumes that the transformations applied to the two basic processes remain constant over time. One natural extension of our modelling approach is to consider time-varying or conditional transformations assuming that the basic process remains the same.

To consider time-varying transformation, we assume that the functional form of the transformation remains fixed over the sample whereas the parameters vary according to some evolution equation. This is in the spirit of

⁵The expression is derived under the assumption that $P(x_t \leq 0) = 0$, which is only approximately true in theory but harmless in practice for reasons discussed above. See appendix for details of the proof.

Hansen (1994) "autoregressive conditional density". An alternative to this approach may be to allow also for time variation in the functional form using a regime switching model. The latter requires knowledge about at least a couple of reasonable transformation functions that may be applied to a basic process, either OU or CIR.

For both OU-CEV and CIR-CEV processes, the transformation functions depend on a single parameter γ . We call it the *transformation parameter*. A natural extension is to specify an equation describing the evolution over time of the parameter γ_t . We propose the following evolution equations.

To reflect the $\gamma < 1$ case in the time-homogenous transformation design, we can define

$$\gamma_t = \Lambda \left(\omega + \sum_{i=1}^p \alpha_j \gamma_{t-i} + \sum_{j=1}^q \beta_i x_{t-j} \right) \quad (18)$$

where $\Lambda(x) \equiv (1 + e^{-x})^{-1}$ is the logistic transformation, used to keep γ_t in $(0, 1)$ all the time. Similarly, we can let

$$\gamma_t = \tilde{\Lambda} \left(\omega + \sum_{i=1}^p \alpha_j \gamma_{t-i} + \sum_{j=1}^q \beta_i x_{t-j} \right) \quad (19)$$

where $\tilde{\Lambda}(x) \equiv (1 + e^{-x})$ is the reciprocal of the the logistic transformation, used to keep γ_t inside the range $(1, \infty)$ all the time. In practice, the choice of p and q can be decided by some model selection procedure. Obviously, significance in either α_j or β_i or both is an indication of the time-varying effect in the transformation.

5 Bivariate Modeling of US and UK Short Term Interest Rates

In this section, we apply the theory developed in previous sections to both US and UK short term interest rates data and compare, in terms of goodness of fit, the performance of our models to existing models that have closed form likelihood functions.

Interest rates between the two countries are known to be strongly related for various social, political and economic reasons. In addition to modeling individual short rate dynamics, it is essential to analyse the dynamic comovements between the two rates. This is done via a time-dependent copula function. Details on modeling the marginal distributions and the selection of copula functions are discussed below.

5.1 Data

We measure the US and UK short term interest rates by 1-Month Eurodollar Rate (EDR) and 1-Month London Interbank Offered Rate (LIBOR) in British Stirling. Two different frequencies, monthly and weekly, of the two rates are employed in this study. The EDR data are collected from the H.15 release of the Federal Reserve website and the data of LIBOR are obtained from BBA (British Banking Association) database. For each of the two rates, we use the longest sample period for which data are available.

Table 2: Descriptive Statistics of EDR and LIBOR Data

	Eurodollar Rate		LIBOR	
	1971.01-2007.12		1986.01-2007.12	
Sample period				
Frequency	Monthly	Weekly	Monthly	Weekly
Sample size	444	1930	264	1148
Mean	6.778	6.781	7.434	7.413
Std. Dev.	3.541	3.550	3.237	3.214
Skewness	1.081	1.099	1.013	1.003
Kurtosis	4.719	4.809	2.875	2.872
Jarque-Bera Statistic	141.149*	651.979*	45.330*	193.280*
ρ_1	0.978	0.996	0.987	0.997
ρ_2	0.944	0.990	0.971	0.995
ρ_3	0.916	0.983	0.955	0.992
ρ_4	0.891	0.975	0.940	0.989
ρ_5	0.869	0.967	0.927	0.986
ρ_6	0.851	0.959	0.909	0.982

An asterisk (*) indicates a rejection of the null hypothesis at the 0.01 level.

Table 2 provides some summary statistics of the data. The EDR data starts from January 1971 to December 2007, yielding 444 monthly observations and 1930 weekly observations. The LIBOR data are available from January 1986 to December 2007, which gives 264 monthly observations and 1148 weekly observations, respectively. The sample mean of the two rates suggest that the UK short rates are on average higher than its US counterpart, whereas the standard deviations indicate the US rates are more variable than UK rates. Both rates exhibit clear departures from normality. While both rates show positive skewness, it is interesting to note that the EDR is leptokurtic whereas the LIBOR is platykurtic. The departure from normality is confirmed by the significance of the Jarque-Bera test of normality of the unconditional distribution.

Time series plots of the data at both frequencies are provided in Figure 1. As expected, the plots of the data at different frequencies appear to be very similar, except that the plot for weekly observations is less smooth than that of the monthly data. Neither of the two rates show discernible trend over its sample period. For the EDRs, due to the shift in monetary policy,

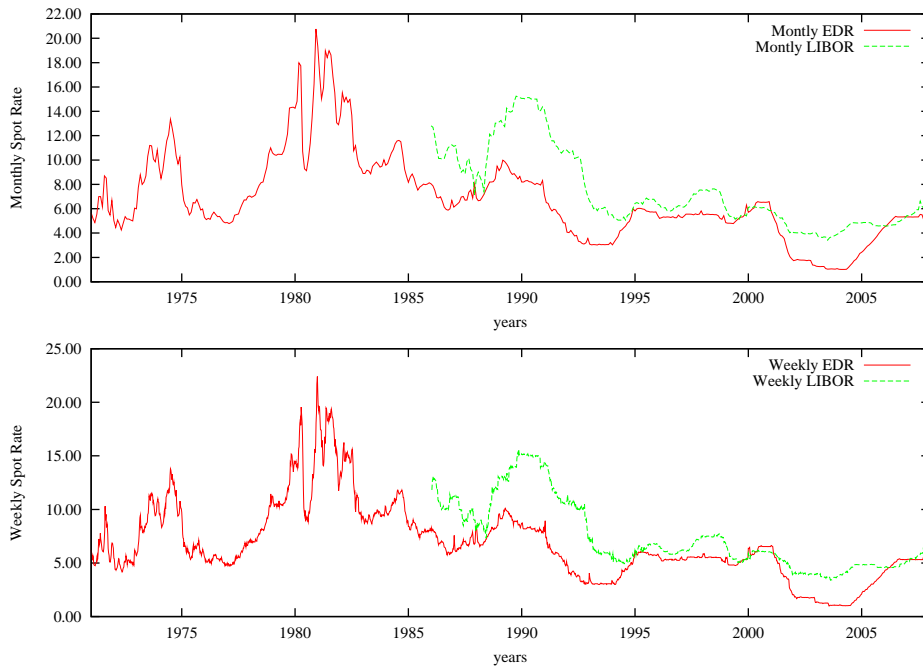


Figure 1: Time Series of Monthly and Weekly Eurodollar Rate and LIBOR

the 1980 to 1982 years are characterised by substantially higher levels than the rest of sample period. For the UK short rates, the LIBORs are highest around year 1990 as a result of the monetary policy followed at the time. Both rates reached their lowest levels during the years 2004 and 2005. The overlapping part of the sample confirm that the UK short rates are often higher than their US counterparts, and the evolution of the two rates shows strong correlations between the two rates.

5.2 Modeling Univariate Interest Rates

For each of the four data series (monthly EDR, monthly LIBOR, weekly EDR and weekly LIBOR), a total of five alternative parametric models are considered in this application. In addition to the two new models proposed in the paper (OU-CEV and CIR-CEV), we also include the OU model, the CIR model, the Ahn and Gao (1999) (AG) model. All of the five models considered here allow for mean reversion in the process and have closed form likelihood functions. We estimate all five models using maximum likelihood. Results based on both monthly and weekly observations for the two rates are obtained. *Tables 3* and *4* report the ML estimates of the parameters together with the AIC and BIC values for each model. The standard error of each parameter estimate is given in the parenthesis under each estimate.

Table 3: Maximum Likelihood Estimates of Different Interest Rate Models
(Monthly data)

		OU	CIR	AG	OU-CEV	CIR-CEV
EDR	a_1	-0.259 (0.120)	-0.148 (0.090)	-0.105 (0.076)	-0.090 (0.070)	-0.100 (0.074)
	a_2	1.745 (0.916)	0.992 (0.509)	0.023 (0.013)	0.355 (0.308)	0.389 (0.382)
	b	2.555 (0.087)	0.782 (0.026)	0.117 (0.004)	0.193 (0.017)	0.193 (0.017)
	γ	0.0	0.5	1.5	1.184 (0.047)	1.186 (0.047)
	AIC	983.82	718.99	599.88	554.35	554.13
	BIC	996.11	731.28	612.17	570.73	570.51
	LIBOR	a_1	-0.142 (0.087)	-0.174 (0.091)	-0.098 (0.081)	-0.124 (0.083)
a_2		0.740 (0.707)	0.981 (0.625)	0.020 (0.012)	0.086 (0.066)	0.012 (0.014)
b		1.315 (0.058)	0.442 (0.019)	0.058 (0.003)	0.051 (0.013)	0.051 (0.013)
γ		0.0	0.5	1.5	1.565* (0.132)	1.565* (0.133)
AIC		239.90	170.95	109.54	111.08	111.31
BIC		250.63	181.67	120.27	125.38	125.61

Only cases where the hypothesis $\gamma = 1.5$ cannot be rejected. In all other cases, $\gamma = 0, 0.5,$ or 1.5 are all rejected.

We begin by analysing the relative performances of the three existing models. Firstly, we find that the CIR model outperforms the OU substantially in terms of both AIC and BIC. This is not surprising as the OU process is indeed the simplest specification which assumes a constant diffusion function. The improved performance from the CIR process is due to the introduction of a non-linear diffusion function which is in proportion to the square-root of the value of the state variable. Both OU and CIR model assume a simple linear drift function. Secondly, we find that the AG model, which is the inverse of the CIR process, outperforms both OU and CIR models substantially by the same criteria. Since the three models have the same number of parameters, this is equivalent to comparing the likelihood of the models. The AG model has non-linearity in both the drift and diffusion function. Ahn and Gao (1999) provide a detailed description of the shape of the drift and diffusion functions of their model. It is claimed that the diffusion function given by $br_t^{1.5}$ is of special interest in the literature. This specification of the diffusion is the same as that estimated parametrically by Chan et al. (1992), and it also reflects the non-parametric estimation result of Ait-Sahalia (1996b) and Stanton (1997). It is claimed that a rapidly

Table 4: Maximum Likelihood Estimates of Different Interest Rate Models
(Weekly data)

		OU	CIR	AG	OU-CEV	CIR-CEV
EDR	a_1	-0.197 (0.104)	-0.120 (0.080)	-0.080 (0.067)	-0.070 (0.062)	-0.079 (0.066)
	a_2	1.301 (0.796)	0.782 (0.452)	0.018 (0.012)	0.226 (0.200)	0.206 (0.172)
	b	2.245 (0.036)	0.698 (0.011)	0.104 (0.002)	0.164 (0.007)	0.163 (0.007)
	γ	0.0	0.5	1.5	1.218 (0.024)	1.219 (0.024)
	AIC	971.78	-131.20	-693.36	-841.98	-842.18
	BIC	988.47	-114.51	-676.67	-819.72	-819.92
	LIBOR	a_1	-0.115 (0.083)	-0.147 (0.087)	-0.090 (0.080)	-0.113 (0.081)
a_2		0.590 (0.674)	0.829 (0.596)	0.018 (0.012)	0.198 (0.143)	0.075 (0.067)
b		1.257 (0.026)	0.423 (0.009)	0.057 (0.001)	0.079 (0.009)	0.079 (0.009)
γ		0.0	0.5	1.5	1.331 (0.060)	1.330 (0.060)
AIC		-748.07	-1047.58	-1234.46	-1240.55	-1240.39
BIC		-732.94	-1032.44	-1219.33	-1220.37	-1220.21

Only cases where the hypothesis $\gamma = 1.5$ cannot be rejected. In all other cases, $\gamma = 0, 0.5, \text{ or } 1.5$ are all rejected.

increasing diffusion function would generate a dense population of large values, which can explain the slowly decaying unconditional distribution of the interest rate documented in Ait-Sahalia (1996b). It is also documented that the shape of non-linear drift is consistent with the empirical findings of Ait-Sahalia (1996b) and Stanton (1997). Both suggest that the linearity of the drift imposed in the affine-class models or Chan et al. (1992) appears to be the main source of misspecification. Our empirical results seem to support their claims, and the superiority of the AG model over both OU and CIR is found for both EDR and LIBOR rates for both sampling frequency considered in this paper.

We now turn our attention to the two new models proposed here. Firstly we find that the goodness of fit of the OU-CEV model in terms of AIC and BIC, or equivalently the likelihood is very close to that of the CIR-CEV model. It is not surprising as it can be shown that the conditional distribution for the OU-CEV is a transformation of a distribution that is akin to a non-central χ^2 with one degree of freedom, whereas the conditional distribution of the CIR-CEV is the same transformation of an exact non-central χ^2 with fractional degree of freedom which is greater than unity. Nevertheless,

the two models do not nest each other. In addition, the two non-centrality parameters of the two distribution are generally different. Secondly, the estimated diffusion functions in the two models are almost identical, including the standard errors for the two parameters. The estimates of a_1 in the two models are also close but different in principle. The most prominent difference is in the estimation of a_2 . This is actually expected. We can see that even in the simple OU and CIR models the estimates of the two parameters are quite different due to different specifications of the diffusion. Besides, the interpretations of the two parameters are different after the transformations. These findings are once again similar for both US and UK rates and invariant to sampling frequencies.

An important feature of the newly proposed OU-CEV and CIR-CEV models is that the basic OU and CIR as well as the AG models are nested within them. Specifically, the OU process is a special case of the OU-CEV when the transformation parameter $\gamma = 1$ and the CIR and AG processes are special cases of the CIR-CEV process when $\gamma = 0.5$ and 1.5 respectively. For this reason, the relative simple specifications can be tested under the more general framework proposed here. Test of the specification of the simple OU model against the more general OU-CEV is a test of the null hypothesis $\gamma = 0$. Note that the hypothesis is on the boundary of the parameter space. Since the model is estimated by ML, the boundary problem in the testing of coefficient may be avoided by using a Lagrange Multiplier (LM) test based only on the restricted OU model. The same test can be used to test the specification of the simple CIR model against the more general CIR-CEV. The results confirm that both OU and CIR are strongly rejected across all cases. The AG specification corresponds to $\gamma = 1.5$ in the CIR-CEV model. A simple likelihood ratio test can be used. The AG specification is strongly rejected at 1% significance level in three out of the four data series. The only case where the AG model is not rejected is the monthly LIBOR case which yields a p -value of 0.627. The AIC and BIC criteria also suggest the AG model is preferred to all other models in the monthly LIBOR case, whereas in all the rest of cases the more general models are preferred. Specifically, the CIR-CEV model is preferred for EDR data at both frequencies and the OU-CEV is preferred for weekly LIBOR rates. These results suggest that despite some evidence in favour of the relatively simple AG model, most of the time existing models that have closed form likelihood functions are too simple to describe the movement of interest rates. This is indeed one of the fundamental motivation of this study.

Both OU-CEV and CIR-CEV models assume time-invariant transformation function. It would be interesting to see whether conditional (time-varying) transformation is indeed a necessary extension. Since the fitted values of γ in the time-homogenous version of the model are all greater than *unity*, we focus on the specification given in (19). To avoid over parameterisations, a more restricted version of (19) is used in this application. We

assume that the dynamics of the conditional transformation parameter is given by

$$\gamma_t = \tilde{\Lambda} \left(\omega + \alpha\gamma_{t-1} + \beta \left(\frac{1}{q} \sum_{j=1}^q r_{t-j} \right) \right). \quad (20)$$

On one hand, using the average of lags of r_t as the only forcing variable rather than including individual lags avoids over-parametrisation of the model and alleviate difficulties in numerical optimisation. On the other hand, unreported results suggest that including more lags of γ_t did not improve the likelihood significantly. The resulting time-varying models are denoted as TV-OU-CEV(1, q) and TV-CIR-CEV(1, q), respectively. Strictly speaking, the resulting processes no longer have constant elasticity diffusion as the γ_t is time-dependent. However, in any given time period ($t-1, t$) the process has constant elasticity diffusion. The value of q is selected by optimising the resulting likelihood of the model. In practice, we estimate the model for $q = 1, 2, \dots, 12$ and chose the one that gives the highest likelihood. This is equivalent to using AIC or BIC, as the number of parameters is always the same.

The results from time-varying models are presented in *Table 5*. Once again, in terms of likelihood the goodness of fit of the time-varying model based on OU-CEV is very similar to that of the model based on CIR-CEV. This is expected because in any given time period from $t-1$ to t the TV-OU-CEV process is simply an OU-CEV and the TV-CIR-CEV is a CIR-CEV. The latter two are shown to have provided similar goodness of fit to the data in the time-homogenous cases. We found $q = 1$ to be optimal for both monthly and weekly observations of the EDR data, and the optimal choices for monthly and weekly LIBOR rates are $q = 5$ and $q = 8$, respectively. The introduction of the conditional transformation has clearly improved the goodness of fit to the data as the values of AIC and BIC have decreased despite the increased number of parameters. However, a formal test is required to confirm the existence of time-varying effects in the transformation parameter. We therefore consider a likelihood ratio test of the restricted time-homogenous model we considered earlier against the unrestricted time-varying model estimated here. Under the null hypothesis of no time-varying effect, which means $\alpha = \beta = 0$, the likelihood ratio test statistic has an asymptotic χ^2 distribution with two degrees of freedom. The test that there is no time-varying effect is strongly rejected as the p -values for all of the eight reported cases are less than 0.0001, especially for the weekly data the p -values of these tests are all practically zero, showing an overwhelming evidence of time-varying effect in the transformation parameter of the proposed models. The use of conditional transformation is therefore justified.

Table 5: ML Estimates of Time-varying Transformation Models

	Monthly		Weekly	
	TV-OU-CEV	TV-CIR-CEV	TV-OU-CEV	TV-CIR-CEV
EDR	$(p = 1)$	$(p = 1)$	$(p = 1)$	$(p = 1)$
a_1	-0.100 (0.074)	-0.114 (0.079)	-0.091 (0.070)	-0.107 (0.076)
a_2	-0.398 (0.295)	0.441 (0.300)	-0.282 (0.218)	0.258 (0.178)
b	0.187 (0.006)	0.186 (0.006)	0.159 (0.003)	0.159 (0.003)
ω	-2.105 (0.697)	-2.087 (0.693)	3.522 (0.468)	3.520 (0.465)
α	3.184 (0.592)	3.158 (0.587)	-1.667 (0.383)	-1.666 (0.380)
β	0.004 (0.001)	0.004 (0.001)	0.003 (0.000)	0.003 (0.000)
AIC	526.90	526.67*	-930.19	-930.49*
BIC	551.48	551.25*	-896.80	-897.10*
LIBOR	$(p = 5)$	$(p = 5)$	$(p = 8)$	$(p = 8)$
a_1	-0.228 (0.141)	-0.223 (0.146)	-0.208 (0.122)	-0.198 (0.123)
a_2	-0.124 (0.072)	0.017 (0.009)	-0.336 (0.190)	0.131 (0.074)
b	0.046 (0.002)	0.046 (0.002)	0.077 (0.002)	0.077 (0.002)
ω	-1.768 (0.310)	-1.766 (0.311)	1.924 (1.566)	1.903 (1.715)
α	1.288 (0.201)	1.286 (0.202)	-0.662 (1.154)	-0.645 (1.264)
β	0.025 (0.006)	0.026 (0.006)	0.006 (0.003)	0.006 (0.003)
AIC	95.06	95.00*	-1264.98*	-1264.95
BIC	116.51	116.46*	-1234.71*	-1234.67

5.3 Bivariate Modeling through Copulas

In addition to greatly simplifying the estimation of the model by breaking the estimation problem into smaller problems, the multi-stage estimation procedure allows us to deal with situations where the time series of variables under consideration have different lengths, as is the case in our data. Patton (2006b) proposed to use a multi-stage maximum likelihood estimator (MSMLE) based on all available data rather than the usual one-stage maximum likelihood estimator (1SMLE) based only on the overlapping part of the data. He provided conditions under which the MSMLE is no less

asymptotically efficient than the 1SMLE and examined the small sample efficiency of the estimators via simulations.

In order to choose the best fitting copula model, we attempted a number of widely used parametric copula specifications including nine constant copulas and three time-varying copulas⁶. Our results (not reported here for economy of space) show that the time-varying SJC copula, proposed by Patton (2006a), is preferred to all the rest of the models in terms of likelihood and two information criteria (AIC and BIC). We therefore chose to base our empirical analysis on the time-varying SJC copula. The details of the time-varying SJC copula and statistical inferences of the model are discussed below.

The SJC copula is a modification of the Joe-Clayton (JC) copula (the “BB7” copula of Joe (1997)).

$$\begin{aligned} & C_{JC}(u, v | \tau^U, \tau^L) \\ &= 1 - \left(1 - \left\{ [1 - (1 - u)^\kappa]^{-\gamma} + [1 - (1 - v)^\kappa]^{-\gamma} - 1 \right\}^{-1/\gamma} \right)^{1/\kappa}, \end{aligned}$$

where $\kappa = 1 / \log_2(2 - \tau^U)$, $\gamma = -1 / \log_2(\tau^L)$, and $\tau^U \in (0, 1)$, $\tau^L \in (0, 1)$ are measures of the upper and lower tail dependencies. For a given copula function C , the upper and lower tail dependencies are defined as

$$\begin{aligned} \tau^U &= \lim_{\delta \rightarrow 1} (1 - 2\delta + C(\delta, \delta)) / (1 - \delta) \text{ and} \\ \tau^L &= \lim_{\varepsilon \rightarrow 0} C(\varepsilon, \varepsilon) / \varepsilon, \end{aligned}$$

respectively. Tail dependence captures the behavior of the random variables during extreme events. A major drawback of the JC copula is that it is always asymmetric even when the two tail dependence measures are equal, i.e., $\tau^U = \tau^L$. A more desirable model would have the tail dependence measures completely determining the presence or absence of asymmetry. For this reason, Patton (2006a) advocates the use of the SJC copula given by:

$$\begin{aligned} & C_{SJC}(u, v | \tau^U, \tau^L) \\ &= \frac{1}{2} [C_{JC}(u, v | \tau^U, \tau^L) + C_{JC}(1 - u, 1 - v | \tau^L, \tau^U) + u + v - 1]. \end{aligned}$$

The SJC copula is by construction symmetric when $\tau^U = \tau^L$ and remains asymmetric otherwise. Therefore the presence or absence of asymmetry of

⁶The nine constant copula models are Nomal copula, Clayton copula, Rotated Clayton copula, Plackett copula, Frank copula, Gumbel copula, Rotated Gumbel copula, Student’s t copula, Symmetrised Joe-Clayton (SJC) copula. The three time varying copula models are Time-varying normal Copula, Time-varying rotated Gumbel copula, Time-varying SJC copula.

the copula is completely determined by the tail dependencies. The fact that the SJC copula nests symmetry as a special case makes it a more flexible specification.

The concept of conditional (time-varying) copula was formalised in Theorem 1 of Patton (2006a). Basically, it states that if $F_{X|W}(\cdot|w)$ and $F_{Y|W}(\cdot|w)$ are the conditional distributions of $X|W = w$ and $Y|W = w$, respectively, and $C(\cdot|w)$ is a conditional copula that is measurable in w , then $F_{XY|W}(x, y|w) = C(F_{X|W}(\cdot|w), F_{Y|W}(\cdot|w)|w)$ is a conditional bivariate distribution function with conditional marginal distributions $F_{X|W}(\cdot|w)$ and $F_{Y|W}(\cdot|w)$. Patton (2006a) proposed to use the following specification for the conditional SJC copula function.

$$\begin{aligned}\tau_t^U &= \Lambda \left(\omega_U + \alpha_U \tau_{t-1}^U + \beta_U \left[\frac{1}{p} \sum_{i=1}^p |u_{t-i} - v_{t-i}| \right] \right) \\ \tau_t^L &= \Lambda \left(\omega_L + \alpha_L \tau_{t-1}^L + \beta_L \left[\frac{1}{p} \sum_{i=1}^p |u_{t-i} - v_{t-i}| \right] \right),\end{aligned}$$

where $\Lambda(\cdot)$ is again the *logistic* transformation to keep τ_t^U and τ_t^L in the range $(0, 1)$. The dynamics of the tail dependence parameters are akin to a restricted ARMA(1, q) process. Patton (2006a) chooses to use a fixed value of $p = 10$. In our study, however, we estimate the time-varying SJC copula with a number of different values of p up to $p = 12$ and choose the best fitting model in terms of the likelihood. This is equivalent to using AIC and BIC as the number of parameters in the model remains the same.

5.4 Copula results

For space of economy, we estimate our model only for the weekly data series. We therefore report the empirical results employing 1148 observations in the overlapping series. The estimation results from the conditional copula model are presented in *Table 6*. We note from the estimated unrestricted time-varying SJC model that the autoregressive parameter in both conditional tail dependence specifications of the SJC copula model, α_U and α_L , are not significantly different from zero. In an attempt to see whether both autoregressive terms can be discarded from the specification, we re-estimate the model imposing the joint restriction $\alpha_U = \alpha_L = 0$. This makes this model for the conditional tail dependence parameter resembles an ARCH model. A simple likelihood ratio test on this joint restriction yields a statistic 1.5214. According to a chi-square distribution with 2 degrees of freedoms, this corresponds to a p -value of 0.4673. Therefore, this joint restriction cannot be rejected and we proceed to estimate the model without the autoregressive terms.

Results for the restricted model are also provided in *Table 5*. Testing for the significance of time variation in such a model can be achieved via

Table 6: ML Estimates of Time-varying SJC Copula

	Time-varying SJC copula model ($p = q = 8$)	
	Unrestricted model	Restricted model
ω_U	3.646 (1.554)	4.464 (0.540)
α_U	0.628 (1.303)	
β_U	-35.565 (9.030)	-40.197 (7.208)
ω_L	3.126 (1.282)	1.845 (1.261)
α_L	-2.512 (1.838)	
β_L	-31.991 (8.723)	-26.447 (8.511)
<i>AIC</i>	-123.96	-126.44
<i>BIC</i>	-93.70	-106.26

a similar likelihood ratio test on the joint restriction that $\beta_U = \beta_L = 0$ in the restricted model. Following such a procedure yields a p -value that is practically zero, indicating strong rejection of the hypothesis of no time variation in the conditional copula.

In Figure 2 we plot the estimated conditional upper and lower tail dependencies from the time-varying SJC copula model. We can see substantial time variation in tail dependencies, ranging from near zero to 0.957 within the sample period. This indicates that the probability of these joint extreme movements can range from near zero to about 96 per cent. When compared with the original plot of the over-lapping part of the two interest rate series, the evolution of the conditional tail dependencies appear to coincide with that of interest rates themselves. That is, the tail dependencies tend to be higher when the interest rates are relatively high, and lower in the opposite situation. In particular, the tail dependencies are lowest when the two rates are at their local bottoms.

We also note that the plots of the two tail dependencies are similar in patterns with difference only in their scales. As we used the same forcing variable in the evolution equations for both upper and lower dependence, we can formally test for the significance of asymmetry in the conditional copula by testing whether the parameters of the upper tail dependence coefficient equal the parameters of the lower tail dependence coefficient in the restricted model. The p -value for such a likelihood ratio test is 0.005, indicating a rejection of symmetry in the conditional tail dependencies of the distributions of the two interest rates series.

In Figure 3 we plot the time path of the conditional linear correlation

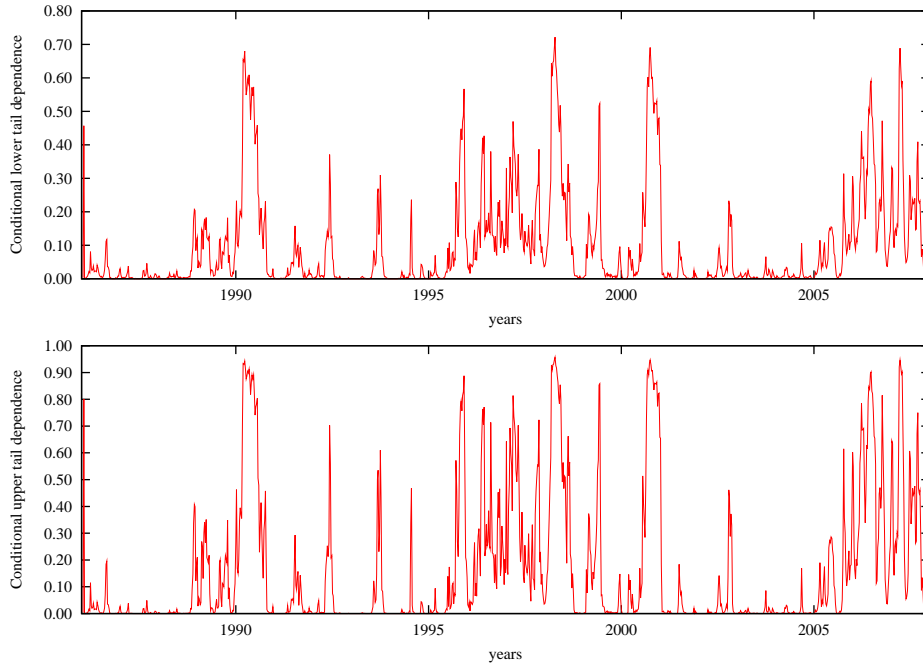


Figure 2: Conditional Upper and Lower Tail Dependencies in the SJC Copula

implied by the time-varying SJC copula model. These conditional correlations are obtained via simulation based on the estimated copula model and the two marginal models estimated earlier. Specifically, at each time period t we generate 10000 pairs of bivariate uniform random variables (F_X, F_Y) from the conditional SJC copula model. These pairs of uniformly distributed (F_X, F_Y) are transformed to random variables of interest (X, Y) according to the conditional marginal distributions implied by TV-CIR-CEV(1,1) and TV-OU-CEV(1,8), respectively. The conditional linear correlation is obtained based on the 10000 simulated (X, Y) pairs. Not surprisingly, the conditional correlations also reveal substantial time variation in the linear dependence between the two rates, ranging from 0.023 zero to 0.959 within the sample period. Similar relationship between the level of dependence and the level of interest rates is also found. The evolution of the conditional linear correlation appear to coincide with that of interest rates themselves. The correlation tends to be stronger when the rates are comparatively high, and weaker when they are relatively low.

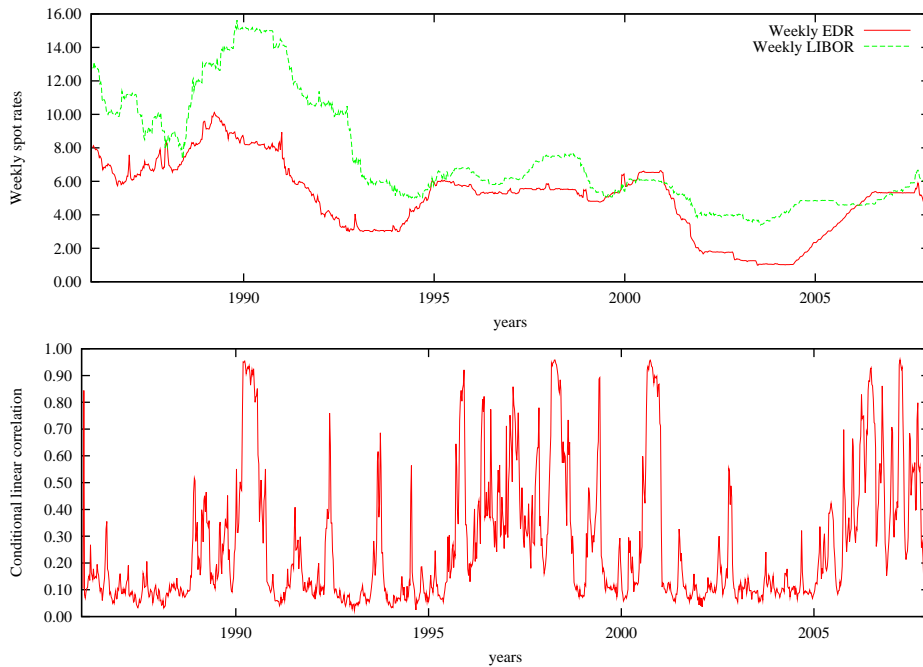


Figure 3: Overlapping Series and Conditional Linear Correlations in the Time-varying SJC Copula

6 Conclusion

In this paper, we develop a copula-based non-linear multivariate interest rates models that account simultaneously for observed non-linearities and correlation across short-term interest rates. The dynamics of the marginal processes are governed by a special type of SDEs, called reducible SDEs. The use of reducible SDEs for modeling financial variables, such as the short term interest rates, has a number of advantages. These SDEs usually lead to exact discretisation and closed form transition density functions. As a result, maximum likelihood can be easily implemented as a major tool for statistical analysis where various types of likelihood-based inferences can be used. These closed form conditional distributions functions are useful tools for statistical analyses based on probability integral transforms (PIT), such as the copula based multivariate modeling. Moreover, it follows straightforwardly from a mapping theorem that the discrete random sequence generated by the non-linear SDEs would also have stationarity and ergodicity properties under minor conditions if the basic processes are stationary and ergodic. These are useful properties for asymptotic analysis.

We consider SDEs that can be reduced to either an OU process or a CIR process, since these two processes have all the desired statistical properties

mentioned above and are widely used in Finance. We focused our attention on a couple of CEV models, the OU-CEV and CIR-CEV models, respectively. We showed that even these relatively simple specifications encompass most existing parametric models that have closed form likelihood functions. These include OU, CIR as well as the Ahn and Gao (1999) model. The statistical properties of the two processes are investigated. In particular, we provide expressions of both conditional and unconditional moments of the two processes. These expressions make the method of moments based inference convenient in case the users intend to impose conditions only on a number of moments instead of the whole distribution of the model. In our empirical studies of monthly and weekly US and UK short term interest rates, we found that simple parametric models like OU and CIR are strongly rejected by the data under their more general CEV frameworks. The AG model is also rejected by all but the monthly LIBOR data, and the only exception is likely to be due to insufficient number of observations. Hence, our new models outperform, in most cases, existing parametric models endowed with closed form likelihood functions. To generate more flexible dynamics, we extended our theory to allow for conditioning variables in the transformation functions. We found that in all four cases the time-varying effects of the transformation parameter are significant.

The dependence of the US and UK short rates were studied via a conditional copula. We found that the time-varying effect in the conditional SJC copula is significant. Also significant is the asymmetry in the tail dependence implied by the copula. From the fitted tail dependence coefficients, we found that the evolution of the conditional tail dependencies appear to coincide with that of interest rates themselves. That is, the tail dependencies tend to be higher when the interest rates are relatively high, and lower in the opposite situation. Similar relationship is also found in the conditional linear correlation coefficients implied by the conditional copula.

References

- [1] Abadir, K. M. (1999). An Introduction to Hypergeometric Functions for Economists. *Econometric Reviews*, 18(3), 287–330.
- [2] Abadir, K. M., and Rockinger, M. (2003). Density Functionals, with An Option-pricing Application. *Econometric Theory*, 19, 778–811.
- [3] Ahn, D. H., and Gao, B. (1999). A Parametric Non-linear Model of Term Structure Dynamics. *Review of Financial Studies*, 12, 721–762.
- [4] Aït-Sahalia, Y. (1996a). Nonparametric Pricing of Interest Rate Derivatives. *Econometrica*, 64, 527–560.

- [5] Aït-Sahalia, Y. (1996b). Testing Continuous-Time Models of the Spot Interest Rate. *Review of Financial Studies*, 9, 385-426.
- [6] Aït-Sahalia, Y. (1999). Transition Densities for Interest Rate and Other Non-linear Diffusions. *The Journal of Finance*, 54(4), 1361-1395.
- [7] Aït-Sahalia, Y. (2002). Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-form Approximation Approach. *Econometrica*, 70, 223-262.
- [8] Aït-Sahalia, Y. (2008). Closed-Form Likelihood Expansions for Multivariate Diffusions. *Annals of Statistics*, 36, 906-937.
- [9] Black, F., and Scholes, M., (1973). The Pricing of Options and Corporate Liabilities, *Journal of Political Economy*, 81, 637-654.
- [10] Brennan, M. and Schwartz, E., (1979). A Continuous Time Approach to the Pricing of Bonds. *Journal of Banking and Finance*, 3, 133-155.
- [11] Bu, R., and Hadri, K., (2007). Estimating Option Implied Risk-neutral Densities using Spline and Hypergeometric Functions. *Econometrics Journal*, 10, 216-244.
- [12] Chan, K. C., Karolyi, A., Longstaff, F. and Sanders, A. (1992). An Empirical Comparison of Alternative Models of the Short-Term Interest Rate. *Journal of Finance*, 47, 1209-1227.
- [13] Cherubini, U., Luciano, E., and Vecchiato, W., (2004). *Copula Methods in Finance*. John Wiley&Sons, England.
- [14] Conley, T. L., Hansen, L., Luttmer, E. and Scheinkman, J. (1997). Short Term Interest Rates as Subordinated Diffusion. *Review of Financial Studies*, 10, 525-578.
- [15] Constantinides, G. M., and Ingersoll, J., (1984). Optimal Bond Trading with Personal Taxes. *Journal of Financial Economics*, 13, 299-335.
- [16] Constantinides, G. M., (1992). A Theory of the Nominal Term Structure of Interest Rates. *Review of Financial Studies* 5, 531-522.
- [17] Courtadon, G., (1982). The Pricing of Options on Default-free Bonds. *Journal of Financial and Quantitative Analysis*, 17, 75-100.
- [18] Cox, J., (1975). Notes on Option Pricing I: Constant Elasticity of Variance Diffusions. Working paper, Stanford University.
- [19] Cox, J., Ingersoll, J. and Ross, S., (1980). An Analysis of Variable Rate Loan Contracts. *Journal of Finance* 35, 389-403.
- [20] Cox, J., Ingersoll, J. and Ross, S., (1985). In Intertemporal General Equilibrium Model of Asset Prices. *Econometrica*, 53, 363-384.
- [21] Cox, J., and Ross, S., (1976). The Valuation of Options for Alternative Stochastic Processes. *Journal of Financial Economics*, 3, 145-166.

- [22] Dothan, U., (1978). On the Term Structure of Interest Rates. *Journal of Financial Economics*, 6, 59-69.
- [23] Duffee, D., and Kan, R., (1996). A Yield Factor Model of Interest Rates. *Mathematical Finance*, 6, 379-406.
- [24] Durham, G., and Gallant, R., (2002). Numerical Techniques for Maximum Likelihood Estimation of Continuous-Time Diffusion Processes. *Journal of Business and Economic Statistics*. 20(3) 297-338.
- [25] Elerian, O., Chib, S., and Shephard, N. (2001). Likelihood Inference for Discretely Observed Non-Linear Diffusions. *Econometrica*, 69, 959–993.
- [26] Embrechts, P., Höing, A., and Juri, A., (2003). Using Copula to Bound the Value-at-Risk for Functions of Dependent Risks. *Finance & Stochastics*, 7, 145-167.
- [27] Gallant, R. and Tauchen, G. (1998). Reprojecting Partially Observed Systems with Application to Interest Rate Diffusions. *Journal of American Statistical Association*, 93, 10-24.
- [28] Hamilton, J. (1989). A New Approach to Economic Analysis of Non-stationary Time Series. *Econometrica*, 57, 357-384.
- [29] Joe, H., (1997). *Multivariate Models and Dependence Concepts*. Chapman & Hall, London.
- [30] Junker, M., Szimayer, Alexander., and Wagner, N., (2006). Non-linear Term Structure Dependence: Copula Functions, Empirics, and Risk Implications. *Journal of Banking and Finance*, 30(4), 1171-1199.
- [31] Karlin, S., and Taylor, H. (2001). *A Second Course in Stochastic Processes*. Academic Press, New York.
- [32] Kessler, M. (1997). Estimation of an Ergodic Diffusion From Discrete Observations. *Scandinavian Journal of Statistics*, 24, 211–229.
- [33] Kloeden, P., and Platen, E. (1992). *Numerical Solution of Stochastic Differential Equations, Applications of Mathematics*. New York: Springer-Verlag.
- [34] Lo, A. W., (1988). Maximum Likelihood Estimation of Generalized Ito Processes with Discretely Sampled Data. *Econometric Theory*, 4, 231–247.
- [35] Marsh, T., and Rosenfeld, E., (1983). Stochastic Processes for Interest Rates and Equilibrium Bond Prices. *Journal of Finance*, 38, 635-646.
- [36] Merton, R. C., (1973). Theory of Rational Option Pricing. *Bell Journal of Economics and Management Science*, 4, 141-183.
- [37] Nelsen, R. B., (1999). *An Introduction to Copulas*. Springer-Verlag, New York.

- [38] Ozun, A., and Cifter, Atilla., (2007). Portfolio Value-At-Risk with Time-Varying Copula: Evidence from Latin America. *Journal of Applied Science*, 7(14), 1916-1923.
- [39] Patton, A., (2006a). Modelling Asymmetric Exchange Rate Dependence. *International Economic Review*, 47(2), 527-556.
- [40] Patton, A., (2006b). Estimation of Multivariate Models for Time Series of Possibly Different Lengths. *Journal of Applied Econometrics*, 21, 147-173.
- [41] Pedersen, A. R., (1995). A New Approach to Maximum-likelihood Estimation for Stochastic Differential Equations Based on Discrete Observations. *Scandinavian Journal of Statistics*, 22, 55–71.
- [42] Pfann, G., Schotman, P., and Tschernig, R., (1996). Non-linear Interest Rate Dynamics and Implications for the Term Structure. *Journal of Econometrics*, 74, 149-176.
- [43] Brandt, M., and Santa-Clara, P., (2002). Simulated Likelihood Estimation of Diffusions with An Application to Exchange Rate Dynamics in Incomplete Markets. *Journal of Financial Econometrics*, 63, 161-210.
- [44] Sklar, A., (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l’Institut Statistique de l’Université de Paris*, 8, 229–231.
- [45] Shoji, I., and Ozaki, T. (1998). Estimation for Non-linear Stochastic Differential Equations by a Local Linearization Method. *Stochastic Analysis and Applications*, 16, 733–752.
- [46] Stanton, R., (1997). A Nonparametric Model of Term Structure Dynamics and the Market Price of Interest Rate Risk. *Journal of Finance*, 52, 1973-2002.
- [47] Vasicek, O., (1977). An Equilibrium Characterization of the Term Structure, *Journal of Financial Economics*, 5, 177–188.
- [48] White, H., (2001). *Asymptotic Theory for Econometricians*. Academic Press, San Diego.

7 Appendix

This appendix provides an analysis of the distributions of r_t implied by the CIR-CEV process and the OU-CEV process defined in Section 3.

7.1 CIR-CEV Process

7.1.1 Stationarity

We examine, with reference to Assumption (i)-(iii) in Ait-Sahalia (1999b), the behaviour of the *speed measure* S and *scale measure* M near both boundaries⁷. When $\gamma > 1$, we note that near zero

$$\mu(u, \phi) / \sigma^2(u, \phi) \propto [a_1 / (2 - 2\gamma) b^2] u^{2\gamma-1}$$

so the speed and scale density functions

$$s(v, \phi) \propto \exp \left\{ \left[-a_1 / 2b^2 (\gamma - 1)^2 \right] v^{2-2\gamma} \right\} \quad (21)$$

$$m(v, \phi) \propto (1 / v^{2\gamma}) \exp \left\{ \left[a_1 / 2b^2 (\gamma - 1)^2 \right] v^{2-2\gamma} \right\} \quad (22)$$

Therefore, $a_1 < 0$ guarantees the divergence of S and the convergence of M . Near infinity

$$\mu(u, \phi) / \sigma^2(u, \phi) \propto [2a_2(1 - \gamma) + b^2(2\gamma - 1) / 2] / b^2 u$$

so

$$s(v, \phi) \propto v^{-[4a_2(1-\gamma)+b^2(2\gamma-1)]/b^2} \quad (23)$$

$$m(v, \phi) \propto v^{[4a_2(1-\gamma)+b^2(2\gamma-1)]/b^2 - 2\gamma} \quad (24)$$

We require

$$4a_2 / b^2 > (2\gamma - 1) / (\gamma - 1) \quad (25)$$

Meanwhile, $a_1 < 0$ ensures $\lim_{r_t \downarrow 0} \mu(r_t, \phi) > 0$ and (25) ensures $\lim_{r_t \uparrow +\infty} \mu(r_t, \phi) < 0$.

When $\gamma < 1$, near zero we obtain the same expressions as in (23) and (24) for the speed and scale density functions except that we have $\gamma < 1$. Therefore, we require

$$4a_2 / b^2 \geq 1 / (1 - \gamma)$$

which also ensures $\lim_{r_t \downarrow 0} \mu(r_t, \phi) > 0$. Near infinity we obtain the same expressions as in (21) and (22) for the two densities. Therefore, $a_1 < 0$ will suffice and also guarantees $\lim_{r_t \uparrow +\infty} \mu(r_t, \phi) < 0$.

7.1.2 Transition Density

Since x_t follows the well-known CIR process, the transition density function for x_t given $x_{t-\Delta}$ is

$$f(x_t | x_{t-\Delta}) = ce^{-u-v} \left(\frac{v}{u} \right)^{q/2} I_q \left[2(uv)^{1/2} \right]$$

⁷See Ait-Sahalia (1999b) for definitions of speed and scale densities and speed and scale measures, and more extensive analysis regarding the conditions for a general process.

where

$$c = \frac{-2a_1}{b^2(1 - e^{a_1\Delta})}, u = cx_{t-\Delta}e^{a_1\Delta}, v = cx_t, q = \frac{2a_2}{b^2} - 1.$$

Since the Jacobian of the transformation is given by $|\partial x_t / \partial r_t| = r_t^{1-2\gamma} / (2|1 - \gamma|)$, standard transformation method of distributions yields

$$f(r_t | r_{t-\Delta}) = \frac{1}{2|1 - \gamma|} r_t^{1-2\gamma} c e^{-u-v} \left(\frac{v}{u}\right)^{q/2} I_q \left[2(uv)^{1/2}\right] \quad (26)$$

where

$$c = \frac{-2a_1}{b^2(1 - e^{a_1\Delta})}, u = c \frac{1}{4} \left[\frac{r_{t-\Delta}^{1-\gamma}}{1-\gamma}\right]^2 e^{a_1\Delta}, v = c \frac{1}{4} \left[\frac{r_t^{1-\gamma}}{1-\gamma}\right]^2, q = \frac{2a_2}{b^2} - 1.$$

7.1.3 Conditional distribution function

It follows from (26) and (13) that the conditional distribution function of r_t can be computed as

$$\begin{aligned} F(r_t | r_{t-\Delta}) &= \int_0^{r_t} \frac{1}{2} \frac{r_t^{1-2\gamma}}{(1-\gamma)} c e^{-u-v} \left(\frac{v}{u}\right)^{q/2} I_q \left[2(uv)^{1/2}\right] dr_t \\ &= \int_0^{x_t} c e^{-u-v} \left(\frac{v}{u}\right)^{q/2} I_q \left[2(uv)^{1/2}\right] dx_t \\ &= D(2cx_t; 2q + 2, 2u) \end{aligned}$$

when $\gamma < 1$, and

$$\begin{aligned} F(r_t | r_{t-\Delta}) &= - \int_0^{r_t} \frac{1}{2} \frac{r_t^{1-2\gamma}}{(1-\gamma)} c e^{-u-v} \left(\frac{v}{u}\right)^{q/2} I_q \left[2(uv)^{1/2}\right] dr_t \\ &= - \int_\infty^{x_t} c e^{-u-v} \left(\frac{v}{u}\right)^{q/2} I_q \left[2(uv)^{1/2}\right] dx_t \\ &= \int_{x_t}^\infty c e^{-u-v} \left(\frac{v}{u}\right)^{q/2} I_q \left[2(uv)^{1/2}\right] dx_t \\ &= 1 - D(2cx_t; 2q + 2, 2u) \end{aligned}$$

when $\gamma > 1$.

7.1.4 Conditional moments

To derive the conditional moments for r_t , we note from (13) that

$$\begin{aligned} r_t &= [2|1 - \gamma|]^{1-\gamma} c^{-\frac{1}{2(1-\gamma)}} v^{\frac{1}{2(1-\gamma)}} \\ dr_t &= [2|1 - \gamma|]^{1-\gamma} c^{-\frac{1}{2(1-\gamma)}} v^{\frac{1}{2(1-\gamma)}-1} [2(1 - \gamma)]^{-1} dv. \end{aligned}$$

The conditional moments for r_t can then be computed as

$$\begin{aligned}
& E[r_t^m | r_{t-\Delta}] \\
&= \int_0^\infty r_t^m f(r_t | r_{t-\Delta}) dr_t \\
&= \int_0^\infty r_t^m \frac{1}{2} \frac{r_t^{1-2\gamma}}{|1-\gamma|} c e^{-u-v} \left(\frac{v}{u}\right)^{q/2} I_q \left[2(uv)^{1/2}\right] dr_t \\
&= \int_0^\infty r_t^{1-2\gamma+m} c e^{-u-v} \left(\frac{v}{u}\right)^{q/2} I_q \left[2(uv)^{1/2}\right] [2|1-\gamma|]^{\frac{2\gamma-1}{1-\gamma}} c^{-\frac{1}{2(1-\gamma)}} v^{\frac{1}{2(1-\gamma)}-1} dv \\
&= [2|1-\gamma|]^{\frac{m}{1-\gamma}} c^{-\frac{m}{2(1-\gamma)}} \int_0^\infty v^{\frac{m}{2(1-\gamma)}} e^{-u-v} \left(\frac{v}{u}\right)^{q/2} I_q \left[2(uv)^{1/2}\right] dv.
\end{aligned}$$

To evaluate the integral, we note that

$$I_q \left[2(uv)^{1/2}\right] = (uv)^{q/2} \sum_{p=0}^\infty \frac{(uv)^p}{\Gamma(p+1)\Gamma(p+1+q)}. \quad (27)$$

Therefore, we obtain

$$\begin{aligned}
& \int_0^\infty v^{\frac{m}{2(1-\gamma)}} e^{-u-v} \left(\frac{v}{u}\right)^{q/2} I_q \left(2(uv)^{1/2}\right) dv \\
&= \int_0^\infty e^{-u-v} v^{q/2 + \frac{m}{2(1-\gamma)}} u^{-q/2} (uv)^{q/2} \sum_{p=0}^\infty \frac{(uv)^p}{\Gamma(p+1)\Gamma(p+1+q)} dv \\
&= e^{-u} \int_0^\infty \sum_{p=0}^\infty \frac{e^{-v} v^{p+q + \frac{m}{2(1-\gamma)}} u^p}{\Gamma(p+1)\Gamma(p+1+q)} dv \\
&= e^{-u} \sum_{p=0}^\infty \frac{u^p}{\Gamma(p+1)\Gamma(p+1+q)} \int_0^\infty e^{-v} v^{p+q + \frac{m}{2(1-\gamma)}} dv \\
&= e^{-u} \sum_{p=0}^\infty \frac{u^p}{\Gamma(p+1)\Gamma(p+1+q)} \Gamma\left(p+q + \frac{m}{2(1-\gamma)} + 1\right) \\
&= e^{-u} \sum_{p=0}^\infty \frac{u^p \Gamma\left(q + \frac{m}{2(1-\gamma)} + 1\right) \left(q + \frac{m}{2(1-\gamma)} + 1\right)_p}{\Gamma(1) (1)_p \Gamma(1+q) (1+q)_p} \\
&= e^{-u} \frac{\Gamma\left(q + \frac{m}{2(1-\gamma)} + 1\right)}{\Gamma(1+q)} \sum_{p=0}^\infty \frac{u^p \left(q + \frac{m}{2(1-\gamma)} + 1\right)_p}{p! (1+q)_p} \\
&= e^{-u} \frac{\Gamma\left(q + \frac{m}{2(1-\gamma)} + 1\right)}{\Gamma(1+q)} {}_1F_1\left(q + \frac{m}{2(1-\gamma)} + 1, 1+q, u\right).
\end{aligned}$$

Finally, we get

$$\begin{aligned} & E[r_t^m | r_{t-\Delta}] \\ &= [2|1-\gamma|]^{\frac{m}{1-\gamma}} c^{-\frac{m}{2(1-\gamma)}} e^{-u} \frac{\Gamma\left(q + \frac{m}{2(1-\gamma)} + 1\right)}{\Gamma(1+q)} {}_1F_1\left(q + \frac{m}{2(1-\gamma)} + 1, 1+q, u\right). \end{aligned}$$

7.1.5 Marginal density

Using the result in (27), the transitional density can be written as

$$f(r_t | r_{t-\Delta}) = \frac{1}{2} \frac{r_t^{1-2\gamma}}{|1-\gamma|} c e^{-u-v} v^q \sum_{p=0}^{\infty} \frac{(uv)^p}{\Gamma(p+1)\Gamma(p+1+q)}.$$

Note that as $\Delta \rightarrow \infty$

$$c \rightarrow -\frac{2a_1}{b^2}, u \rightarrow 0, v \rightarrow -\frac{2a_1}{b^2} x_t.$$

Consequently, the steady-state density function is

$$\begin{aligned} \pi(r_t) &= \lim_{\Delta \rightarrow \infty} f(r_t | r_{t-\Delta}) \\ &= \frac{1}{2} \frac{r_t^{1-2\gamma}}{|1-\gamma|} c e^{-v} v^q \frac{1}{\Gamma(1+q)} \\ &= \frac{1}{2} \frac{r_t^{1-2\gamma}}{|1-\gamma|} \frac{\left(-\frac{2a_1}{b^2}\right)^{\frac{2a_2}{b^2}}}{\Gamma\left(\frac{2a_2}{b^2}\right)} x_t^{\frac{2a_2}{b^2}-1} \exp\left(\frac{2a_1}{b^2} x_t\right). \end{aligned} \quad (28)$$

7.1.6 Unconditional moments

We note from (13) that

$$\begin{aligned} r_t &= \left[|1-\gamma|(4x_t)^{\frac{1}{2}}\right]^{\frac{1}{1-\gamma}} = [2|1-\gamma|]^{\frac{1}{1-\gamma}} x_t^{\frac{1}{2(1-\gamma)}} \\ dr_t &= [2|1-\gamma|]^{\frac{1}{1-\gamma}} x_t^{\frac{1}{2(1-\gamma)}-1} \frac{1}{2(1-\gamma)} dx_t. \end{aligned}$$

Then it follows from (28) that

$$\begin{aligned}
E[r_t^m] &= \int_0^\infty r_t^m \pi(r_t) dr_t \\
&= \int_0^\infty r_t^m \pi(r_t) [2|1-\gamma|]^{1-\gamma} x_t^{\frac{1}{2(1-\gamma)}-1} dx_t \\
&= [2|1-\gamma|]^{1-\gamma} \frac{\left(-\frac{2a_1}{b^2}\right)^{\frac{2a_2}{b^2}}}{\Gamma\left(\frac{2a_2}{b^2}\right)} \int_0^\infty x_t^{\frac{m-2+2\gamma}{2(1-\gamma)}+\frac{2a_2}{b^2}} \exp\left(\frac{2a_1}{b^2}x_t\right) dx_t \\
&= [2|1-\gamma|]^{1-\gamma} \frac{\left(-\frac{2a_1}{b^2}\right)^{-\frac{m}{2(1-\gamma)}}}{\Gamma\left(\frac{2a_2}{b^2}\right)} \Gamma\left(\frac{m-2+2\gamma}{2(1-\gamma)}+\frac{2a_2}{b^2}+1\right) \\
&= [2|1-\gamma|]^{1-\gamma} \left(-\frac{2a_1}{b^2}\right)^{-\frac{m}{2(1-\gamma)}} \frac{\Gamma\left(\frac{2a_2}{b^2}+\frac{m}{2(1-\gamma)}\right)}{\Gamma\left(\frac{2a_2}{b^2}\right)}.
\end{aligned}$$

7.2 OU-CEV Process

7.2.1 Transition density

Since x_t follows the well known OU process, it has a normal transition density function with mean and variance given by

$$\begin{aligned}
\mu_{ou} &= e^{a_1\Delta}x_{t-\Delta} - \frac{a_2}{a_1}(1 - e^{a_1\Delta}) \\
\sigma_{ou}^2 &= \frac{b^2(e^{2a_1\Delta} - 1)}{2a_1}.
\end{aligned}$$

Since the Jacobian of the transformation is $r_t^{-\gamma}$, the transition density function for the process r_t is given by

$$f(r_t|r_{t-\Delta}) = r_t^{-\gamma} \frac{1}{\sqrt{2\pi\sigma_{ou}^2}} \exp\left[-\frac{1}{2}\left(\frac{x_t - \mu_{ou}}{\sigma_{ou}}\right)^2\right]. \quad (29)$$

7.2.2 Conditional distribution function

For $\gamma < 1$, it follows from (10) and (29) that the conditional distribution function of r_t given $r_{t-\Delta}$ can be computed as

$$\begin{aligned}
F(r_t|r_{t-\Delta}) &= \int_0^{r_t} r_t^{-\gamma} \frac{1}{\sqrt{2\pi\sigma_{ou}^2}} \exp\left[-\frac{1}{2}\left(\frac{x_t - \mu_{ou}}{\sigma_{ou}}\right)^2\right] dr_t \\
&= \int_0^{x_t} \frac{1}{\sqrt{2\pi\sigma_{ou}^2}} \exp\left[-\frac{1}{2}\left(\frac{x_t - \mu_{ou}}{\sigma_{ou}}\right)^2\right] dx_t \\
&= \Phi(x_t) - \Phi(0).
\end{aligned}$$

For $\gamma > 1$, (11) and (29) yields

$$\begin{aligned}
F(r_t|r_{t-\Delta}) &= \int_0^{r_t} r_t^{-\gamma} \frac{1}{\sqrt{2\pi\sigma_{ou}^2}} \exp\left[-\frac{1}{2}\left(\frac{x_t - \mu_{ou}}{\sigma_{ou}}\right)^2\right] dr_t \\
&= -\int_{\infty}^{x_t} \frac{1}{\sqrt{2\pi\sigma_{ou}^2}} \exp\left[-\frac{1}{2}\left(\frac{x_t - \mu_{ou}}{\sigma_{ou}}\right)^2\right] dx_t \\
&= \int_{x_t}^{\infty} \frac{1}{\sqrt{2\pi\sigma_{ou}^2}} \exp\left[-\frac{1}{2}\left(\frac{x_t - \mu_{ou}}{\sigma_{ou}}\right)^2\right] dx_t \\
&= 1 - \Phi(x_t).
\end{aligned}$$

7.2.3 Conditional moments

For $\gamma < 1$, it follows from (10) that

$$\begin{aligned}
r_t &= [(1 - \gamma)x_t]^{\frac{1}{1-\gamma}} = (1 - \gamma)^{\frac{1}{1-\gamma}} x_t^{\frac{1}{1-\gamma}} \\
dr_t &= (1 - \gamma)^{\frac{\gamma}{1-\gamma}} x_t^{\frac{\gamma}{1-\gamma}} dx_t
\end{aligned}$$

The conditional moments for r_t can be then be written as

$$\begin{aligned}
E(r_t^m|r_{t-\Delta}) &= \int_0^{\infty} r_t^m r_t^{-\gamma} \frac{1}{\sqrt{2\pi\sigma_{ou}^2}} \exp\left[-\frac{1}{2}\left(\frac{x_t - \mu_{ou}}{\sigma_{ou}}\right)^2\right] dr_t \\
&= (1 - \gamma)^{\frac{m}{1-\gamma}} \int_0^{\infty} x_t^{\frac{m}{1-\gamma}} \frac{1}{\sqrt{2\pi\sigma_{ou}^2}} \exp\left[-\frac{1}{2}\left(\frac{x_t - \mu_{ou}}{\sigma_{ou}}\right)^2\right] dx_t
\end{aligned}$$

The integral above is by definition the *fractional moments of a truncated normal distribution*. If we let $w = \mu_{ou}/\sqrt{2}\sigma_{ou}$ and $z = x_t/\sqrt{2}\sigma_{ou}$, the conditional moments become

$$E(r_t^m|r_{t-\Delta}) = \frac{[\sqrt{2}\sigma_{ou}(1 - \gamma)]^{\frac{m}{1-\gamma}}}{\sqrt{\pi}} \int_0^{\infty} z^{\frac{m}{1-\gamma}} \exp[-(z - w)^2] dz$$

For $\gamma > 1$ with transformation being (11), similar reasoning yields

$$E(r_t^m|r_{t-\Delta}) = \frac{[\sqrt{2}\sigma_{ou}(\gamma - 1)]^{\frac{m}{1-\gamma}}}{\sqrt{\pi}} \int_0^{\infty} z^{\frac{m}{1-\gamma}} \exp[-(z - w)^2] dz$$

To evaluate the integral, we let $v = m/1 - \gamma$. Note that

$$\begin{aligned}
& \int_0^\infty z^v e^{-(z-w)^2} dz \\
&= e^{-w^2} \int_0^\infty z^v e^{-z^2} e^{2wz} dz \\
&= e^{-w^2} \int_0^\infty z^v e^{-z^2} \sum_{j=0}^\infty \frac{(2wz)^j}{j!} dz \\
&= e^{-w^2} \sum_{j=0}^\infty \frac{(2w)^j}{j!} \int_0^\infty e^{-z^2} z^{j+v} dz \\
&= \frac{1}{2} e^{-w^2} \sum_{j=0}^\infty \frac{(2w)^j}{j!} \Gamma\left(\frac{j+v+1}{2}\right)
\end{aligned}$$

Focusing on the summation, we note that by regrouping terms in terms of whether j is odd or even we obtain

$$\begin{aligned}
& \sum_{j=0}^\infty \frac{(2w)^j}{j!} \Gamma\left(\frac{j+v+1}{2}\right) \\
&= \sum_{j=0}^\infty \frac{(2w)^{2j}}{(2j)!} \Gamma\left(\frac{2j+v+1}{2}\right) + \sum_{j=0}^\infty \frac{(2w)^{(2j+1)}}{(2j+1)!} \Gamma\left(\frac{(2j+1)+v+1}{2}\right) \\
&= \sum_{j=0}^\infty \frac{(2w)^{2j}}{(2j)!} \Gamma\left(\frac{v+1}{2}\right) \left(\frac{v+1}{2}\right)_j + \sum_{j=0}^\infty \frac{(2w)^{(2j+1)}}{(2j+1)!} \Gamma\left(1+\frac{v}{2}\right) \left(1+\frac{v}{2}\right)_j \\
&= \Gamma\left(\frac{v+1}{2}\right) \sum_{j=0}^\infty \frac{(2w)^{2j}}{(2j)!} \left(\frac{v+1}{2}\right)_j + \Gamma\left(1+\frac{v}{2}\right) \sum_{j=0}^\infty \frac{(2w)^{(2j+1)}}{(2j+1)!} \left(1+\frac{v}{2}\right)_j \\
&= \Gamma\left(\frac{v+1}{2}\right) \sum_{j=0}^\infty \frac{\left(\frac{v+1}{2}\right)_j (4w^2)^j}{(2j)!} + wv\Gamma\left(\frac{v}{2}\right) \sum_{j=0}^\infty \frac{\left(1+\frac{v}{2}\right)_j (4w^2)^j}{(2j+1)!} \\
&= \Gamma\left(\frac{v+1}{2}\right) \sum_{j=0}^\infty \frac{\left(\frac{v+1}{2}\right)_j (w^2)^j}{\left(\frac{1}{2}\right)_j j!} + wv\Gamma\left(\frac{v}{2}\right) \left[\sum_{j=0}^\infty \frac{\left(1+\frac{v}{2}\right)_j (w^2)^j}{\left(\frac{3}{2}\right)_j j!} \right] \\
&= \Gamma\left(\frac{v+1}{2}\right) {}_1F_1\left(\frac{v+1}{2}, \frac{1}{2}, w^2\right) + wv\Gamma\left(\frac{v}{2}\right) {}_1F_1\left(1+\frac{v}{2}, \frac{3}{2}, w^2\right)
\end{aligned}$$

Finally, we get

$$\begin{aligned}
E(r_t^m | r_{t-\Delta}) &= \frac{[\sqrt{2}\sigma_{ou} |1-\gamma|]^{1-\gamma}}{\sqrt{\pi}} \frac{1}{2} e^{-w^2} \left[\Gamma\left(\frac{v+1}{2}\right) {}_1F_1\left(\frac{v+1}{2}, \frac{1}{2}, w^2\right) \right. \\
&\quad \left. + wv\Gamma\left(\frac{v}{2}\right) {}_1F_1\left(1+\frac{v}{2}, \frac{3}{2}, w^2\right) \right]
\end{aligned}$$

where $v = m/1 - \gamma$ and $w = \mu_{ou} / \sqrt{2}\sigma_{ou}$.

7.2.4 Marginal density and unconditional moments

To get the marginal density and unconditional moments we simply have to replace the conditional mean and variance in (16) and (17) by their corresponding limits $-a_2/a_1$ and $-b^2/2a_1$ as $\Delta \rightarrow \infty$.