

# Generalized Affine Models

Bruno Feunou\*

Nour Meddahi†

Université de Montréal and CREST

Imperial College London

November 15, 2007

## Abstract

Affine models are very popular in modeling financial time series as they allow for analytical calculation of prices of financial derivatives like treasury bonds and options. The main property of affine models is that the conditional cumulant function, defined as the logarithmic of the conditional characteristic function, is affine in the state variable. Consequently, an affine model is Markovian, like an autoregressive process, which is an empirical limitation. The paper generalizes affine models by adding in the current conditional cumulant function the past conditional cumulant function. Hence, generalized affine models are non-Markovian, such as ARMA and GARCH processes, allowing one to disentangle the short term and long-run dynamics of the process. Importantly, the new model keeps the tractability of prices of financial derivatives. This paper studies the statistical properties of the new model, derives its conditional and unconditional moments, as well as the conditional cumulant function of future aggregated values of the state variable which is critical for pricing financial derivatives. It derives the analytical formulas of the term structure of interest rates and option prices. Different estimating methods are discussed (MLE, QML, GMM, and characteristic function based estimation methods). Three empirical applications developed in companion papers are presented. The first one based on Feunou (2007) presents a no-arbitrage VARMA term structure model with macroeconomic variables and shows the empirical importance of the inclusion of the MA component. The second application based on Feunou and Meddahi (2007a) models jointly the high-frequency realized variance and the daily asset return and provides the term structure of risk measures such as the Value-at-Risk, and highlights the powerful use of generalized affine models. The third application based on Feunou, Christoffersen, Jacobs and Meddahi (2007) uses the model developed in Feunou and Meddahi (2007a) to price options theoretically and empirically.

Keywords: Affine models; cumulant function; option pricing; term structure of interest rates.

JEL codes: C12; C13; C14; C22; C51; E43; G12.

---

\*Corresponding author. Email: bruno.feunou.kamkui@umontreal.ca.

†Email: n.meddahi@imperial.ac.uk.

# 1 Introduction

Affine models are often used when one models the short term of interest rates because they lead to closed form of the bond prices and yields whatever the maturity. In addition, these yields are linear functions of the state variables, often the short term interest rate when one considers a one-factor model, which makes the pricing and the statistical inference simple. This approach has been introduced in continuous time by Vasicek (1977) where the short term interest rate is assumed to follow a Gaussian autoregressive process of order one and extended by Duffie and Kan (1997) to more non-negative models. Discrete time versions of affine models are studied theoretically in Darolles, Gourieroux, Jasiak (2006) and Gourieroux, Monfort, and Polimenis (2002) among others while several papers, including Piazzesi (2005) and Ang and Piazzesi (2003), used them to characterize the term structure of interest rates and its interaction with macroeconomic variables; see Piazzesi (2003) for a survey on affine term structure models. Likewise, several authors used the affine processes for modeling the stochastic volatility of asset returns and characterized analytically the formulas of option prices; see Heston (1993) and Duffie, Pan and Singleton (2000) in continuous time and Heston and Nandi (2001) in discrete time.

A discrete time process  $x_t$  is called affine when its conditional cumulant function, denoted  $\psi_t(u)$ , and defined as the logarithmic of the moment generating function,<sup>1</sup> i.e.,

$$\psi_t(u) \equiv \log[E[\exp(ux_{t+1}) \mid x_\tau, \tau \leq t]],$$

is given by

$$\psi_t(u) = \omega(u) + \alpha(u)x_t. \tag{1.1}$$

Any autoregressive process of order one, AR(1), with i.i.d. innovations is affine. A consequence of (1.1) is that an affine process is Markovian, which could be a limitation for modeling some financial data. It is well known that financial data, like volatility of asset returns, exhibit serial correlation that the Markov ARCH models of Engle (1982) do not describe well, which leads to the introduction of the GARCH models in Bollerslev (1986). Likewise, we do know that

---

<sup>1</sup>Instead of considering the moment generating function, one could use the characteristic function which exists for any random variable while the moment generating function does not exist for some random variables. The theory developed in this paper holds for characteristic functions. However, we decided to use the moment generating function for convenience and due to its familiarity with researchers in financial economics.

allowing for non-Markovian components in a model, like moving average (MA) components, allows one to disentangle the short-term and the long-run dynamics of the variable of interest, which could be important for some financial data like volatility of asset returns and short term of interest rates (Andersen and Lund (1997)).

Several generalizations of affine models have been introduced in order to include more memory in the basic model (1.1) and maintaining the tractability of affine models, i.e. by maintaining closed forms for the yields. Dai and Singleton (2003) and Dai, Singleton and Yang (2006) assumed that the coefficients that drive the affine model follow a Markov switching model. The authors show the empirical usefulness of this approach although filtering techniques are needed to price and estimate the model. Darolles, Gourieroux and Jasiak (2006) added lags of  $x_t$  in (1.1), i.e. they proposed an affine model of order  $p > 1$ . Monfort and Pegoraro (2007a) successfully applied this approach to the term structure of interest rates, although one could need many lags leading to the estimation of many parameters. In a different paper, Monfort and Pegoraro (2007b) combined the two approaches describe above, i.e. they added lags and assumed that some parameters are driven by a Markov switching model. Again, such a method needs filtering techniques for both pricing and estimating the model.

In this paper, we follow a more traditional approach by including MA component in the model. The following examples highlight our approach. Assume that the process  $x_t$  is an ARMA(1,1) given by

$$x_t = a + bx_{t-1} + \varepsilon_t - c\varepsilon_{t-1}, \quad \varepsilon_t \text{ i.i.d.}, \quad |b| < 1, \quad |c| < 1,$$

where the cumulant function of  $\varepsilon$  is denoted  $\psi_\varepsilon(\cdot)$ . One can show (see Section 2) that

$$\psi_t(u) = (ua + (1 - c)\psi_\varepsilon(u)) + u(b - c)x_t + c\psi_{t-1}(u),$$

which suggests the following extension of (1.1)

$$\psi_t(u) = \omega(u) + \alpha(u)x_t + \beta\psi_{t-1}(u). \tag{1.2}$$

One could view the new model as an extension of AR models to ARMA ones. Actually, it is much more than that because Eq. (1.2) implies that any power function of  $x_t$  is an ARMA process. This is the case because the conditional cumulant function of  $x_t$  is an auto-regression.

Assume that the process  $x_t$  is an VARMA(1,1) given by

$$x_t = A + Bx_{t-1} + \varepsilon_t - C\varepsilon_{t-1}, \quad \varepsilon_t \text{ i.i.d.}, \quad |B| < 1, \quad |C| < 1,$$

where the cumulant function of  $\varepsilon$  is denoted  $\psi_\varepsilon(\cdot)$ . One can show (see Section 2) that

$$\psi_t(u) = (u'A + \psi_\varepsilon(u) - \psi_\varepsilon(C'u)) + u'(B - C)x_t + \psi_{t-1}(C'u),$$

which suggests the following extension of (1.1)

$$\psi_t(u) = \omega(u) + \alpha(u)x_t + \psi_{t-1}(\theta u). \quad (1.3)$$

Our approaches have several advantages. They involve less parameters than the approach in Darolles, Gouriéroux and Jasiak (2006). The pricing and estimation procedures of the model are simpler than those of a model with Markov switching factors like Dai, Singleton and Yang (2006). Another important advantage of the approach is to allow one to disentangle the short term dynamics of  $x_t$  from its long-run ones. When one considers an affine model (1.1), the function  $\alpha(u)$  does the two jobs, which could be restrictive. We do know from the volatility literature that GARCH models allow for more persistence than ARCH models and that this is empirically important. Our empirical examples highlight this advantage.

Several dynamic term structure models with macroeconomic variables introduce latent variables in the affine state variable; see Ang and Piazzesi (2003). Such approach is often done because current values of the macroeconomic variables do not fully explain the term structure of interest. However, it is always hard to understand what exactly these latent variables are. It is well known from the time series literature that AR models with latent variables, called structural models, imply reduced form ARMA representations for the observable variables. Consequently, one could interpret the new model as a reduced form of affine models with latent factors.

We introduce a slightly more general model than (1.2) by allowing the coefficient in front of  $\psi_{t-1}(u)$  to be a function of  $u$ , i.e., we study the model defined by

$$\psi_t(u) = \omega(u) + \alpha(u)x_t + \beta(u)\psi_{t-1}(u). \quad (1.4)$$

We denote this model as the generalized affine model of type I. Similar to type I, we introduce a slightly more general model than (1.3) by allowing the argument of  $\psi_{t-1}(\cdot)$  to be a function of  $u$  which is not necessarily linear and by allowing a coefficient in front of  $\psi_{t-1}(\theta(u))$ , i.e., we study the model defined by

$$\psi_t(u) = \omega(u) + \alpha(u)x_t + \beta\psi_{t-1}(\theta(u)). \quad (1.5)$$

These extensions are theoretically important because Eq. (1.2) implies that the vector  $(x_t, m_t)$ , where  $m_t = E[x_{t+1} | x_\tau, \tau \leq t]$ , is a bivariate affine model while it is not the case for models

defined by Eq. (1.4) and (1.5). Likewise, we allow for several lags of  $x_t$  and  $\psi_{t-1}(u)$  in Eq. (1.4) and (1.5), i.e., we consider ARMA(p,q) type models.

The paper has several contributions. First of all, we study the statistical properties of the models and derived several conditional and unconditional moments and cumulants. We also derive the conditional cumulant function of the vector  $(x_{t+1}, x_{t+2}, \dots, x_{t+h})$ . This function is critical when one wants to derive analytical formulas of yields and option prices. We then derive the Treasury yields when assuming that short term of interest rate is given by (1.4) or (1.5) under the risk neutral measure or the physical measure (the latter needs the specification of the price of the risk). Likewise, we derive the formulas of options prices when assuming a stochastic volatility where the dynamics of the stochastic variance is given by Eq. (1.4).

One can use several methods to estimate to model. Sometimes one could characterize the likelihood of the model as in our empirical analysis. Otherwise, one could follow Singleton (2001) by using the characteristic function of the process  $x_t$  and the instrumental variable approach of Hansen (1982). Actually, an efficient use of the whole characteristic function leads to an efficient estimation of the parameters comparable to the maximum likelihood estimators; see Carrasco and Florens (2001, 2006) and Carrasco, Chernov, Florens, and Ghysels (2006). It would also be possible to use the conditional mean and variance of the process  $x_t$  combined with the Gaussian quasi-maximum likelihood approach to consistently estimate the parameters.

Three empirical applications developed in companion papers are presented. The first one based on Feunou (2007) presents a no-arbitrage VARMA term structure model with macroeconomic variables and shows the empirical importance of the inclusion of the MA component. The second application based on Feunou and Meddahi (2007a) models jointly the high-frequency realized variance and the daily asset return and provides the term structure of risk measures such as the Value-at-Risk, and highlights the powerful use of generalized affine models. The third application based on Feunou, Christoffersen, Jacobs and Meddahi (2007) uses the model developed in Feunou and Meddahi (2007a) to price options theoretically and empirically.

The rest of the paper is organized as follows. Section 2 provides the simple generalized affine model and provides its statistical properties. Section 3 provides the analytical formulas of the term structure of interest rates when the short term of interest rates is a generalized affine process under the physical or the risk-neutral measure. Likewise, section 3 provides the formulas of the option prices when the volatility of the stock returns is a generalized affine process under the physical or the risk-neutral measure. We discuss several estimation methods

in section 4. Section 5 provides three empirical examples while Section 6 concludes. Appendix A provides an example where the function  $\beta(\cdot)$  is not constant. The proofs of Sections 2 and 3 are provided in Appendix B.

## 2 Generalized Affine Models

This section introduces and studies the two simple models defined in the previous section.

**Definition: Generalized Affine Process.** A process  $x_t$  is called a generalized affine process of type I and order (1,1) when the conditional cumulant function of  $x_{t+1}$  given its lagged values  $x_t, x_{t-1}, \dots$ , is characterized by

$$\psi_t(u) \equiv \log E[\exp(ux_{t+1}) \mid x_\tau, \tau \leq t] = \omega(u) + \alpha(u)x_t + \beta(u)\psi_{t-1}(u). \quad (2.1)$$

A process  $x_t$  is called a generalized affine process of type II and order (1,1) when the conditional cumulant function of  $x_{t+1}$  given its lagged values  $x_t, x_{t-1}, \dots$ , is characterized by

$$\psi_t(u) \equiv \log E[\exp(ux_{t+1}) \mid x_\tau, \tau \leq t] = \omega(u) + \alpha(u)x_t + \beta\psi_{t-1}(\theta(u)). \quad (2.2)$$

To simplify the exposure, we combine the two types in the same framework and we provide the following general definition of the generalized affine model.

A process  $x_t$  is called a generalized affine process of order (1,1) when the conditional cumulant function of  $x_{t+1}$  given its lagged values  $x_t, x_{t-1}, \dots$ , is characterized by

$$\psi_t(u) \equiv \log E[\exp(ux_{t+1}) \mid x_\tau, \tau \leq t] = \omega(u) + \alpha(u)x_t + \beta(u)\psi_{t-1}(\theta(u)). \quad (2.3)$$

### 2.1 Examples

Several well know examples in the time series and financial literatures are generalized affine. Obviously, affine models correspond to the case  $\beta(u) = 0$  or  $\theta(u) = 0$ . Other examples are given below.

#### 2.1.1 Linear and Non-Linear ARMA(1,1) Models

Assume that  $x_t$  follows a linear ARMA(1,1) where the innovation process is i.i.d., i.e.

$$x_t = a + bx_{t-1} + \varepsilon_t - c\varepsilon_{t-1}, \quad \varepsilon_t \text{ i.i.d., } |b| < 1, |c| < 1,$$

where the cumulant function of  $\varepsilon$  is denoted  $\psi_\varepsilon(\cdot)$ . Denote the conditional mean of  $x_{t+1}$  by  $m_t$ , i.e.,

$$m_t = a + bx_t - c\varepsilon_t.$$

Observe that

$$m_t = a + (b - c)x_t + cm_{t-1}.$$

Hence,

$$\begin{aligned} \psi_t(u) &= \log E_t[\exp(ux_{t+1})] = um_t + \psi_\varepsilon(u) = u(a + (b - c)x_t) + \psi_\varepsilon(u) + ucm_{t-1} \\ &= u(a + (b - c)x_t) + \psi_\varepsilon(u) + c(\psi_{t-1}(u) - \psi_\varepsilon(u)) \\ &= (ua + (1 - c)\psi_\varepsilon(u)) + u(b - c)x_t + c\psi_{t-1}(u), \end{aligned}$$

i.e. any ARMA(1,1) process with i.i.d. innovations defined in (2.1.1) is a generalized affine process given in (2.1) where

$$\omega(u) = ua + (1 - c)\psi_\varepsilon(u), \quad \alpha(u) = u(b - c), \quad \beta(u) = c.$$

Let us now assume that the conditional mean of  $x_t$  is non-linear but still has an MA(1) structure, i.e.

$$x_t = f(x_{t-1}) + \varepsilon_t - c\varepsilon_{t-1}, \quad \varepsilon_t \text{ i.i.d.}, \quad |c| < 1.$$

The condition mean of  $x_{t+1}$  denoted  $m_t$  is given by

$$m_t = f(x_t) - c\varepsilon_t = f(x_t) - cx_t + cm_{t-1}.$$

Hence,

$$\begin{aligned} \psi_t(u) &= \log E_t[\exp(ux_{t+1})] = um_t + \psi_\varepsilon(u) = u(f(x_t) - cx_t) + \psi_\varepsilon(u) + ucm_{t-1} \\ &= (1 - c)\psi_\varepsilon(u) + u(f(x_t) - cx_t) + c\psi_{t-1}(u). \end{aligned}$$

Consequently, a non-linear ARMA(1) process with i.i.d. innovations is not a generalized affine process but belongs to the family defined by

$$\psi_t(u) = \omega(u) + \alpha(u, x_t) + \beta(u)\psi_{t-1}(u). \tag{2.4}$$

This family is studied in Feunou and Meddahi (2007b) and called generalized non-affine models.

### 2.1.2 Linear VARMA(1,1) Models

Assume that  $x_t$  follows a linear VARMA(1,1) where the innovation process is i.i.d., i.e.

$$x_t = A + Bx_{t-1} + \varepsilon_t - C\varepsilon_{t-1}, \quad \varepsilon_t \text{ i.i.d.}, \quad |B| < 1, \quad |C| < 1,$$

where the cumulant function of  $\varepsilon$  is denoted  $\psi_\varepsilon(\cdot)$ . Denote the conditional mean of  $x_{t+1}$  by  $m_t$ , i.e.,

$$m_t = A + Bx_t - C\varepsilon_t.$$

Observe that

$$m_t = A + (B - C)x_t + Cm_{t-1}.$$

Hence,

$$\begin{aligned} \psi_t(u) &= \log E_t[\exp(u'x_{t+1})] = u'm_t + \psi_\varepsilon(u) = u'(A + (B - C)x_t) + \psi_\varepsilon(u) + u'Cm_{t-1} \\ &= u'(A + (B - C)x_t) + \psi_\varepsilon(u) + \psi_{t-1}(C'u) - \psi_\varepsilon(C'u) \\ &= (u'A + \psi_\varepsilon(u) - \psi_\varepsilon(C'u)) + u'(B - C)x_t + \psi_{t-1}(C'u), \end{aligned}$$

i.e. any VARMA(1,1) process with i.i.d. innovations defined in (2.1.2) is a generalized affine process of type 2 given in (2.2) where

$$\omega(u) = u'A + \psi_\varepsilon(u) - \psi_\varepsilon(C'u), \quad \alpha(u) = u'(B - C), \quad \beta = 1, \quad \theta(u) = C'u$$

### 2.1.3 GARCH(1,1) Type Models

We start the analysis by considering the model introduced in Bollerslev (1986), i.e.,

$$x_t = \mu + \varepsilon_t = \mu + \sqrt{h_{t-1}}z_t, \quad z_t \text{ i.i.d. } \mathcal{N}(0, 1), \quad h_t = \omega + \alpha\varepsilon_t^2 + \beta h_{t-1},$$

with  $\alpha \geq 0, \beta \geq 0, \alpha + \beta < 1$ . By doing the same calculations as in the ARMA example, one gets

$$\psi_t(u) = \left( (1 - \beta)\mu u + \omega \frac{u^2}{2} \right) + \frac{1}{2}\alpha u^2 (x_t - \mu)^2 + \beta \psi_{t-1}(u). \quad (2.5)$$

In other words, the GARCH(1,1) is not a generalized affine process as (2.1) but a generalized non-affine process given by (2.4).

It is well known that the GARCH(1,1) does not lead to closed forms of option prices. Heston and Nandi (2000) proposed a different specification for  $h_t$  that solved the problem

where  $h_t$  is given by<sup>2</sup>

$$h_t = \omega + \alpha(z_t - \gamma\sqrt{h_{t-1}})^2 + \beta h_{t-1}.$$

Likewise, one can show that the conditional cumulant function of  $x_{t+1}$  is given by

$$\psi_t(u) = \left( u\mu(1 - \beta - \alpha\gamma^2) + \frac{u^2}{2} \left( \omega + \alpha \frac{(x_t - \mu)^2}{h_{t-1}} - 2\gamma(x_t - \mu) \right) \right) + (\beta + \alpha\gamma^2)\psi_{t-1}(u). \quad (2.6)$$

Consequently, the Heston and Nandi (2000) model is a generalized non-linear model defined by (2.4) where the function  $\alpha(x_t, u)$  depends  $x_t$  and  $h_{t-1}$ , i.e., the whole past of  $x_t$ .

Eq. (2.6) looks more non-linear than Eq. (2.5), which is puzzling given that the Heston and Nandi (2000) model leads to analytical formulas for option prices while the Bollerslev (1986) does not. As already mentioned, affine models lead to closed form of prices of derivatives. It turns out that the variance process  $h_t$  is affine when one considers the Heston and Nandi (2000) while it is not the case for the traditional GARCH model. More precisely, one has

$$\text{Heston \& Nandi : } \log E[\exp(uh_{t+1}) | h_\tau, \tau \leq t] = u\omega + \psi_{\chi^2(1)}(\alpha u) + ((\beta + \alpha\gamma^2)u - 2\gamma^2 u^2)h_t$$

$$\text{Bollerslev : } \log E[\exp(uh_{t+1}) | h_\tau, \tau \leq t] = u\omega + \psi_{\chi^2(1)}(\alpha u h_t) + \beta u h_t,$$

where  $\psi_{\chi^2(1)}(\cdot)$  denotes the cumulant function of the  $\chi^2(1)$  distribution. We will consider again the Heston and Nandi model when we will derive the option pricing formulas of generalized affine models.

#### 2.1.4 ACD(1,1) type models

Engle and Russell (1997) introduced the autoregressive conditional duration (ACD) model where the duration  $x_i$  between two consecutive trades follows the process

$$x_i = \eta_{i-1} v_i, \quad v_i \text{ i.i.d., } v_i > 0, \quad E[v_i] = 1, \quad \eta_i = \omega + \alpha x_i + \beta \eta_{i-1}.$$

If one assume that  $v_i$  follows an exponential distribution whose density function is  $f_v(v) = \exp(-v)$ , then one gets

$$\psi_i = E_i[\exp(ux_{i+1})] = \frac{1}{1 - u\eta_i}, \quad u < \frac{1}{\eta_i},$$

which is not a generalized affine model. However, it is the case for the logarithmic duration model of Bauwens and Giot (2000) defined by

$$x_i = \exp(\eta_{i-1}) v_i, \quad v_i \text{ i.i.d., } v_i > 0, \quad E[v_i] = 1, \quad \log(\eta_i) = \omega + \alpha \log(x_i) + \beta \eta_{i-1}.$$

For this model,  $\log(x_i)$  is an ARMA(1,1) and therefore a generalized affine process.

---

<sup>2</sup>There is an additional coefficient  $\gamma$  that appears in (2.1.3) which captures the leverage effect. One could easily add such term in the Bollerslev's GARCH equation.

### 2.1.5 The generalized autoregressive gamma process

The autoregressive gamma process (ARG) defined in Gouriéroux and Jasiak (2006) is built throughout the following state space representation:

$$\begin{aligned}\frac{x_{t+1}}{\mu} | U_{t+1}, I_t &\sim \gamma(\nu + U_{t+1}) \\ U_{t+1} | I_t &\sim P\left(\frac{\rho x_t}{\mu}\right)\end{aligned}$$

with  $I_t$  defined as the sigma algebra generated by  $(x_s, s \leq t)$ ,  $\gamma$  means the standard gamma distribution and  $P$  the Poisson distribution. Gouriéroux and Jasiak (2006) show that  $\psi_t(u) = \omega(u) + \alpha(u)x_t$  with  $\psi_t = \ln E_t[\exp(ux_{t+1})]$ ,  $\omega(u) = -\nu \log(1 - u\mu)$ , and  $\alpha(u) = \frac{\rho u}{1 - u\mu}$ .

We generalize the ARG process to the GARG process which is built throughout the following state space representation:

$$x_{t+1} = \beta^t \theta^t \bar{Z}_{t+1} + \sum_{j=0}^{t-1} Z_{t+1}^{(j)}$$

with  $\bar{Z}_{t+1}$ ,  $Z_{t+1}^{(j)}$  for  $j = 0, \dots, t-1$ ,  $t+1$  conditionally (conditional on  $I_t$ ) independent random variables,  $\bar{Z}_{t+1} \sim \psi_0(u)$ , and

$$\begin{aligned}\frac{Z_{t+1}^{(j)}}{\mu_j} | U_{t+1}^{(j)}, I_t &\sim IG\left(\nu_j + U_{t+1}^{(j)}\right) \\ U_{t+1}^{(j)} | I_t &\sim P\left(\frac{\rho_j x_{t-j}}{\mu_j}\right)\end{aligned}$$

where

$$\nu_j = \nu \beta^j, \quad \mu_j = \mu \theta^j, \quad \rho_j = \rho \beta^j \theta^j$$

We can show that

$$\psi_t(u) = \omega(u) + \alpha(u)x_t + \beta \psi_{t-1}(\theta u)$$

Then  $x_t$  follows a generalized affine model of type 2 given in (2.2).

## 2.2 Existence of Generalized Affine Models

### 2.2.1 The Function $\beta(\cdot)$ is Constant.

Generalized affine models are defined recursively by their conditional cumulant function in (2.3). Therefore, one needs to show that the function  $\psi_t(\cdot)$  in (2.3) is a proper cumulant function. The rest of this subsection focuses on the case where  $\beta(\cdot)$  is constant, the other case being studied in section 2.2.2.

The first important property on cumulant function is that the sum of cumulant functions is a cumulant function. Consequently, when  $\omega(u)$ ,  $\alpha(u)x_t$ , and  $\beta(u)\psi_{t-1}(\theta(u))$  are cumulant functions, the function  $\psi_t(u)$  defined in (2.3) is a cumulant function. Observe that often, as in our empirical examples,  $\omega(u) + \alpha(u)x_t$  is the cumulant function of an affine model. Therefore, the generalized affine model is well defined when  $\beta(u)\psi_{t-1}(\theta(u))$  is a cumulant function.

The second important property on cumulant functions is related to infinitely divisible random variables. A random variable  $z$  whose cumulant function is denoted  $\psi_z(u)$ , is called infinitely divisible when for any positive number  $c$ ,  $c\psi_z(u)$  is a cumulant function. Observe that a consequence of this definition is that  $c\psi_z(u)$  is the cumulant function of an infinitely divisible random variable. Such variables appear in central limit theorems; examples of infinitely divisible random variables include normal, Poisson, and Gamma random variables. The first version of Darolles et al. (2006) provided sufficient conditions such that an affine process is infinitely divisible. In particular, popular affine models in finance, i.e. the Gaussian and the square root processes are infinitely divisible.

The third important property on cumulant functions is still related to infinitely divisible random variables. For a given positive random variable  $z$  whose cumulant function is denoted  $\psi_z(u)$ , and an infinitely divisible cumulant function  $\theta(u)$ ,  $\psi_z(\theta(u))$  is a well defined cumulant function. In the characteristic function literature, this type of construction is recognized as the subordination of processes.

The second and third properties on cumulant functions are quite important for our purpose. By expanding recursively  $\psi_t(u)$  given in (2.3), one gets

$$\psi_t(u) = \sum_{i=0}^{t-1} \beta^i (\omega(\theta^{\circ i}(u)) + \alpha(\theta^{\circ i}(u))x_{t-i}) + \beta^t \psi_0(\theta^{\circ t}(u))$$

where  $\psi_0(u)$  is the unconditional cumulant function of  $x_1$  and  $\theta^{\circ i}(u)$  denotes function  $\theta(\cdot)$  compounded  $i$  times with itself. Consequently, when  $\beta > 0$ ,  $\theta(\cdot)$  a cumulant of an infinitely

divisible positive random variable (which implies that  $\omega(\theta^{\circ i}(u)) + \alpha(\theta^{\circ i}(u))x_{t-i}$  is the cumulant function of an indivisible random variable like some affine models derived in Darolles et al. (2006)),  $\beta^i(\omega(\theta^{\circ i}(u)) + \alpha(\theta^{\circ i}(u))x_{t-i})$  is a cumulant function of an infinitely divisible random variable. The definition of infinitely divisible random variables implies that the sum of infinitely divisible random variables is also an infinitely divisible random variable. Therefore,  $\sum_{i=0}^t \beta^i(\omega(\theta^{\circ i}(u)) + \alpha(\theta^{\circ i}(u))x_{t-i})$  is the cumulant function of an infinitely divisible random variable. Consequently,  $\psi_t(u)$  is the cumulant function of an infinitely divisible random variable when one assumes that this is the case for  $\psi_0(u)$ . In other words, sufficient conditions to guarantee that  $\psi_t(u)$  defined in (2.3) is a proper cumulant function are:  $\beta \geq 0$ ,  $\theta(\cdot)$  an infinitely divisible cumulant function of a positive random variable,  $\omega(u) + \alpha(u)x$  and  $\psi_0(u)$  are cumulant functions of indivisible random variables.

The previous reasoning handles the case of positive random variable. To analyze random processes which have the whole real set as support, we will treat the two types of generalized affine models differently. In the type I case, in general  $\beta$  need not to be positive for processes which have the whole real line as support (the Gaussian Vector Autoregressive Process for instance). The only needed conditions are  $|\beta| < 1$  and function  $\omega(u)$  being an infinitely divisible cumulant generating function. Indeed we can write  $\psi_t(u) = \frac{\omega(u)}{1-\beta} + \sum_{i=0}^{\infty} \alpha(u)\beta^i x_{t-i}$ . Since  $\omega(u)$  an infinitely divisible cumulant generating function and  $\frac{1}{1-\beta} > 0$  then  $\frac{\omega(u)}{1-\beta}$  is a cumulant generating function (c.f.g). By denoting  $y_{t-i} = \beta^i x_{t-i}$ , we have that for each  $i$   $\alpha(u)y_{t-i}$  is a well defined c.f.g. Using the fact that the sum of c.f.g is a c.f.g, we get that  $\psi_t(u)$  is a c.f.g.

In the case of type II, we consider the following restrictions  $\beta > 0$  and  $\theta(u) = \theta u$  (which is the case of VARMA models). Then, (2.3) is well defined for processes with the whole real line as support. Indeed if  $\omega(u)$ ,  $\alpha(u)$  and  $\psi_0(u)$  are infinitely divisible cumulant functions, we can show that  $\omega(cu)$ ,  $\alpha(cu)$  and  $\psi_0(cu)$  are infinitely divisible cumulant function whatever the sign of  $c$ . Therefore for each  $i$ ,  $\omega(\theta^i u) + \alpha(\theta^i u)x_{t-i}$  is a cumulant function of an infinitely divisible function. This implies that (2.3) is well defined.

Another question not studied here is the existence of a stationary solution of (2.1). As usual, such a question is very difficult for discrete time non-linear models like GARCH models and it is left for future research. In the sequel of the paper, we assume such existence.

### 2.2.2 The Function $\beta(\cdot)$ is Varying.

It is worth noting that in all the examples discussed in section 2.1, the function  $\beta(u)$  given in (2.1) does not depend on  $u$ . In this section we provide two flexible classes of generalized affine models that generalize any positive affine model  $\omega(u) + \alpha(u)x_t$  where  $\omega(u)$  is a cumulant generating function of a positive random variable and  $\alpha(u) = q(f(u) - 1)$  with  $f(\cdot)$  the moment generating function of a positive random variable. We provide more details on the construction of such processes in Appendix A.

**Proposition 2.1** *We show in Appendix A that the following model is a well defined generalized affine model of order (1,1).*

$$\psi_t(u) = \omega(u) + \alpha(u)x_t + \beta(u)\psi_{t-1}(u)$$

with  $\psi(u) = \ln[E_t(\exp(ux_{t+1}))]$ ,  $\beta(u) = \frac{\mu g(u)}{1-u\mu}$ ,  $\alpha(u) = u\mu + g(u) - 1$  and  $\omega(u) = u\rho - \frac{\rho}{1-\mu} \left( \frac{\beta(u)}{\mu} - 1 \right) + h(u)(1-\beta(u))$ . where  $g(u)$  a Laplace transform of a positive random variable,  $h(u)$  a cumulant of a positive random variable and  $\mu > 0$ .

Notice that this family of generalized affine processes of order (1,1) is very flexible in the sense that it is defined up to unspecified functions  $g(\cdot)$  and  $h(\cdot)$ . The only required conditions are that  $g(\cdot)$  and  $h(\cdot)$  are respectively moment and log-moment generating functions of positive random variable.

Given the properties of functions  $g(\cdot)$ ,  $h(\cdot)$  and  $(\cdot)$ , it is straightforward to show that  $\frac{\beta(\cdot)}{\mu}$  is a moment generating function of a positive random variable, and  $\alpha(\cdot)$  is a log-moment generating function of an infinitely divisible positive random variable. But it is hard to conclude that  $\omega(\cdot)$  is a log-moment generating function of a positive random variable, unless we impose that  $h(u) = \frac{q\mu \left( \frac{\beta(u)}{\mu} - 1 \right)}{1-\beta(u)}$  with  $q > \max(0, \frac{\rho}{\mu(1-\mu)})$  and  $\mu < 1$ . In that case  $\omega(\cdot)$  is a log-moment generating function, and we can reformulate  $\omega(\cdot)$  as follows

$$\omega(u) = u\rho + \frac{q^*\alpha(u)}{1-u\mu}$$

with  $q^* = q - \frac{\rho}{\mu(1-\mu)}$ .

The following proposition gives a generalization of proposition (2.1) to order (2,1).

**Proposition 2.2** *We show in Appendix A that the following model is a well defined generalized affine model of order (2,1).*

$$\psi_t(u) = \omega(u) + \alpha_1(u)x_t + \alpha_2(u)x_{t-1} + \beta(u)\psi_{t-1}(u)$$

with  $\psi(u) = \ln [E_t(\exp(ux_{t+1}))]$ ,  $\beta(u) = \mu f(u)$ ,  $\alpha_1(u) = u\mu + g(u) - 1$ ,  $\alpha_2(u) = \mu((1 - \mu u)f(u) - g(u))$  and  $\omega(u) = (1 - \beta(u)) \left( u\rho + \frac{\rho}{1-\mu} + h(u) \right) - \rho g(u)$ .

where  $f(u)$  a moment generating function of a positive infinitely divisible random,  $g(u)$  a Laplace transform of a positive random variable,  $h(u)$  a cumulant of a positive random variable and  $\mu > 0$ .

The order (1,1) example built in proposition (2.1) is obtained by solving  $\alpha_2(u) = 0$ . The following affine process of order (1,1) is obtained by imposing  $\mu = \rho = 0$

$$\psi_t(u) = h(u) + (g(u) - 1)x_t$$

We have then proved the following general theorem,

**Theorem 2.1** *The following affine model  $\psi_t(u) = h(u) + q(g(u) - 1)x_t$  (with  $q \geq 0$ ,  $h(u)$  a positive c.f.g,  $g(u)$  a positive m.g.f) defined on positive process  $x_t$  can be extended to a generalized affine model (of order (1,1) or (2,1)) with non-constant  $\beta(\cdot)$  as shown in propositions (2.1) and (2.2). The corresponding generalized model is*

$$\psi_t(u) = \omega(u) + \alpha_1(u)x_t + \alpha_2(u)x_{t-1} + \beta(u)\psi_{t-1}(u)$$

with  $\beta(u) = \mu f(u)$ ,  $\alpha_1(u) = uq\mu + q(g(u) - 1)$ ,  $\alpha_2(u) = q\mu((1 - \mu u)f(u) - g(u))$  and  $\omega(u) = (1 - \beta(u)) \left( u\rho + \frac{\rho}{1-\mu} + h(u) \right) - \rho g(u)$ , where  $\mu, \rho \geq 0$ .

This family of affine processes which have been generalized is well known in the literature as the family of compound processes with positive value, and it contains the autoregressive gamma process of Gourieroux and Jasiak (2006) (which is in fact the discrete time counterpart of the Cox-Ingersoll-Ross diffusion process (Cox, Ingersoll, Ross (1985))) and the autoregressive inverse gaussian process.

### 2.3 Cumulant and Moment Structures

We now derive some conditional and unconditional moments (and cumulants as well) implied by the generalized affine model.

### 2.3.1 Conditional Cumulants and Moments

Given that the process  $x_t$  is defined by its conditional cumulant function, it is more convenient to derive the conditional cumulant of  $x_{t+1}$  and then the conditional moments. The conditional cumulant of  $x_{t+1}$  of order  $n$  denoted  $\kappa_{n,t}$ , is given by

$$\kappa_{n,t} = \psi_t^{(n)}(0),$$

where  $f^{(n)}(\cdot)$  denotes the  $n$ -th derivative function of a function  $f(\cdot)$ . We will also use the notation

$$\bar{\kappa}_{n,t} \equiv (\kappa_{1,t}, \kappa_{2,t}, \dots, \kappa_{n,t})^\top. \quad (2.7)$$

From the following paragraph to the end of this paper, we will use the following definition of function  $\beta_{n,i}(u)$ .

$$\beta_{n,i}(u) = \sum_{j=0}^{n-i} \binom{n}{j} \beta^{(j)}(u) B_{n-j,i}(\theta^{(1)}(u), \theta^{(2)}(u), \dots, \theta^{(n-j-i+1)}(u))$$

where  $B_{n,k}(a_1, a_2, \dots, a_{n-k+1})$  denote the Bell polynomials

**Proposition 2.3** *Let  $x_t$  be a generalized affine process defined by (2.3). Then,*

$$\kappa_{n,t} = \omega^{(n)}(0) + \alpha^{(n)}(0)x_t + \sum_{i=1}^n \beta_{n,i}(0) \kappa_{i,t-1} \quad (2.8)$$

and

$$\bar{\kappa}_{n,t} = \bar{\omega}_n + \bar{\alpha}_n x_t + \bar{\beta}_n \bar{\kappa}_{n,t-1} \quad (2.9)$$

where

$$\bar{\omega}_n = \begin{pmatrix} \omega^{(1)}(0) \\ \omega^{(2)}(0) \\ \vdots \\ \omega^{(n)}(0) \end{pmatrix}, \quad \bar{\alpha}_n = \begin{pmatrix} \alpha^{(1)}(0) \\ \alpha^{(2)}(0) \\ \vdots \\ \alpha^{(n)}(0) \end{pmatrix}$$

and

$$\bar{\beta}_n = \begin{pmatrix} \beta_{1,1}(0) & 0 & 0 & \dots & 0 \\ \beta_{2,1}(0) & \beta_{2,2}(0) & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \beta_{n,1}(0) & \beta_{n,2}(0) & \dots & \beta_{n,n-1}(0) & \beta_{n,n}(0) \end{pmatrix}$$

A direct consequence of proposition (2.3) is that  $\bar{\kappa}_{n,t}$  is a  $VAR(1)$ . Indeed using relation (2.9), it can be established that

$$E_{t-1}[\bar{\kappa}_{n,t}] = \bar{\omega}_n + \bar{\rho}_n \bar{\kappa}_{n,t-1} \quad (2.10)$$

where

$$\bar{\rho}_n = \begin{pmatrix} \alpha^{(1)}(0) + \beta_{1,1}(0) & 0 & 0 & \cdots & 0 \\ \alpha^{(2)}(0) + \beta_{2,1}(0) & \beta_{2,2}(0) & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \alpha^{(n)}(0) + \beta_{n,1}(0) & \beta_{n,2}(0) & \cdots & \beta_{n,n-1}(0) & \beta_{n,n}(0) \end{pmatrix}$$

An important implication of proposition (2.3) is that any conditional cumulant  $\kappa_{n,t}$  is a linear combination of  $x_t$  and its lagged values. This property is a characteristic of generalized affine type models. One has different forms when one considers generalized non-affine models defined in (2.4). Another consequence of the VAR representation is that when  $\beta(\cdot)$  is not constant or  $\theta(\cdot)$  is not linear, a cumulant admits an ARMA representation of higher order. However, on the one hand when  $\beta(\cdot)$  is constant and  $\theta(u) = u$ , one has a GARCH(1,1) type equation for  $\kappa_{n,t}$

$$\kappa_{n,t} = \omega^{(n)}(0) + \alpha^{(n)}(0)x_t + \beta\kappa_{n,t-1}$$

on the other hand when  $\beta(\cdot) = 1$  and  $\theta(u) = \theta u$ , one has the following GARCH(1,1) type equation for  $\kappa_{n,t}$

$$\kappa_{n,t} = \omega^{(n)}(0) + \alpha^{(n)}(0)x_t + \theta^n \kappa_{n,t-1}$$

We will show below that  $x_t$  admits an ARMA(1,1) representation. Hence,  $\kappa_{n,t}$  admits an ARMA(2,1) representation.

There is a mapping between the cumulants and the moments of a random variable, which allows one to derive the conditional moments of  $x_{t+1}$  from its conditional cumulants. More precisely, by denoting the conditional moments by  $m_{n,t}$ , i.e.,

$$m_{n,t} = E[x_{t+1}^n \mid x_\tau, \tau \leq t],$$

In general the conditional moment  $m_{n,t}$  is related to conditional cumulant  $\kappa_{1,t}, \dots, \kappa_{n,t}$  through-out the complete Bell polynomial. Indeed one has

$$m_{n,t} = B_n(\kappa_{1,t}, \dots, \kappa_{n,t}) \quad (2.11)$$

where  $Bn(a_1; \dots; a_n)$  is the the nth complete Bell polynomial. Using (2.11) we can deduced the following relations for the first 6 moments.

$$\begin{aligned}
m_{1,t} &= \kappa_{1,t} \\
m_{2,t} &= \kappa_{2,t} + \kappa_{1,t}^2 \\
m_{3,t} &= \kappa_{3,t} + 3\kappa_{2,t}\kappa_{1,t} + \kappa_{1,t}^3 \\
m_{4,t} &= \kappa_{4,t} + 4\kappa_{3,t}\kappa_{1,t} + 3\kappa_{2,t}^2 + 6\kappa_{2,t}\kappa_{1,t}^2 + \kappa_{1,t}^4 \\
m_{5,t} &= \kappa_{5,t} + 5\kappa_{4,t}\kappa_{1,t} + 10\kappa_{3,t}\kappa_{2,t} + 10\kappa_{3,t}\kappa_{1,t}^2 + 15\kappa_{2,t}^2\kappa_{1,t} + 10\kappa_{2,t}\kappa_{1,t}^3 + \kappa_{1,t}^5 \\
m_{6,t} &= \kappa_{6,t} + 6\kappa_{5,t}\kappa_{1,t} + 15\kappa_{4,t}\kappa_{2,t} + 15\kappa_{4,t}\kappa_{1,t}^2 + 10\kappa_{3,t}^2 + 60\kappa_{3,t}\kappa_{2,t}\kappa_{1,t} \\
&\quad + 20\kappa_{3,t}\kappa_{1,t}^3 + 15\kappa_{2,t}^3 + 45\kappa_{2,t}^2\kappa_{1,t}^2 + 15\kappa_{2,t}\kappa_{1,t}^4 + \kappa_{1,t}^6.
\end{aligned}$$

Therefore, by using the results of Proposition 2.3, one gets the conditional moments of  $x_{t+1}$ .

### 2.3.2 Unconditional first and second moments of process $x_t$ and cumulant $\bar{\kappa}_{n,t}$

As in affine models, we can compute unconditional moments which are useful to understand the dynamics of the model and to estimate unknown parameters. We start by focusing on the covariance structure of the process  $x_t$  which will allow us to show that  $x_t$  is an ARMA(1,1) (with possibly heteroskedastic innovations).

Let  $x_t$  be a generalized affine process of order (1,1) defined in (2.3). Using relation (2.9) for  $n = 1$  and taking the unconditional expectation of both side of the equation we have,

$$E(x_t) = E(\kappa_{1,t}) = \frac{\bar{\omega}_1}{1 - \bar{\alpha}_1 - \bar{\beta}_1}$$

In general the unconditional expectation of the vector of cumulant  $\bar{\kappa}_{n,t}$  is straightforwardly computed using relation (2.10) and taking the unconditional expectation of both sides of the equation, we have

$$E(\bar{\kappa}_{n,t}) = (I_n - \bar{\rho}_n)^{-1} \bar{\omega}_n \quad (2.12)$$

For the second order unconditional moments, we first look at the unconditional variance. Let consider relation (2.9) for  $n = 1$ , and square the two sides of the equation then take the unconditional expectation. Let denote

$$\phi = \bar{\alpha}_1 + \bar{\beta}_1$$

we have:

$$E(\kappa_{1,t}^2) = \frac{\bar{\omega}_1^2 + \bar{\alpha}_1^2 E(\kappa_{2,t}) + 2\bar{\omega}_1 \phi E(\kappa_{1,t})}{1 - \phi^2}$$

We can then compute  $E(x_t^2)$  and  $Var(x_t)$  using the following relations  $E(x_t^2) = E(\kappa_{1,t}^2) + E(\kappa_{2,t})$  and  $Var(x_t) = E(x_t^2) - E(x_t)^2$ .

In general unconditional variance of the vector of cumulant  $(\bar{\kappa}_{n,t})$  is obtained by taking the unconditional variance of the two sides of relation (2.9). one gets

$$Var(\bar{\kappa}_{n,t}) = Var(x_t)\bar{\alpha}_n\bar{\alpha}_n^\top + \bar{\beta}_n var(\bar{\kappa}_{n,t})\bar{\beta}_n^\top + \bar{\alpha}_n cov(x_t, \bar{\kappa}_{n,t-1})\bar{\beta}_n^\top$$

which implies that  $Var(\bar{\kappa}_{n,t})$  is the solution to the following quaternion Matrix equation

$$Var(\bar{\kappa}_{n,t}) - \bar{\beta}_n V(\bar{\kappa}_{n,t})\bar{\beta}_n^\top = \theta_n$$

where

$$\theta_n = Var(x_t) \left[ \bar{\alpha}_n\bar{\alpha}_n^\top + \bar{\alpha}_n\bar{\alpha}_n^\top \left( I_n - \bar{\beta}_n^\top \right)^{-1} \bar{\beta}_n^\top + \bar{\beta}_n \left( I_n - \bar{\beta}_n \right)^{-1} \bar{\alpha}_n\bar{\alpha}_n^\top \right]$$

By using the formula

$$vec(AXB) = (B^\top \otimes A) vec(X)$$

one gets the unconditional variance-covariance matrix of  $\bar{\kappa}_{n,t}$

$$vec[Var(\bar{\kappa}_{n,t})] = (I_n - \bar{\beta}_n \otimes \bar{\beta}_n)^{-1} vec(\theta_n)$$

The unconditional autocovariance  $cov(x_{t+h}, x_t)$  is obtained using the following recursion  $cov(x_{t+h}, x_t) = \phi cov(x_{t+h-1}, x_t)$  which is a consequence of relation (2.9) applied at  $n = 1$ . This implies that

$$cov(x_{t+h}, x_t) = \phi^h Var(x_t)$$

hence,  $x_t$  is an ARMA(1,1) whose autoregressive root equals  $\phi$ . As it was the case for unconditional mean and variance, this result on unconditional covariance of process  $x_t$  can be generalized to cumulant  $\bar{\kappa}_{n,t}$ .

$$\begin{aligned} Cov[\bar{\kappa}_{n,t}, \bar{\kappa}_{n,t+h}] &= Cov[\bar{\kappa}_{n,t}, \bar{\kappa}_{n,t+h-1}]\bar{\beta}_n^\top + cov[\bar{\kappa}_{n,t}, \kappa_{1,t+h-1}]\bar{\alpha}_n^\top \\ &= Cov[\bar{\kappa}_{n,t}, \bar{\kappa}_{n,t+h-1}]\bar{\beta}_n^\top + Var(x_t)\phi^{h-1}(I_n - \bar{\beta}_n)^{-1}\bar{\alpha}_n\bar{\alpha}_n^\top \\ &= Var(\bar{\kappa}_{n,t})\left(\bar{\beta}_n^\top\right)^h + Var(x_t)\phi^{h-1}(I_n - \bar{\beta}_n)^{-1}\bar{\alpha}_n\bar{\alpha}_n^\top\left(I_n - \bar{\beta}_n^\top\right)^{-1}\left[I_n - \left(\bar{\beta}_n^\top\right)^h\right] \end{aligned}$$

We can also evaluate the unconditional covariance between  $x_t$  and  $\bar{\kappa}_{n,t+h}$ . Indeed using one more time relation (2.9),  $E(x_t\bar{\kappa}_{n,t+h})$  is computed recursively as following

$$E(x_t\bar{\kappa}_{n,t+h}) = E(x_t)\bar{\omega}_n + E(x_t^2)\bar{\alpha}_n + \bar{\beta}_n E(x_t\bar{\kappa}_{n,t+h-1}) \quad (2.13)$$

with the first term  $E(x_t\bar{\kappa}_{n,t}) = E(x_t)\bar{\omega}_n + E(x_t^2)\bar{\alpha}_n + \bar{\beta}_n E(\kappa_{1,t}\bar{\kappa}_{n,t})$  where  $E(\kappa_{1,t}\bar{\kappa}_{n,t})$  is deduced from the variance-covariance matrix of  $\bar{\kappa}_{n,t}$ .

### 2.3.3 Higher order covariance

In this subsection, we use results on first and second unconditional moments of  $x_t$  and  $\bar{\kappa}_{n,t}$  derived in the previous subsection to compute the third and fourth unconditional moments of  $x_t$ . We begin by deriving the third moments.

#### Third order moments of $x_t$

Using the relation between moments and cumulants, we have

$$E(x_t^3) = E(m_{3,t}) = E(\kappa_{3,t}) + 3E(\kappa_{2,t}\kappa_{1,t}) + E(\kappa_{1,t}^3) \quad (2.14)$$

From the previous subsection, All the terms in the right hand side of Eq 2.14 are known except  $E(\kappa_{1,t}^3)$ .  $E(\kappa_{1,t}^3)$  is obtained by cubing the two sides of Eq. (2.9) for  $n = 1$ , it implies that

$$\begin{aligned} (1 - \phi^3)E(\kappa_{1,t}^3) &= \bar{\omega}_1^3 + \bar{\alpha}_1^3 E(\kappa_{3,t}) + 3\bar{\alpha}_1^2 \phi E(\kappa_{1,t}\kappa_{2,t}) + 3\bar{\omega}_1^2 \phi E(\kappa_{1,t}) \\ &+ 3\bar{\omega}_1(\bar{\alpha}_1^2 + \bar{\beta}_1^2)E(\kappa_{1,t}^2) + 3\bar{\omega}_1\bar{\alpha}_1^2 E(\kappa_{2,t}) \end{aligned}$$

consequently we can deduce the closed form expression of  $E(x_t^3)$ .

We now consider  $E(x_t x_{t+h}^2)$ . Using the relation between moments and cumulants, we have

$$E(x_t x_{t+h}^2) = E(x_t \kappa_{2,t+h-1}) + E(x_t \kappa_{1,t+h-1}^2) \quad (2.15)$$

$E(x_t \kappa_{2,t+h-1})$  is deduced from Eq. (2.13). We need to evaluate  $E(x_t \kappa_{1,t+h-1}^2)$ . By squaring the two sides of Eq. (2.9) for  $n = 1$  and multiply by  $x_t$ , we show that

$$\begin{aligned} E(x_t \kappa_{1,t}^2) &= \bar{\omega}_1^2 E(x_t) + 2\bar{\omega}_1 \bar{\alpha}_1 E(x_t^2) + \bar{\alpha}_1^2 E(x_t^3) + (\bar{\beta}_1^2 + 2\bar{\omega}_1 \bar{\beta}_1) E(\kappa_{1,t}^2) \\ &+ 2\bar{\alpha}_1 \bar{\beta}_1 E(\kappa_{1,t} \kappa_{2,t}) + 2\bar{\alpha}_1 \bar{\beta}_1 E(\kappa_{1,t}^3) \end{aligned}$$

In general, for any integer  $h$ , given  $E(x_t \kappa_{1,t+h-1}^2)$ ,  $E(x_t \kappa_{1,t+h}^2)$  is computed as follows:

$$E(x_t \kappa_{1,t+h}^2) = \bar{\omega}_1^2 E(x_t) + \bar{\alpha}_1^2 E(x_t \kappa_{2,t+h-1}) + 2\bar{\omega}_1 \phi E(x_t x_{t+h}) + \phi^2 E(x_t \kappa_{1,t+h-1}^2) \quad (2.16)$$

All the terms of the right hand side of Eq. (2.16) are known. consequently we can deduce the closed form expression of  $E(x_t x_{t+h}^2)$ .

We will conclude on this section on third order moments by evaluating  $E(x_t^2 \bar{\kappa}_{n,t+h})$ . To achieve this purpose, we need  $E(\kappa_{1,t}^2 \bar{\kappa}_{n,t})$ . using the relation (2.9), we have:

$$\begin{aligned} (I_n - \phi^2 \bar{\beta}_n) E(\kappa_{1,t}^2 \bar{\kappa}_{n,t}) &= \bar{\omega}_1^2 E(\bar{\kappa}_{n,t}) + \phi^2 E(\kappa_{1,t}^2) \bar{\omega}_n + \bar{\alpha}_1^2 E(\kappa_{2,t}) \bar{\omega}_n + \bar{\alpha}_1^2 E(\kappa_{3,t}) \bar{\alpha}_n + \bar{\alpha}_1^2 \bar{\beta}_n E(\kappa_{2,t} \bar{\kappa}_{n,t}) \\ &+ \phi^2 E(\kappa_{1,t}^3) \bar{\alpha}_n + (3\bar{\alpha}_1^2 + \bar{\alpha}_1 \bar{\beta}_1) E(\kappa_{2,t} \kappa_{1,t}) \bar{\alpha}_n + 2\bar{\omega}_1 \bar{\alpha}_1 E(x_t \bar{\kappa}_{n,t}) + 2\bar{\omega}_1 \bar{\beta}_1 E(\kappa_{1,t-1} \bar{\kappa}_{n,t}) \end{aligned}$$

we deduce

$$E(x_t^2 \bar{\kappa}_{n,t}) = E(x_t^2) \bar{\omega}_n + E(x_t^3) \bar{\alpha}_n + \bar{\beta}_n E(\kappa_{2,t} \bar{\kappa}_{n,t}) + \bar{\beta}_n E(\kappa_{1,t}^2 \bar{\kappa}_{n,t})$$

We are now able to compute  $E(x_t^2 \bar{\kappa}_{n,t+h})$  using the following recursion:

$$E(x_t^2 \bar{\kappa}_{n,t+h}) = E(x_t^2) \bar{\omega}_n + E(x_t^3) \bar{\alpha}_n + \bar{\beta}_n E(x_t^2 \bar{\kappa}_{n,t+h-1})$$

#### Fourth order moments of $x_t$

Using the relation between moments and cumulants, we have

$$E(x_t^4) = E(m_{4,t}) = E(\kappa_{4,t}) + 4E(\kappa_{3,t} \kappa_{1,t}) + 3E(\kappa_{2,t}^2) + 6E(\kappa_{2,t} \kappa_{1,t}^2) + E(\kappa_{1,t}^4)$$

Except  $E(\kappa_{1,t}^4)$ , all the terms in the right hand side are known.  $E(\kappa_{1,t}^4)$  is obtained by raising the two sides of Eq. (2.9) for  $n = 1$  to the 4th power, it implies that

$$\begin{aligned} (1 - \phi^4)E(\kappa_{1,t}^4) &= \bar{\omega}_1^4 + 4\bar{\omega}_1^3 \phi E(\kappa_{1,t}) + 6\bar{\omega}_1^2 \bar{\alpha}_1^2 E(\kappa_{2,t}) + 4\bar{\alpha}_1^3 \bar{\omega}_1 E(\kappa_{3,t}) + 6\bar{\omega}_1^2 \phi^2 E(\kappa_{1,t}^2) \\ &+ 12\bar{\alpha}_1^2 \bar{\omega}_1 \phi E(\kappa_{1,t} \kappa_{2,t}) + 4\bar{\alpha}_1^3 \phi E(\kappa_{3,t} \kappa_{1,t}) + 3\bar{\alpha}_1^4 E(\kappa_{2,t}^2) + 6\bar{\alpha}_1^2 \phi^2 E(\kappa_{2,t} \kappa_{1,t}^2) + 4\bar{\omega}_1 \phi^3 E(\kappa_{1,t}^3) \end{aligned}$$

consequently we can deduce the closed form expression of  $E(x_t^4)$ .

Let consider the autocorrelation of  $x_t^2$ .

$$E(x_t^2 x_{t+h}^2) = E(x_t^2 \kappa_{2,t+h-1}) + E(x_t^2 \kappa_{1,t+h-1}^2)$$

From the results derived in the third order moments  $E(x_t^2 \kappa_{2,t+h-1})$  is known, we can show that  $E(x_t^2 \kappa_{1,t+h-1}^2)$  is computed recursively as follows:

$$E(x_t^2 \kappa_{1,t+h-1}^2) = \bar{\omega}_1 E(x_t^2) + \bar{\alpha}_1^2 E(x_t^2 \kappa_{2,t+h-2}) + \phi^2 E(x_t^2 \kappa_{1,t+h-2}) + 2\bar{\omega}_1 \phi E(x_t^2 x_{t+h-2})$$

with the first term

$$E(x_t^2 \kappa_{1,t-1}^2) = E(\kappa_{1,t}^4) + E(\kappa_{1,t}^2 \kappa_{2,t})$$

Then we obtain the closed form expression of the autocorrelation of  $x_t^2$

## 2.4 Cumulant Function of Future Aggregated Returns

An important formula used in the analytical calculation of the term structure of interest rates and option prices is the conditional distribution function of  $\sum_{i=1}^h u_i x_{t+i}$  for given real numbers  $u_i$ . Affine models allow one to derive the conditional cumulant function of  $(x_{t+1}, x_{t+2}, \dots, x_{t+h})$

and consequently the one of  $\sum_{i=1}^n u_i x_{t+i}$ . It turns out that this is the case for generalized affine models. Let denote  $\bar{u}_h = (u_1, u_2, \dots, u_h)^\top$  and  $V_{t,h}(\bar{u}_h) = \log E_t \left[ \exp \left( \sum_{i=1}^h u_i x_{t+i} \right) \right]$ , if the process  $x_t$  is generated by (2.3) then we establish the following result (see appendix B for proof)

**Proposition 2.4** *The conditional cumulant function of  $(x_{t+1}, x_{t+2}, \dots, x_{t+h})$  is given by*

$$V_{t,h}(\bar{u}_h) = a_h(\bar{u}_h) x_t + b_h(\bar{u}_h) + \sum_{k=1}^h \bar{\beta}_k(d_k) \psi_{t-1}(\theta_k(d_k)) \quad (2.17)$$

with

$$\begin{aligned} a_h(\bar{u}_h) &= \sum_{k=1}^h \bar{\beta}_{k-1}(d_k) \alpha(\theta_{k-1}(d_k)) \\ b_h(\bar{u}_h) &= \sum_{k=1}^h \left[ \sum_{j=1}^k \bar{\beta}_{j-1}(d_k) \omega(\theta_{j-1}(d_k)) \right] \end{aligned}$$

The sequence  $(d_k)$  is defined recursively as follows:  $d_h = u_h$  and for  $k \leq h-1$

$$d_k = u_k + \sum_{j=1}^{h-k} \bar{\beta}_{j-1}(d_{k+j}) \alpha(\theta_{j-1}(d_{k+j}))$$

where functions  $\bar{\beta}_j(\cdot)$  and  $\theta_j(\cdot)$  are related to  $\beta(\cdot)$  and  $\theta(\cdot)$  as follows

$$\begin{aligned} \theta_j(u) &= \theta^{\circ j}(u) \\ \bar{\beta}_0(u) &= 1 \\ \bar{\beta}_j(u) &= \prod_{k=1}^j \beta(\theta_{k-1}(u)), \quad \text{for } j \geq 1 \end{aligned}$$

We will often use (2.17) in the next section when we derive yields and option prices. When  $\beta(u) = 0$  or  $\theta(u) = 0$ , we obtain the well known affine model's result which stipulates that the conditional log-moment generating function of  $(x_{t+1}, x_{t+2}, \dots, x_{t+h})$  is affine in  $x_t$ .

### 3 Analytical Formulas of Prices of Financial Derivatives

This section analytically characterizes the yields and option prices when one assumes a generalized affine model for the interest rate and the stochastic volatility respectively. For each of them, we follow two approaches. We assume the generalized affine model under the physical

measure (P-measure) and specify the price of risk and then derive the price of the financial derivatives (bonds or options). The second approach consists on specifying the generalized affine model under the risk neutral measure (Q-measure) and then derive the prices of the financial derivatives. We start the analysis by studying the term structure model

### 3.1 The Term Structure of Interest Rates

#### 3.1.1 Generalized Affine Model Under the P-Measure

We assume that under the P-measure, the short term of interest rate denoted  $r_t$  follows a generalized affine process given in (2.1), i.e.,

$$\begin{aligned}\psi_t^P(u) &\equiv \ln E_t^P [\exp(ur_{t+1})] \\ &= \omega^P(u) + \alpha^P(u)r_t + \beta^P(u)\psi_{t-1}^P(\theta^P(u))\end{aligned}$$

When  $\beta^P(u) = 0$  or  $\theta^P(u) = 0$ , one gets affine modes like those of Vasicek (1977), CIR (1985) and Duffie-Kan (1996) who derived the term structure of interest rates.

In order to derive the dynamics of  $r_t$  under the Q-measure and the yield curve, one needs to specify the stochastic discount factor denoted here  $M_{t,t+1}$  or the price of risk. We follow the general approach of Gourieroux and Monfort (2007) who proposed the following formulation:

$$M_{t,t+1} = \exp(\gamma r_{t+1} + \theta_t). \quad (3.1)$$

Given the restriction

$$\exp(-r_t) = E_t^P [M_{t,t+1}], \quad (3.2)$$

one gets

$$\theta_t = -r_t - \psi_t(\gamma) \quad \text{and} \quad M_{t,t+1} = \exp(\gamma r_{t+1} - r_t - \psi_t(\gamma)). \quad (3.3)$$

In the sequel, we define  $B(t, h)$  and  $r_{t,h}$  as

$$B(t, h) = E_t^P \left[ \prod_{i=1}^h M_{t+i-1, t+i} \right], \quad r_{t,h} = -\frac{\log(B(t, h))}{h}. \quad (3.4)$$

We are now able to derive the term structure of interest rates, i.e., the formula of  $r_{t,h}$  when  $h$  varies.

#### Proposition 3.1

$$r_{t,h} = d_h + c_{h,0}r_t + z_{t-1}^{(h)}$$

with

$$\begin{aligned}
d_h &= \frac{1}{h} \sum_{k=1}^{h-1} \left( \sum_{i=1}^k \bar{\beta}_{i-1}^P(\gamma) \omega^P(\theta_{i-1}^P(\gamma)) - \bar{\beta}_{i-1}(d_k) \omega(\theta_{i-1}(d_k)) \right) \\
c_{h,0} &= \frac{1}{h} \left[ 1 + \sum_{k=1}^{h-1} (\bar{\beta}_{k-1}(\gamma) \alpha^P(\theta_{k-1}^P(\gamma)) - \bar{\beta}_{k-1}(d_k) \alpha(\theta_{k-1}(d_k))) \right] \\
z_{t-1}^{(h)} &= \sum_{k=1}^{h-1} \left[ \frac{\bar{\beta}_k(\gamma)}{h} \psi_{t-1}^P(\theta_k^P(\gamma)) - \frac{\bar{\beta}_k(d_k)}{h} \psi_{t-1}^P(\theta_k^P(d_k)) \right]
\end{aligned}$$

The sequence  $d_k$  for  $k \leq h-1$  satisfy the following backward recursion :

$$d_{h-1} = \gamma - 1 \text{ and for } k \leq h-2$$

$$d_k = \gamma - 1 + \sum_{j=1}^{h-1-k} [\bar{\beta}_{j-1}(d_{k+j}) \alpha(\theta_{j-1}(d_{k+j})) - \bar{\beta}_{j-1}(\gamma) \alpha(\theta_{j-1}(\gamma))]$$

### Special case

We study the case of constant function  $\beta(\cdot)$  and linear function  $\theta(\cdot)$ , this implies the following model for the short term rate:

$$\psi_t^P(u) = \omega(u) + \alpha(u)r_t + \beta\psi_{t-1}^P(\theta u)$$

where  $\beta, \theta \geq 0, .$  Since  $r_t$  is a positive random variable,  $\omega(u)$  and  $\alpha(u)$  are log-moment generating function of positive random variables. The yield to maturity  $n$  formula is the follows:

$$y_t^{(n)} = d_{n,t} + \sum_{j=0}^{t-1} c_{n,j} r_{t-j}$$

where

$$\begin{aligned}
d_{n,t} &= \frac{1}{n} \sum_{k=1}^{n-1} \left( \beta^{t+k-1} [\psi_0^P(\theta^{t+k-1}\gamma) - \psi_0^P(\theta^{t+k-1}c_k)] + \sum_{i=1}^{t+k-1} \beta^{i-1} (\omega(\theta^{i-1}\gamma) - \omega(\theta^{i-1}c_k)) \right) \\
c_{n,0} &= \frac{1}{n} \left[ 1 + \sum_{k=1}^{n-1} \beta^{k-1} (\alpha(\theta^{k-1}\gamma) - \alpha(\theta^{k-1}c_k)) \right]
\end{aligned}$$

and for  $j \geq 1$

$$c_{n,j} = \sum_{k=1}^{n-1} \frac{\beta^{j+k-1}}{n} (\alpha(\theta^{j+k-1}\gamma) - \alpha(\theta^{j+k-1}c_k))$$

$c_{n-1} = \gamma - 1$  and for  $0 \leq k \leq n - 2$

$$c_k = \gamma - 1 + \sum_{j=1}^{n-1-k} \beta^{j-1} (\alpha(\theta^{j-1}c_{k+j}) - \alpha(\theta^{j-1}\gamma))$$

One issue of interest is the positivity of yield, in this simple particular case we are able to specify some conditions under which yields generated by our model are always positive. Function  $\omega(u)$  and  $\alpha(u)$  are defined on  $u \leq 0$  and are absolutely monotone, in particular  $\omega'(u)$ ,  $\alpha'(u) \geq 0$ . This implies that  $\omega(u)$  and  $\alpha(u)$  are increasing functions. The model is well defined if we choose the price of risk  $\gamma \leq 0$ , it implies that  $c_{n-1} < \gamma \leq 0$ . Since  $\alpha(u)$  is an increasing function, it implies that  $c_{n-2} < \gamma \leq 0$ . Using the same reasoning, we deduce that  $c_k < \gamma \leq 0$  for  $k \leq n - 1$ .

Since  $c_k < \gamma \leq 0$  for  $k \leq n - 1$ , it follows that  $c_{n,0} \geq 0$ ,  $d_n \geq 0$  and

$$\sum_{k=1}^{n-1} \frac{\beta^k}{n} \left[ \psi_{t-1}^P(\theta^k \gamma) - \psi_{t-1}^P(\theta^k c_k) \right] \geq 0$$

Consequently our model is coherent because it generates positive yield  $y_t^{(n)}$  at any given maturity.

### Characterization of the dynamics of $r_t$ under the Q-measure

Another question of interest is the characterization of the dynamics of  $r_t$  under the Q-measure. We denote  $\psi_t^Q(u)$  the conditional cumulant function of  $r_{t+1}$  under the Q-measure, i.e.,

$$\psi_t^Q(u) \equiv \log E_t^Q[\exp(ur_{t+1})]. \quad (3.5)$$

We will restrict our analysis to the generalized affine of type I (i.e we restrict function  $\theta^P(u) = u$ ).

**Proposition 3.2 Dynamics of  $r_t$  under the Q-measure.** *One has*

$$\psi_t^Q(u) = \psi_t^P(u + \gamma) - \psi_t^P(\gamma). \quad (3.6)$$

Hence if  $\theta^P(u) = u$ ,

$$\psi_t^Q(u) = \omega^Q(u) + \alpha^Q(u)r_t + \beta^Q(u)\psi_{t-1}^Q(u) + [\beta(u + \gamma) - \beta(\gamma)]\psi_t(\gamma), \quad (3.7)$$

where

$$\omega^Q(u) = \omega^P(u + \gamma) - \omega^P(\gamma), \quad \alpha^Q(u) = \alpha^P(u + \gamma) - \alpha^P(\gamma), \quad \beta^Q(u) = \beta^P(u + \gamma). \quad (3.8)$$

Eq. (3.6) is model free, i.e. it does not depend on our generalized affine specification. In particular, the same equation appears in affine models; see Gourieroux and Monfort (2007) and Monfort and Pegoraro (2006a). An additional term appears in (3.7) which vanishes when  $\beta(\cdot)$  is constant, as in our empirical examples. When this term does not vanish, the short term of interest rate is not a generalized affine under the Q-measure. However, the following proposition characterizes the conditional cumulant of  $(r_{t+1}, \psi_{t+1}^P(\gamma))$  which will allow us to understand the dynamics of  $r_{t+1}$  under the Q-measure. In the sequel,  $\psi_{r, \psi(\gamma), t}^Q(u, v)$  denotes the conditional cumulant function of  $(r_{t+1}, \psi_{t+1}^P(\gamma))$  under the Q-measure.

**Proposition 3.3**  $\theta^P(u) = u$  implies

$$\begin{aligned} \psi_{r, \psi(\gamma), t}^Q(u, v) &= \omega_1^Q(u, v) + \left( \alpha_1^Q(u, v) r_t + \alpha_2^Q(u, v) \psi_t^P(\gamma) \right) + \beta_1^Q(u, v) \psi_{r, \psi(\gamma), t-1}^Q(u, v) \\ &\quad - \alpha_2^Q(u, v) \beta_1^Q(u, v) \psi_{t-1}^Q(\gamma) \end{aligned} \quad (3.9)$$

where

$$\begin{aligned} \omega_1^Q(u, v) &= v\omega^P(\gamma) [1 - \beta^P(u + v\alpha^P(\gamma) + \gamma)] + \omega^P(u + v\alpha^P(\gamma) + \gamma) \\ \alpha_1^Q(u, v) &= \alpha^P(u + v\alpha^P(\gamma) + \gamma), \quad \alpha_2^Q(u, v) = v\beta^P(\gamma) - 1 \\ \beta_1^Q(u, v) &= \beta^P(u + v\alpha^P(\gamma) + \gamma) \end{aligned}$$

While the definition of generalized affine models (2.1) is given for univariate processes and of order (1,1), the extensions to multivariate and higher order is not very difficult. Eq. (3.9) means that the bivariate vector  $(r_{t+1}, \psi_{t+1}^P(\gamma))$  is a generalized affine process of order (2,1). Consequently, one can characterize formulas of financial derivatives, including yields, by using the generalized affine dynamics of  $(r_{t+1}, \psi_{t+1}^P(\gamma))$  under the Q-measure.

### 3.1.2 Generalized Affine Model Under the Q-Measure

We now assume that the short term of interest rate  $r_t$  follows a generalized affine process given in (2.3) under the Q-measure, i.e.

$$\log E_t^Q[\exp(ur_{t+1})] \equiv \psi_t^Q(u) = \omega^Q(u) + \alpha^Q(u)r_t + \beta^Q(u)\psi_{t-1}^Q(\theta^Q(u)).$$

The following proposition provides the formula of the yield curve.

**Proposition 3.4** *The yield at horizon  $n$  is*

$$r_{t,n} = b_n + a_n r_t - \frac{\sum_{k=1}^{n-1} \bar{\beta}_k(d_k) \psi_{t-1}^Q(\theta_k(d_k))}{n} \quad (3.10)$$

where

$$\begin{aligned}
a_n &= \frac{1 - \sum_{k=1}^{n-1} \bar{\beta}_{k-1}(d_k) \alpha^Q(\theta_{k-1}(d_k))}{n} \\
b_{n-1} &= \frac{-\sum_{k=1}^{n-1} \left[ \sum_{j=1}^k \bar{\beta}_{j-1}(d_k) \omega^Q(\theta_{j-1}(d_k)) \right]}{n} \\
d_k &= -1 + \sum_{j=1}^{n-1-k} \bar{\beta}_{j-1}(d_{k+j}) \alpha^Q(\theta_{j-1}(d_{k+j})) \text{ , for } k < n-1, \quad d_{n-1} = -1
\end{aligned}$$

One could also characterize the dynamics of  $r_t$  under the P-measure if one assumes a stochastic discount factor. We assume again that the stochastic discount factor is given by<sup>3</sup> (3.1). Hence, one gets

$$\psi_t^p(u) = \psi_t^Q(u - \gamma) - \psi_t^Q(-\gamma). \quad (3.11)$$

Again, this equation is model free and appears in affine models (Gourieroux and Monfort (2007), Monfort and Pegoraro (2006a)). Likewise,  $r_{t+1}$  is not a generalized affine process under the P-measure. However in the case of generalized affine model of type I, the vector  $(r_{t+1}, \psi_{t+1}(-\gamma))$  is a generalized affine process of order (2,1) as shown in the following proposition. In the sequel,  $\psi_{r,\psi(-\gamma),t}^P$  denotes the conditional cumulant function of  $(r_{t+1}, \psi_{t+1}(-\gamma))$  under the P-measure:

**Proposition 3.5**  $\theta^P(u) = u$  implies that

$$\begin{aligned}
\psi_{r,\psi(-\gamma),t}^P(u, v) &= \omega_1^P(u, v) + (\alpha_1^P(u, v) r_t + \alpha_2^P(u, v) \psi_t^p(-\gamma)) + \beta_1^P(u, v) \psi_{r,\psi(-\gamma),t-1}^Q(u, v) \\
&\quad - \alpha_2^Q(u, v) \beta_1^P(u, v) \psi_{t-1}^P(-\gamma)
\end{aligned} \quad (3.12)$$

where

$$\begin{aligned}
\omega_1^P(u, v) &= v \omega^Q(-\gamma) [1 - \beta^Q(u + v \alpha^Q(-\gamma) - \gamma)] + \omega^Q(u + v \alpha^Q(-\gamma) - \gamma) \\
\alpha_1^P(u, v) &= \alpha^Q(u + v \alpha^Q(-\gamma) - \gamma), \quad \alpha_2^P(u, v) = v \beta^Q(-\gamma) - 1 \\
\beta_1^P(u, v) &= \beta^Q(u + v \alpha^Q(-\gamma) - \gamma).
\end{aligned}$$

---

<sup>3</sup>Observe that when one specifies the dynamics of  $r_t$  under the Q-measure as a generalized affine process, one could allow  $\gamma$  in (3.1) to be time-varying and adapted to the information available at time t. A consequence is that the short term of interest rate will not be a generalized affine process under the P-measure; see Gourieroux and Monfort (2007) for the same discussion about affine models.

## 3.2 Option Pricing

We now consider models of stock returns where we assume that the conditional variance of the returns is time-varying and is generalized affine. In what follows  $r_t$  denotes the log-returns of the stock price, i.e.

$$r_{t+1} = \ln \left( \frac{S_{t+1}}{S_t} \right)$$

The key approach behind the analytical calculations of Heston (1993), Duffie, Pan, and Singleton (2000), and Heston and Nandi (2000) is the possibility to write the joint process  $(r_t, h_t)$  as an affine process, where  $h_{t+1}$  is the conditional variance of  $r_{t+1}$  given an information set that contains  $r_t$  and its lagged values and possibly another variable, latent or not, like in stochastic volatility models. In what follows, we will allow for both cases. We will write the joint model of  $(r_{t+1}, h_{t+1})$ . The variable  $h_{t+1}$  could be the conditional variance of  $r_{t+2}$  given  $\{r_\tau, h_\tau, \tau \leq t\}$  (including GARCH type models). The variable  $h_{t+1}$  could be an observable variable like in our empirical example where it equals the high frequency realized volatility. In the rest of this section the information set  $I_t$  is the sigma algebra generated by  $\{r_\tau, h_\tau, \tau \leq t\}$ . The conditional expectation operator  $E[\cdot | I_t]$  will be denoted  $E_t[\cdot]$ .

In this section, we restrict ourself to the generalized affine model of type I.

### 3.2.1 Generalized Affine process under the P-Measure

We denote the conditional cumulant function of  $(r_{t+1}, h_{t+1})$  under the P measure by  $\psi_t^P(u, v)$ :

$$\psi_t^P(u, v) = \log E_t^P[\exp(ur_{t+1} + vh_{t+1})] = \omega^P(u, v) + \alpha^P(u, v)h_t + \beta^P(u, v)\psi_{t-1}^P(u, v)$$

When one assumes that  $h_t$  is exactly the conditional variance of  $r_{t+1}$ , one needs to impose the following restrictions on the cumulant function in order to guarantee this assumption:

$$\frac{\partial^2 \omega^P}{\partial u^2}(0, 0) = 0, \quad \frac{\partial^2 \alpha^P}{\partial u^2}(0, 0) = 1, \quad \frac{\partial \beta^P}{\partial u}(0, 0) = 0, \quad \beta^P(0, 0) = 0 \quad (3.13)$$

which implies

$$\frac{\partial^2 \psi_t}{\partial u^2}(0, 0) = \text{Var}_t^P[r_{t+1}] = h_t.$$

We denote by  $r$  the short term interest rate supposed constant for simplicity. We consider the following stochastic discount factor

$$M_{t,t+1} = \exp(\gamma r_{t+1} + \lambda h_{t+1} + \theta_t). \quad (3.14)$$

Observe that both Heston and Nandi (2000) and Christoffersen et al. (2006) assumed that  $\lambda = 0$ . There is no theoretical foundation for such assumption other than simplicity. In other words, we allow the volatility to be priced.

In addition, one needs to impose restrictions in order to guarantee that  $M_{t,t+1}$  is a stochastic discount factor, which implies that prices under the Q-measure are martingales. This is the purpose of the following proposition.

**Proposition 3.6** *The parameters  $\gamma$  and  $\lambda$  are restricted by the following system of equations*

$$\begin{aligned}\omega(1 + \gamma, \lambda) - \omega(\gamma, \lambda) &= r(1 - \beta(\gamma, \lambda)) \\ \alpha(1 + \gamma, \lambda) &= \alpha(\gamma, \lambda) \\ \beta(1 + \gamma, \lambda) &= \beta(\gamma, \lambda).\end{aligned}$$

Observe that when  $\beta(\cdot)$  is a constant function, the third equation in the previous system holds, which leads to a fully identified system.

We will now characterize the dynamics of  $(r_{t+1}, h_{t+1})$  under the Q-measure by deriving its conditional cumulant function denoted  $\psi_t^Q(u, v)$ .

**Proposition 3.7** *We have*

$$\Psi_t^Q(u, v) = \Psi_t^P(u + \gamma, v + \lambda) - \Psi_t^P(\gamma, \lambda) \quad (3.15)$$

and

$$\begin{aligned}\Psi_t^Q(u, v) &= (\omega^p(u + \gamma, v + \lambda) - \omega^p(\gamma, \lambda)) \\ &\quad + (\alpha^p(u + \gamma, v + \lambda) - \alpha^p(\gamma, \lambda))h_t + \beta(u + \gamma, v + \lambda)\Psi_{t-1}^Q(u, v) \\ &\quad + (\beta(u + \gamma, v + \lambda) - \beta(\gamma, \lambda))\psi_{t-1}(\gamma, \lambda).\end{aligned} \quad (3.16)$$

Several remarks are in order. Similarly to the term structure of interest rates, an additional term appears in (3.16), implying that the process  $(r_{t+1}, h_{t+1})$  is not generalized affine process under the Q-measure. Likewise, this additional term vanishes when the function  $\beta(\cdot)$  is constant. Again, one can still prove that a particular vector is a generalized affine model of higher order, which will allow us to derive option prices. More precisely, one can show that the vector  $(r_{t+1}, h_{t+1}, \psi_{t+1}(\gamma, \lambda))$  is a generalized affine process of order. We now provide the formula of the option prices.

**Proposition 3.8** *The price at time  $t$  of a European call option with payoff  $(S_{t+h} - X)^+$  at time  $t + h$  is given by*

$$C_t = \exp(-rh)S_t C_{1,t} - \exp(-rh)X C_{2,t} \quad (3.17)$$

where

$$\begin{aligned} C_{1,t} &= \frac{\exp(rh)}{2} + \int_0^{+\infty} \frac{1}{\pi u} \text{Im} \left[ \exp \left( \Psi_{t,t+h}^Q (1 + iu) - iu \ln \left( \frac{X}{S_t} \right) \right) \right] du \\ C_{2,t} &= \frac{1}{2} + \int_0^{+\infty} \frac{1}{\pi u} \text{Im} \left[ \exp \left( -iu \ln \left( \frac{X}{S_t} \right) + \Psi_{t,t+h}^Q(iu) \right) \right] du \end{aligned}$$

and

$$\Psi_{t,t+h}^Q(u) = a_h(u)h_t + b_h(u) + \sum_{k=1}^h \left( \beta^P(d_k)^k \psi_{t-1}^P(d_k) - \beta^P(\gamma, \lambda)^k \psi_{t-1}^P(\gamma, \lambda) \right)$$

with

$$\begin{aligned} a_h(u) &= \sum_{k=1}^h \left( \beta^P(d_k)^k \alpha^P(d_k) - \beta^P(\gamma, \lambda)^k \alpha^P(\gamma, \lambda) \right) \\ b_h(u) &= \sum_{k=1}^h \left( \frac{1 - \beta^P(d_k)^k}{1 - \beta^P(d_k)} \omega^P(d_k) - \frac{1 - \beta^P(\gamma, \lambda)^k}{1 - \beta^P(\gamma, \lambda)} \omega^P(\gamma, \lambda) \right) \end{aligned}$$

and

$$\begin{aligned} d_k &= (u + u_k, v_k) + \sum_{j=k}^{h-1} \beta(d_{j+1})^{j-k} (0, \alpha(d_{j+1})) \text{ for } k \leq h-1 \\ d_h &= (u + u_h, v_h) \end{aligned} \quad (3.18)$$

where

$$\begin{aligned} u_h &= \gamma, \quad v_h = \lambda \\ u_j &= \gamma - \alpha(\gamma, \lambda) \frac{1 - \beta(\gamma, \lambda)^{h-j}}{1 - \beta(\gamma, \lambda)} \text{ for } 1 \leq j < h \\ v_j &= \lambda - \alpha(\gamma, \lambda) \frac{1 - \beta(\gamma, \lambda)^{h-j}}{1 - \beta(\gamma, \lambda)} \text{ for } 1 \leq j < h \end{aligned}$$

$$\begin{aligned} \psi_{t-1}^P(d_k) &= \omega^P(d_k) + \alpha^P(d_k)h_{t-1} + \beta^P(d_k)\psi_{t-2}^P(d_k) \\ &= \frac{\omega^P(d_k)}{1 - \beta^P(d_k)} + \alpha^P(d_k) \left( \sum_{i=0}^{\infty} \beta^P(d_k)^i h_{t-1-i} \right) \end{aligned}$$

This proposition uses Fourier transforms, which is a traditional approach in affine models. It is important to notice that, for this purpose, we had to use the logarithmic of the characteristic function instead of the logarithmic of the moment generating function. A simple modification of the notation is sufficient to do this change.

### 3.3 Generalized Affine process under the Q-Measure

This subsection specifies the dynamics of  $(r_{t+1}, h_{t+1})$  under the Q-measure,  $\Psi_t^Q(u, v)$ , and derives the option prices. We assume that

$$\Psi_{t+1}^Q(u, v) = \omega(u, v) + \alpha(u, v) h_{t+1} + \beta(u, v) \Psi_t^Q(u, v) \quad (3.19)$$

A well defined risk-neutral distribution for log-returns must satisfy

$$\exp(r) = E^Q[\exp(r_{t+1}) | I_t]$$

where  $r$  is the risk-free rate. Thus  $\Psi_t^Q(1, 0)$  must satisfy

$$\Psi_t^Q(1, 0) = r$$

**Proposition 3.9** *Eq. (3.19) is a valid risk-neutral model if and only if*

$$\begin{aligned} \frac{\omega(1, 0)}{1 - \beta(1, 0)} &= r \\ \alpha(1, 0) &= 0 \end{aligned} \quad (3.20)$$

The result is an implication of the following representation of expression of the model:

$$\Psi_{t+1}^Q(u, v) = \frac{\omega(u, v)}{1 - \beta(u, v)} + \alpha(u, v) \sum_{i=0}^{\infty} \beta(u, v)^i h_{t-i+1}$$

We are now able to characterize the option prices.

**Proposition 3.10** *When (3.20) holds, the price at time  $t$  of European call option with payoff  $(S_{t+h} - X)^+$  at time  $t + h$ :*

$$C_t = \exp(-rh)C_{1,t} - \exp(-rh)XC_{2,t}$$

where

$$\begin{aligned} C_{1,t} &= \frac{\exp(rh)}{2} + \int_0^{+\infty} \frac{1}{\pi u} \text{Im} \left[ \exp \left( \Psi_{t,t+h}^Q(1 + iu) - iu \ln \left( \frac{X}{S_t} \right) \right) \right] du \\ C_{2,t} &= \frac{1}{2} + \int_0^{+\infty} \frac{1}{\pi u} \text{Im} \left[ \exp \left( -iu \ln \left( \frac{X}{S_t} \right) + \Psi_{t,t+h}^Q(iu) \right) \right] du \end{aligned}$$

$$\Psi_{t,t+h}^Q(u) = a_h(u)h_t + b_h(u) + \sum_{k=1}^h \left( \beta(d_k)^k \psi_{t-1}(d_k) \right)$$

with  $a_h(u) = \sum_{k=1}^h (\beta(d_k)^{k-1} \alpha(d_k))$ ,  $b_h(u) = \sum_{k=1}^h \left( \frac{1-\beta(d_k)^k}{1-\beta(d_k)} \omega(d_k) \right)$ ,  $\psi_{t-1}(d_k) = \frac{\omega(d_k)}{1-\beta(d_k)} + \alpha(d_k) \left( \sum_{i=0}^{\infty} \beta(d_k)^i h_{t-1-i} \right)$  and the sequence  $(d_k)_{1 \leq k \leq h}$  is defined as follows:

$$\begin{aligned} d_k &= (u, 0) + \sum_{j=k}^{h-1} \beta(d_{j+1})^{j-k} (0, \alpha(d_{j+1})) \text{ for } k \leq h-1 \\ d_h &= (u, 0) \end{aligned} \tag{3.21}$$

We will use these formulas in the empirical section.

## 4 Estimation of generalized affine models

In general we can distinguish between parametric and non-parametric estimation methods. In the early version (2002) of Darolles et al.(2006) a detailed discussion have been made on non-parametric estimation of discrete time affine models. In this section we focus on parametric methods, this implies that we consider the general affine model defined by (2.3) where functions  $\omega(\cdot)$ ,  $\alpha(\cdot)$ ,  $\beta(\cdot)$  and  $\theta(\cdot)$  are specified up to unknown vector of parameters (the dimension of the vector must be finite) which we aim to estimate. Among parametric methods we will consider empirical characteristic function method (ECF) which have been introduced in finance by Singleton (2001), the generalized method of moment (GMM), the quasi-maximum likelihood and the maximum likelihood.

### 4.1 Empirical characteristic function

Our modeling strategy differs fundamentally from the "classic" approach which consists in writing down an equation for each component (it could be the mean, the variance, the intensity of the jump component...) of the conditional distribution of the process of interest. We model directly the conditional characteristic function of the process. For this reason the ECF seems to be the most natural approach to estimate efficiently our generalized affine models. The basic idea behind ECF is to match the theoretical characteristic function given by the model and the empirical characteristic function obtained from data. This approach has several advantages, it avoids difficulties inherent in calculating and maximizing the likelihood. Although the likelihood function can be unbounded, its Fourier transform (which is the characteristic function) is always bounded. The important theorem which establishes the one-one correspondence between characteristic function (CF) and the cumulative distribution function

(CDF) suggests that estimation and inference via the ECF should work as the likelihood-based approaches.

Let first recall the ECF principle in the case of affine models. It means that we are considering  $\beta(u) = 0$  or  $\theta(u) = 0$ . Let denote  $\lambda_0$  the unknown parameter to estimate. We can rewrite the affine model (1.1) as follows:

$$E[\exp(ux_{t+1}) - \exp(\omega(u; \lambda_0) + \alpha(u; \lambda_0)x_t) | x_s, s \leq t] = 0 \quad \forall u. \quad (4.1)$$

This implies that for any weighting function,  $w$  (often termed instruments in the GMM literature), we have:

$$E[(\exp(ux_{t+1}) - \exp(\omega(u; \lambda_0) + \alpha(u; \lambda_0)x_t)) w(x_t, v)] = 0 \quad \forall u, v. \quad (4.2)$$

This leads to continuum of moments restrictions; hence we can estimate by applying the GMM to a continuum of moments restrictions (see Carrasco, Florens (2000)). The ML efficiency is achieved by choosing Carrasco et al. (2002) weighting function, i.e  $w(x_t, v) = e^{vx_t}$ .

Let Consider now the generalized affine model (2.3), the model can be rewrite as follow:

$$\psi_t(u) = \bar{\beta}_j(u) \psi_{t-j}(\theta_j(u)) + \sum_{k=1}^j \bar{\beta}_{k-1}(u) [\omega(\theta_{k-1}(u)) + \alpha(\theta_{k-1}(u))x_{t-k+1}] \quad (4.3)$$

This implies that for any weighting function,  $w(\cdot, \cdot)$

$$E \left[ \left( e^{ux_{t+1}} - e^{\bar{\beta}_t(u; \lambda_0) \psi_0(\theta_t(u; \lambda_0); \lambda_0) + \sum_{k=1}^t \bar{\beta}_{k-1}(u; \lambda_0) [\omega(\theta_{k-1}(u; \lambda_0); \lambda_0) + \alpha(\theta_{k-1}(u; \lambda_0); \lambda_0) x_{t-k+1}]} \right) w(\underline{x}_t, v) \right] = 0 \quad \forall u, v. \quad (4.4)$$

where  $\underline{x}_t = (x_{t-q+1}, \dots, x_t)'$  and  $v = (v_1, \dots, v_q)'$ . Similarly to the affine model results, we specify the following weighting function  $w(\underline{x}_t, v) = e^{\underline{x}_t' v}$ .

## 4.2 Generalized method of moments

One of the advantages of the discrete time affine models is the fact that we can compute unconditional moments of any component of the process of interest. This point have been studied in detail in Feunou and Tedongap (2007). This is an important result for estimation purpose because even when there are some unobserved components in the process of interest, we can still compute the moments of observed components and use them to implement a GMM estimation routine. It turns out that we keep this advantage in the generalized affine model as shown in section 2. In the case of observable variable of interest, there is no need to compute

the unconditional moments. We can use the conditional moments equations which have been derived in section 2. The derivation of conditional moments equations has nothing to do with the affine structure of the model, but it is the consequence of the fact that we have written a model directly on conditional characteristic function. This means that in case of generalized non-affine models (2.4) we can still compute conditional moments equations.

The following moment conditions have been used in the literature by Bollerslev and Zhou (2002) to estimate one factor and two factors stochastic volatility models by means of conditional moments of realized variance:

$$\begin{aligned}
E[x_{t+1} - \mu_{1,0}] &= 0, & E[x_{t+1}^2 - \mu_{2,0}] &= 0 \\
E[(x_{t+1} - \mu_{1,t})x_t] &= 0, & E[(x_{t+1}^2 - \mu_{2,t})x_t] &= 0 \\
E[(x_{t+1} - \mu_{1,t})x_t^2] &= 0, & E[(x_{t+1}^2 - \mu_{2,t})x_t^2] &= 0 \\
E[(x_{t+1} - \mu_{1,t})x_{t-1}] &= 0, & E[(x_{t+1}^2 - \mu_{2,t})x_{t-1}] &= 0 \\
E[(x_{t+1} - \mu_{1,t})x_{t-1}^2] &= 0, & E[(x_{t+1}^2 - \mu_{2,t})x_{t-1}^2] &= 0
\end{aligned}$$

where  $\mu_{1,0} = E(x_{t+1})$ ,  $\mu_{2,0} = E(x_{t+1}^2)$ ,  $\mu_{1,t} = E_t[x_{t+1}]$  and  $\mu_{2,t} = E_t[x_{t+1}^2]$ . We simulate the generalized autoregressive gamma (GARG) model built section (2.1.5) with  $\mu = 2.784E - 05$ ,  $\nu = 0.1394$ ,  $\rho = 0.1125$ ,  $\beta = 0.9227$  and  $\theta = 0.9066$ . These parameters have been obtained by estimating the GARG on realized variance data. For different sample size ( $T$ ) and number of replication ( $N$ ) we estimate the GARG and report in table 1 different statistics (mean, median and root mean square errors (RMSE)) across different replication size. The GMM does well if we consider the longest sample size (4000) and the biggest number of sample (1000).

### 4.3 Maximum and quasi-maximum likelihood method

In general the conditional likelihood can be obtained from the conditional characteristic function using the following inversion formula.

$$\begin{aligned}
f(x_{t+1}|I_t) &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp(-iux_{t+1} + \psi_t(iu)) du \\
&= \frac{1}{\pi} \int_0^{+\infty} \text{Re}[\exp(-iux_{t+1} + \psi_t(iu))] du
\end{aligned} \tag{4.5}$$

Most of the time, we don't have the likelihood in closed form, except for some specific generalized affine processes like VARMA, GARG with  $\theta(u) = u$ . When  $\theta(u) \neq u$ , we estimate the GARG on interest rate and realized variance data by maximizing (4.6) (using numerical integration tools), we obtain significant estimators of parameters  $\beta$  and  $\theta$ . We did not report

results in this paper. To circumvent the numerical integration difficulties, we can derive the first two conditional moments equations (see section 2 for details on conditional moments structure) and then implement a quasi-maximum likelihood. In the case of processes with positive support, we can use the standard gamma density (with two parameters), a gaussian density being used when the support is the whole real line. Finally we run the same Monte-Carlo exercise as in section 4.2, but we use the QMLE method (with gamma density) instead of the GMM. Results are summarized in table 2. Compared to GMM, the QMLE have some difficulties in estimating parameters  $\nu$  and  $\mu$ .

## 5 Three Empirical Examples

This section provides three empirical examples developed in companion papers.

### 5.1 The Term Structure of Realized Risk

This example is studied in Feunou and Meddahi (2007a). There are two goals. The first one is to model the joint dynamics of the returns and the realized variance. The second goal is to compute the term structure of the value-at-risk, i.e. to characterize the quantile function of the aggregated returns,  $\sum_{i=1}^h r_{t+i}$ , when  $h$  varies.

We consider the daily realized variance computed as the sum of squared intra-daily returns, five-minutes and thirty-minutes returns in our empirical application. The recent literature on volatility shows the importance of such measures. The basic theory on realized volatility assumes that the underlying process is in continuous time and shows that the realized variance converges to the integrated variance when the length of intra-day returns goes to zero. In our empirical analysis, we specify the model in discrete time and we do not make the formal connection between the realized variance and the daily returns. We will specify discrete models, affine or generalized affine, and allow the data to select the best model. We will, however, use some insights from continuous time when we specify the discrete model. In what follows the conditioning information is  $I_t = \sigma(r_\tau, RV_\tau, \tau \leq t)$  where  $r_t$  is the daily returns.

We start our analysis by modeling the realized variance as either an affine process or a generalized one. Consider the affine model given by

$$\psi_t(u) = \log E_t[\exp(uRV_{t+1})] = \omega(u) + \alpha(u)RV_t. \quad (5.1)$$

Given the non-negativity of the realized variance process, we will consider two examples. The

first one corresponds to the Inverse Gaussian case while the second is the Gamma case, which corresponds to the exact discretization of the square-root process, studied in Gouriéroux and Jasiak (2006)):

$$\text{Inverse Gaussian : } \omega(u) = \nu(1 - \sqrt{1 - 2u\mu}), \quad \alpha(u) = \frac{\rho}{\mu}(\exp(1 - \sqrt{1 - 2u\mu}) - 1) \quad (5.2)$$

$$\text{Gamma : } \omega(u) = -\nu \log(1 - u\mu), \quad \alpha(u) = \frac{\rho u}{1 - u\mu} \quad (5.3)$$

When we extend our analysis to the generalized affine case, i.e.,

$$\psi_t(u) = \log E_t[\exp(uRV_{t+1})] = \omega(u) + \alpha(u)RV_t + \beta\psi_{t-1}(u), \quad (5.4)$$

we still consider the same two examples of Inverse-Gaussian and Gamma. We prove in Section 2 that this leads to a proper cumulant function.

We use the maximum likelihood method to estimate the four models (two models on realized variance only, and two on joint realized variance and returns). Joint estimation or estimation on realized variance only yield quite the same estimator for the realized variance dynamic. Also, inverse gaussian and gamma model provide very similar results. For this reason in this paper we report only the estimation of the dynamic of returns conditional on realized variance and the dynamic of realized variance given by the maximization of the joint likelihood of returns and realized variance in the inverse gaussian case. These empirical results are provided in Table 3. The main empirical result is that the coefficient  $\beta$  is non-zero whatever the model or the realized volatility measure (based on five-minutes or thirty-minutes). In particular, the increase of the log-likelihood is substantial when one allows  $\beta$  to be non-zero. Another interesting result is that the inverse Gaussian model describes better the data for the two frequencies.

We now want to specify a joint model for the returns and the realized variance. When one considers a continuous time stochastic volatility model

$$d \log p_u = (a + b\sigma_u^2)du + \sigma_u dW_u$$

and assumes that there is no leverage effect, one gets that the daily return  $r_{t+1} = \log(p_{t+1}) - \log(p_t)$  has the following distribution:

$$r_t | \sigma(p_\tau, \sigma_s, \tau \leq t, s \leq t+1) \sim \mathcal{N}(a + bIV_{t+1}, IV_{t+1}),$$

which suggests the following discrete time model that we study:

$$r_{t+1} | \sigma(r_\tau, RV_\tau, RV_{t+1}, \tau \leq t) \sim \mathcal{N}(a + bRV_{t+1}, c + dRV_{t+1}). \quad (5.5)$$

We assume that  $RV_{t+1}$  follows (5.4) where  $\alpha(u)$  follows either (5.2) or (5.3). By denoting the joint cumulant function of  $(r_{t+1}, RV_{t+1})$  as  $\psi_{r,RV;t}(v, u)$  defined by

$$\psi_{r,RV;t}(v, u) \equiv \log E_t[\exp(vr_{t+1} + uRV_{t+1})],$$

one gets

$$\psi_{r,RV;t}(v, u) = (va + v^2c/2) + \psi_t(vb + v^2d/c + u).$$

Hence, the joint process  $r_t, RV_t$  is indeed a generalized affine process because one has

$$\psi_{r,RV;t}(v, u) = \tilde{\omega}(v, u) + \tilde{\alpha}(v, u)RV_t + \beta\psi_{r,RV;t-1}(v, u), \quad (5.6)$$

where

$$\tilde{\omega}(v, u) = (va + v^2c/2)(1 - \beta) + \omega(vb + v^2d/2 + u) \quad (5.7)$$

$$\tilde{\alpha}(v, u) = \alpha(vb + v^2d/2 + u). \quad (5.8)$$

We compute the term structure of the Value-at-Risk, i.e., we compute the 5%-quantile of

$$\bar{r}_{t+1:t+h} \equiv \frac{1}{\sqrt{h}} \sum_{i=1}^h r_{t+i}.$$

For this purpose, we derive the conditional characteristic function of  $\bar{r}_{t+1:t+h}$  and then we invert it to get the cumulative distribution function. This approach has been used in the affine case and continuous time by Duffie and Pan (2001).

In practice, the value at risk of  $\bar{r}_{t+1:t+h}$  will depend on  $RV_t$  its lagged values. In order to graphically present the results, one needs to choose  $RV_t$ . We proceed by taking from the data three values for  $RV_t$ : a small value (low case), a median one (median case) and a large one (high case). Then, we use the lagged values of each of them to plot the term structure of the value-at-risk (VaR).

Figures 1 to 5 present and compare Affine and Generalized affine term structure of the value-at-risk. Figure 3 shows that in a low variance day, the VaR increases with the maturity and that the affine model overestimates the VaR. In contrast, in a high or median volatility day, affine model overestimates the VaR for lowest maturity and underestimates it for longer maturities. Underestimation of the VaR could lead to important risk management problems; see Feunou and Meddahi (2007a) for more discussions. Likewise, we show in Feunou and Meddahi (2007a) that it is useful to consider realized variances, i.e., we did the same approach with the Heston and Nandi (2000) daily model and show that the model with realized volatility is the best one. We also provide in Feunou and Meddahi (2007a) the term structure of another risk measure called the expected shortfall.

## 5.2 No-Arbitrage VARMA Term Structure Models

This example hinges on Feunou (2007) where no-arbitrage VARMA term structure models with macroeconomic variables are studied in detail. Ang and Piazzesi (2003) argued that macroeconomic variables add something new in traditional three factors term structure models. One motivation of generalized affine models studied in this paper is the possibility that they offer to reduce the dimension of the state vector usually used in affine models. In particular we argue that an affine model (of order (1,1)) on a vector of dimension  $n + 1$  is less parsimonious and would yield poor out-of sample forecast compared to a generalized affine model (of order (1,1)) on  $n$  components of the same vector. In the context of this example, we will compare for instance a VARMA model (recall that a VARMA is a generalized affine model, see section 1 and 2 for full details) on three observed yields to a VAR on the same three observed yields and one macroeconomic variable. We will run an out-of-sample forecasting exercise to compare the considered models. In the literature the "Nelson-Siegel with AR(1) factor dynamic" is known as one of the best performer in forecasting the entire yield curve at different horizons (see Diebold and Li (2006) for more details). Our competitors are then the "Nelson-Siegel with AR(1) factor dynamic" and the random walk model (see Duffee(2002) where the random walk model is shown to provide better out-of-sample forecasting results compared to a bunch of affine models).

The state vector is denoted by  $Z_{t+1}$ , we consider the following VARMA(1,1) dynamic under the historical probability measure  $P$ .

$$Z_{t+1} = \mu + \phi Z_t + \Sigma (\varepsilon_{t+1} - \Theta \varepsilon_t), \quad \varepsilon \sim i.i.d. \mathcal{N}(0, I).$$

As shown in section 1 and 2, this model is part of the generalized affine class. Indeed the conditional cumulant function of  $Z_{t+1}$  denoted  $\psi_t(u)$  satisfies the following recursive relation:

$$\psi_t(u) = \omega(u) + \alpha(u)' Z_t + \psi_{t-1}(\theta(u))$$

with

$$\omega(u) = u' \mu + \frac{1}{2} u' \Sigma (I_4 - \Theta \Theta') \Sigma' u, \quad \alpha(u) = u' (\phi - \Sigma \Theta \Sigma^{-1}), \quad \theta(u) = (\Sigma \Theta \Sigma^{-1})' u$$

The specification of the pricing kernel is similar the one used in Ang and Piazzesi (2003),

$$M_{t+1} = \exp \left( -y_t^{(1)} - \frac{1}{2} \lambda_t^\Gamma \lambda_t - \lambda_t^\Gamma \varepsilon_{t+1} \right)$$

To maintain the tractability of the model, affine price of risk ( $\lambda_t$ ) is often used  $\lambda_t = \lambda_0 + \lambda_1 Z_t$ , (again see Ang and Piazzesi (2003)). In the context of our VARMA model, Feunou(2007) discuss this specification and add the conditional expectation of state vector  $E_t(Z_{t+1})$ , i.e.,

$$\lambda_t = \lambda_0 + \lambda_1 Z_t + \lambda_2 E_t(Z_{t+1}) \quad (5.9)$$

Several interpretation can be given to equation (5.9). First, we can reformulate it as follows

$$\begin{aligned} \lambda_t &= \lambda_0 + \lambda_1 Z_t + \lambda_2 (\mu + (\phi - \Sigma\Theta\Sigma^{-1}) Z_t + \Sigma\Theta\Sigma^{-1} E_{t-1}(Z_t)) \\ &= \lambda_0^* + \lambda_1^* Z_t + \lambda_2^* E_{t-1}(Z_t) \end{aligned}$$

where

$$\lambda_0^* = \lambda_0 + \lambda_2 \mu, \quad \lambda_1^* = \lambda_1 + \lambda_2 (\phi - \Sigma\Theta\Sigma^{-1}), \quad \text{and} \quad \lambda_2^* = \lambda_2 \Sigma\Theta\Sigma^{-1}.$$

Thus, the parameter  $\lambda_2$  captures the past information impact on the current market price of risk. Another way of rewriting the price of risk is to express it in terms of the expected variable  $E_{t-1}(Z_t)$  and the unexpected news  $\Sigma\varepsilon_t$ :

$$\lambda_t = \lambda_0^* + \lambda_1^* \Sigma\varepsilon_t + (\lambda_1^* + \lambda_2^*) E_{t-1}(Z_t)$$

Feunou (2007) shows that bond yields (with maturity  $n$ ) are no longer affine of the state variable  $Z_t$ , but are rather affine function of the state variable  $Z_t$  and it past conditional expectation  $E_{t-1}(Z_t)$ , i.e.

$$y_t^{(n)} = a_n + b_{1,n}^\top Z_t + b_{2,n}^\top E_{t-1}(Z_t). \quad (5.10)$$

Another representation derived in Feunou (2007) is

$$y_t^{(n)} = a_n + (b_{1,n} + b_{2,n})^\top E_{t-1}(Z_t) + b_{1,n}^\top \Sigma\varepsilon_t, \quad (5.11)$$

where the coefficients  $a_n$ ,  $b_{1,n}$ , and  $b_{2,n}$  are given in Feunou (2007).

The estimation of the unknown parameters (parameters of the historical distribution and parameters of the price of risk) is done in two steps. The first step estimates the parameters of the historical distribution of the state vector by using the the maximum likelihood method. By taking the parameters of the historical dynamic to their estimated values (results of the first step), we estimate in the second step the pricing kernel's parameters by minimizing the squared difference between the model implied yields and the observed yields (in practice, the maturities are 3, 12, 36, 60 and 120 months). Since there are observed yields in the state vector, Feunou (2007) (see also Ang et al (2006)) argued that the second step is a constrained

optimization problem. The yields data used have been obtained from unsmoothed Fama-Bliss forward rates (see Diebold and Li(2006) for full details on construction and description of these yields data) and they are the same used in Diebold and li(2006). The Macroeconomic Data are the 2 factors termed "inflation" and "real activity" used in Ang and Piazzesi (2006). We estimate and forecast recursively, using data from 1985:1 to the time that the forecast is made, beginning in 1994:1 and extending through 2000:12. In Tables (4) to (6) we compare h-month-ahead out-of sample forecasting results from "Nelson–Siegel with AR(1) factor dynamic" to no arbitrage VAR and VARMA models, for maturities of 3, 12, 36, 60 and 120 months, and forecast horizons of  $h = 1, 6$  and 12 months.

To summarize the RMSE results, the best model in DL(2006) ("Nelson–Siegel with AR(1) factor dynamic model") performs better only at horizon 1 for 3 and 120 months yield. In general the best performer is the VARMA(1,1) model on 1, 24, and 96 months yield, except for the 10 years yield at horizon 6 months where the VAR(1) model on 1, 24, and 96 months yield and inflation performs better.

In conclusion, by using observable variables in either a no-arbitrage VAR or VARMA model we can do better than "Nelson–Siegel with AR(1) factor dynamic" in forecasting the entire yield curve at any horizon. Macro-economic factors add new information which are not contained into affine yields only model, but we can cope with these macro-economic factors by implementing a no arbitrage generalized affine model (here the VARMA) on the classic three factors model.

### 5.3 Realized Option Pricing model

This subsection hinges on Feunou, Christoffersen, Jacobs and Meddahi (2007). We used the model developed in the first empirical example and used the option pricing formulas derived in Section 3.2 where  $h_t$  equals the realized variance  $RV_t$ . We model jointly the dynamics of the return  $r_t$  and realized variance  $RV_t$  in the same way as in section 5.1, with a slight modification of the distribution of the stock log-returns  $r_t$  conditional on realized variance  $RV_t$ . Following Christoffersen et al (2006), Feunou (2006) and Feunou and Tedongap (2007), we used a skewed inverse Gaussian distribution, which nests the normal distribution. This extension is empirically important.

The model is given by

$$r_{t+1} | \sigma(r_\tau, RV_\tau, RV_{t+1}, \tau \leq t) \sim a + bRV_{t+1} - \eta(c + dRV_{t+1}) + \frac{1}{\eta}y_{t+1}. \quad (5.12)$$

with  $y_{t+1} \sim \mathcal{IG}(\eta^2(c + dRV_{t+1}))$ .  $\mathcal{IG}$  means the standard inverse gaussian distribution. The conditional cumulant function of the return  $r_{t+1}$  conditional on  $I_t$  and  $RV_{t+1}$  is given by

$$E[\exp(ur_{t+1})|RV_{t+1}, I_t] = \exp(\omega_0(u) + \alpha_0(u)RV_{t+1})$$

with

$$\omega_0(u) = u(a - c\eta) + c\eta^2 \left(1 - \sqrt{1 - \frac{2u}{\eta}}\right), \text{ and } \alpha_0(u) = u(b - d\eta) + d\eta^2 \left(1 - \sqrt{1 - \frac{2u}{\eta}}\right).$$

In the affine case, the conditional cumulant function of  $RV_{t+1}$  given  $I_t$  is given by (5.1) where  $\omega(u)$  and  $\alpha(u)$  are defined either by (5.2) for the inverse gaussian case or by (5.3) for the gamma case. We extend this affine case to the generalized affine of order (1,2) as follows

$$\psi_t(u) = \log E_t[\exp(uRV_{t+1})] = \omega(u) + \alpha(u)RV_t + \beta_1\psi_{t-1}(u) + \beta_2\psi_{t-2}(u) \quad (5.13)$$

Consequently, the joint cumulant function of  $(r_{t+1}, RV_{t+1})$  given  $I_t$  is

$$\psi_{r,RV;t}(v, u) = \omega_0(v) + \psi_t(u + \alpha_0(v)).$$

Eq. (5.13) implies that the joint process  $(r_t, RV_t)$  is a generalized affine process

$$\psi_{r,RV;t}(v, u) = \tilde{\omega}(v, u) + \tilde{\alpha}(v, u)RV_t + \beta_1\psi_{r,RV;t-1}(v, u) + \beta_2\psi_{r,RV;t-2}(v, u) \quad (5.14)$$

with  $\tilde{\omega}(v, u) = \omega_0(v)(1 - \beta_1 - \beta_2) + \omega(u + \alpha_0(v))$  and  $\tilde{\alpha}(v, u) = \alpha(u + \alpha_0(v))$ .

We assume that the generalized affine model is defined under the risk-neutral probability measure. The estimation is done by minimizing the MSE of the implied Black-Scholes volatility from the option (IVMSE) defined as

$$IVMSE = \frac{1}{n} \sum_{i=1}^n (\sigma_i - \sigma_i(\theta))^2$$

where the implied volatilities are obtained as

$$\sigma_i = BS^{-1}(C_i, T_i, X_i, S, r) \quad \text{and} \quad \sigma_i(\theta) = BS^{-1}(C_i(\theta), T_i, X_i, S, r).$$

with  $BS^{-1}$  being the inverse of the Black-Scholes formula,  $T_i$  the time to maturity,  $X_i$  the strike price,  $S$  the price of the underlying stocks and  $r$  the riskless interest rate. Figures 6, 7 and 8 represent the daily implied volatility bias, option price bias and implied volatility RMSE. The generalized affine model clearly outperforms the affine model in terms of pricing errors. This result holds whatever the maturity of the moneyness; see Tables 7 and 8.

## 6 Conclusion

The paper extends affine models by introducing moving average type components in the conditional cumulant function. The extension is important theoretically because important models like ARMA are not affine, beside that we show how we can build parsimonious infinite order affine models. The extension is also empirically important as shown in three empirical examples. In particular the term structure exercise shows that a generalized affine model on traditional three factors term structure model forecasts better the entire yield curve than an affine model on three factors and macroeconomic variables. There is an alternative approach that leads to non-Markov affine processes. It uses the conditional Laplace transform of the process  $x_t$  defined as  $\mathcal{L}_t(u) = \exp(\psi_t(u))$  instead of the cumulant function. The traditional affine models are characterized by

$$\mathcal{L}_t(u) = \exp(\omega(u) + \alpha(u)x_t).$$

In a companion paper, we are currently studying the process defined by

$$\mathcal{L}_t(u) = \gamma(u) + \exp(\omega(u) + \alpha(u)x_t) + \beta(u)\mathcal{L}_{t-1}(u).$$

## References

- [1] Ang, A. and M., Piazzesi (2003). A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. *Journal of Monetary Economics* 50 745-787.
- [2] Ang, A., M., Piazzesi and M., Wei (2006). What does the yield curve tell us about GDP growth. *Journal of Econometrics* 131 359-403.
- [3] Barndorff-Nielsen OE. (1997). Normal inverse Gaussian distributions and stochastic volatility modelling. *Scandinavian Journal of Statistics* 24, 1-13.
- [4] Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroscedasticity. *Journal of Econometrics* 31, 307-327.
- [5] Bollerslev, T. and L., Forsberg (2002). Bridging the gap between the distribution of realized (ECU) and ARCH modelling (of the euro): The GARCH-NIG model. *Journal of Applied Econometrics* 17, 535-548.
- [6] Bollerslev, T. and Zhou, H. (2002). Estimating stochastic volatility diffusion using conditional moments of integrated volatility. *Journal of Econometrics, Elsevier, vol. 109(1), pages 33-65, July*
- [7] Carrasco, M. and J-P Florens (2002). Efficient GMM Estimation Using the Empirical Characteristic Function. *Working Paper; Departement of Economics: University of Rochester.*
- [8] Carrasco, M., and J-P., Florens(2000). *Generalization of GMM to a continuum of moment conditions.*, *Econometric Theory*, 16, 2000, 797-834.
- [9] Carrasco, M., M., Chernov, J-P., Florens, and E., Ghysels (2007). Efficient estimation of jump diffusions and general dynamic models with a continuum of moment conditions. *Journal of Econometrics*, 140 529-573
- [10] Christoffersen, P. (1998) valuating interval forecasts. *E International Economic Review* 39, 841.862.
- [11] Christoffersen, P., R., Elkamhi, B., Feunou and K., Jacobs (2006). Option Valuation with Conditional Heteroskedasticity and Non-Normality. *Working Paper, McGill University.*
- [12] Christoffersen, P., S., Heston, and K., Jacobs (2006). Option valuation with conditional skewness. *Journal of Econometrics* 131 253-284.
- [13] Christoffersen, P., and K. Jacobs (2004). Which GARCH Model for Option Valuation?. *Management Sci.* 50. 1204-1221.
- [14] Christoffersen, P., K., Jacobs and Y., Wang (2005). Volatility Components: Affine Restrictions and Non-normal Innovations. *Working Paper; Faculty of Management, McGill University.*

- [15] Cox, J.C., J.E., Ingersoll and S.A. Ross (1985). A theory of the term structure of interest rates. *Econometrica* 53, 385-407.
- [16] Dai, Q. and K., Singleton, (2000). Specification Analysis of affine term structure models. *Journal of Finance* 55, 1943-1978.
- [17] Darolles, S., C., Gouriou, and J., Jasiak (2006). Structural Laplace Transform and Compound Autoregressive Models. *Journal of time serie analysis Vol. 27, No. 4.*
- [18] Diebold, F., M., Piazzesi, and D., Rudebusch (2005). Modeling Bond Yields in Finance and Macroeconomics. *American Economic Review P&P Volume 95, Issue 2, May 2005, pp. 415-420.*
- [19] Diebold, F., D., Rudebusch, and B., Aruoba (2006). The macroeconomy and the yield curve: a dynamic latent factor approach. *Journal of Econometrics* 131 309-338.
- [20] Diebold, F. X., T. A., Gunther and T. S., Tay (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review* 39, 863. 883.
- [21] Diebold, F and C., Li (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics* 130 337-364.
- [22] Duan, J. C. (1995). The GARCH option pricing model. *Math. Finance* 5, 13-32.
- [23] Duffee, G. R. (2002). Term Premia and Interest Rate Forecasts in Affine Models. *The Journal of Finance, Vol. 57, No. 1. pp. 405-443.*
- [24] Duffie, D., J., Pan and K., Singleton (2000). Transform Analysis and Asset Pricing for affine jump diffusions. *Econometrica.* 68, 1343-1376.
- [25] Duffie, D., and , R., Kan(1996). *A yield-Factor model of interest rates* Mathematical Finance. 6, 379-406.
- [26] Engle, R. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation. *Econometrica* 50 987-1008.
- [27] Feller, W. (1968). An introduction to probability theory and its applications. *John Wiley & Sons.*
- [28] Feuerverger, A. and R.A., Mureika (1977). The Empirical Characteristic Function and its Applications. *The annals of Statistics.* 5, 88-97.
- [29] Feuerverger, A., and P., McDunnough (1981). On the Efficiency of Empirical Characteristic Function Procedure. *Journal of the Royal Statistical Society. Series B.* 43, 20-27.
- [30] Feunou, B. (2007). No-arbitrage VARMA term structure models with macroeconomic variables. *Working Paper, Université de Montréal.*
- [31] Feunou, B., and N., Meddahi (2007). Realized term structure of risk. *Work in Progress.*

- [32] Feunou, B., P., Christoffersen, K., Jacobs, and N., Meddahi (2007). Realized Option pricing models. *Work in progress*.
- [33] Feunou, B., and R. Tedongap (2007). Affine Stochastic Skewness Models. *Working Paper, Université de Montréal*.
- [34] Gouriéroux, C. (2006). Continuous Time Wishart Process for Stochastic Risk. *25, Issue 2 & 3, pages 177 - 217*.
- [35] Gouriéroux, C., and J.Jasiak (2002). Value at Risk . *Handbook of financial Econometric*.
- [36] Gouriéroux, C., and J., Jasiak (2006). Autoregressive Gamma Processes. *Journal of Forecasting. 25, 129-152*.
- [37] Gouriéroux, C., A., Monfort and V., Polimenis (2002). Affine term structure models. *CREST DP*.
- [38] Gouriéroux, C. and A., Monfort (2006). Econometric Specifications of Stochastic Discount Factor Models. *Journal of Econometrics 136 509-530*.
- [39] Hansen, L. P.(1982). Large Sample properties of generalized method of moment estimators. *Econometrica 50, 1029-1053*.
- [40] Heston, S. (1993). A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. *Review of Financial Studies, 6, 327-343*.
- [41] Heston, S. and S. Nandi (2000). A Closed-Form GARCH Option Pricing Model. *Review of Financial Studies, 13, 585-626*.
- [42] Johnson, N.L., S., Kotz and N., Balakrishnan. Continuous Univariate distributions. *Volume1. Second Edition. Wiley series in probability and mathematical statistics*.
- [43] Monfort, A. and F. Pegoraro (2006). Switching VARMA Term Structure Models. *Journal of Financial Econometrics 5(1): 105-153*.
- [44] Monfort, A. and F., Pegoraro (2006). Multi-Lag Term Structure Models with Stochastic Risk Premia. Working paper, CREST.
- [45] Lukacs, E. (1970). Characteristic Functions. *Griffin 1820 London*.
- [46] Singleton, K. (2001). Estimation of affine asset pricing models using the empirical characteristic function. *Journal of Econometrics 102, 111-141*.
- [47] Vasicek, O. (1977). An equilibrium characterization of term structure. *Journal of Financial Economics 5, 177-188*.
- [48] Yu, J. (2004). Empirical Characteristic Function Estimation and its applications. *Econometric Reviews, 23 (2) :93-123*.

## Appendix A

In this appendix, we build a generalized affine model of type I where the function  $\beta(\cdot)$  varies. Let us consider a positive process  $X_t$  with conditional cumulant function  $\Psi_t$ .

$$E_t [\exp (u X_{t+1})] = \exp (\Psi_t (u))$$

We define  $\Psi_t (u)$  recursively as follows:

$$\begin{aligned} \Psi_0 (u) &= \omega (u) + a_0 (u, X_0) \\ \Psi_1 (u) &= \omega (u) + a_0 (u, X_1) + a_1 (u, X_0) \\ \Psi_2 (u) &= \omega (u) + a_0 (u, X_2) + a_1 (u, X_1) + a_2 (u, X_0) \end{aligned}$$

and generally, we have

$$\Psi_t (u) = \omega (u) + \sum_{i=0}^t a_i (u, X_{t-i}) \tag{A.1}$$

The first issue is to give some conditions on sequence functions  $a_i(u, x)$  and  $\omega(u)$  such that  $\Psi_t(u)$  is a well defined cumulant function.

If  $\omega(u)$  and  $a_i(u, x)$  are cumulant functions  $\forall i$ , then  $\Psi_t(u)$  is a well defined cumulant function. Indeed, the sum of cumulant function is a cumulant function.

Consequently we will choose  $\omega(u)$  and  $a_i(u, x)$  such that they will be always cumulant functions. Another consequence is the fact that we can write  $X_{t+1}$  as follows

$$X_{t+1} = \eta_{t+1} + \sum_{i=0}^t Z_{i,t+1}$$

where  $\eta_{t+1}$  and  $Z_{i,t+1}$  are mutually conditionally independent with cumulant function  $\omega(u)$  and  $a_i(u, X_{t-i})$ . This give us a simple approach to simulate  $X_{t+1}$ .

The final goal is to rewrite definition of  $\Psi_t(u)$  given by (A.1) recursively. To achieve this goal the following expression is given to cumulant function  $a_i(u, x)$

$$\begin{aligned} a_i (u, x) &= P_i (x) [\exp (a (u) i + b (u)) - 1] \\ P_i (x) &= \exp (\lambda_0 + \lambda_1 i) x \end{aligned} \tag{A.2}$$

As it was the case with  $\Psi_t(u)$ , we need to make sure that (A.2) is a valid cumulant function. This is done using Lemma 5.4.1 of Lukacs (1970) (page 111) where it is shown that  $p(g(u) - 1)$  is an infinitely divisible cumulant function whenever  $g(u)$  is a characteristic function and  $p > 0$ .

Thus if  $a(u)$  and  $b(u)$  are cumulant functions and  $X$  a positive process, then  $a_i(u, x)$  is a cumulant function. Since process  $X_t$  is built using cumulant generating function, it is hard to simulate. We give an answer in the following lines. Proposition 6.1 shows how a random variable with cumulant function  $a_i(u, x)$  can be simulated.

**Proposition 6.1**  $p(g(u) - 1)$  is the cumulant function of  $Z$  iff

$$Z = \sum_{n=0}^N Y_n$$

where random variables  $N$  and  $Y_n$  are mutually independent,  $N$  follows Poisson distribution of parameter  $p$  and the moment generating function of  $Y_n$  is  $g(u)$ .

Since  $\Psi_t(u)$  is the conditional cumulant function of  $X_{t+1}$  (which is a positive random variable), we must then choose  $\omega(u)$ ,  $a(u)$  and  $b(u)$  such that  $\Psi_t(u)$  is a cumulant function of a positive random variable. The following proposition addresses this issue.

**Proposition 6.2** If  $a(u)$ ,  $b(u)$  and  $\omega(u)$  are cumulant functions of positive random variable, then  $\Psi_t(u)$  is a well defined conditional cumulant function of positive random variable  $X_{t+1}$

We are now ready to write  $\Psi_t(u)$  recursively.

$$\begin{aligned} \Psi_t(u) &= \omega(u) + \sum_{i=0}^t P_i(X_{t-i}) [\exp(a(u)i + b(u)) - 1] \\ &= \omega(u) + \sum_{i=0}^t P_i(X_{t-i}) \exp(a(u)i + b(u)) - \sum_{i=0}^t P_i(X_{t-i}) \\ &= \omega(u) + \sum_{i=0}^t \exp((a(u) + \lambda_1)i + \lambda_0 + b(u)) X_{t-i} - \sum_{i=0}^t \exp(\lambda_0 + \lambda_1 i) X_{t-i} \end{aligned}$$

**Proposition 6.3**  $\omega(u)$  can always be reformulated as following

$$\omega(u) = \frac{c(u)}{1 - \exp(a(u) + \lambda_1)} - \frac{c(0)}{1 - \exp(\lambda_1)}$$

As shown below, the proof of Proposition 6.3 is a direct consequence of the fact that  $\omega(u)$  is a cumulant function.

We can then rewrite  $\Psi_t(u)$  as following.

$$\Psi_t(u) = f_t(u) - f_t(0)$$

with

$$\begin{aligned} f_t(u) &= \frac{c(u)}{1 - \exp(a(u) + \lambda_1)} + \sum_{i=0}^t \exp((a(u) + \lambda_1)i + \lambda_0 + b(u)) X_{t-i} \\ &= \frac{c(u)}{1 - \exp(a(u) + \lambda_1)} + \exp(\lambda_0 + b(u)) \sum_{i=0}^t \exp(a(u) + \lambda_1)^i X_{t-i} \end{aligned}$$

**Proposition 6.4**  $f_{t+1}(u)$  evolves recursively as follows:

$$f_{t+1}(u) = c(u) + \exp(\lambda_0 + b(u)) X_{t+1} + \exp(a(u) + \lambda_1) f_t(u)$$

An immediate consequence of Proposition 6.4 is the recursive formulation of  $\Psi_t(u)$  given by proposition 6.5.

**Proposition 6.5**

$$\Psi_t(u) = \omega_0(u) + \alpha_1(u) f_t(0) + \alpha_2(u) f_{t-1}(0) + \beta(u) \Psi_{t-1}(u) \quad (\text{A.3})$$

where

$$\begin{aligned} \omega_0(u) &= c(u) - c(0)e^{b(u)} \\ \alpha_1(u) &= e^{b(u)} - 1 \\ \alpha_2(u) &= e^{\lambda_1} [e^{a(u)} - e^{b(u)}] \\ \beta(u) &= e^{\lambda_1 + a(u)} \end{aligned}$$

Note that in the right hand side of equation (A.3), we have  $f_t(0)$  instead of  $X_t$ . For this reason the conditional cumulant generating function of  $f_t(0)$  is evaluated. The joint conditional moment generating function of  $X_{t+1}$  and  $f_{t+1}(0)$  is:

$$\begin{aligned} E_t [\exp(uX_{t+1} + vf_{t+1}(0))] &= E_t [\exp(uX_{t+1} + v(c(0) + e^{\lambda_0} X_{t+1} + e^{\lambda_1} f_t(0)))] \\ &= \exp(vc(0) + ve^{\lambda_1} f_t(0) + \Psi_t(u + ve^{\lambda_0})) \end{aligned}$$

Thus if we denote  $\Psi_t^c(u, v) = \ln(E_t [\exp(uX_{t+1} + vf_{t+1}(0))])$ , we have

$$\begin{aligned} \Psi_t^c(u, v) &= vc(0) + ve^{\lambda_1} f_t(0) + \Psi_t(u + ve^{\lambda_0}) \\ &= vc(0) + ve^{\lambda_1} f_t(0) + \omega_0(u + ve^{\lambda_0}) + \alpha_1(u + ve^{\lambda_0}) f_t(0) \\ &\quad + \alpha_2(u + ve^{\lambda_0}) f_{t-1}(0) + \beta(u + ve^{\lambda_0}) \Psi_{t-1}(u + ve^{\lambda_0}) \\ &= vc(0) + ve^{\lambda_1} f_t(0) + \omega_0(u + ve^{\lambda_0}) + \alpha_1(u + ve^{\lambda_0}) f_t(0) \\ &\quad + \alpha_2(u + ve^{\lambda_0}) f_{t-1}(0) + \beta(u + ve^{\lambda_0}) [\Psi_{t-1}^c(u, v) - vc(0) - ve^{\lambda_1} f_{t-1}(0)] \end{aligned}$$

The whole expression of  $\Psi_t^c(u, v)$  is summarized in the following proposition.

**Proposition 6.6**

$$\Psi_t^c(u, v) = W(u, v) + A_1(u, v) f_t(0) + A_2(u, v) f_{t-1}(0) + B(u, v) \Psi_{t-1}^c(u, v)$$

where

$$\begin{aligned} W(u, v) &= vc(0) (1 - \beta(u + ve^{\lambda_0})) + \omega_0(u + ve^{\lambda_0}) \\ A_1(u, v) &= ve^{\lambda_1} + \alpha_1(u + ve^{\lambda_0}) \\ A_2(u, v) &= \alpha_2(u + ve^{\lambda_0}) - ve^{\lambda_1} \beta(u + ve^{\lambda_0}) \\ B(u, v) &= \beta(u + ve^{\lambda_0}) \end{aligned}$$

In conclusion the vector  $(X_{t+1}, f_{t+1}(0))$  is a generalized affine of order (2,1), implying a univariate generalized affine for  $f_{t+1}(0)$  as stated in the following corollary.

**Corollary 1** *Notice that by imposing  $u = 0$  we have a generalized affine model of order (2,1) for  $f_t(0)$ .*

Indeed

$$\begin{aligned} E_t [\exp (v f_t (0))] &\equiv \exp \left( \Psi_t^f (v) \right) \\ &= \exp \left( \Psi_t^c (0, v) \right) \end{aligned}$$

with

$$\Psi_t^c (0, v) = W (0, v) + A_1 (0, v) f_t(0) + A_2 (0, v) f_{t-1}(0) + B (0, v) \Psi_{t-1}^c (0, v)$$

$$\psi_t^f (u) = \omega^f (u) + \alpha_1^f (u) f_t + \alpha_2^f (u) f_{t-1} + \beta^f (u) \psi_{t-1}^f (u)$$

where

$$\begin{aligned} \psi^f (u) &= \ln [E_t (\exp (u f_{t+1}))] \\ \beta^f (u) &= \mu f (u) \\ \alpha_1^f (u) &= u \mu + g (u) - 1 \\ \alpha_2^f (u) &= \mu ((1 - \mu u) f (u) - g (u)) \\ \omega^f (u) &= (1 - \beta^f (u)) \left( u \rho + \frac{\rho}{1 - \mu} + h (u) \right) - \rho g (u) \end{aligned}$$

with  $\mu = e^{\lambda_1}$ ,  $f(u) = e^{a(ue^{\lambda_0})}$  a moment generating function of a positive infinitely divisible random,  $g(u) = e^{b(ue^{\lambda_0})}$  a moment generating function of a positive random variable and  $h(u) = \omega(ue^{\lambda_0})$  a cumulant generating function of a positive random variable.

Hence we get the result stated in proposition (2.2).

**Proposition 6.7** *Generally, for any given  $s$ ,  $(f_t(0), f_t(s))$  is a generalized affine of order (2,1)*

We can restrict  $f_t(0)$  to be positive by just imposing  $c(0)$  to be positive and considering positive initial value  $f_0(0)$ . On the other hand  $f_t(0)$  can take any sign if any restriction is made on  $c(0)$  and  $f_0(0)$ . All these assertions are consequences of the recursive definition of  $f_{t+1}(0)$

$$f_{t+1}(0) = c(0) + \exp(\lambda_0) X_{t+1} + \exp(\lambda_1) f_t(0)$$

Since  $X_{t+1}$  is a positive random variable, if  $f_t(0) \geq 0$  and  $c(0) \geq 0$ , then  $f_{t+1}(0) \geq 0$ .  $c(0)$  is an undetermined parameter with undetermined sign. This implies that if the sign of  $c(0)$  and  $f_t(0)$  are undetermined then  $f_{t+1}(0)$ 's sign is also undetermined.

Generalized affine of order (1,1) (for  $f_t(0)$ ) can be obtained by restricting functions  $a$  and  $b$  to satisfy  $A_2(0, v) = 0$ . Solving  $A_2(0, v) = 0$  implies

$$b(v) = a(v) + \ln(1 - ve^{\lambda_1 - \lambda_0})$$

which in fact is equivalent to imposing the following restriction to the moment generating function  $f$ .

$$f(u) = \frac{g(u)}{1 - \mu u}$$

We then get the result of proposition (2.1).

**Proof of Proposition 6.1:** The proof is quite easy, in fact it is done by realizing that if  $G$  is the distribution function corresponding to characteristic function  $g$  (or moment generating function), then  $F = e^{-p} \sum_0^\infty \frac{p^n}{n!} G^{n*}$  is the distribution function corresponding to characteristic function (or moment generating function  $\exp(p(g(u) - 1))$ ). In this expression  $G^{n*}$  means the convolution of  $n$  identical distribution function  $G$ . The simulation of random variable corresponding to distribution function  $F$  is also easy to deal with. Let consider a sequence of *iid* random variable  $(Y_i)_{i=1,2,\dots}$ , and a discrete random variable  $N$  which is independent to  $(Y_i)_{i=1,2,\dots}$  and which follows a Poisson distribution with parameter  $p$ . The following random variable  $X$  has  $F$  as distribution function:

$$Z = \sum_{n=0}^N Y_n$$

where  $Y_0$  is a constant.

**Proof of Proposition 6.2:** The result is the consequence of the fact that  $p(g(u) - 1)$  is the cumulant function of positive random variable when  $g(u)$  is the moment generating function of a positive random variable. This result is deduced from the previous Proposition, indeed since  $p(g(u) - 1)$  is the cumulant function of  $Z = \sum_{n=0}^N Y_n$ , and  $g(u)$  the moment generating function of  $Y_n$ .  $Y_n \geq 0 \Rightarrow Z \geq 0$

**Proof of Proposition 6.3:** In fact, for any given choice of a cumulant function of positive random variable  $\omega(u)$ , choose  $c(u)$  as follows

$$c(u) = (1 - \exp(a(u) + \lambda_1)) \left[ \omega(u) + \frac{\delta}{1 - \exp(\lambda_1)} \right]$$

for any real  $\delta$ . Since  $\omega(u)$  and  $a(u)$  are a cumulant functions, thus  $\omega(0) = a(0) = 0$ , which implies that

$$c(0) = \delta$$

**Proof of Proposition 6.4:** Indeed

$$\begin{aligned}
& c(u) + \exp(\lambda_0 + b(u)) X_{t+1} + \exp(a(u) + \lambda_1) f_t(u) \\
= & c(u) + \exp(\lambda_0 + b(u)) X_{t+1} + \frac{\exp(a(u) + \lambda_1) c(u)}{1 - \exp(a(u) + \lambda_1)} + \exp(\lambda_0 + b(u)) \sum_{i=0}^t \exp(a(u) + \lambda_1)^{i+1} X_{t-i} \\
= & \frac{c(u)}{1 - \exp(a(u) + \lambda_1)} + \exp(\lambda_0 + b(u)) \sum_{i=0}^{t+1} \exp(a(u) + \lambda_1)^i X_{t+1-i} \\
= & f_{t+1}(u)
\end{aligned}$$

**Proof of Proposition 6.5:**

$$f_{t+1}(u) = c(u) + \exp(\lambda_0 + b(u)) X_{t+1} + \exp(a(u) + \lambda_1) f_t(u)$$

and

$$\Psi_t(u) = f_t(u) - f_t(0)$$

imply that

$$\begin{aligned}
\Psi_t(u) + f_t(0) &= c(u) + e^{b(u)} [f_t(0) - c(0) - e^{\lambda_1} f_{t-1}(0)] \\
&\quad + e^{\lambda_1 + a(u)} [\Psi_{t-1}(u) + f_{t-1}(0)]
\end{aligned}$$

**Proof of Proposition 6.7:** Indeed

$$\begin{aligned}
E_t [\exp(uf_{t+1}(0) + vf_{t+1}(s))] &= E_t \left[ \exp \left( \begin{array}{l} u(c(0) + e^{\lambda_0} X_{t+1} + e^{\lambda_1} f_t(0)) \\ +v(c(s) + e^{\lambda_0 + b(s)} X_{t+1} + e^{\lambda_1 + a(s)} f_t(s)) \end{array} \right) \right] \\
&= \exp \left( \begin{array}{l} uc(0) + vc(s) + ue^{\lambda_1} f_t(0) \\ +ve^{\lambda_1 + a(s)} f_t(s) + \Psi_t(ue^{\lambda_0} + ve^{\lambda_0 + b(s)}) \end{array} \right) \\
&\equiv \exp(\Psi_t^{c,s}(u, v))
\end{aligned}$$

$$\begin{aligned}
\Psi_t^{c,s}(u, v) &= uc(0) + vc(s) + ue^{\lambda_1} f_t(0) + ve^{\lambda_1 + a(s)} f_t(s) + \Psi_t(ue^{\lambda_0} + ve^{\lambda_0 + b(s)}) \\
&= uc(0) + vc(s) + ue^{\lambda_1} f_t(0) + ve^{\lambda_1 + a(s)} f_t(s) + \omega_0 (ue^{\lambda_0} + ve^{\lambda_0 + b(s)}) \\
&\quad + \alpha_1 (ue^{\lambda_0} + ve^{\lambda_0 + b(s)}) f_t(0) + \alpha_2 (ue^{\lambda_0} + ve^{\lambda_0 + b(s)}) f_{t-1}(0) \\
&\quad + \beta (ue^{\lambda_0} + ve^{\lambda_0 + b(s)}) \Psi_{t-1}(ue^{\lambda_0} + ve^{\lambda_0 + b(s)}) \\
&= uc(0) + vc(s) + ue^{\lambda_1} f_t(0) + ve^{\lambda_1 + a(s)} f_t(s) + \omega_0 (ue^{\lambda_0} + ve^{\lambda_0 + b(s)}) \\
&\quad + \alpha_1 (ue^{\lambda_0} + ve^{\lambda_0 + b(s)}) f_t(0) + \alpha_2 (ue^{\lambda_0} + ve^{\lambda_0 + b(s)}) f_{t-1}(0) \\
&\quad + \beta (ue^{\lambda_0} + ve^{\lambda_0 + b(s)}) \left[ \Psi_{t-1}^{c,s}(u, v) - uc(0) - vc(s) - ue^{\lambda_1} f_{t-1}(0) - ve^{\lambda_1 + a(s)} f_{t-1}(s) \right]
\end{aligned}$$

Hence

$$\Psi_t^{c,s}(u, v) = W^s(u, v) + A_1^s(u, v)' f_t(0, s) + A_2^s(u, v)' f_{t-1}(0, s) + B^s(u, v) \Psi_{t-1}^{c,s}(u, v)$$

where

$$f_t(0, s) = \begin{pmatrix} f_t(0) \\ f_t(s) \end{pmatrix}$$

$$W^s(u, v) = (uc(0) + vc(s)) \left(1 - \beta \left(ue^{\lambda_0} + ve^{\lambda_0+b(s)}\right)\right) + \omega_0 \left(ue^{\lambda_0} + ve^{\lambda_0+b(s)}\right)$$

$$A_1^s(u, v) = \begin{pmatrix} ue^{\lambda_1} + \alpha_1 \left(ue^{\lambda_0} + ve^{\lambda_0+b(s)}\right) \\ ve^{\lambda_1+a(s)} \end{pmatrix}$$

$$A_2^s(u, v) = \begin{pmatrix} \alpha_2 \left(ue^{\lambda_0} + ve^{\lambda_0+b(s)}\right) - ue^{\lambda_1} \beta \left(ue^{\lambda_0} + ve^{\lambda_0+b(s)}\right) \\ -ve^{\lambda_1+a(s)} \beta \left(ue^{\lambda_0} + ve^{\lambda_0+b(s)}\right) \end{pmatrix}$$

$$B^s(u, v) = \beta \left(ue^{\lambda_0} + ve^{\lambda_0+b(s)}\right)$$

## Appendix B

This appendix provides the proofs of Section 2 and Section 3. **Proof of Proposition 2.4:**

From (2.3) we have

$$\psi_t(u) = \bar{\beta}_j(u) \psi_{t-j}(\theta_j(u)) + \sum_{k=1}^j \bar{\beta}_{k-1}(u) [\omega(\theta_{k-1}(u)) + \alpha(\theta_{k-1}(u))x_{t-k+1}]$$

$$\begin{aligned} \exp(V_{t,h}(\bar{u}_h)) &= E_t \left[ \exp \left( \sum_{i=1}^h u_i x_{t+i} \right) \right] \\ &= E_t \left[ \exp \left( \sum_{i=1}^{h-1} u_i x_{t+i} + \psi_{t+h-1}(u_h) \right) \right] \\ &= E_t \left[ \exp \left( \sum_{i=1}^{h-1} u_i x_{t+i} \right) \times \right. \\ &\quad \left. \exp \left( \bar{\beta}_h(u_h) \psi_{t-1}(\theta_h(u_h)) \right) \right. \\ &\quad \left. \times \exp \left( + \sum_{k=1}^h \bar{\beta}_{k-1}(u_h) [\omega(\theta_{k-1}(u_h)) + \alpha(\theta_{k-1}(u_h))x_{t+h-k}] \right) \right] \\ &= \exp \left( \bar{\beta}_h(u_h) \psi_{t-1}(\theta_h(u_h)) + \bar{\beta}_{h-1}(u_h) \alpha(\theta_{h-1}(u_h))x_t \right) \times \\ &\quad \times E_t \left[ \exp \left( \sum_{i=1}^{h-1} u_i^{(h-1)} x_{t+i} \right) \right] \end{aligned}$$

thus

$$\begin{aligned} V_{t,h}(\bar{u}_h) &= \bar{\beta}_h(u_h^{(h)}) \psi_{t-1}(\theta_h(u_h^{(h)})) + \bar{\beta}_{h-1}(u_h^{(h)}) \alpha(\theta_{h-1}(u_h^{(h)})) x_t \\ &\quad + \sum_{k=1}^h \bar{\beta}_{k-1}(u_h^{(h)}) \omega(\theta_{k-1}(u_h^{(h)})) + V_{t,h-1}(\bar{u}_{h-1}) \end{aligned}$$

where

$$\begin{aligned}\bar{u}_{h-1} &= \left( u_1^{(h-1)}, u_2^{(h-1)}, \dots, u_{h-1}^{(h-1)} \right)^\top \\ u_j^{(h-1)} &= u_j^{(h)} + \bar{\beta}_{h-1-j} \left( u_h^{(h)} \right) \alpha \left( \theta_{h-1-j} \left( u_h^{(h)} \right) \right) \\ u_j^{(h)} &= u_j\end{aligned}$$

Let denote

$$d_k = u_k^{(k)}$$

we have  $d_h = u_h$  and for  $k \leq h-1$

$$\begin{aligned}d_k &= u_k^{(k)} = u_k^{(k+1)} + \bar{\beta}_0(d_{k+1}) \alpha(\theta_0(d_{k+1})) \\ &= u_k^{(k+2)} + \bar{\beta}_1(d_{k+2}) \alpha(\theta_1(d_{k+2})) + \bar{\beta}_0(d_{k+1}) \alpha(\theta_0(d_{k+1})) \\ &= u_k^{(h)} + \sum_{j=1}^{h-k} \bar{\beta}_{j-1}(d_{k+j}) \alpha(\theta_{j-1}(d_{k+j})) \\ &= u_k + \sum_{j=1}^{h-k} \bar{\beta}_{j-1}(d_{k+j}) \alpha(\theta_{j-1}(d_{k+j}))\end{aligned}$$

Thus

$$\begin{aligned}V_{t,h}(\bar{u}_h) &= \bar{\beta}_h(d_h) \psi_{t-1}(\theta_h(d_h)) + \bar{\beta}_{h-1}(d_h) \alpha(\theta_{h-1}(d_h)) x_t + \\ &\quad \sum_{k=1}^h \bar{\beta}_{k-1}(d_h) \omega(\theta_{k-1}(d_h)) + V_{t,h-1}(\bar{u}_{h-1}) \\ &= \bar{\beta}_h(d_h) \psi_{t-1}(\theta_h(d_h)) + \bar{\beta}_{h-1}(d_{h-1}) \psi_{t-1}(\omega(\theta_{h-1}(d_{h-1}))) \\ &\quad + [\bar{\beta}_{h-1}(d_h) \alpha(\theta_{h-1}(d_h)) + \bar{\beta}_{h-2}(d_{h-1}) \alpha(\theta_{h-2}(d_{h-1}))] x_t \\ &\quad + \sum_{k=1}^h \bar{\beta}_{k-1}(d_h) \omega(\theta_{k-1}(d_h)) \\ &\quad + \sum_{k=1}^{h-1} \bar{\beta}_{k-1}(d_{h-1}) \omega(\theta_{k-1}(d_{h-1})) + V_{t,h-2}(\bar{u}_{h-2}) \\ &= a_h(\bar{u}_h) x_t + b_h(\bar{u}_h) + \sum_{k=1}^h \bar{\beta}_k(d_k) \psi_{t-1}(\theta_k(d_k))\end{aligned}$$

where

$$\begin{aligned}a_h(\bar{u}_h) &= \sum_{k=1}^h \bar{\beta}_{k-1}(d_k) \alpha(\theta_{k-1}(d_k)) \\ b_h(\bar{u}_h) &= \sum_{k=1}^h \left[ \sum_{j=1}^k \bar{\beta}_{j-1}(d_k) \omega(\theta_{j-1}(d_k)) \right]\end{aligned}$$

**Proof of Proposition 3.4.** The yield to maturity  $n$   $\left( y_t^{(n)} \right)$  is related to the short term rate as following

for  $n = 1$  we have  $y_t^{(1)} = r_t$ , for  $n \geq 2$

$$\begin{aligned}
y_t^{(n)} &= -\frac{1}{n} \ln \left\{ E_t^Q \left[ \exp \left( - \sum_{k=0}^{n-1} y_{t+k}^{(1)} \right) \right] \right\} \\
&= -\frac{1}{n} \ln \left\{ E_t^P \left[ \left[ \prod_{k=0}^{n-2} \frac{M_{t+k,t+k+1}}{E_{t+k}^P [M_{t+k,t+k+1}]} \right] \exp \left( - \sum_{k=0}^{n-1} y_{t+k}^{(1)} \right) \right] \right\} \\
&= -\frac{1}{n} \ln \left\{ E_t^P \left[ \exp \left( - \sum_{k=0}^{n-1} y_{t+k}^{(1)} + \sum_{k=0}^{n-2} \left( \gamma y_{t+k+1}^{(1)} - \psi_{t+k}^P(\gamma) \right) \right) \right] \right\} \\
&= -\frac{1}{n} \ln \left\{ E_t^P \left[ \exp \left( - \sum_{k=0}^{n-1} y_{t+k}^{(1)} + \sum_{k=1}^{n-1} \left( \gamma y_{t+k}^{(1)} - \psi_{t+k-1}^P(\gamma) \right) \right) \right] \right\} \\
&= \frac{y_t^{(1)} + \psi_t^P(\gamma)}{n} - \frac{1}{n} \ln \left\{ E_t^P \left[ \exp \left( \sum_{k=1}^{n-1} (\gamma - 1) y_{t+k}^{(1)} - \sum_{k=1}^{n-2} \psi_{t+k}^P(\gamma) \right) \right] \right\} \\
&= E_t^P \left[ \exp \left( \sum_{k=1}^{n-1} (\gamma - 1) y_{t+k}^{(1)} - \sum_{k=1}^{n-2} \psi_{t+k}^P(\gamma) \right) \right] \\
&= E_t^P \left[ \exp \left( \sum_{k=1}^{n-1} (\gamma - 1) y_{t+k}^{(1)} - \sum_{k=1}^{n-2} \left( \sum_{i=1}^k \bar{\beta}_{i-1}^P(\gamma) \left[ \omega^P(\theta_{i-1}^P(\gamma)) + \alpha^P(\theta_{i-1}^P(\gamma)) y_{t+k-i+1}^{(1)} \right] \right) \right) \right] \\
&= \exp \left( - \sum_{k=1}^{n-2} \left[ \bar{\beta}_k^P(\gamma) \psi_t^P(\theta_k^P(\gamma)) + \sum_{i=1}^k \bar{\beta}_{i-1}^P(\gamma) \omega^P(\theta_{i-1}^P(\gamma)) \right] \right) \times \\
&E_t^P \left[ \exp \left( \sum_{k=1}^{n-1} (\gamma - 1) y_{t+k}^{(1)} - \sum_{k=1}^{n-2} \left( \sum_{i=1}^k \bar{\beta}_{i-1}^P(\gamma) \alpha^P(\theta_{i-1}^P(\gamma)) y_{t+k-i+1}^{(1)} \right) \right) \right] \\
&= E_t^P \left[ \exp \left( \sum_{k=1}^{n-1} (\gamma - 1) y_{t+k}^{(1)} - \sum_{k=1}^{n-2} \left( \sum_{i=1}^k \bar{\beta}_{i-1}^P(\gamma) \alpha^P(\theta_{i-1}^P(\gamma)) y_{t+k-i+1}^{(1)} \right) \right) \right] \\
&= E_t^P \left[ \exp \left( \sum_{k=1}^{n-1} (\gamma - 1) y_{t+k}^{(1)} - \sum_{k=1}^{n-2} \left( \sum_{j=1}^k \bar{\beta}_{k-j}^P(\gamma) \alpha^P(\theta_{k-j}^P(\gamma)) y_{t+j}^{(1)} \right) \right) \right] \\
&= E_t^P \left[ \exp \left( \sum_{k=1}^{n-1} (\gamma - 1) y_{t+k}^{(1)} - \sum_{j=1}^{n-2} \left( \sum_{k=j}^{n-2} \bar{\beta}_{k-j}^P(\gamma) \alpha^P(\theta_{k-j}^P(\gamma)) \right) y_{t+j}^{(1)} \right) \right] \\
&= E_t^P \left[ \exp \left( \sum_{k=1}^{n-1} u_k y_{t+k}^{(1)} \right) \right]
\end{aligned}$$

where

$$\begin{aligned}
u_{n-1} &= \gamma - 1 \\
u_j &= \gamma - 1 - \sum_{k=j}^{n-2} \bar{\beta}_{k-j}^P(\gamma) \alpha^P(\theta_{k-j}^P(\gamma)) \text{ for } j < n - 1
\end{aligned}$$

$$\begin{aligned}
y_t^{(n)} &= \frac{y_t^{(1)} + \psi_t^P(\gamma) + \sum_{k=1}^{n-2} \left[ \bar{\beta}_k^P(\gamma) \psi_t^P(\theta_k^P(\gamma)) + \sum_{i=1}^k \bar{\beta}_{i-1}^P(\gamma) \omega^P(\theta_{i-1}^P(\gamma)) \right]}{n} \\
&\quad - \frac{a_{n-1}(\bar{u}_{n-1}) y_t^{(1)} + b_{n-1}(\bar{u}_{n-1}) + \sum_{k=1}^{n-1} \bar{\beta}_k(d_k) \psi_{t-1}^P(\theta_k^P(d_k))}{n} \\
&= \frac{y_t^{(1)} + \sum_{k=1}^{n-1} \left( \sum_{i=1}^k \bar{\beta}_{i-1}^P(\gamma) \omega^P(\theta_{i-1}^P(\gamma)) \right) + \sum_{k=1}^{n-1} \bar{\beta}_k(\gamma) \psi_{t-1}^P(\theta_k^P(\gamma))}{n} \\
&\quad + \frac{y_t^{(1)} \sum_{k=1}^{n-1} \bar{\beta}_{k-1}(\gamma) \alpha^P(\theta_{k-1}^P(\gamma))}{n} \\
&\quad - \frac{a_{n-1}(\bar{u}_{n-1}) y_t^{(1)} + b_{n-1}(\bar{u}_{n-1}) + \sum_{k=1}^{n-1} \bar{\beta}_k(d_k) \psi_{t-1}^P(\theta_k^P(d_k))}{n} \\
&= d_n + c_{n,0} y_t^{(1)} + \sum_{k=1}^{n-1} \left[ \frac{\bar{\beta}_k(\gamma)}{n} \psi_{t-1}^P(\theta_k^P(\gamma)) - \frac{\bar{\beta}_k(d_k)}{n} \psi_{t-1}^P(\theta_k^P(d_k)) \right] \\
&= d_n + c_{n,0} y_t^{(1)} + z_{t-1}^{(n)}
\end{aligned}$$

where

$$\begin{aligned}
d_n &= \frac{1}{n} \sum_{k=1}^{n-1} \left( \sum_{i=1}^k \bar{\beta}_{i-1}^P(\gamma) \omega^P(\theta_{i-1}^P(\gamma)) - \bar{\beta}_{i-1}(d_k) \omega(\theta_{i-1}(d_k)) \right) \\
c_{n,0} &= \frac{1}{n} \left[ 1 + \sum_{k=1}^{n-1} \left( \bar{\beta}_{k-1}(\gamma) \alpha^P(\theta_{k-1}^P(\gamma)) - \bar{\beta}_{k-1}(d_k) \alpha(\theta_{k-1}(d_k)) \right) \right]
\end{aligned}$$

$d_{n-1} = \gamma - 1$  and for  $k \leq n - 2$

$$d_k = \gamma - 1 + \sum_{j=1}^{n-1-k} \left[ \bar{\beta}_{j-1}(d_{k+j}) \alpha(\theta_{j-1}(d_{k+j})) - \bar{\beta}_{j-1}(\gamma) \alpha(\theta_{j-1}(\gamma)) \right]$$

**Proof of Proposition 3.6.** One has

$$\begin{aligned}
E_t[M_{t+1}] &= \exp(-r) \\
E_t[M_{t+1} \exp(r_{t+1})] &= 1,
\end{aligned}$$

which leads to

$$\begin{aligned}
\theta_t + \Psi_t(\gamma, \lambda) &= -r \\
\theta_t + \Psi_t(1 + \gamma, \lambda) &= 0.
\end{aligned}$$

Hence,

$$\begin{aligned}
\theta_t &= -r - \Psi_t(\gamma, \lambda) \\
\Psi_t(1 + \gamma, \lambda) - \Psi_t(\gamma, \lambda) &= r
\end{aligned}$$

By using the following expression of the model:

$$\Psi_{t+1}(u, v) = \frac{\omega(u, v)}{1 - \beta(u, v)} + \alpha(u, v) \sum_{i=0}^{\infty} \beta(u, v)^i h_{t-i+1}$$

one gets,

$$\frac{\omega(1 + \gamma, \lambda)}{1 - \beta(1 + \gamma, \lambda)} - \frac{\omega(\gamma, \lambda)}{1 - \beta(\gamma, \lambda)} + \sum_{i=0}^{\infty} \left[ \beta(1 + \gamma, \lambda)^i \alpha(1 + \gamma, \lambda) - \beta(\gamma, \lambda)^i \alpha(\gamma, \lambda) \right] h_{t-i} = r$$

which implies

$$\begin{aligned} \frac{\omega(1 + \gamma, \lambda)}{1 - \beta(1 + \gamma, \lambda)} - \frac{\omega(\gamma, \lambda)}{1 - \beta(\gamma, \lambda)} &= r \\ \beta(1 + \gamma, \lambda)^i \alpha(1 + \gamma, \lambda) - \beta(\gamma, \lambda)^i \alpha(\gamma, \lambda) &= 0, \quad \forall i \geq 0. \end{aligned}$$

Therefore,

$$\begin{aligned} \frac{\omega(1 + \gamma, \lambda)}{1 - \beta(1 + \gamma, \lambda)} - \frac{\omega(\gamma, \lambda)}{1 - \beta(\gamma, \lambda)} &= r \\ \beta(1 + \gamma, \lambda) &= \beta(\gamma, \lambda) \\ \alpha(1 + \gamma, \lambda) &= \alpha(\gamma, \lambda). \end{aligned}$$

### Proof of Proposition 3.7.

$$E_t^Q [\exp(ur_{t+1} + v\Psi_{t+1}(\gamma))] = E_t^Q [\exp(ur_{t+1} + v(\omega(\gamma) + \alpha(\gamma)r_{t+1} + \beta(\gamma)\Psi_t(\gamma)))]$$

thus

$$\begin{aligned} \Psi_t^*(u, v) &= v\omega(\gamma) + v\beta(\gamma)\Psi_t(\gamma) + \Psi_t^Q(u^*) \\ &= v\omega(\gamma) + v\beta(\gamma)\Psi_t(\gamma) + \Psi_t(u^* + \gamma) - \Psi_t(\gamma) \end{aligned}$$

where

$$u^* = u + v\alpha(\gamma)$$

this implied that

$$\begin{aligned} \Psi_{t+1}^*(u, v) &= v\omega(\gamma) + (v\beta(\gamma) - 1)\Psi_{t+1}(\gamma) + \Psi_{t+1}(u^* + \gamma) \\ &= v\omega(\gamma) + (v\beta(\gamma) - 1)\Psi_{t+1}(\gamma) + \omega(u^* + \gamma) \\ &\quad + \alpha(u^* + \gamma)r_{t+1} + \beta(u^* + \gamma)\Psi_t(u^* + \gamma) \\ &= v\omega(\gamma) + (v\beta(\gamma) - 1)\Psi_{t+1}(\gamma) + \omega(u^* + \gamma) + \\ &\quad + \alpha(u^* + \gamma)r_{t+1} + \beta(u^* + \gamma)[\Psi_t^*(u, v) - v\omega(\gamma) - (v\beta(\gamma) - 1)\Psi_t(\gamma)] \end{aligned}$$

Table 1: **Monte carlo exercise for the GMM.**

Mean, Median and RMSE of parameters estimates across  $N$  samples (of  $T$  observations). Parameters used in the simulation of GARG model are  $\mu = 2.78\text{E-}05$ ,  $\nu = 0.139$ ,  $\rho = 0.112$ ,  $\beta = 0.922$  and  $\theta = 0.906$

Par	Mean	Median	RMSE	Mean	Median	RMSE	Mean	Median	RMSE
	N=1000, T=250			N=1000, T=500			N=1000, T=1000		
$\mu$	2.58E-05	3.66E-05	1.68E-05	2.61E-05	3.66E-05	1.66E-05	2.64E-05	3.66E-05	1.64E-05
$\nu$	0.077	0.109	0.079	0.077	0.109	0.078	0.078	0.109	0.077
$\rho$	0.104	0.115	0.018	0.104	0.115	0.018	0.105	0.115	0.018
$\beta$	0.695	0.946	0.437	0.702	0.946	0.430	0.710	0.946	0.423
$\theta$	1.318	0.879	0.780	1.307	0.879	0.770	1.293	0.879	0.755
	N=2000, T=250			N=2000, T=500			N=2000, T=1000		
$\mu$	2.80E-05	3.67E-05	1.55E-05	2.90E-05	3.66E-05	1.49E-05	2.97E-05	3.66E-05	1.44E-05
$\nu$	0.083	0.109	0.072	0.086	0.109	0.069	0.088	0.109	0.066
$\rho$	0.106	0.115	0.017	0.107	0.115	0.016	0.108	0.115	0.015
$\beta$	0.744	0.946	0.395	0.766	0.946	0.372	0.784	0.946	0.353
$\theta$	1.248	0.879	0.733	1.208	0.878	0.693	1.175	0.878	0.657
	N=4000, T=250			N=4000, T=500			N=4000, T=1000		
$\mu$	3.20E-05	3.66E-05	1.28E-05	3.24E-05	3.66E-05	1.25E-05	3.27E-05	3.66E-05	1.24E-05
$\nu$	0.095	0.109	0.056	0.096	0.109	0.054	0.097	0.109	0.053
$\rho$	0.110	0.115	0.012	0.111	0.115	0.012	0.111	0.115	0.011
$\beta$	0.837	0.946	0.291	0.847	0.946	0.278	0.852	0.946	0.270
$\theta$	1.084	0.879	0.557	1.066	0.878	0.533	1.056	0.879	0.517

Table 2: **Monte carlo exercise for the QMLE.**

Mean, Median and RMSE of parameters estimates across  $N$  samples (of  $T$  observations). Parameters used in the simulation of GARG model are  $\mu = 2.78\text{E-}05$ ,  $\nu = 0.139$ ,  $\rho = 0.112$ ,  $\beta = 0.922$  and  $\theta = 0.906$

Par	Mean	Median	RMSE	Mean	Median	RMSE	Mean	Median	RMSE
	N=1000, T=250			N=1000, T=500			N=1000, T=1000		
$\mu$	6.00E-05	5.13E-05	5.37E-05	5.97E-05	5.20E-05	5.19E-05	6.13E-05	5.31E-05	5.43E-05
$\nu$	0.142	0.051	0.453	0.154	0.051	0.415	0.148	0.050	0.427
$\rho$	0.110	0.109	0.029	0.108	0.109	0.027	0.108	0.106	0.028
$\beta$	1.028	0.959	0.356	1.048	0.960	0.415	1.066	0.964	0.488
$\theta$	0.879	0.909	0.189	0.875	0.908	0.204	0.870	0.909	0.206
	N=2000, T=250			N=2000, T=500			N=2000, T=1000		
$\mu$	5.45E-05	5.02E-05	4.20E-05	5.46E-05	4.99E-05	4.12E-05	5.56E-05	5.15E-05	4.23E-05
$\nu$	0.216	0.057	0.963	0.220	0.055	0.915	0.179	0.052	0.711
$\rho$	0.109	0.109	0.023	0.107	0.107	0.023	0.107	0.107	0.024
$\beta$	0.998	0.937	0.270	0.993	0.941	0.243	1.001	0.947	0.255
$\theta$	0.891	0.918	0.164	0.893	0.918	0.158	0.888	0.915	0.161
	N=4000, T=250			N=4000, T=500			N=4000, T=1000		
$\mu$	5.30E-05	5.07E-05	3.82E-05	5.26E-05	5.17E-05	3.57E-05	5.30E-05	5.31E-05	3.57E-05
$\nu$	0.244	0.048	0.942	0.373	0.049	3.038	0.275	0.050	2.347
$\rho$	0.108	0.109	0.022	0.108	0.107	0.021	0.107	0.106	0.021
$\beta$	0.969	0.940	0.168	0.963	0.941	0.146	0.966	0.964	0.147
$\theta$	0.901	0.914	0.126	0.905	0.913	0.121	0.903	0.909	0.122

Table 3: **MLE Estimation of Generalized- Autoregressive Normal Inverse Gaussian Process on returns and realized variance Data.**

The data is the Deutsche mark (DM) / US dollar (USD) exchange rate returns and realized variance. Sample period is 1986:12:01 to 1996:12:01 with a total of 2449 observations

par	30 min				5 min			
	Affine		G-Affine		Affine		G-Affine	
	Est	STD	Est	STD	Est	STD	Est	STD
$\beta$			0.6111	0.0396			0.5449	0.0419
$\rho$	0.3255	0.0203	0.1754	0.0179	0.3444	0.0193	0.2150	0.0192
$\mu$	0.2341	0.0114	0.1834	0.0087	0.1642	0.0071	0.1328	0.0058
$\nu$	1.2565	0.0398	0.5045	0.0545	2.0818	0.0647	0.9380	0.0961
$a$	0.0063	0.0139	0.0063	0.0140	0.0064	0.0180	0.0064	0.0180
$c$	-0.0214	0.0448	-0.0214	0.0449	-0.0180	0.0433	-0.0180	0.0434
$b$	1.74E-08	5.709E-06	1.44E-08	6.024E-06	4.98E-08	1.093E-05	1.54E-08	5.777E-06
$d$	0.9282	0.0290	0.9282	0.0290	0.7551	0.0236	0.7551	0.0236
LIK	-1600.0932		-1547.4719		-1838.6743		-1790.0531	
BIC	0.8069		0.7850		0.9234		0.9034	

Table 4: **Out-of-sample 1-month-ahead forecasting results.**

We present the results of out-of-sample 1-month-ahead forecasting using eight models, as described in detail in the text. We estimate all models recursively from 1985:1 to the time that the forecast is made, beginning in 1994:1 and extending through 2000:12. We define forecast errors at  $t+1$  as  $y_{t+1}(\tau) - \hat{y}_{t+1}(\tau)$ , and we report the mean, standard deviation and root mean squared errors of the forecast errors, as well as their first and 12th sample autocorrelation coefficients.

Maturity	Mean	Std. Dev.	RMSE	$\hat{\rho}(1)$	$\hat{\rho}(12)$	Mean	Std. Dev.	RMSE	$\hat{\rho}(1)$	$\hat{\rho}(12)$
	Random walk					Nelson–Siegel with AR(1) factor dynamics				
3 months	0.033	0.176	0.179	0.220	0.053	-0.045	0.170	0.176	0.247	0.017
1 year	0.021	0.240	0.241	0.340	-0.153	0.023	0.235	0.236	0.425	-0.213
3 years	0.007	0.279	0.279	0.341	-0.133	-0.056	0.273	0.279	0.332	-0.117
5 years	-0.003	0.276	0.276	0.275	-0.131	-0.091	0.277	0.292	0.333	-0.116
10 years	-0.011	0.254	0.254	0.215	-0.145	-0.062	0.252	0.260	0.259	-0.115
	VAR(1) on 1, 24, 96 months yields					VARMA(1,1) on 1, 24, 96 months yields				
3 months	-0.043	0.196	0.200	0.126	0.320	-0.021	0.231	0.230	0.014	0.299
1 year	-0.011	0.235	0.234	0.380	-0.139	0.004	0.230	0.229	0.055	-0.006
3 years	0.018	0.269	0.268	0.358	-0.153	0.023	0.250	0.249	0.143	-0.095
5 years	-0.014	0.281	0.280	0.375	-0.154	-0.007	0.267	0.265	0.266	-0.100
10 years	-0.163	0.274	0.318	0.386	-0.094	-0.152	0.264	0.304	0.348	-0.066
	VAR(1) on 1,24, 96 months yields and inflation					VAR(1) on 1,24, 96 months yields and real activity				
3 months	0.078	0.196	0.210	0.120	0.295	-0.082	0.202	0.217	0.143	0.316
1 year	0.088	0.227	0.243	0.328	-0.144	-0.057	0.239	0.245	0.411	-0.149
3 years	0.134	0.261	0.292	0.294	-0.099	0.024	0.265	0.264	0.332	-0.139
5 years	0.052	0.273	0.276	0.315	-0.107	-0.031	0.278	0.278	0.351	-0.136
10 years	-0.050	0.269	0.272	0.354	-0.100	-0.107	0.277	0.296	0.403	-0.092

Table 5: **Out-of-sample 6-months-ahead forecasting results.**

We present the results of out-of-sample 6-months-ahead forecasting using eight models, as described in detail in the text. We estimate all models recursively from 1985:1 to the time that the forecast is made, beginning in 1994:1 and extending through 2000:12. We define forecast errors at  $t+6$  as  $y_{t+6}(\tau) - \hat{y}_{t+6/t}(\tau)$ , and we report the mean, standard deviation and root mean squared errors of the forecast errors, as well as their sixth and eighteenth sample autocorrelation coefficients.

Maturity	Mean	Std. Dev.	RMSE	$\hat{\rho}(6)$	$\hat{\rho}(18)$	Mean	Std. Dev.	RMSE	$\hat{\rho}(6)$	$\hat{\rho}(18)$
	Random walk					Nelson–Siegel with AR(1) factor dynamics				
3 months	0.220	0.564	0.605	0.381	-0.214	0.083	0.510	0.517	0.301	-0.190
1 year	0.181	0.758	0.779	0.139	-0.150	0.131	0.656	0.669	0.168	-0.174
3 years	0.099	0.873	0.879	0.018	-0.211	-0.052	0.748	0.750	0.049	-0.189
5 years	0.048	0.860	0.861	0.008	-0.249	-0.173	0.758	0.777	0.069	-0.273
10 years	-0.020	0.758	0.758	0.019	-0.271	-0.251	0.676	0.721	0.058	-0.288
	VAR(1) on 1, 24, 96 months yields					VARMA(1,1) on 1, 24, 96 months yields				
3 months	-0.074	0.494	0.496	0.193	-0.109	0.001	0.531	0.528	0.312	-0.163
1 year	-0.040	0.696	0.693	0.085	-0.142	0.015	0.665	0.661	0.208	-0.168
3 years	-0.089	0.777	0.777	-0.014	-0.197	-0.028	0.725	0.721	0.047	-0.200
5 years	-0.180	0.789	0.805	-0.006	-0.220	-0.103	0.735	0.738	0.049	-0.223
10 years	-0.388	0.735	0.827	-0.019	-0.186	-0.297	0.681	0.739	0.018	-0.198
	VAR(1) on 1,24, 96 months yields and inflation					VAR(1) on 1,24, 96 months yields and real activity				
3 months	0.590	0.464	0.750	0.143	0.116	-0.184	0.508	0.537	0.156	-0.042
1 year	0.562	0.671	0.872	0.028	-0.025	-0.151	0.712	0.723	0.067	-0.113
3 years	0.396	0.769	0.861	-0.045	-0.135	-0.116	0.787	0.791	-0.020	-0.191
5 years	0.186	0.789	0.806	-0.023	-0.178	-0.218	0.798	0.822	-0.007	-0.219
10 years	-0.040	0.727	0.723	-0.060	-0.187	-0.341	0.739	0.810	-0.019	-0.187

Table 6: **Out-of-sample 12-months-ahead forecasting results.**

We present the results of out-of-sample 12-months-ahead forecasting using eight models, as described in detail in the text. We estimate all models recursively from 1985:1 to the time that the forecast is made, beginning in 1994:1 and extending through 2000:12. We define forecast errors at  $t+12$  as  $y_{t+12}(\tau) - \hat{y}_{t+12/t}(\tau)$ , and we report the mean, standard deviation and root mean squared errors of the forecast errors, as well as their 12th and 24th sample autocorrelation coefficients.

Maturity	Mean	Std. Dev.	RMSE	$\hat{\rho}(12)$	$\hat{\rho}(24)$	Mean	Std. Dev.	RMSE	$\hat{\rho}(12)$	$\hat{\rho}(24)$
	Random walk					Nelson–Siegel with AR(1) factor dynamics				
3 months	0.150	0.724	0.739	-0.288	0.001	0.150	0.724	0.739	-0.288	0.001
1 year	0.173	0.823	0.841	-0.332	-0.004	0.173	0.823	0.841	-0.332	-0.004
3 years	-0.123	0.910	0.918	-0.408	0.015	-0.123	0.910	0.918	-0.408	0.015
5 years	-0.337	0.918	0.978	-0.412	0.003	-0.337	0.918	0.978	-0.412	0.003
10 years	-0.531	0.825	0.981	-0.433	-0.003	-0.531	0.825	0.981	-0.433	-0.003
	VAR(1) on 1, 24, 96 months yields					VARMA(1,1) on 1, 24, 96 months yields				
3 months	-0.152	0.792	0.801	-0.214	-0.076	-0.137	0.708	0.716	-0.059	-0.123
1 year	-0.188	0.913	0.926	-0.307	-0.027	-0.170	0.794	0.807	-0.143	-0.076
3 years	-0.325	0.953	1.001	-0.393	0.001	-0.267	0.832	0.868	-0.269	-0.043
5 years	-0.459	0.956	1.055	-0.413	-0.006	-0.371	0.849	0.921	-0.305	-0.049
10 years	-0.710	0.875	1.123	-0.440	-0.006	-0.596	0.775	0.974	-0.357	-0.053
	VAR(1) on 1,24, 96 months yields and inflation					VAR(1) on 1,24, 96 months yields and real activity				
3 months	0.854	0.851	1.202	0.020	-0.045	-0.329	0.852	0.908	-0.245	-0.081
1 year	0.743	1.002	1.242	-0.137	-0.005	-0.359	0.994	1.051	-0.344	-0.030
3 years	0.436	1.044	1.125	-0.286	0.016	-0.390	1.018	1.084	-0.420	-0.008
5 years	0.153	1.050	1.054	-0.323	0.001	-0.528	1.007	1.131	-0.430	-0.014
10 years	-0.143	0.933	0.937	-0.414	-0.001	-0.682	0.901	1.125	-0.455	-0.011

Table 7: **Implied volatility Root mean squared error by maturity, Implied volatility bias and option price bias by Moneyness**

We estimate the models on a total of 16, 506 contracts with an average call price of 46.05 and average implied volatility of 20.26. The estimation have been done by minimizing the Black-Scholes IVRMSE

Moneyness					
	S/X<0.975	0.975<S/X<1	1<S/X<1.025	1.025<S/X	All
Model	IVRMSE (%)				
Affine	3.8809	4.2988	4.4313	5.0642	4.3768
G-Affine	2.9471	2.9476	3.2201	3.7181	3.1915
Model	IV bias (%)				
Affine	0.2134	-0.0661	0.0346	-0.3556	-0.0166
G-Affine	0.0211	0.1873	0.3723	-0.2216	0.0694
Model	Option price bias				
Affine	0.4357	-0.3601	-0.4654	-1.2721	-0.3124
G-Affine	0.1809	0.0281	0.0990	-0.9638	-0.1342

Table 8: **Implied volatility Root mean squared error by maturity, Implied volatility bias and option price bias by Moneyness**

We estimate the models on a total of 16, 506 contracts with an average call price of 46.05 and average implied volatility of 20.26. The estimation have been done by minimizing the Black-Scholes IVRMSE

Maturity					
	DTM<30	30<DTM<90	90<DTM<180	180<DTM	All
Model	IVRMSE (%)				
Affine	5.4750	4.3179	3.9963	3.8295	4.3768
G-Affine	3.9103	3.1432	2.8824	2.9301	3.1915
Model	IV bias (%)				
Affine	0.5038	-0.2423	-0.0693	0.1737	-0.0166
G-Affine	0.8447	0.1338	-0.2344	-0.4328	0.0694
Model	Option price bias				
Affine	0.3352	-0.5526	-0.5843	0.0199	-0.3124
G-Affine	0.8570	0.3845	-0.5387	-1.8517	-0.1342

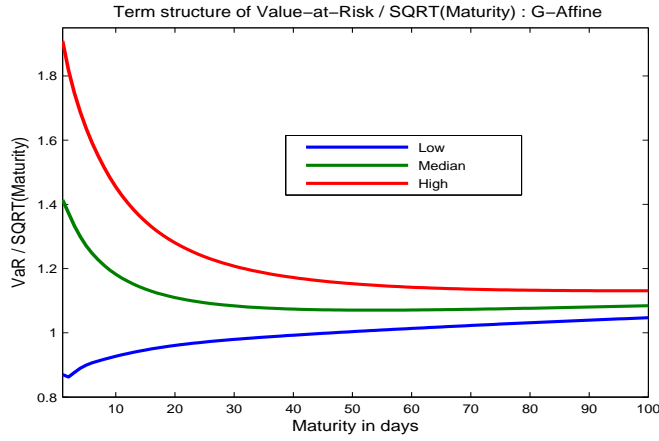


Figure 1: **Generalized Autoregressive Normal Inverse gaussian term structure of value-at-risk**

We use parameters estimated from the MLE to compute the term structure of value at risk. Several cases have been considered depending on the day where the term structure is evaluated. The cases are High volatility day (day with higher realized variance), Median volatility day and Low volatility day

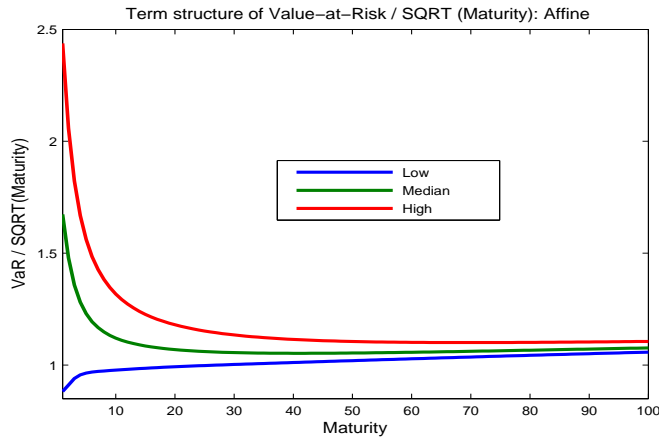


Figure 2: **Autoregressive Normal Inverse gaussian term structure of value-at-risk**

We use parameters estimated from the MLE to compute the term structure of value at risk. Several cases have been considered depending on the day where the term structure is evaluated. The cases are High volatility day (day with higher realized variance), Median volatility day and Low volatility day

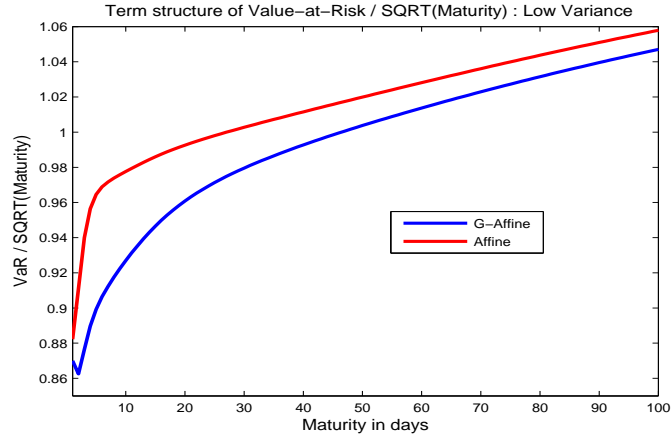


Figure 3: **Term structure of Value-at-risk Conditional on a low variance day**  
 We use parameters estimated from the MLE to compute the term structure of value at risk. We compared Affine and Generalized affine term structure

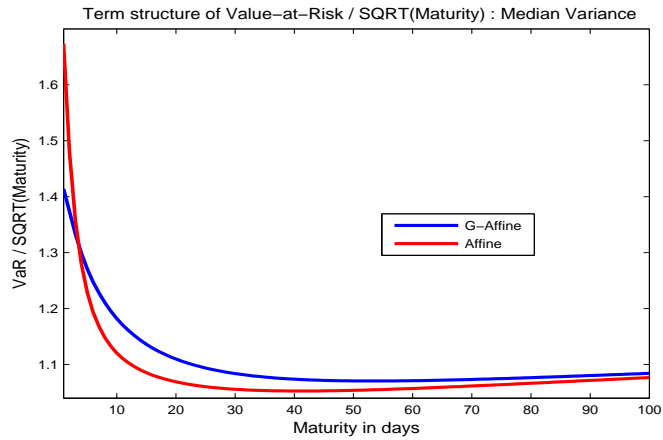


Figure 4: **Term structure of Value-at-risk Conditional on a median variance day**  
 We use parameters estimated from the MLE to compute the term structure of value at risk. We compared Affine and Generalized affine term structure

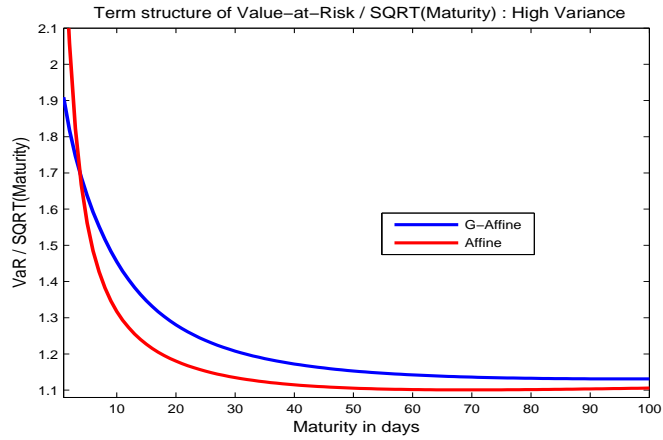


Figure 5: **Term structure of Value-at-risk Conditional on a high variance day**  
 We use parameters estimated from the MLE to compute the term structure of value at risk. We compared Affine and Generalized affine term structure.

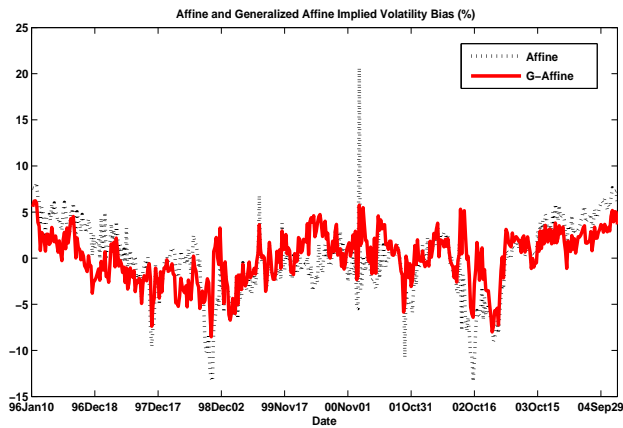


Figure 6: **Implied Volatility Bias**  
 The figure displays implied volatility bias as a function of the day at which option is priced. Implied volatility bias is the difference between model and observed black scholes implied volatility. For each day we compute average available Implied volatility bias

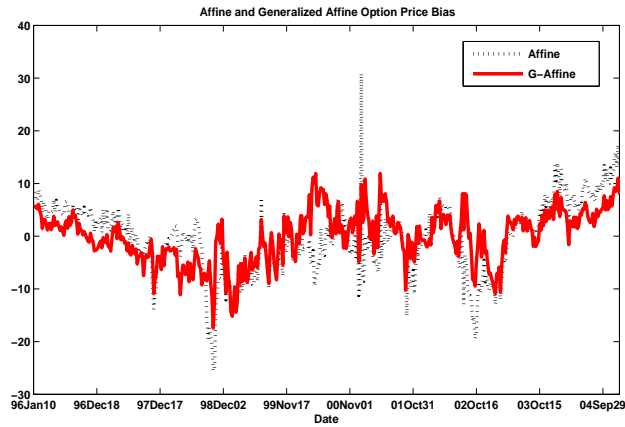


Figure 7: Option price Bias

The figure displays Option price bias as a function of the day at which option is priced. Option price bias is the difference between model and observed Option price. For each day we compute average available option bias

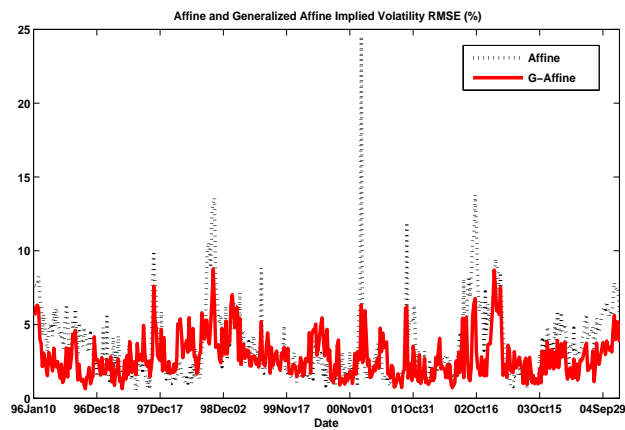


Figure 8: Implied Volatility Root mean squared error (IVRMSE)

The figure displays IVRMSE as a function of day at which option is priced. IVRMSE is the square-root of the average squared difference between model and observed black scholes implied volatility.