Switching VARMA Term Structure Models

Alain MONFORT [(1)] Fulvio PEGORARO [(2)]

First version : February, 2005
This version : September, 2006

Abstract
Switching VARMA Term Structure Models

The purpose of this paper is to propose a global discrete-time modeling of the term structure of interest rates which is able to capture simultaneously the following important features : (i) an historical dynamics of the factor driving term structure shapes involving several lagged values, and switching regimes; (ii) a specification of the stochastic discount factor (SDF) with time-varying and regime-dependent risk-premia; (iii) explicit or quasi explicit formulas for zero-coupon bond and interest rate derivative prices; (iv) positiveness of the yields at each maturity. We develop the Switching Autoregressive Normal (SARN) and the Switching Vector Autoregressive Normal (SVARN) Term Structure Models of order p and the Switching Autoregressive Gamma (SARG) Term Structure Model of order p. The factor is considered as a latent variable or an observable variable : in the second case the factor is a vector of several yields. Regime shifts are described by a Markov chain with (historical) non-homogeneous transition probabilities. An empirical analysis of bivariate single-regime and regime-switching models in the Gaussian family, using monthly observations of the U.S. term structure of interest rates, and a goodness-of-fit and Expectation Hypothesis Puzzle comparison with competing models in the literature, show the determinant role played by the observable nature of the factor, lags and switching regimes in the term structure modeling.

Keywords : Affine Term Structure Models, Stochastic Discount Factor, Car(p) processes, Switching Regimes, VARMA processes, Lags, Positiveness, Derivative Pricing, Expectation Hypothesis Puzzle.

JEL number : C1, C5, G1

[1]CNAM [E-mail: monfort@cnam.fr] and CREST, Laboratoire de Finance-Assurance [E-mail: monfort@ensae.fr].
[2]CREST, Laboratoire de Finance-Assurance [E-mail: pegoraro@ensae.fr].
[3]We are grateful to Luc Bauwens, René Garcia, Eric Ghysels, Christian Gourieroux, Michel Lubrano, Nour Meddahi, Monika Piazzesi, Eric Renault, Olivier Scaillet, Ken Singleton, Hao Zhou, two anonymous referees and to all the participants at the Journal of Applied Econometrics Annual Lecture 2005, CORE Econometrics Seminar 2005, CREST Financial Econometrics Seminar 2005, GREQAM Econometrics and Statistics Seminar 2005, EC2 Conference 2006, CIREQ Financial Econometrics Conference 2006, Konstanz International Conference on High Frequency Finance 2006, CREST Financial Econometrics Conference 2006, Econometric Society 2006 Summer Meeting for their helpful suggestions.

1 Introduction

In this paper we propose a global discrete-time modeling of the term structure of interest rates, which is able to capture simultaneously the following important features :

- an historical dynamics of the factor driving term structure shapes involving several lagged values, and switching regimes;
- a specification of the stochastic discount factor (SDF) with time-varying and regime-dependent risk-premia;
- explicit or quasi explicit formulas for zero-coupon bond and interest rate derivative prices;
- positiveness of the yields at each maturity.

It is well known in the literature that interest rates show an historical dynamics involving lagged values *and* switching regimes [see, among the others, Hamilton (1988), Cai (1994), Driffill and Sola (1994), Garcia and Perron (1996), Gray (1996), Boudoukh, Richardson, Smith, and Whitelaw (1999), Ang and Bekaert (2002a, 2002b), Christiansen (2002), Christiansen and Lund (2005), Cochrane and Piazzesi (2005)]; indeed, changes in the business cycle conditions or monetary policy may affect real rates and expected inflation and cause interest rates to behave quite differently in different time periods, both in terms of level and volatility. In addition, there is a large empirical literature on bond yields, based in general on the class of Affine Term Structure Models (ATSMs)[4], suggesting that regime switching models describe the term structure of interest rates better than single-regime models [see, for example, Bansal and Zhou (2002), Driffill, Kenc and Sola (2003), Evans (2003), Ang and Bekaert (2005), Dai Singleton and Yang (2005)].

This results lead us to propose dynamic term structure models (DTSMs) where the yield curve is driven by a univariate or multivariate factor (x_t) which depends on its p most recent lagged values [X_t, say] and for which all the sensitivity coefficients depend on the present and past values of a latent J-states non-homogeneous Markov Chain (z_t) [Z_t, say] describing different regimes in the economy. Consequently, the joint dynamics of (x_t, z_t) is not a Compound Autoregressive (Car) process[5] under the historical probability, and thus allows for nonlinearities already documented by the literature [see Ait-Sahalia (1996), Stanton (1997), Ang and Bekaert (2002b)]. The factor (x_t) is considered as a latent variable or an observable variable: in the second case the factor is a vector of several yields.

We consider an exponential-affine SDF with time-varying and regime-dependent risk correction coefficients which are defined as functions of the present and past values of the factor (x_t) and the regime indicator function (z_t). In our models, both factor risk and regime-shift risk are priced, and this done by taking into account not just the information at date t, that is (x_t, z_t), but a larger information given by (X_t, Z_t). This specification leads to stochastic and regime-dependent

[4]The Affine family of dynamic term structure models (DTSMs) is characterized by the fact that the zero-coupon bond yields are affine functions of Markovian state variables, and it gives closed-form expressions for zero-coupon bond prices which greatly facilitates pricing and econometric implementation [see Duffie and Kan (1996), Duffie and Singleton (2000, 2003) and Piazzesi (2003)]. Observe that the Affine Term Structure family is much larger that it has been considered in the literature : indeed, it has been observed recently that the family of Quadratic Term Structure Models (QTSMs) [see Beaglehole and Tenney (1991), Ahn, Dittmar and Gallant (2002), and Leippold and Wu (2002)] is a special case of the Affine class obtained by stacking the factor values and their squares [see Gourieroux and Sufana (2003), Cheng and Scaillet (2005)].

[5]A Car (discrete-time affine) process is a Markovian process with an exponential-affine conditional Laplace transform [see Darolles, Gourieroux, Jasiak (2006) for details].

risk premia, and is coherent with recent empirical literature suggesting to define risk correction coefficients as functions of both the factors and their volatilities. Such specification is helpful in order to replicate correclty the observed temporal variation of one-period expected excess returns on zero-coupon bonds [see Ahn, Dittmar and Gallant (2002), Dai and Singleton (2002), Duffee (2002), Duarte (2004), Cheridito, Filipovic and Kimmel (2005), Dai, Singleton and Yang (2005)].

At the same time, we want to exploit the tractability of Car models, and obtain explicit or quasi explicit formulas for zero-coupon bond and interest rate derivative prices. This result is achieved by matching the historical distribution and the SDF in order to get a Car risk-neutral joint dynamics for (x_t, z_t), and by using the property of the Car family of processes of being able to incorporate lags and switching regimes. It is now well known [see Gourieroux, Monfort and Polimenis (2005), and Darolles, Gourieroux, Jasiak (2006)] that the class of discrete-time affine (Car) models is much larger than the discrete-time counterparts of the continuous-time affine processes [see Duffie and Kan (1996), Dai and Singleton (2000), and Duffie, Filipovic and Schachermayer (2003)].

We develop the Switching Autoregressive Normal (SARN) and the Switching Vector Autoregressive Normal (SVARN) Factor-Based Term Structure Models of order p. Ang and Bekaert (2005) also propose a discrete time regime-switching Gaussian term structure model (to identify the real and expected inflation components of nominal interest rates). In their model, the historical dynamics of the tridimensional factor (x_t) driving term structure shapes is described by a regime-switching VAR(1) process with a constant autoregressive matrix. The regime indicator function (z_t) is driven, under the historical probability, by a homogeneous Markov chain and regime-shift risk is not priced. Bansal and Zhou (2002) propose a bivariate (approximate) discrete-time Cox-Ingersoll-Ross term structure model with regime shifts. In their modeling, (z_t) is a homogeneous Markov chain under the historical probability; the associated risk correction coefficient is assumed equal to zero, and the provided term structure formula is based on a log-linear approximation applied on the fundamental asset pricing equation. Our SVARN(p) Factor-Based Term Structure Model relax all of these assumptions.

Dai, Singleton and Yang (2005) propose a Gaussian discrete time model where the historical dynamics of the latent factor (x_t) is described by a trivariate SVARN(1) process with non-homogeneous regime-switching. They price regime-shift risk, and their factor risk correction coefficient generalizes to the case of multiple regimes the essentially affine specification of Duffee (2002). In our approach, the historical dynamics of (x_t) depends on several lagged values and on several past non-homogeneous regime-indicators (z_t) [the SVARN(p) process], we price regime-shift risk and our specification of the factor risk correction coefficient extends to the case of multiple lags that of Dai, Singleton and Yang (2005). Moreover, in the empirical analysis of SVARN(p) Factor-Based Term Structure Models, we overcome their admissibility and identification problems given that the factor (x_t) will be observable (yields at different maturities). In this general setting, we are able to derive formulas, as well as for the yield curve and for the price of derivatives, with simple analytical or quasi explicit representations.

We develop also the Switching Autoregressive Gamma[6] (SARG) Factor-Based Term Structure Model of order p [see Pegoraro (2006) for a presentation of the Switching Vector Autoregressive Gamma (SVARG) Factor-Based Term Structure Model of order p]. This approach implies the positiveness of the yields for each time to maturity, and regardless the latent or observable nature of the factor (x_t). The univariate SARG(p) [and the multivariate SVARG(p)] Factor-Based Term Structure Models are able to replicate complex nonlinear (historical and risk-neutral) factor dynamics and provide explicit or tractable formulas for zero-coupon bond and derivative prices. We

[6] The Autoregressive Gamma (ARG) process is a Car process, and the ARG(1) specification is the discrete-time counterpart of the Cox-Ingersoll-Ross process [see Gourieroux and Jasiak (2006), Cox, Ingersoll, and Ross (1985)].

extend the Bansal and Zhou (2002) framework in several directions; we use the exact discrete-time equivalent of the CIR process (with switching regimes) generalized to an autoregressive order p larger than one; we allow for a non-homogeneous historical transition matrix for (z_t); we price the regime-shift risk, and we provide an exact yield-to-maturity formula.

The plan of the paper is as follows. In Section 2, we present the Index-Car(p) processes. This family of processes is developed in the univariate and multivariate case, with and without Switching Regimes. In particular, we study the (scalar and vector) Autoregressive Gaussian of order p models and the (scalar and vector) Autoregressive Gamma of order p models, under single-regime and regime-switching specifications. Then, this class of processes is used, following the SDF modeling principle, to derive our multi-lag regime-switching term structure models. In Section 3 we study the SARN(p) and the SVARN(p) Factor-Based Term Structure Models, we derive the Generalized Linear Term Structure formulas and we specify the historical and risk-neutral dynamics of the yield curve processes. These results are given for a latent or an observable factor. Moreover, we discuss the propagation of shocks on the interest rate surface. Section 4 deals with the SARG(p) Factor-Based Term Structure Models. Here, regardless the observable or latent nature of the factor (x_t), we derive the Generalized Linear Term Structure formulas and the yield curve processes, and we guarantee the positiveness of the yields for each time to maturity. Finally, the pricing methodology proposed in sections 3 and 4, for zero-coupon bonds, is generalized in Section 5 to the case of interest rate derivatives. In Section 6 we present an empirical analysis of (single-regime) VARN(p) and SVARN(p) Factor-Based Term Structure Models, using monthly observations on the U.S. term structure of interest rates, from June 1964 to December 1995. In Section 6.1 we justify the observable factor approach, in Section 6.2 we describe the data set used in the empirical analysis, while Section 6.3 presents the estimated models and the estimation methods. Sections 6.4 and 6.5 give the estimation results (under the historical and risk-neutral probability) of VARN(p) and SVARN(p) Factor-Based Term Structure Models, for $p \in \{1, 2\}$, and show, in particular, the crucial role played by lags and non-homogeneous switching regimes to explain interest rates autocorrelation [Ljung-Box and Portmanteau tests]. In Section 6.6 we compare the goodness-of-fit of our multi-lag single-regime and regime-switching term structure models, measured in terms of absolute pricing errors, with those of other competing models like the 2-Factor (approximate) discrete-time regime-switching CIR model of Bansal and Zhou (2002) and other models studied in their paper [2-Factor and 3-Factor square-root (CIR) models, and the 3-Factor Affine model that Dai and Singleton (2000) labeled as $\mathbb{A}_1(3)$]. This comparison highlights the importance of observable lagged factor values in the yield-to-maturity formula. In Section 6.7 we test the ability of the SVARN(2) Factor-Based Term Structure Model to replicate the observed violations of the Expectations Hypothesis, and its better performance in comparison with the above mentioned competing models. Section 6.8 compares the performances (goodness-of-fit, Campbell-Shiller regressions and Portmanteau test) of the SVARN(2) model with those obtained when we do not price the risk associated to the second lagged factor value, or the regime-shift risk. Section 7 concludes and appendices gather the proofs.

2 Laplace Transforms, Car(p) Processes and Switching Regimes

It is now well documented [see e.g. Darolles, Gourieroux and Jasiak (2006), Gourieroux and Monfort (2006), Gourieroux, Monfort and Polimenis (2003, 2006), Pegoraro (2006), Polimenis (2001)] that the Laplace transform (or moment generating function) is a very convenient mathematical tool in many financial domains. It is, in particular, a crucial notion in the theory of Car(p) processes [see Darolles, Gourieroux and Jasiak (2006) for details].

2.1 Definition of a Car(p) Process

Definition 1 [Car(p) process]: A n-dimensional process $\tilde{x} = (\tilde{x}_t,\, t \geq 0)$ is a compound autoregressive process of order p [Car(p)] if the distribution of $\tilde{x}_{t+1}$ given the past values $\underline{\tilde{x}_t} = (\tilde{x}_t, \tilde{x}_{t-1}, \ldots)$ admits a real Laplace transform of the following type:

$$\begin{aligned} E\left[\exp(u'\tilde{x}_{t+1}) \,|\, \underline{\tilde{x}_t}\right] &= E_t[\exp(u'\tilde{x}_{t+1})] \\ &= \exp\left[\tilde{a}_1(u)'\tilde{x}_t + \ldots + \tilde{a}_p(u)'\tilde{x}_{t+1-p} + \tilde{b}(u)\right], \quad u \in \mathbb{R}^n, \end{aligned} \tag{1}$$

where $a_i(u)$, $i \in \{1, \ldots, p\}$, and $b(u)$ are nonlinear functions, and where $a_p(u) \neq 0$, $\forall\, u \in \mathbb{R}^n$. The existence of this Laplace transform in a neighborhood of $u = 0$, implies that all the conditional moments exist, and that the conditional expectations and variance-covariance matrices (and all conditional cumulants) are affine functions of $(\tilde{x}'_t, \tilde{x}'_{t-1}, \ldots, \tilde{x}'_{t+1-p})$.

2.2 Univariate Index-Car(p) Process

An important class of Car(p) processes is the class Index-Car(p) processes, which are built from a Car(1) process. In this section we consider a univariate process x_t and the multivariate case will be considered in sections 2.6 and 2.7.

Definition 2 [Univariate Index-Car(p) process]: Let $\exp[a(u)y_t + b(u)]$ be the conditional Laplace transform of a univariate Car(1) process y_t, the process x_t admitting a conditional Laplace transform defined by:

$$E\left[\exp(ux_{t+1}) \,|\, \underline{x_t}\right] = \exp\left[a(u)(\beta_1 x_t + \ldots + \beta_p x_{t+1-p}) + b(u)\right], \quad u \in \mathbb{R}, \tag{2}$$

is called an Univariate Index-Car(p) process.

Note that, if y_t is a positive process and if the parameters $\beta_1, \ldots, \beta_p$ are positive, the process x_t will be positive.

Using the notation $\beta = (\beta_1, \ldots, \beta_p)'$ and $X_t = (x_t, x_{t-1}, \ldots, x_{t+1-p})'$, the Laplace transform (2) can be written as:

$$E\left[\exp(ux_{t+1}) \,|\, \underline{x_t}\right] = \exp\left[a(u)\beta' X_t + b(u)\right]. \tag{3}$$

2.3 Examples of Univariate Index-Car(p) Processes

a. Gaussian model

If y_t is a Gaussian AR(1) process defined by:

$$y_{t+1} = \nu + \rho y_t + \varepsilon_{t+1}$$

where ε_{t+1} is a gaussian white noise distributed as $\mathcal{N}(0, \sigma^2)$, the conditional Laplace transform of y_{t+1} given $\underline{y_t}$ is:

$$E\left[\exp(uy_{t+1}) \,|\, \underline{y_t}\right] = \exp\left[u\rho y_t + u\nu + \frac{\sigma^2}{2}u^2\right].$$

The process is Car(1) with $a(u) = u\rho$ and $b(u) = u\nu + \frac{\sigma^2}{2}u^2$. The associated Index-Car(p) process has a conditional Laplace transform defined by:

$$E\left[\exp(ux_{t+1}) \,|\, \underline{x_t}\right] = \exp\left[u\rho(\beta_1 x_t + \ldots + \beta_p x_{t+1-p}) + u\nu + \frac{\sigma^2}{2}u^2\right];$$

so, using the notation $\varphi_i = \rho\beta_i$, we see that x_{t+1} is the Gaussian AR(p) process defined by:

$$x_{t+1} = \nu \, + \, \varphi_1 \, x_t \, + \, \ldots \, + \, \varphi_p \, x_{t+1-p} \, + \, \varepsilon_{t+1} \tag{4}$$

and its conditional Laplace transform becomes:

$$E\left[\exp(ux_{t+1}) \,|\, \underline{x_t}\right] = \exp\left[u\varphi' X_t + u\nu + \tfrac{\sigma^2}{2}u^2\right] \,, \tag{5}$$

where $\varphi = (\varphi_1, \ldots, \varphi_p)'$.

b. Gamma model

Let us now consider an autoregressive gamma of order one [ARG(1)] process y_t. The conditional Laplace transform is [see Gourieroux and Jasiak (2005) for details]:

$$E\left[\exp(uy_{t+1}) \,|\, \underline{y_t}\right] = \exp\left[\tfrac{\rho u}{1-u\mu}\, y_t - \nu \log(1-u\mu)\right] \,, \quad \rho > 0 \,, \mu > 0 \,, \nu > 0 \,,$$

and it is well known that, given y_t, y_{t+1} can be obtained by first drawing a latent variable U_{t+1} in the Poisson distribution $\mathcal{P}(\frac{\rho y_t}{\mu})$ and, then, drawing $\frac{y_{t+1}}{\mu}$ in the gamma distribution $\gamma(\nu + U_{t+1})$. The process y_{t+1} is positive and the associated Index-Car(p) process x_{t+1} is also positive. The conditional Laplace transform of this process is:

$$E\left[\exp(ux_{t+1}) \,|\, \underline{x_t}\right] = \exp\left[\tfrac{\rho u}{1-u\mu}(\beta_1 x_t + \ldots + \beta_p x_{t+1-p}) - \nu \log(1-u\mu)\right] \,,$$

with $\beta_i \geq 0$, for $i \in \{1, \ldots, p\}$, or using the same notation as above:

$$E\left[\exp(ux_{t+1}) \,|\, \underline{x_t}\right] = \exp\left[\tfrac{u}{1-u\mu}\varphi' X_t - \nu \log(1-u\mu)\right] \,. \tag{6}$$

Similarly, given X_t, x_{t+1} can be obtained by drawing U_{t+1} in $\mathcal{P}(\frac{\varphi' X_t}{\mu})$ and $\frac{x_{t+1}}{\mu}$ in $\gamma(\nu + U_{t+1})$. It easily seen that the conditional mean and variance of x_{t+1}, given $\underline{x_t}$, are respectively given by $\nu\mu + \varphi' X_t$ and $\nu\mu^2 + 2\mu\varphi' X_t$; so, the process x_{t+1} has the weak AR(p) representation:

$$x_{t+1} = \nu\mu + \varphi' X_t + \varepsilon_{t+1} \,, \tag{7}$$

where ε_{t+1} is a conditionally heteroscedastic martingale difference, whose conditional variance is $\nu\mu^2 + 2\mu\varphi' X_t$; the process is stationary if and only if $\varphi' e < 1$ [where $e = (1, \ldots, 1) \in \mathbb{R}^p$] and, in this case, the process ε_{t+1} has finite unconditional variance given by $\nu\mu^2 + 2\nu\mu^2 \frac{\varphi' e}{1-\varphi' e}$. The unconditional mean of x_{t+1} is given by $\frac{\nu\mu}{1-\varphi' e}$.

2.4 Univariate Switching Regimes Car(p) Process

Let us first consider a J-states homogeneous Markov Chain z_{t+1}, which can take the values $e_j \in \mathbb{R}^J$, $j \in \{1, \ldots, J\}$, where e_j is the j^{th} column of the $(J \times J)$ identity matrix. The transition probability, from state e_i to state e_j is $\pi(e_i, e_j) = Pr(z_{t+1} = e_j \,|\, z_t = e_i)$. It is first worth noting that z_{t+1} is a Car(1) process.

Proposition 1 : The Markov chain process z_{t+1} is a Car(1) process with a conditional Laplace transform given by:

$$E[\exp(v' z_{t+1}) |\, \underline{z_t}] = \exp(a_z(v, \pi)' z_t) \,, \tag{8}$$

where

$$a_z(v,\pi) = \left[\log\left(\sum_{j=1}^{J}\exp(v'e_j)\pi(e_1,\, e_j)\right),\ldots,\log\left(\sum_{j=1}^{J}\exp(v'e_j)\pi(e_J,\, e_j)\right)\right]'.$$

[Proof : straightforward.]

Let us now consider a univariate Index-Car(p) process with a conditional Laplace transform given by $\exp\left[a(u)\beta' X_t + b(u)\right]$, and let us assume that $b(u)$ can be written:

$$\begin{aligned} b(u) &= \tilde{b}(u)'\lambda \quad \text{where} \\ \tilde{b}(u) &= (b_1(u),\ldots,b_m(u))' \text{ and } \lambda = (\lambda_1,\ldots,\lambda_m)'. \end{aligned} \tag{9}$$

We can generalize this model by assuming that the parameters λ_i are stochastic and linear functions of $Z_t = (z'_t,\ldots,z'_{t-p})'$. More precisely, we assume that the conditional distribution of x_{t+1} given $\underline{x_t}$ and $\underline{z_{t+1}}$ has a Laplace transform given by:

$$E[\exp(ux_{t+1})\,|\,\underline{x_t}\,,\underline{z_{t+1}}] = \exp\left[a(u)\beta' X_t + \tilde{b}(u)'\Lambda Z_t\right], \tag{10}$$

where Λ is a $[m,(p+1)J]$ matrix. Note that we assume no instantaneous causality between x_{t+1} and z_{t+1} and we admit one more lag in Z_t that in X_t [examples given in Section 2.5 show that this assumption may be convenient]; if the process z_t is not observed by the econometrician the no instantaneous causality assumption is not really important at the estimation stage since we could rename z_t as z_{t+1}, however it will be useful at the pricing level in order to obtain simple pricing procedures [Dai, Singleton and Yang (2005) also make this kind of assumption]. The joint process $(x_{t+1}, z'_{t+1})'$ is easily seen to be a Car($p+1$) process.

Proposition 2 : The conditional Laplace transform of $(x_{t+1}, z'_{t+1})'$ given $\underline{x_t}\,,\underline{z_t}$ has the following form:

$$E\left[\exp(ux_{t+1} + v'z_{t+1})\,|\,\underline{z_t},\,\underline{x_t}\,\right] = \exp\left\{a(u)\beta' X_t + \left[e'_1 \otimes a_z(v,\pi)' + \tilde{b}(u)'\Lambda\right] Z_t\right\}, \tag{11}$$

where e_1 is the first component of the canonical basis in $\mathbb{R}^{p+1}$, and where $\otimes$ denotes the Kronecker product. [Proof : straightforward.]

2.5 Examples of Univariate Switching Regimes Car(p) Processes

a. Gaussian case

Let us start from the AR(p) model (4). Its conditional Laplace transform is given by (5):

$$E\left[\exp(ux_{t+1})\,|\,\underline{x_t}\right] = \exp\left[u\varphi' X_t + u\nu + \tfrac{\sigma^2}{2}u^2\right],$$

and the function $b(u)$ has the form (9) with $\tilde{b}(u)' = \left(u,\, \frac{u^2}{2}\right)$ and $\lambda' = (\nu, \sigma^2)$.

If λ is replaced by ΛZ_t, the joint process $(x_{t+1}, z'_{t+1})'$ is Car($p+1$) with a conditional Laplace transform given by:

$$E\left[\exp(ux_{t+1} + v'z_{t+1})\,|\,\underline{z_t},\,\underline{x_t}\,\right] = \exp\left[u\varphi' X_t + \left(u,\, \tfrac{u^2}{2}\right)\Lambda Z_t + a_z(v,\pi)z_t\right]. \tag{12}$$

More precisely, the dynamics is given by [using the notation $\Lambda = \begin{pmatrix}\lambda_1' \\ \lambda_2'\end{pmatrix}$]:

$$x_{t+1} = \lambda_1' Z_t + \varphi' X_t + (\lambda_2' Z_t)^{1/2} \varepsilon_{t+1} \,, \tag{13}$$

where ε_{t+1} is a gaussian white noise distributed as $\mathcal{N}(0, \sigma^2)$, $Z_t = (z_t', \ldots, z_{t-p}')'$ and z_t is a Markov chain such that $Pr(z_{t+1} = e_j \,|\, z_t = e_i) = \pi(e_i,\, e_j)$.

In particular, let us consider the case:

$$\Lambda = \begin{bmatrix} (1, -\varphi_1, \ldots, -\varphi_p) \otimes \nu^{*'} \\ e_1' \otimes \sigma^{*2'} \end{bmatrix} \tag{14}$$

and $\nu^{*'} = (\nu_1^*, \ldots, \nu_J^*)$, $\sigma^{*2'} = (\sigma_1^{*2}, \ldots, \sigma_J^{*2})$, the conditional distribution of x_{t+1} given $\underline{x_t}$ and $\underline{z_{t+1}}$ is the one corresponding to the switching AR(p) model defined by:

$$x_{t+1} - \nu^{*'} z_t = \varphi_1 \left(x_t - \nu^{*'} z_{t-1}\right) \,+\, \ldots \,+\, \varphi_p \left(x_{t+1-p} - \nu^{*'} z_{t-p}\right) \,+\, (\sigma^{*'} z_t)\varepsilon_{t+1} \,. \tag{15}$$

b. Gamma case

Let us now start from the ARG(p) process associated with the conditional Laplace transform (6):

$$E\left[\exp(u x_{t+1}) \,|\, \underline{x_t}\right] = \exp\left[\tfrac{u}{1-u\mu} \varphi' X_t - \nu \log(1 - u\mu)\right] \,.$$

Here we have $\tilde{b}(u) = -\log(1 - u\mu)$ and $\lambda = \nu$. If ν is replaced by ΛZ_t, where $\Lambda Z_t > 0$, the process x_t has, conditionally to the process z_t, a weak AR(p) representation given by:

$$x_{t+1} = \mu \Lambda Z_t + \varphi_1 \, x_t \,+\, \ldots \,+\, \varphi_p \, x_{t+1-p} \,+\, \zeta_{t+1} \,, \tag{16}$$

where ζ_{t+1} is a conditionally heteroscedastic martingale difference. For instance, if we take :

$$\Lambda = e_1' \otimes \frac{\tilde{\nu}\,'}{\mu} \,, \tag{17}$$

where $\tilde{\nu}\,' = (\tilde{\nu}_1, \ldots, \tilde{\nu}_J)$, $\tilde{\nu}_j \geq 0$, we have $\Lambda Z_t = \frac{\tilde{\nu}\,'}{\mu} z_t$ and, conditionally to the process z_t, the process x_t has a weak AR(p) representation given by:

$$x_{t+1} = \tilde{\nu}\,' z_t + \varphi_1 \, x_t \,+\, \ldots \,+\, \varphi_p \, x_{t+1-p} \,+\, \zeta_{t+1} \,. \tag{18}$$

It is also possible to consider a Λ of the form $(1, -\varphi_1, \ldots, -\varphi_p) \otimes \frac{\tilde{\nu}\,'}{\mu}$ if $\min(\tilde{\nu}_i) > \max(\tilde{\nu}_i) \sum_{i=1}^{J} \varphi_j$, since in this case $\Lambda Z_t = \frac{1}{\mu}\left(\tilde{\nu}\,' z_t - \sum_{i=1}^{J} \varphi_j \, \tilde{\nu}\,' z_{t-i}\right) \geq 0$. The weak conditional AR(p) representation is then given by:

$$x_{t+1} - \tilde{\nu}\,' z_t = \varphi_1 \left(x_t - \tilde{\nu}\,' z_{t-1}\right) \,+\, \ldots \,+\, \varphi_p \left(x_{t+1-p} - \tilde{\nu}\,' z_{t-p}\right) \,+\, \zeta_{t+1} \,. \tag{19}$$

2.6 Specification of Multivariate Car(1) Processes

In order to have simple notations we will consider the bivariate case, but all the results are easily extended to the general case. A bivariate Car(1) process $y_t = (y_{1,t}, y_{2,t})'$ will be defined in a recursive way. We consider two univariate exponential affine Laplace transforms :

$$\begin{aligned} &\exp\left[a_1(u_1) w_{1,t} + b_1(u_1)\right] \,, \\ \text{and} \qquad &\exp\left[a_2(u_2) w_{2,t} + b_2(u_2)\right] \,. \end{aligned} \tag{20}$$

Then, we assume that the conditional distribution of $y_{1,t+1}$ given $(y_{2,t+1}, \underline{y_{1,t}}, \underline{y_{2,t}})$ has a Laplace transform given by :

$$E_t[\exp(u_1 y_{1,t+1}) \,|\, y_{2,t+1}, \underline{y_{1t}}, \underline{y_{2t}}] = \exp\left[a_1(u_1)(\beta_o y_{2,t+1} + \beta_{11} y_{1,t} + \beta_{12} y_{2,t}) + b_1(u_1)\right] \tag{21}$$

and the conditional distribution of $y_{2,t+1}$, given $(\underline{y_{1,t}}, \underline{y_{2,t}})$, has a Laplace transform given by

$$E_t[\exp(u_2 y_{2,t+1}) \,|\, \underline{y_{1,t}}, \underline{y_{2,t}}] = \exp\left[a_2(u_2)(\beta_{21} y_{1,t} + \beta_{22} y_{2,t}) + b_2(u_2)\right] . \tag{22}$$

Note that, if the Laplace transforms (20) correspond to positive variables and if the parameters $\beta_o, \beta_{11}, \beta_{12}, \beta_{21}, \beta_{22}$ are positive the bivariate process y_t has positive components. Moreover, we have the following result :

Proposition 3 : The bivariate process y_t defined by the conditional dynamics (21), (22) is a bivariate Car(1) process with a conditional Laplace transform given by :

$$\begin{aligned} E[\exp(u_1 y_{1,t+1} + u_2 y_{2,t+1}) | \underline{y_{1,t}}, \underline{y_{2,t}}] \;=\; & \exp\{[a_1(u_1)\beta_{11} + a_2(u_2 + a_1(u_1)\beta_o)\beta_{21}] y_{1,t} \\ & + [a_1(u_1)\beta_{12} + a_2(u_2 + a_1(u_1)\beta_o)\beta_{22}] y_{2,t} \\ & + b_1(u_1) + b_2(u_2 + a_1(u_1)\beta_o)\} . \end{aligned} \tag{23}$$

[Proof : see Appendix 1.]

2.7 Specification of Multivariate Index-Car(p) Processes

We consider a bivariate process $\tilde{x}_t = (x_{1,t}, x_{2,t})'$ and we introduce the notations: $X_{1t} = (x_{1,t}, \ldots, x_{1,t+1-p})'$, $X_{2t} = (x_{2,t}, \ldots, x_{2,t+1-p})'$. Given univariate Laplace transforms like (20), a bivariate Index-Car(p) is defined in the following way.

Definition 3 : A bivariate Index-Car(p) dynamics is defined by the conditional Laplace transforms:

$$\begin{aligned} & E_t[\exp(u_1 x_{1,t+1}) \,|\, x_{2,t+1}, \underline{x_{1,t}}, \underline{x_{2,t}}] \\ & = \exp\left[a_1(u_1)(\beta_o x_{2,t+1} + \beta'_{11} X_{1t} + \beta'_{12} X_{2t}) + b_1(u_1)\right] , \\ & E_t[\exp(u_2 x_{2,t+1}) \,|\, \underline{x_{1,t}}, \underline{x_{2,t}}] = \exp\left[a_2(u_2)(\beta'_{21} X_{1t} + \beta'_{22} X_{2t}) + b_2(u_2)\right] , \end{aligned} \tag{24}$$

where the β_{ij} are p-vectors. It is easily seen that the process $\tilde{x}_t$ is a Car(p) process with a conditional Laplace transform given by (23) in which $y_{1,t}$ is replaced by X_{1t} and $y_{2,t}$ by X_{2t} and the β_{ij} by the β'_{ij}. Let us consider two important particular cases:

a) Normal VAR(p) or VARN(p) Processes

In this case the conditional distributions defined by (20) are gaussian, with affine expectations and fixed variances. In other words:

$$\begin{aligned} a_1(u_1) = \rho_1 u_1 \, , \; b_1(u_1) = \nu_1 u_1 + \tfrac{\sigma_1^2 u_1^2}{2} \\ a_2(u_2) = \rho_2 u_2 \, , \; b_2(u_2) = \nu_2 u_2 + \tfrac{\sigma_2^2 u_2^2}{2} \, . \end{aligned} \tag{25}$$

Using the notations $\varphi_o = \rho_1\beta_o$, $\varphi_{11} = \rho_1\beta_{11}$, $\varphi_{12} = \rho_1\beta_{12}$, $\varphi_{21} = \rho_2\beta_{21}$, $\varphi_{22} = \rho_2\beta_{22}$, we have the following strong VAR(p) recursive representation for the process $\tilde{x}_t = (x_{1,t}, x_{2,t})'$:

$$\begin{cases} x_{1,t+1} &= \nu_1 + \varphi_o x_{2,t+1} + \varphi_{11}' X_{1t} + \varphi_{12}' X_{2t} + \sigma_1 \eta_{1,t+1} \\ x_{2,t+1} &= \nu_2 + \varphi_{21}' X_{1t} + \varphi_{22}' X_{2t} + \sigma_2 \eta_{2,t+1} \,, \end{cases} \tag{26}$$

where $\eta_t = (\eta_{1,t}, \eta_{2,t})'$ is a bivariate gaussian white noise distributed as $\mathcal{N}(0, I_2)$, where I_2 denotes the (2×2) identity matrix.

b) Gamma VAR(p) or VARG(p) Processes

In this case we have:

$$\begin{aligned} a_1(u_1) &= \tfrac{\rho_1 u_1}{1-u_1\mu_1} \,,\; b_1(u_1) = -\nu_1 \log(1 - u_1\mu_1) \\ a_2(u_2) &= \tfrac{\rho_2 u_2}{1-u_2\mu_2} \,,\; b_2(u_2) = -\nu_2 \log(1 - u_2\mu_2) \,, \end{aligned} \tag{27}$$

and the process $\tilde{x}_t = (x_{1,t}, x_{2,t})'$ has the following weak VAR(p) representation (using the same notation as above, and where all the parameters are positive):

$$\begin{cases} x_{1,t+1} &= \nu_1\mu_1 + \varphi_o x_{2,t+1} + \varphi_{11}' X_{1t} + \varphi_{12}' X_{2t} + \xi_{1,t+1} \\ x_{2,t+1} &= \nu_2\mu_2 + \varphi_{21}' X_{1t} + \varphi_{22}' X_{2t} + \xi_{2,t+1} \,, \end{cases} \tag{28}$$

where $\xi_{1,t}$ and $\xi_{2,t}$ are non correlated, conditionally heteroscedastic, martingale differences. It is important to stress that the components of this VARG(p) process are positive[7].

Switching regimes can also be introduced in a multivariate Index-Car(p) model using a method extending the one retained in the univariate case [see Pegoraro (2006) for a detailed presentation].

3 Switching Autoregressive Normal (SARN) Factor-Based Term Structure Model of order p

We first consider the case of a univariate latent factor (x_t); the case of an observable factor and the multivariate cases will be discussed, respectively, in sections 3.7 and 3.8.

3.1 The Historical Dynamics

The first set of assumptions of a SARN(p) Term Structure model deals with the historical dynamics. We assume that the historical dynamics of the exogenous factor x_t is given by

$$x_{t+1} \;=\; \nu(Z_t) + \varphi_1(Z_t)x_t + \ldots + \varphi_p(Z_t)x_{t+1-p} + \sigma(Z_t)\varepsilon_{t+1} \,, \tag{29}$$

where ε_{t+1} is a gaussian white noise with $\mathcal{N}(0,1)$ distribution, $Z_t = (z_t', \ldots, z_{t-p}')'$, and z_t is a J-states non-homogeneous Markov chain such that $P(z_{t+1} = e_j \,|\, z_t = e_i; \underline{x_t}) = \pi(e_i, e_j\,; X_t)$ (e_i is the i^{th} column of the identity matrix I_J). Equation (29) will be also written

$$x_{t+1} \;=\; \nu(Z_t) + \varphi(Z_t)' X_t + \sigma(Z_t)\varepsilon_{t+1} \,, \tag{30}$$

[7] In a recent paper Dai, Le and Singleton (2006) propose a multivariate conditionally Gaussian term structure model where nonlinearities are introduced in the (latent) state-factor (historical and risk-neutral) dynamics by means of stochastic volatility factors; the joint risk-neutral dynamics of these volatility factors is described by a particular VARG(1) process with conditionally independent components [$\varphi_o = 0$ in our system (28) notation].

where $X_t = (x_t, \ldots, x_{t+1-p})'$, $\varphi(Z_t) = (\varphi_1(Z_t), \ldots, \varphi_p(Z_t))'$. This model can also be rewritten in the following vectorial form:

$$X_{t+1} = \Phi(Z_t)X_t + [\nu(Z_t) + \sigma(Z_t)\varepsilon_{t+1}]\, e_1 \tag{31}$$

where

$$\Phi(Z_t) = \begin{bmatrix} \varphi_1(Z_t) & \ldots & \ldots & \varphi_p(Z_t) \\ 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \ldots & 1 & 0 \end{bmatrix}$$

is a $(p \times p)$-matrix, and where e_1 is the first column of the identity matrix I_p. Note that, since the coefficients φ_i are allowed to depend on Z_t and since the Markov chain z_t may not be homogeneous, the dynamics of (x_t, z_t) is not Car in general.

3.2 The Stochastic Discount Factor

The second element of a SARN(p) modeling is the SDF. We denote by $M_{t,t+1}$ the stochastic discount factor (SDF) between the date t and $t+1$ and in order to get time-varying risk-premia we specify it as an exponential affine function of the variables $(x_{t+1},\ z_{t+1})$ but with coefficients depending on the information at time t. More precisely we assume that:

$$M_{t,t+1} = \exp\left[-c'X_t - d'Z_t + \Gamma(Z_t, X_t)\,\varepsilon_{t+1} - \tfrac{1}{2}\Gamma(Z_t, X_t)^2 - \delta(Z_t, X_t)'z_{t+1}\right], \tag{32}$$

where $\Gamma(Z_t, X_t) = \gamma(Z_t) + \tilde{\gamma}'(Z_t)X_t$ and $\delta(Z_t, X_t) = [\,\delta_1(Z_t, X_t), \ldots, \delta_J(Z_t, X_t)]'$. Our specification of the factor risk correction coefficient $\Gamma(Z_t, X_t)$ extends to the multi-lag case the regime-switching essentially affine specification proposed by Dai, Singleton and Yang (2005). Bansal and Zhou (2002) assume a market price of factor risk proportional to factor volatilities (completely affine specification)[8]. Duffee (2002) and Dai and Singleton (2002) show that, among single-regime continuous time term structure models, essentially affine specifications for the market price of factor risk explain dynamic properties of yield curves better than the completely affine specifications of multifactor CIR models. In sections 6.6 and 6.7 we will find confirmation of this result and, in Section 6.8, we will see how pricing the risk associated to the second lagged factor value is crucial in explaining the long horizon Expectation Hypothesis Puzzle. Naik and Lee (1997), Bansal and Zhou (2002) and Ang and Bekaert (2005) consider the j^{th}-regime risk correction coefficient $\delta_j(Z_t, X_t) = 0$ for each $j \in \{1, \ldots, J\}$, while, the fact to price regime-shift risk, gives to our approach the possibility to better fit interest rates dynamics [see Section 6.8].

It is well known that the existence of a positive stochastic discount factor is equivalent to the absence of arbitrage opportunity condition and that the price p_t at t of a payoff W_{t+1} at $t+1$ is given by:

$$\begin{aligned} p_t &= E[M_{t,t+1}W_{t+1} \mid I_t] \\ &= E_t[M_{t,t+1}W_{t+1}], \end{aligned}$$

where the information I_t, available for the investors at the date t, is given by $(\underline{x_t}, \underline{z_t})$. More generally, the price $p_{t,h}$ at t of an asset paying W_{t+h} at $t+h$ is:

$$p_{t,h} = E_t\left[M_{t,t+1} \cdot \ldots \cdot M_{t+h-1,t+h}W_{t+h}\right].$$

[8] A market price of factor risk is said to be essentially affine when it is proportional to both factor volatilities and state-factors [see Duffee (2002), Dai and Singleton (2003)].

Using the absence of arbitrage assumption for the short-term interest rate between t and $t+1$, denoted by r_{t+1} and known at t, we get:

$$\begin{aligned}\exp(-r_{t+1}) &= E_t\left(M_{t,t+1}\right) \\ &= \exp\left[-c'X_t - d'Z_t\right] \times \textstyle\sum_{j=1}^{J} \pi\left(e_i, e_j\,; X_t\right) \exp\left[-\delta(Z_t, X_t)'e_j\right] \,,\end{aligned}$$

and assuming the normalization condition:

$$\textstyle\sum_{j=1}^{J} \pi\left(e_i, e_j\,; X_t\right) \exp\left[-\delta(Z_t, X_t)'e_j\right] = 1 \quad \forall Z_t, X_t\,, \tag{33}$$

we obtain:

$$r_{t+1} = c'X_t + d'Z_t\,. \tag{34}$$

3.3 Risk Premia

In this paper we will follow the definition of risk premium proposed by Dai, Singleton and Yang (2005). More precisely, if we denote with p_t the price of a given asset at time t, the risk premium of this asset between t and $t+1$ is $\omega_t = \log(E_t p_{t+1}) - \log p_t - r_{t+1}$. In particular, the risk premium between t and $t+1$ of an asset providing the payoff $\exp(-\theta x_{t+1})$ at $t+1$ is :

$$\omega_t(\theta) = \theta\Gamma(X_t, Z_t)\sigma(Z_t)\,. \tag{35}$$

and, if we consider a digital asset providing one money unit at $t+1$ if $z_{t+1} = e_j$, its risk premium between t and $t+1$ is given by :

$$\omega_t(\theta) = \delta_j(X_t, Z_t)\,. \tag{36}$$

We observe that, in general, the magnitude of the risk premium $\omega_t(\theta)$ is not just depending on the currently observed values x_t and z_t [as in Dai, Singleton and Yang (2005)], but it reflects the present and past values of both factors, that is, it is a function of the larger information represented by X_t and Z_t.

3.4 The Risk-Neutral Dynamics

The assumptions on the historical dynamics and on the SDF imply a risk-neutral dynamics. The probability density function of the one-period conditional risk-neutral probability with respect to the corresponding historical probability is $\frac{M_{t,t+1}}{E_t(M_{t,t+1})} = \exp(r_{t+1})M_{t,t+1}$. Note that using $E_t^{\mathbb{Q}}$ as the conditional expectation with respect to this risk-neutral distribution, the risk-premium ω_t can be written $\log(E_t p_{t+1}) - \log(E_t^{\mathbb{Q}} p_{t+1})$.

Proposition 4 : The risk-neutral dynamics of the process (x_t, z_t) is given by:

$$x_{t+1} \overset{\mathbb{Q}}{=} \nu(Z_t) + \gamma(Z_t)\sigma(Z_t) + [\varphi(Z_t) + \tilde{\gamma}(Z_t)\sigma(Z_t)]'X_t + \sigma(Z_t)\xi_{t+1}\,, \tag{37}$$

where $\overset{\mathbb{Q}}{=}$ denotes the equality in distribution (associated to the probability $\mathbb{Q}$), ξ_{t+1} is (under $\mathbb{Q}$) a gaussian white noise with $\mathcal{N}(0,1)$ distribution, and z_t is a Markov chain such that:

$$\mathbb{Q}(z_{t+1} = e_j \mid \underline{z_t}; \underline{x_t}) = \pi\left(z_t, e_j\,; X_t\right) \exp\left[(-\delta(Z_t, X_t))'e_j\right]\,.$$

Note that, from (33), these probabilities add to one. [Proof : see Appendix 2.]

In order to get a generalized linear term structure we impose that the risk-neutral dynamics is switching regime gaussian Car(p). Using (13), this impose that the dynamics has to satisfy the following specification :

$$x_{t+1} \overset{\mathbb{Q}}{=} \nu^{*\prime} Z_t + \varphi^{*\prime} X_t + (\sigma^{*\prime} Z_t)\xi_{t+1} \,, \tag{38}$$

where z_t is a J-states Markov chain such that[9]

$$\mathbb{Q}(z_{t+1} = e_j \,|\, z_t = e_i) = \pi^* \left(e_i, e_j\right) \,. \tag{39}$$

From Proposition 4, this implies the following restrictions on the historical dynamics and on the SDF:

i) $\sigma(Z_t) = \sigma^{*\prime} Z_t$: the historical stochastic volatility must be linear in Z_t;

ii)

$$\gamma(Z_t) \;=\; \frac{\nu^{*\prime} Z_t - \nu(Z_t)}{\sigma^{*\prime} Z_t} \;:$$

for given historical stochastic drift $\nu(Z_t)$ and stochastic volatility $\sigma^{*\prime} Z_t$, the coefficient $\gamma(Z_t)$ belongs to the previous family indexed by the free parameter vector ν^*.

iii)

$$\tilde{\gamma}(Z_t) \;=\; \frac{\varphi^* - \varphi(Z_t)}{\sigma^{*\prime} Z_t} \;:$$

for given historical stochastic slope parameter $\varphi(Z_t)$ and stochastic volatility $\sigma^{*\prime} Z_t$ the vector $\tilde{\gamma}(Z_t)$ belongs to the previous family indexed by the free parameter vector φ^*.

iv)

$$\delta_j(X_t, Z_t) \;=\; \log\left[\frac{\pi\left(z_t, e_j\,;X_t\right)}{\pi^*(z_t, e_j)}\right] \;:$$

for a given historical transition matrix $\pi\left(z_t, e_j\,;X_t\right)$, the coefficient $\delta_j(X_t, Z_t)$ depend on z_t only and belongs to the previous family indexed by the entries $\pi^*(z_t, e_j)$ of a transition matrix.

Note that condition *iv*) implies that the risk premia coefficients δ_j, $j \in \{1, \ldots, J\}$, cannot be all positive [or all negative] since this would imply $\pi\left(z_t, e_j\,;X_t\right) > \pi^*(z_t, e_j)$, $\forall j$ [or $\pi\left(z_t, e_j\,;X_t\right) < \pi^*(z_t, e_j)$, $\forall j$], which is impossible since $\sum_{j=1}^{J} \pi\left(z_t, e_j\,;X_t\right) = \sum_{j=1}^{J} \pi^*(z_t, e_j) = 1$. Also note that condition *iv*) implies the normalization condition (33).

[9] Ang and Bekaert (2005) also assume their counterpart to φ^* to be constant over time, and a homogeneous (risk-neutral) transition matrix for (z_t). Bansal and Zhou (2002) consider $\mathbb{Q}(z_{t+1} = e_j \,|\, z_t = e_i) = \pi^*\left(e_i, e_j\right)$, but they allow the risk-neutral autoregressive matrix to switch over time [$\varphi^*(z_{t+1})$, in our notation] : this feature leads their approach to use a log-linear approximation in order to find an explicit (approximate) pricing formula. This kind of log-linear approximation, in a general equilibrium square-root term structure model, is also used by Wu and Zeng (2005).

3.5 The Generalized Linear Term Structure

We have seen in the previous section that the risk-neutral dynamics is defined by relations (38), (39); relation (38) can be rewritten:

$$X_{t+1} \overset{\mathbb{Q}}{=} \Phi^* X_t + \left[\nu^{*'} Z_t + (\sigma^{*'} Z_t)\xi_{t+1}\right] e_1 \tag{40}$$

where

$$\Phi^* = \begin{bmatrix} \varphi_1^* & \dots & \dots & \varphi_p^* \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \dots & 1 & 0 \end{bmatrix} \text{ is a } (p \times p) - \text{matrix},$$

$$X_t = (x_t, \dots, x_{t+1-p})',$$

and where e_1 is the first column of the identity matrix I_p.

Denoting by $B(t,h)$ the price at t of a zero-coupon with residual maturity h, we have the following result.

Proposition 5 : In the univariate SARN(p) Term Structure model the price at date t of the zero-coupon bond with residual maturity h is :

$$B(t,h) = \exp\left(C_h' X_t + D_h' Z_t\right), \text{ for } h \geq 1, \tag{41}$$

where the vectors C_h and D_h satisfy the following recursive equations :

$$\begin{cases} C_h = \Phi^{*'} C_{h-1} - c \\ \\ D_h = -d + C_{1,h-1}\,\nu^* + \frac{1}{2} C_{1,h-1}^2\,\sigma^{*2} + \tilde{D}_{h-1} + F(D_{1,h-1}), \end{cases} \tag{42}$$

where $C_{1,h-1}$ denotes the first component of the p-dimensional vector C_{h-1}, $D_{1,h-1}$ and $D_{2,h-1}$ are, respectively, the first J-dimensional component and the remaining (pJ)-dimensional component of D_{h-1}, i.e. $D_{h-1} = (D_{1,h-1}', D_{2,h-1}')'$, $\tilde{D}_{h-1} = (D_{2,h-1}', 0)'$, and where $F(D_{1,h-1}) = e_1 \otimes a_z(D_{1,h-1}, \pi^*)$, e_1 being the vector $(1, 0, \dots, 0)'$ of size $(p+1)$ and a_z is the J-vector given in proposition 1; σ^{*2} is the vector whose components are the squares of the entries of σ^*. The initial conditions are $C_0 = 0$, $D_0 = 0$ (or $C_1 = -c$, $D_1 = -d$). [Proof : see Appendix 3.]
For clarity we give again the expression of $a_z(D_{1,h-1}, \pi^*)$:

$$a_z(D_{1,h-1}, \pi^*)$$

$$= \left[\log\left(\sum_{j=1}^{J} \exp(D_{1,h-1}' e_j)\pi^*(e_1,\, e_j)\right), \dots, \log\left(\sum_{j=1}^{J} \exp(D_{1,h-1}' e_j)\pi^*(e_J,\, e_j)\right)\right]'.$$

From Proposition 5 we see that the yields to maturity are:

$$\begin{aligned} R(t,h) &= -\frac{1}{h}\log B(t,h) \\ &= -\frac{C_h'}{h} X_t - \frac{D_h'}{h} Z_t, \quad h \geq 1. \end{aligned} \tag{43}$$

So, they are linear functions of the p-dimensional vector X_t and of the $(p+1)J$-dimensional vector Z_t. This means that, the term structure at date t depends on the present and past values of x_t and z_t, and not just on their values in t. Moreover, we observe that there is, in general, instantaneous causality between $R(t,h)$ and z_t.

3.6 The Switching VARMA Yield Curve Process

The result presented in Proposition 5 describes, conditionally to X_t and Z_t, the yields as a deterministic function of the time to maturity h, for a fixed date t. Nevertheless, in many financial and economic contexts one needs, for instance, also to study the effects, of a shock in the state variables, on the yield curve at different future times and for several maturities (e.g.: a Central Bank that needs to set a monetary policy). This means that we are interested in the dynamics of the process $R_{\mathcal{H}} = [\,R(t,h),\, 0 \leq t < T,\, h \in \mathcal{H}\,]$, for a given set of residual time to maturities $\mathcal{H} = (1,\ldots,H)$.

If we consider a fixed h, the process $R = [\,R(t,h),\, 0 \leq t < T\,]$ can be described by the following proposition.

Proposition 6 : For a fixed time to maturity h, the process $R = [\,R(t,h),\, 0 \leq t < T\,]$ is, under the historical probability, a Switching ARMA(p, $p-1$) process of the following type :

$$\Psi(L,Z_t)\,R(t+1,h) \;=\; D_h(L)\,\Psi(L,Z_t)\,z_{t+1} + C_h(L)\,\nu(Z_t) + C_h(L)[(\sigma^{*'}Z_t)\,\varepsilon_{t+1}]\,. \tag{44}$$

where

$$\begin{aligned} C_h(L) &= -\frac{1}{h}\,(C_{1,h} + C_{2,h}L + \ldots + C_{p,h}L^{p-1}) \\ D_h(L) &= -\frac{1}{h}\,(D_{1,h} + D_{2,h}L + \ldots + D_{p+1,h}L^{p}) \\ \Psi(L,Z_t) &= 1 - \varphi_1(Z_t)L - \ldots - \varphi_p(Z_t)L^p\,, \end{aligned}$$

are lag polynomials in the lag operator L, and where the AR polynomial $\Psi(L,Z_t)$ applies to t. [Proof : see Appendix 4].

Proposition 7 : For a given set of residual time to maturities $\mathcal{H} = (1,\ldots,H)$, the stochastic evolution of the yield curve process $R_{\mathcal{H}} = [\,R(t,h),\; 0 \leq t < T,\, h \in \mathcal{H}\,]$ takes the following particular Switching H-variate VARMA(p, $p-1$) representation:

$$\Psi(L,Z_t)\begin{pmatrix} R(t+1,1) \\ R(t+1,2) \\ \vdots \\ R(t+1,H) \end{pmatrix} = \begin{pmatrix} C_1(L) \\ C_2(L) \\ \vdots \\ C_H(L) \end{pmatrix}(\sigma^{*'}Z_t)\varepsilon_{t+1} + \begin{pmatrix} D_1(L) \\ D_2(L) \\ \vdots \\ D_H(L) \end{pmatrix}\Psi(L,Z_t)\,z_{t+1} + \begin{pmatrix} C_1(L) \\ C_2(L) \\ \vdots \\ C_H(L) \end{pmatrix}\nu(Z_t)\,. \tag{45}$$

Similar results are easily obtained in the risk-neutral world.

3.7 The Case of an Observable Factor

In the previous sections the factor x_t was latent. It is often assumed, in term structure models, that the factor (x_t) is the short rate process (r_{t+1}). In this case the previous results remain valid, and the only modification comes from the absence of arbitrage opportunity condition for r_{t+1}, which imposes:

$$c = e_1 \, , d = 0 \, , \tag{46}$$

with e_1 the first column of the identity matrix I_p. Consequently, the initial conditions in the recursive equations of Proposition 5 become:

$$C_1 = -e_1 \, , D_1 = 0 \, . \tag{47}$$

Moreover, the Switching ARMA(p, $p-1$) representation (44), or its analogous in the risk-neutral world, could be used to analyse how a shock on ε_t, i.e. on $r_{t+1} = R(t,1)$, is propagated on the surface $[\, R(t+\tau, h), \, \tau \in \mathcal{T}, \, h \in \mathcal{H} \,]$, where $\mathcal{T} = \{0, \ldots, T-t-1\}$ and $\mathcal{H} = (1, \ldots, H)$ (for instance when the process z_t is exogenous).

3.8 Multi-Factor Generalization : the SVARN(p) Factor-Based Term Structure Model

For sake of notational simplicity we consider the two factor case but an extension to more that two factors is straightforward. The historical dynamics of $\tilde{x}_t = (x_{1,t}, x_{2,t})'$ is a bivariate SVARN(p) model given by:

$$\left\{ \begin{array}{lcl} x_{1,t+1} & = & \nu_1(Z_t) + \varphi_o(Z_t) x_{2,t+1} + \varphi_{11}(Z_t)' X_{1t} + \varphi_{12}(Z_t)' X_{2t} + \sigma_1(Z_t) \varepsilon_{1,t+1} \\ \\ x_{2,t+1} & = & \nu_2(Z_t) + \varphi_{21}(Z_t)' X_{1t} + \varphi_{22}(Z_t)' X_{2t} + \sigma_2(Z_t) \varepsilon_{2,t+1} \, , \end{array} \right. \tag{48}$$

where $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ are independent standard normal white noises, $X_{1t} = (x_{1,t}, \ldots, \, x_{1,t+1-p})'$, $X_{2t} = (x_{2,t}, \ldots, x_{2,t+1-p})'$, $Z_t = (z_t', \ldots, z_{t-p}')'$, with z_t a J-states non-homogeneous Markov chain such that $P(z_{t+1} = e_j \,|\, z_t = e_i ; \underline{\tilde{x}_t}) = \pi(e_i, e_j \, ; \tilde{X}_t)$, and where $\tilde{X}_t = (X_{1t}', X_{2t}')'$. The recursive form (48) is equivalent to the canonical form:

$$\left\{ \begin{array}{lcl} x_{1,t+1} & = & \tilde{\nu}_1(Z_t) + \tilde{\varphi}_{11}(Z_t)' X_{1t} + \tilde{\varphi}_{12}(Z_t)' X_{2t} + \sigma_1(Z_t) \varepsilon_{1,t+1} + \varphi_o(Z_t) \sigma_2(Z_t) \varepsilon_{2,t+1} \\ \\ x_{2,t+1} & = & \nu_2(Z_t) + \varphi_{21}(Z_t)' X_{1t} + \varphi_{22}(Z_t)' X_{2t} + \sigma_2(Z_t) \varepsilon_{2,t+1} \, , \end{array} \right. \tag{49}$$

where $\tilde{\nu}_1 = \nu_1 + \varphi_o \nu_2$, $\tilde{\varphi}_{11} = \varphi_{11} + \varphi_o \varphi_{21}$, $\tilde{\varphi}_{12} = \varphi_{12} + \varphi_o \varphi_{22}$ or, with obvious notations:

$$\tilde{x}_{t+1} \;=\; \tilde{\nu}(Z_t) + \tilde{\Phi}(Z_t) \tilde{X}_t + S(Z_t) \varepsilon_{t+1} \, , \tag{50}$$

where

$$S(Z_t) = \begin{bmatrix} \sigma_1(Z_t) & \varphi_o(Z_t) \sigma_2(Z_t) \\ 0 & \sigma_2(Z_t) \end{bmatrix} .$$

Using the notation

$$\Gamma(Z_t, \tilde{X}_t) = \left[\Gamma_1(Z_t, \tilde{X}_t), \, \Gamma_2(Z_t, \tilde{X}_t) \right]' ,$$

where $\Gamma_i(Z_t, \tilde{X}_t) = \gamma_i(Z_t) + \tilde{\gamma}_i(Z_t)' \tilde{X}_t$, $i \in \{1, 2\}$ and $\Gamma(Z_t, \tilde{X}_t) = \gamma(Z_t) + \tilde{\Gamma}(Z_t) \tilde{X}_t$, with $\gamma(Z_t) = [\gamma_1(Z_t), \gamma_2(Z_t)]'$, $\tilde{\Gamma}(Z_t) = [\tilde{\gamma}_1(Z_t)', \tilde{\gamma}_2(Z_t)']'$, the SDF is defined as:

$$M_{t,t+1} \;=\; \exp \left[-c' \tilde{X}_t - d' Z_t + \Gamma(Z_t, \tilde{X}_t)' \, \varepsilon_{t+1} - \tfrac{1}{2} \Gamma(Z_t, \tilde{X}_t)' \Gamma(Z_t, \tilde{X}_t) - \delta(Z_t, \tilde{X}_t)' z_{t+1} \right] . \tag{51}$$

Assuming the normalization condition (33) and the absence of arbitrage opportunity for r_{t+1} we get:

$$r_{t+1} = c'\tilde{X}_t + d'Z_t\,. \tag{52}$$

It is also easily seen that the risk premium for an asset providing the payoff $\exp(-\theta'\tilde{x}_{t+1})$ at $t+1$ is $\omega(\theta) = \theta' S(Z_t)\Gamma(Z_t, \tilde{X}_t)$ and that the risk premium associated with the digital payoff $\mathbb{I}_{(e_j)}(z_{t+1})$ is unchanged.

Proposition 8 : The risk-neutral dynamics of the process $(\tilde{x}_t, z_t)$ is given by:

$$\tilde{x}_{t+1} \overset{\mathbb{Q}}{=} \tilde{\nu}(Z_t) + S(Z_t)\gamma(Z_t) + [\tilde{\Phi}(Z_t) + S(Z_t)\tilde{\Gamma}(Z_t)]\tilde{X}_t + S(Z_t)\xi_{t+1}\,, \tag{53}$$

where $\overset{\mathbb{Q}}{=}$ denotes the equality in distribution (associated to the probability $\mathbb{Q}$), ξ_{t+1} is (under $\mathbb{Q}$) a bivariate gaussian white noise with $\mathcal{N}(0, I_2)$ distribution, and where $Z_t = (z_t', \ldots, z_{t-p}')'$, with z_t a Markov chain such that:

$$\mathbb{Q}(z_{t+1} = e_j \,|\, \underline{z_t}; \underline{\tilde{x}_t}) = \pi(z_t, e_j\,; \tilde{X}_t) \exp\left[(-\delta(Z_t, \tilde{X}_t))' e_j\right]\,.$$

[Proof : see Appendix 5.]

If we want to obtain a Switching bivariate Car process in the risk-neutral world, we must have using similar arguments as in the univariate case :

i)

$$\begin{aligned} \sigma_1(Z_t) &= \sigma_1^{*'} Z_t \\ \sigma_2(Z_t) &= \sigma_2^{*'} Z_t \\ \varphi_o(Z_t) &= \varphi_o^*\,, \end{aligned}$$

and, therefore,

$$S(Z_t) = \begin{bmatrix} \sigma_1^{*'} Z_t & \varphi_o^* \sigma_2^{*'} Z_t \\ 0 & \sigma_2^{*'} Z_t \end{bmatrix} = S^*(Z_t)$$

ii)

$$\gamma(Z_t) = [S(Z_t)]^{-1}[\nu^* Z_t - \tilde{\nu}(Z_t)]\,,$$

where ν^* is a $(2 \times (p+1)J)$-matrix.

iii)

$$\tilde{\Gamma}(Z_t) = [S(Z_t)]^{-1}\left[\Phi^* - \tilde{\Phi}(Z_t)\right]\,,$$

where Φ^* is a $(2 \times 2p)$-matrix.

iv)

$$\delta_j(\tilde{X}_t, Z_t) = \log\left[\frac{\pi(z_t, e_j\,; \tilde{X}_t)}{\pi^*(z_t, e_j)}\right]\,.$$

The risk-neutral dynamics can be written:

$$\begin{cases} x_{1,t+1} & \overset{\mathbb{Q}}{=} & \nu_1^* Z_t + \Phi_1^* \tilde{X}_t + S_1^*(Z_t)\xi_{t+1} \\ \\ x_{2,t+1} & \overset{\mathbb{Q}}{=} & \nu_2^* Z_t + \Phi_2^* \tilde{X}_t + S_2^*(Z_t)\xi_{t+1}\,, \end{cases} \tag{54}$$

where ν_i^*, Φ_i^*, S_i^* are the i^{th} row of ν^*, Φ^*, S^*, with $i \in \{1,2\}$, or

$$\tilde{X}_{t+1} \overset{\mathbb{Q}}{=} \tilde{\Phi}^* \tilde{X}_t + [\nu_1^* Z_t + S_1^*(Z_t)\xi_{t+1}]\, e_1 + [\nu_2^* Z_t + S_2^*(Z_t)\xi_{t+1}]\, e_{p+1}\,,$$

where e_1 (respectively, e_{p+1}) is of size $2p$, with entries equal to zero except the first (respectively, the $(p+1)^{th}$) one which is equal to one, and

$$\tilde{\Phi}^* = \begin{bmatrix} \Phi_{11}^* & \Phi_{12}^* \\ \\ \tilde{I} & \tilde{\mathbf{0}} \\ \\ \Phi_{21}^* & \Phi_{22}^* \\ \\ \tilde{\mathbf{0}} & \tilde{I} \end{bmatrix}$$

where $\Phi_1^* = (\Phi_{11}^*, \Phi_{12}^*)$, $\Phi_2^* = (\Phi_{21}^*, \Phi_{22}^*)$, and where $\tilde{\mathbf{0}}$ is a $[(p-1) \times p]$-matrix of zeros and $\tilde{I}$ is a $[(p-1) \times p]$-matrix equal to $(I_{p-1}, 0)$, where 0 is a vector of size $(p-1)$.
The term structure is given by the following proposition:

Proposition 9 : In the bivariate SVARN(p) Term Structure model the price at date t of the zero-coupon bond with residual maturity h is :

$$B(t,h) = \exp\left(C_h' \tilde{X}_t + D_h' Z_t\right)\,, \text{ for } h \geq 1 \tag{55}$$

where the vectors C_h and D_h satisfy the following recursive equations :

$$\begin{cases} C_h = \tilde{\Phi}^{*'} C_{h-1} - c \\ \\ D_h = -d \,+\, C_{1,h-1}\, \nu_1^{*'} \,+\, C_{p+1,h-1}\, \nu_2^{*'} + \frac{1}{2} C_{1,h-1}^2(\sigma_1^{*2} + \varphi_o^{*2}\sigma_2^{*2}) \\ \\ \qquad + (C_{1,h-1})(C_{p+1,h-1})\varphi_o^{*2}\sigma_2^{*2} + \frac{1}{2} C_{p+1,h-1}^2 \sigma_2^{*2} + \tilde{D}_{h-1} + F(D_{1,h-1})\,, \end{cases} \tag{56}$$

where $\tilde{D}_{h-1}$ and $F(D_{1,h-1})$ have the same meaning as in Proposition 5, and the initial conditions are $C_0 = 0$, $D_0 = 0$ (or $C_1 = -c$, $D_1 = -d$). [Proof : see Appendix 6.]

So, Proposition 9 shows that the yields to maturity are:

$$R(t,h) \;=\; -\frac{C_h'}{h}\,\tilde{X}_t - \frac{D_h'}{h}\,Z_t\,, \quad h \geq 1\,. \tag{57}$$

In the case of observable factor, we can take $x_{1t} = r_{t+1}$, and $x_{2t} = R(t,H)$ for a given time to maturity H. In this case the absence of arbitrage conditions for r_{t+1} and $R(t,H)$ imply:

$$\begin{aligned} (i)\; C_1 &= -e_1,\; D_1 = 0\,,\text{ or } c = e_1,\; d = 0 \\ (ii)\; C_H &= -H\, e_{p+1},\; D_H = 0\,. \end{aligned} \tag{58}$$

Using the notations $C_h = (C_{1,h}, C^*_{1,h}, C_{p+1,h}, C^*_{2,h})'$, $\tilde{C}_{1,h} = (C^{*\prime}_{1,h}, 0)'$, $\tilde{C}_{2,h} = (C^{*\prime}_{2,h}, 0)'$ (where the zeros are scalars), and $\tilde{C}_h = (\tilde{C}'_{1,h}, \tilde{C}'_{2,h})'$, it easily seen that the recursive equation $C_h = \tilde{\Phi}^{*\prime} C_{h-1} - c$ can be written :

$$C_h = \Phi_1^{*\prime} C_{1,h-1} + \Phi_2^{*\prime} C_{p+1,h-1} + \tilde{C}_{h-1} - c\,.$$

Conditions (i) are used as initial values in the recursive procedure of Proposition 9, and conditions (ii) implies restrictions on the parameters $\tilde{\Phi}^*, \nu_1^*, \nu_2^*, \sigma_1^*, \sigma_2^*, \varphi_o^*, \pi^*(z_t, e_j)$ which must be taken into account at the estimation stage.

4 Switching Autoregressive Gamma (SARG) Factor-Based Term Structure Model of order p

For conciseness reasons, we present the model only in the scalar framework. A detailed presentation of the SVARG(p) Factor-Based Term Structure Models is given in Pegoraro (2006).

4.1 The Historical Dynamics

We assume that the Laplace transform of the conditional distribution of x_{t+1}, given $(\underline{x_t}, \underline{z_t})$, is:

$$\begin{aligned} E\left[\exp(u x_{t+1}) \,|\, \underline{x_t}, , \underline{z_t}\right] &= \exp\Big[\tfrac{u}{1-u\mu(X_t,Z_t)}\left[\varphi_1(Z_t)x_t + \ldots + \varphi_p(Z_t)x_{t-p+1}\right] \\ &\qquad - \nu(Z_t)\log(1-u\mu(X_t,Z_t))\Big]\,, \end{aligned} \tag{59}$$

where $Z_t = (z'_t, \ldots, z'_{t-p})'$, with z_t a J-states non-homogeneous Markov chain such that $P(z_{t+1} = e_j \,|\, z_t = e_i; \underline{x_t}) = \pi(e_i, e_j\,; \tilde{X}_t)$, and where $X_t = (x_t, \ldots, x_{t+1-p})'$. Using the notation:

$$\begin{aligned} A[u; \varphi(Z_t), \mu(X_t,Z_t)] &= \tfrac{u}{1-u\mu(X_t,Z_t)}[\varphi_1(Z_t), \ldots, \varphi_p(Z_t)]' = \tfrac{u}{1-u\mu(X_t,Z_t)}\varphi(Z_t) \\ b[u; \nu(Z_t), \mu(X_t,Z_t)] &= -\nu(Z_t)\log(1-u\mu(X_t,Z_t))\,, \end{aligned}$$

relation (59) can be written:

$$\begin{aligned} E\left[\exp(u x_{t+1}) \,|\, \underline{x_t}, , \underline{z_t}\right] &= \exp\{A[u; \varphi(Z_t), \mu(X_t,Z_t)]'X_t \\ &\qquad + b[u; \nu(Z_t), \mu(X_t,Z_t)]\}\,. \end{aligned} \tag{60}$$

The process (x_t) can also be written:

$$\begin{aligned} x_{t+1} &= \nu(Z_t)\mu(X_t,Z_t) + \varphi_1(Z_t)x_t + \ldots + \varphi_p(Z_t)x_{t+1-p} + \varepsilon_{t+1} \\ &= \nu(Z_t)\mu(X_t,Z_t) + \varphi(Z_t)'X_t + \varepsilon_{t+1}\,, \end{aligned} \tag{61}$$

where ε_{t+1} is a martingale difference sequence with conditional Laplace transform given by:

$$\begin{aligned} E\left[\exp(u\varepsilon_{t+1}) \,|\, \underline{x_t}, , \underline{z_t}\right] &= \exp\{-u[\nu(Z_t)\mu(X_t,Z_t) + \varphi(Z_t)'X_t] + A[u; \varphi(Z_t), \mu(X_t,Z_t)]'X_t \\ &\qquad + b[u; \nu(Z_t), \mu(X_t,Z_t)]\} \\ &= \exp\{[A[u; \varphi(Z_t), \mu(X_t,Z_t)] - u\varphi(Z_t)]'X_t \\ &\qquad + b[u; \nu(Z_t), \mu(X_t,Z_t)] - u\,\nu(Z_t)\mu(X_t,Z_t)\}\,. \end{aligned} \tag{62}$$

Note that the dynamics of (x_t, z_t) is in general not Car.

4.2 The Stochastic Discount Factor

In the SARG(p) model the SDF is specified in the following way:

$$\begin{aligned} M_{t,t+1} &= \exp\{-c'X_t - d'Z_t + \Gamma(Z_t, X_t)\varepsilon_{t+1} + \Gamma(Z_t, X_t)\,[\nu(Z_t)\mu(X_t, Z_t) + \varphi(Z_t)'X_t] \\ &\quad - A[\Gamma(Z_t, X_t); \varphi(Z_t), \mu(X_t, Z_t)]'X_t \\ &\quad -b[\Gamma(Z_t, X_t); \nu(Z_t), \mu(X_t, Z_t)] - \delta(Z_t, X_t)'z_{t+1}\}\ , \end{aligned} \tag{63}$$

where $\Gamma(Z_t, X_t) = \gamma(Z_t) + \tilde{\gamma}'(Z_t)X_t$, or, equivalently

$$\begin{aligned} M_{t,t+1} &= \exp\{-c'X_t - d'Z_t + \Gamma(Z_t, X_t)x_{t+1} - A[\Gamma(Z_t, X_t); \varphi(Z_t), \mu(X_t, Z_t)]'X_t \\ &\quad -b[\Gamma(Z_t, X_t); \nu(Z_t), \mu(X_t, Z_t)] - \delta(Z_t, X_t)'z_{t+1}\}\ , \end{aligned} \tag{64}$$

Assuming the normalization condition (33), we get that:

$$r_{t+1} = c'X_t + d'Z_t\,. \tag{65}$$

4.3 Useful Lemmas

In the subsequent sections we will use several times the following lemmas. Let us consider the functions:

$$\tilde{a}(u; \rho, \mu) = \frac{\rho u}{1 - u\mu} \quad \text{and} \quad \tilde{b}(u; \nu, \mu) = -\nu \log(1 - u\mu)\,;$$

we have:

Lemma 1 :

$$\tilde{a}(u + \alpha; \rho, \mu) - \tilde{a}(\alpha; \rho, \mu) = \tilde{a}(u; \rho^*, \mu^*)$$

$$\tilde{b}(u + \alpha; \nu, \mu) - \tilde{b}(\alpha; \nu, \mu) = \tilde{b}(u; \nu, \mu^*)$$

$$\text{with } \rho^* = \frac{\rho}{(1 - \alpha\mu)^2},\ \mu^* = \frac{\mu}{1 - \alpha\mu}\,,$$

[Proof : see Appendix 7.]

Lemma 1 immediately implies lemma 2.

Lemma 2 :

$$A[u + \alpha; \varphi(Z_t), \mu(X_t, Z_t)] - A[\alpha; \varphi(Z_t), \mu(X_t, Z_t)] = A[u; \varphi^*(Z_t), \mu^*(X_t, Z_t)]$$

$$b[u + \alpha; \nu(Z_t), \mu(X_t, Z_t)] - b[\alpha; \nu(Z_t), \mu(X_t, Z_t)] = b[u; \nu(Z_t), \mu^*(Z_t, X_t)]$$

$$\text{with } \varphi^*(Z_t) = \frac{\varphi(Z_t)}{[1 - \alpha\mu(Z_t, X_t)]^2},\ \mu^*(Z_t, X_t) = \frac{\mu(X_t, Z_t)}{1 - \alpha\mu(X_t, Z_t)}\,.$$

4.4 The Risk-Neutral Dynamics

The Laplace transform of the risk-neutral conditional distribution of (x_{t+1}, z_{t+1}) is, using the notation $\Gamma_t = \Gamma(X_t, Z_t)$:

$$\begin{aligned} & E_t^{\mathbb{Q}}[\exp(ux_{t+1} + v'z_{t+1})] \\ = \; & E_t\{\exp\left[(u+\Gamma_t)x_{t+1} - A[\Gamma_t; \varphi(Z_t), \mu(X_t, Z_t)]'X_t \; - b[Z_t; \nu(Z_t), \mu(X_t, Z_t)] \right. \\ & \left. \qquad\qquad\qquad + (v - \delta(X_t, Z_t))'z_{t+1}\right]\} \\ = \; & \exp\{[(A[u+\Gamma_t; \varphi(Z_t), \mu(X_t, Z_t)] - A[\Gamma_t; \varphi(Z_t), \mu(X_t, Z_t)])'X_t \\ & \qquad + b[u+\Gamma_t; \nu(Z_t), \mu(X_t, Z_t)] - b[\Gamma_t; \nu(Z_t), \mu(X_t, Z_t)]]\} \\ & \qquad\qquad \times \textstyle\sum_{j=1}^{J} \pi(z_t, e_j\,; X_t) \exp\left[(v - \delta(Z_t, X_t))'e_j\right] \;, \end{aligned} \tag{66}$$

and, using lemma 2, (67) can be written:

$$\begin{aligned} & E_t^{\mathbb{Q}}[\exp(ux_{t+1} + v'z_{t+1})] \\ = \; & \exp\{A[u; \varphi^*(Z_t), \mu^*(X_t, Z_t)]'X_t + b[u; \nu(Z_t), \mu^*(Z_t, X_t)]\} \\ & \qquad \times \textstyle\sum_{j=1}^{J} \pi(z_t, e_j\,; X_t) \exp\left[(v - \delta(Z_t, X_t))'e_j\right] \;, \end{aligned} \tag{67}$$

with $\varphi^*(Z_t) = \dfrac{\varphi(Z_t)}{[1 - \Gamma_t \mu(Z_t, X_t)]^2}$ and $\mu^*(Z_t, X_t) = \dfrac{\mu(X_t, Z_t)}{1 - \Gamma_t \mu(X_t, Z_t)}$.

So, from (60), we see that the risk-neutral conditional distribution of x_{t+1}, given $(\underline{x_t}, \underline{z_t})$, is in the same class as the historical one and obtained by replacing $\varphi(Z_t)$ with $\varphi^*(Z_t)$, and $\mu(X_t, Z_t)$ with $\mu^*(Z_t, X_t)$.

In order to get a generalized linear term structure we impose that the risk-neutral dynamics is a switching regime Gamma Car(p) process. So, using the results in Section 2.5.b, we see that $\varphi^*(Z_t)$ and $\mu^*(Z_t, X_t)$ must be constant, $\nu(Z_t) = \nu^{*'} Z_t$ and $\pi\,(z_t, e_j\,; X_t) = \pi^*\,(z_t, e_j) \exp\left[(\delta(Z_t, X_t))'e_j\,\right]$. Also note that μ^* must be positive as well as the components of ν^* and φ^*. This implies the following constrains on the historical dynamics and on the SDF:

$$\begin{aligned} \mu(X_t, Z_t) &= \mu^*[1 - \Gamma(X_t, Z_t)\mu(X_t, Z_t)] \\ \varphi(Z_t) &= \varphi^*[1 - \Gamma(X_t, Z_t)\mu(X_t, Z_t)]^2 \\ \nu(Z_t) &= \nu^{*'} Z_t \\ \delta_j(X_t, Z_t) &= \log\left[\tfrac{\pi(z_t, e_j\,; X_t)}{\pi^*(z_t, e_j)}\right] \;. \end{aligned}$$

We see that $\varphi(Z_t) = \frac{\varphi^*}{\mu^{*2}}\, \mu(X_t, Z_t)^2$, so $\mu(X_t, Z_t)$ must depend only on Z_t, and therefore the same is true for $\Gamma(X_t, Z_t)$. Finally, we have the constraint:

i)

$$\mu(Z_t) = \mu^*[1 - \Gamma(Z_t)\mu(Z_t)]$$

ii)

$$\varphi(Z_t) \;=\; \varphi^*[1-\Gamma(Z_t)\mu(Z_t)]^2$$

iii)

$$\nu(Z_t) \;=\; \nu^{*'} Z_t$$

iv)

$$\delta_j(X_t, Z_t) \;=\; \log\left[\frac{\pi\,(z_t, e_j\,; X_t)}{\pi^*(z_t, e_j)}\right]\;;$$

In particular, since $\varphi(Z_t) = \frac{\varphi^*}{\mu^{*2}}\,\mu(Z_t)^2$, the random vector must be proportional to a deterministic vector.

Moreover, it is easily seen that the risk premium corresponding to the payoff $\exp(-\theta x_{t+1})$ at $t+1$ is:

$$\begin{aligned}\omega_t(\theta) \;=\; & \{A[-\theta;\varphi(Z_t),\mu(Z_t)] - A[-\theta;\varphi^*,\mu^*]\}'\,X_t \\ & + b[-\theta;\nu^{*'}Z_t,\mu(Z_t)] - b[-\theta;\nu^{*'}Z_t,\mu^*]\,.\end{aligned}$$

Like in the gaussian case, we obtain an affine function in X_t also depending on Z_t. The risk premium associated with the digital asset providing one money unit at $t+1$ if $z_{t+1} = e_j$, is still given by (36).

4.5 The Generalized Linear Term Structure

Let us introduce the notations:

$$\begin{aligned} A^*(u) \;&=\; A(u;\varphi^*,\mu^*) \\ \tilde{C}_h \;&=\; (C_{2,h},\ldots,C_{p,h},0)'\,. \end{aligned} \tag{68}$$

As usual, $B(t,h)$ is the price at t of a zero-coupon bond with residual maturity h.

Proposition 10 : In the univariate SARG(p) Term Structure model the price at date t of the zero-coupon bond with residual maturity h is :

$$B(t,h) = \exp\left(C_h'\,X_t + D_h'\,Z_t\right)\;,\;\text{for } h \geq 1\,, \tag{69}$$

where the vectors C_h and D_h satisfy the following recursive equations :

$$\left\{\begin{aligned} C_h \;&=\; -c + A^*(C_{1,h-1}) + \tilde{C}_{h-1} \\ D_h \;&=\; -d \;-\; \nu^*\log(1 - C_{1,h-1}\mu^*) + \;\tilde{D}_{h-1} + F(D_{1,h-1})\,, \end{aligned}\right. \tag{70}$$

where $\tilde{D}_{h-1}$ and $F(D_{1,h-1})$ have the same meaning as in Proposition 5; the initial conditions are $C_0 = 0$, $D_0 = 0$ (or $C_1 = -c$, $D_1 = -d$) [Proof : see Appendix 8].

Again, we obtain a generalized linear term structure given by:

$$R(t,h) \;=\; -\frac{C_h'}{h}X_t - \frac{D_h'}{h}Z_t\,,\quad h \geq 1\,, \tag{71}$$

and, in the same spirit of propositions 6 and 7 for the univariate SARN(p) model [see section 3.6], it is easy to verify that the processes $R = [\,R(t,h),\, 0 \leq t < T\,]$ and $R_{\mathcal{H}} = [\,R(t,h),\; 0 \leq t < T,\; h \in \mathcal{H}\,]$ are, respectively, a weak Switching ARMA($p,\; p-1$) process and a weak H-variate Switching VARMA($p,\; p-1$) process. In the case of an observable factor, that is $x_t = r_{t+1}$, the previous results remains valid with $C_1 = -e_1$, $D_1 = 0$.

4.6 Positiveness of the Yields

Since $r_{t+1} = R(t,1) = c'X_t + d'Z_t$, and since the components of X_t are positive, the short term process will be positive as soon as the components of c and d are nonnegative. The positiveness of r_{t+1} implies that of $R(t,h)$, at any date t and time to maturity h, because $R(t,h) = -\frac{1}{h}\log E_t^{\mathbb{Q}}\left[\exp(-r_{t+1} - \ldots\ -r_{t+h})\right]$.

This positiveness can also be observed from the recursive equations of Proposition 10. Indeed, using the fact that μ^* and the components of φ^* and ν^* are positive and that $0 < \pi^*_{ij} < 1$, it easily seen that, for any $u < 0$, the components of $A^*(u)$ and $-\nu^* \log(1 - C_{1,h-1}\mu^*)$ are negative and the result follows.

5 Derivative Pricing

5.1 Generalization of the Recursive Pricing Formula

In the previous sections we have derived recursive formulas for the zero-coupon bond price $B(t,h)$ in various contexts which share the feature that the process $(\tilde{x}_t, z_t)$ is Car in the risk-neutral world. In fact the recursive approach can be generalized to other assets.

Let us consider a class of payoffs $g(\tilde{X}_{t+h}, Z_{t+h})$, (t,h) varying, for a given g function and let us assume that the price at t of this payoff is of the form:

$$P_t(g,h) = \exp\left[C_h(g)'\tilde{X}_t + D_h(g)'Z_t\right] . \tag{72}$$

It is clear that:

$$\begin{aligned}
\exp\left[C_h(g)'\tilde{X}_t + D_h(g)'Z_t\right] &= E_t\left[M_{t,t+1}\exp\left(C_{h-1}(g)'\tilde{X}_{t+1} + D_{h-1}(g)'Z_{t+1}\right)\right] \\
&= \exp(-c'\tilde{X}_t - d'Z_t)E_t^{\mathbb{Q}}\left[\exp\left(C_{h-1}(g)'\tilde{X}_{t+1} + D_{h-1}(g)'Z_{t+1}\right)\right] ;
\end{aligned}$$

so the sequences $C_h(g), D_h(g), h \geq 1$, follow recursive equations which does not depend on g and, therefore, are identical to the case $g = 1$, that is to say to the zero-coupon bond pricing formulas given in the previous sections. The only condition for (72) to be true is to hold for $h = 1$ and, of course, this initial condition depends on g.

Formula (72) is valid for $h = 1$ if $g(\tilde{X}_{t+h}, Z_{t+h}) = \exp(\tilde{u}'\tilde{X}_{t+h} + \tilde{v}'Z_{t+h})$ for some vector $\tilde{u}$ and $\tilde{v}$. Indeed, using the notations

$$\begin{aligned}
\tilde{u}'\tilde{X}_{t+1} &= u_1'\tilde{x}_{t+1} + u_{-1}'\tilde{X}_t \\
\tilde{v}'Z_{t+1} &= v_1'z_{t+1} + v_{-1}'Z_t ,
\end{aligned}$$

with $u_{-1}' = (u_2', \ldots, u_p', 0)$, $v_{-1}' = (v_2', \ldots, v_p', 0)$, we get:

$$\begin{aligned}
P_t(\tilde{u}, \tilde{v}; 1) &= \exp(-c'\tilde{X}_t - d'Z_t + u_{-1}'\tilde{X}_t + v_{-1}'Z_t) \\
&\qquad \times E_t^{\mathbb{Q}}\left[\exp\left(u_1'\tilde{x}_{t+1} + v_1'z_{t+1}\right)\right] ,
\end{aligned} \tag{73}$$

which, using the Car representation of $(\tilde{x}_{t+1}, z_{t+1})$ under the probability $\mathbb{Q}$, has obviously the exponential linear form (72) and provides the initial conditions of the recursive equations. The standard recursive equations provide the price $P_t(\tilde{u}, \tilde{v}; h)$ at date t for the payoff $\exp(\tilde{u}'\tilde{X}_{t+h} + \tilde{v}'Z_{t+h})$. So we have the following proposition.

Proposition 11 : The price $P_t(\tilde{u}, \tilde{v}; h)$ at time t of the payoff $g(\tilde{X}_{t+h}, Z_{t+h}) = \exp(\tilde{u}'\tilde{X}_{t+h} + \tilde{v}'Z_{t+h})$ has the exponential form (72) where $C_h(g)$ and $D_h(g)$ follow the *same* recursive equations as in the zero-coupon bond case with initial values $C_1(g)$ and $D_1(g)$ given by the coefficients of $\tilde{X}_t$ and Z_t in equation (73).

When $\tilde{u}$ and $\tilde{v}$ have complex components, $P_t(\tilde{u}, \tilde{v}; h)$ provides the complex Laplace transform $E_t[M_{t,t+h} \exp(\tilde{u}'\tilde{X}_{t+h} + \tilde{v}'Z_{t+h})]$.

5.2 Explicit and Quasi Explicit Pricing Formulas

The explicit formulas for zero-coupon bond prices also immediately provide explicit formulas for some derivatives like swaps. Moreover, the result of section 5.1, where $\tilde{u}$ and $\tilde{v}$ have complex components, can be used to price payoffs of the form:

$$\left[\exp(\tilde{u}_1'\tilde{X}_{t+h} + \tilde{v}_1'Z_{t+h}) - \exp(\tilde{u}_2'\tilde{X}_{t+h} + \tilde{v}_2'Z_{t+h})\right]^+ ,$$

like caps, floors or options on zero-coupon bonds. Let us consider, for instance, the problem to price, at date t, a European call option on the zero-coupon bond $B(t+h, H-h)$, then the pricing relation is :

$$\begin{aligned} p_t(K,h) &= E_t\left[M_{t,t+h}\left(B(t+h,H-h) - K\right)^+\right] \\ &= E_t\left[M_{t,t+h}\left(\exp[-(H-h)R(t+h,H-h)] - K\right)^+\right] , \end{aligned} \tag{74}$$

and, substituting the yield-to-maturity formula (57), for the SVARN(p) Factor-Based Term Structure Model, or formula (71), for the SARG(p) Factor-Based Term Structure Model, we can write:

$$\begin{aligned} p_t(K,h) &= E_t\left[M_{t,t+h}\left(\exp[C'_{H-h}\tilde{X}_{t+h} + D'_{H-h}Z_{t+h}] - K\right)^+\right] \\ &= E_t\left[M_{t,t+h}\left(\exp[C'_{H-h}\tilde{X}_{t+h} + D'_{H-h}Z_{t+h}] - K\right)\mathbb{I}_{[-C'_{H-h}\tilde{X}_{t+h} - D'_{H-h}Z_{t+h} < -\log K]}\right] \\ &= E_t\left[M_{t,t+h}\left(\exp[C'_{H-h}\tilde{X}_{t+h} + D'_{H-h}Z_{t+h}]\right)\mathbb{I}_{[-C'_{H-h}\tilde{X}_{t+h} - D'_{H-h}Z_{t+h} < -\log K]}\right] \\ &\qquad -KE_t\left[M_{t,t+h}\mathbb{I}_{[-C'_{H-h}\tilde{X}_{t+h} - D'_{H-h}Z_{t+h} < -\log K]}\right] \\ &= G_t(C_{H-h}, D_{H-h}, -C_{H-h}, -D_{H-h}, -\log K; h) \\ &\qquad -KG_t(0, 0, -C_{H-h}, -D_{H-h}, -\log K; h) , \end{aligned} \tag{75}$$

where $\mathbb{I}$ denotes the indicator function, and where

$$G_t(\tilde{u}_0, \tilde{v}_0, \tilde{u}_1, \tilde{v}_1, K; h) = E_t\left[M_{t,t+h}\left(\exp[\tilde{u}_0'\tilde{X}_{t+h} + \tilde{v}_0'Z_{t+h}]\right)\mathbb{I}_{[-\tilde{u}_1'\tilde{X}_{t+h} - \tilde{v}_1'Z_{t+h} < K]}\right]$$

denotes the truncated real Laplace transform that we can deduce from the (untruncated) complex Laplace transform. More precisely, we have the following formula [see Duffie, Pan, Singleton (2000) for details]:

$$G_t(\tilde{u}_0, \tilde{v}_0, \tilde{u}_1, \tilde{v}_1, K; h) = \frac{P_t(\tilde{u}_0, \tilde{v}_0, h)}{2}$$
$$- \frac{1}{\pi} \int_0^{+\infty} \left[\frac{Im[P_t(\tilde{u}_0 + i\tilde{u}_1 y, \tilde{v}_0 + i\tilde{v}_1 y; h)] \exp(-iyK)}{y} \right] dy \qquad (76)$$

where $Im(z)$ denotes the imaginary part of the complex number z. So, formula (75) is quasi explicit since it only requires a simple (one-dimensional) integration to derives the values of G_t.

6 Empirical Analysis

6.1 Observable Factor Approach

The purpose of this section is to propose an empirical analysis of the Gaussian term structure models presented in Section 3, using observations on the U. S. term structure of interest rates.

We have seen that SVARN(p) Factor-Based Term Structure Models can be characterized by a latent or observable factor (x_t). In the present empirical analysis we consider an observable factor (yields at different maturities), because it presents several important advantages. First, thanks to data, we are able to detect stylized facts on interest rates which give us the possibility to justify the autoregressive model with switching regimes we propose for the historical dynamics of (x_t): indeed, a large empirical literature on bond yields show that interest rates have an historical multi-lag dynamics characterized by switching of regimes [see, among the others, Hamilton (1989), Garcia and Perron (1996), Christiansen and Lund (2003), Cochrane and Piazzesi (2005)]. Second, observations about the Gaussian-distributed factor lead to a maximum likelihood estimation of historical parameters and, therefore, we are able to rank the models in terms of various information criteria. Finally, the difference between directly observed and estimated factor values determine model residuals that can be used to derive various diagnostic criteria.

Compared with this (multi-lag regime-switching) endogenous discrete-time approach, the classical continuous-time affine term structure approach *à la* Duffie and Kan (1996) and Dai and Singleton (2000) has some different features. First, the factors are in general assumed as not observable and therefore justifications for the (historical) factors dynamics, along with a precise econometric analysis of model residuals, are not possible. Second, in order to reconstruct a time series of the latent factors, for an exact maximum likelihood estimation, prices of some zero-coupon bonds are assumed to be perfectly observed in order to inverse the pricing equations [see Chen and Scott (1987) and Pearson and Sun (1994)]; this inversion technique depends on the selected zero-coupon bonds and on their parameter values, which are not initially available, and therefore the reconstructed time series are model-sensitive [see also Collin-Dufresne, Goldstein and Jones (2004)]. Third, the class of discrete-time affine (Compound Autoregressive) processes is much larger than the discrete-time counterpart of the continuous-time affine class[10] [see Gourieroux, Monfort and Polimenis (2005), and Darolles, Gourieroux and Jasiak (2006)].

In order to study and to precisely understand the role played by lags and switching regimes in the term structure modeling, we start the empirical analysis by estimating (single-regime) VARN(p) Factor-Based Term Structure Models [see Monfort and Pegoraro (2006) for details], in a bivariate (short rate, and spread between the long and short rate) setting. The historical parameters

[10]For instance, the discrete-time Gaussian VAR(1) process has a continuous-time equivalent if and only if there exists a matrix ϑ such that $\varphi = \exp(-\vartheta)$.

are estimated by exact Maximum Likelihood, while the risk-neutral parameters are estimated by nonlinear least squares (NLLS) applied to the yield-to-maturity formula. The second step of the empirical illustration concerns the estimation of bivariate SVARN(p) Factor-Based Term Structure Models, where the latent variable (z_t) is assumed to be a two-states non-homogeneous Markov chain. As in the single-regime specifications, the factor is given by the short rate and spread. The historical parameters are estimated by maximization of the likelihood function calculated using the Kitagawa-Hamilton filter. The risk-neutral parameters are estimated by constrained NLLS applied on the yield-to-maturity formula, after extraction of the latent variable $Z_t = (z_t, \ldots, z_{t-p})'$ using smoothed probabilities calculated by means of the Kim's Smoothing algorithm.

6.2 Description of the Data

The CRSP data set on the U. S. term structure of interest rates [treasury zero-coupon bond (ZCB) yields], that we consider in the following application, covers the period from June 1964 to December 1995 and contains 379 monthly observations for each of the nine maturities : 1, 3, 6 and 9 months and 1, 2, 3, 4 and 5 years[11]. Summary statistics about the above mentioned (annualized) yields are presented in Table 1 : the term structure is, on average, upward sloping and the yields with larger standard deviation, skewness and kurtosis are those with shorter maturities. Moreover, yields are highly autocorrelated with a persistence which is increasing with the time to maturity.

Maturity	1-m	3-m	6-m	9-m	1-yr	2-yr	3-yr	4-yr	5-yr
Mean	0.0645	0.0672	0.0694	0.0709	0.0713	0.0734	0.0750	0.0762	0.0769
Std. Dev.	0.0265	0.0271	0.0270	0.0269	0.0260	0.0252	0.0244	0.0240	0.0237
Skewness	1.2111	1.2118	1.1518	1.1013	1.0307	0.9778	0.9615	0.9263	0.8791
Kurtosis	4.5902	4.5237	4.3147	4.1605	3.9098	3.6612	3.5897	3.5063	3.3531
Minimum	0.0265	0.0277	0.0287	0.0299	0.0311	0.0366	0.0387	0.0397	0.0398
Maximum	0.1640	0.1612	0.1655	0.1644	0.1581	0.1564	0.1556	0.1582	0.1500
ACF(5)	0.8288	0.8531	0.8579	0.8588	0.8604	0.8783	0.8915	0.8986	0.9053
ACF(10)	0.7278	0.7590	0.7691	0.7699	0.7683	0.7885	0.8021	0.8075	0.8212
ACF(15)	0.5887	0.6164	0.6285	0.6313	0.6395	0.6720	0.6908	0.6987	0.7201
ACF(20)	0.4303	0.4631	0.4880	0.4996	0.5156	0.5742	0.6051	0.6193	0.6431

Table 1 : Summary Statistics on U. S. Monthly Yields from June 1964 to December 1995. ACF(k) indicates the empirical autocorrelation between yields $R(t,h)$ and $R(t,h-k)$, with h and k expressed on a monthly basis.

6.3 Estimated Models and Estimation Methods

6.3.1 Estimated Models

In the present and following sections we present the parameter estimates of alternative single-regime and (non-homogeneous) regime-switching bivariate (short rate and spread) term structure models, presented in the Section 3. We start with the estimation of the following :

VARN(p) Factor-Based Term Structure Models :

$$\begin{aligned} x_{t+1} &= \nu + \varphi_1 \, x_t \, + \ldots + \, \varphi_p \, x_{t+1-p} + \sigma \, \varepsilon_{t+1} \\ &= \nu \, + \, \varphi X_t \, + \, \sigma \varepsilon_{t+1} \, , \end{aligned} \tag{77}$$

[11]The same data set is used in the papers of Longstaff and Schwartz (1992) and Bansal and Zhou (2002). We are grateful to Ravi Bansal and Hao Zhou for providing us the data set.

where ε_{t+1} is an 2-dimensional Gaussian white noise with $\mathcal{N}(0, I)$ distribution [I denotes the (2×2) identity matrix]; σ and φ_j, for each $j \in \{1, \ldots, p\}$, are (2×2) matrices [σ can be chosen, for instance, lower triangular], and $\varphi = [\varphi_1, \ldots, \varphi_p]$ is an $(2 \times 2p)$ matrix; ν is an 2-dimensional vector, $X_t = (x_t', \ldots, x_{t+1-p}')'$ is an $(2p)$-dimensional vector, and $p \in \{1, 2\}$.

Then, we move to the regime-switching setting and we estimate the following :

SVARN$^{nh}_{mv}(p)$ Factor-Based Term Structure Models :

$$\begin{aligned} x_{t+1} &= \nu(z_t) + \varphi_1 x_t + \ldots + \varphi_p x_{t+1-p} + \eta_{t+1} \\ &= \nu(z_t) + \varphi X_t + \eta_{t+1} , \end{aligned} \tag{78}$$

where (η_{t+1}) is a bivariate Gaussian white noise with $\mathcal{N}(0, \Sigma(z_t))$ distribution, $p \in \{1, 2\}$, and (z_t) is a 2-states non-homogeneous Markov chain. We estimate the (upper triangular) Cholesky decomposition $\mathcal{A}(z_t)$ of the variance-covariance matrix $\Sigma(z_t)$, where $\Sigma(z_t) = \mathcal{A}(z_t)'\mathcal{A}(z_t)$. The notation (nh) and (mv) in SVARN$^{nh}_{mv}(p)$ indicate that switching regimes are (under the historical probability) described by a non-homogeneous Markov chain applying to both the conditional mean and volatility of (x_t). The transition probabilities have the following logistic form:

$$\begin{aligned} P(z_{t+1} = e_j | z_t = e_j, x_t) &= \pi(e_j, e_j; x_t) \\ &= \frac{e^{a_j + b_j' x_t}}{1 + e^{a_j + b_j' x_t}} , \; j \in \{1, 2\} . \end{aligned} \tag{79}$$

We will see, in the following sections, that the regime $z_t = e_1$ will be identified with a low (L) volatility regime of the factor (x_t), while $z_t = e_2$ will be identified with an high (H) volatility regime.

6.3.2 Estimation Methods

The methodology we follow to estimate the parameters of both single-regime and regime-switching models is based on a consistent two-step procedure.

As far as the VARN(p) Factor-Based Term Structure Models estimation is concerned, in the first step, thanks to observations on the 2-dimensional endogenous factor (x_t), we estimate the vector of parameters $\theta_{\mathbb{P}} = [\,\nu', vec(\varphi)', \, vech(\sigma\sigma')']'$, characterizing the historical dynamics (x_t), by exact Maximum Likelihood (ML). With regard to the estimation of bivariate SVARN$^{nh}_{mv}(p)$ Factor-Based Term Structure Models, in the first step, using the observations on the endogenous factor (x_t), the vector of historical parameters $\tilde{\theta}_{\mathbb{P}} = [\,\nu(e_1)', \nu(e_2)', vec(\varphi)', vech[\mathcal{A}(e_1)], \, vech[\mathcal{A}(e_2)], a_1, b_1', a_2, b_2']'$ is estimated from the maximization of the likelihood function calculated by means of the Kitagawa-Hamilton filter [see Hamilton (1994)].

In the second step, using observations on yields with maturities different from those used in the first step and for a given estimates, respectively, of $vech(\sigma\sigma')$ in the VARN(p) case, and $vech[\mathcal{A}(z_t)]$ in the regime-switching case, we estimate the vector of risk-neutral parameters $\theta_{\mathbb{Q}}$ by minimizing the sum of squared fitting errors between the observed and theoretical yields. The latent variable $Z_t = (z_t, \ldots, z_{t-p})'$, in the yield-to-maturity formula of the SVARN$^{nh}_{mv}(p)$ model, is extracted using smoothed transition probabilities calculated with the Kim's smoothing algorithm [see Kim (1994) and Appendix 9][12]. Thus, we estimate $\theta_{\mathbb{Q}}$ by constrained nonlinear lest squares (NLLS)

[12]It is important to highlight the fact that, contrary to what is indicated in Hamilton (1994), the Kim's smoothing algorithm is correct even if the historical dynamics of the observable factor x_{t+1} depends from $\underline{x_t} = (x_t, x_{t-1}, \ldots)$

and constrains are imposed to satisfy restrictions (58)-(ii) implied by the absence of arbitrage opportunity on the long rate. The data used for the estimation of historical and risk-neutral parameters are monthly observations of annualized yields divided by 12 (yields expressed on a monthly basis), and we consider as short rate the one-month yield to maturity in order to satisfy the associated arbitrage restriction. Thus, the yield-to-maturity formula is applied with a one month unit for the time to maturity h $[r_{t+1} = R(t,1)]$. In our empirical illustration the factor is given by:

$$x_t = [R(t,1), R(t,60) - R(t,1)]' ,$$

where $[R(t,60) - R(t,1)]$ is the spread at date t between the sixty-months (five-years) and one-month yield to maturity [see Ang and Bekaert (2002), and Ang, Piazzesi and Wei (2005) for similar specifications].

Given the complete set of nine maturities of our data base, and given a number $n = 2$ of yields used to estimate the vector of historical parameters $\theta_{\mathbb{P}}$, we denote by H_2^* the set and the number of remaining maturities used to estimate the vector of risk-neutral parameters $\theta_{\mathbb{Q}}$. At the end of the first estimation step, models are ranked in terms of the Akaike Information Criterion (AIC), while, after the second estimation step models are ranked in terms of root mean square errors (RMSE) and the absolute errors (Section 6.6).

The constrained NLLS estimator is given by :

$$\begin{cases} \hat{\theta}_{\mathbb{Q}} = Arg\min_{\theta_{\mathbb{Q}}} S^2(\theta_{\mathbb{Q}}) \\[2ex] S^2(\theta_{\mathbb{Q}}) = \displaystyle\sum_{t=p}^{T} \sum_{h \in H_2^*} [\tilde{R}(t,h) - R(t,h)]^2, \\[2ex] \text{s. t.} \ \displaystyle\sum_{t=p}^{T} [\tilde{R}(t,60) - R(t,60)]^2 = 0 \,, \end{cases} \tag{80}$$

where $R(t,h)$ is the theoretical yield determined by formula (57) [see Monfort and Pegoraro (2006) for details about the single-regime yield-to-maturity formula]. In the yield-to-maturity formula, the historical parameters $vech(\sigma\sigma')'$, for the VARN(p) model, and $vech[\mathcal{A}(z_t)]$ for the SVARN$_{mv}^{nh}(p)$ model, have been replaced by their ML estimates. The constraint in the minimization program (80) guarantees the absence of arbitrage opportunity on the five-year yield to maturity.

6.4 Estimation Results for VARN(p) Factor-Based Term Structure Models

6.4.1 Historical Parameter Estimates

We present the maximum value of the mean Log-Likelihood function and the values of the estimated vector of parameters $\theta_{\mathbb{P}} = [\nu', \, vec(\varphi)', vech(\sigma\sigma')']'$ of the bivariate VARN(p) Factor-Based Term Structure models (for an AR order $p = 1$ and $p = 2$) in Tables 2 and 3 [the t-values are given in parenthesis][13].

and Z_t, and the transition probability of the (latent) Markov chain z_{t+1} depends on $(z_t, \underline{x_t})$. Hamilton (1994) describes this algorithm as correct only in the case of a homogeneous Markov chain and an historical dynamics of x_{t+1} depending from $\underline{x_t}$ and z_t [see the Appendix 9 for the proof, and Billio and Monfort (1998) for an equivalent result in a Kalman filtering setting].

[13] We have also estimated the historical parameters of the above mentioned bivariate Gaussian VAR(p) models, for p larger than 2, but the AIC criterion has indicated the first two AR orders as the preferred ones.

	VARN(1)	VARN(2)
$mlogL$	12.6429	12.6864
AIC	25.2381	25.3038
ν_1	0.65 [0.5856]	1.32 [1.2262]
ν_2	0.80 [0.8157]	0.26 [0.2701]
σ_1^2	0.0059^{**} [5.94750]	0.0036^{**} [6.02614]
σ_{21}	-0.0042^{**} [-6.0995]	-0.0026^{**} [-6.2100]
σ_2^2	0.0030^{**} [7.6713]	0.0028^{**} [8.0731]

Table 2 : VARN(p) Factor-Based Term Structure Models. Maximum value of the mean Log-Likelihood function, AIC and parameter estimates of (ν_1, ν_2) and $(\sigma_1^2, \sigma_{21}, \sigma_2^2)$. The short rate and long rate observations are expressed on a monthly basis. Parameter estimates are multiplied by 10^4. We denote with $mlogL$ the mean log-Likelihood of the VAR(p) model : $mlogL = logL(\theta_{\mathbb{P}}|x_{p+1}, \ldots, x_T)/(T-p)$. (**) denotes a parameter significant at 0.05; (*) denotes a parameter significant at 0.1. The Akaike Information Criterion (AIC) is given by $2mlogL - (2k/(T-p))$, with k denoting the dimension of $\theta_{\mathbb{P}}$.

	VARN(1)		VARN(2)	
φ_1	0.9742^{**} [59.8835]	0.0719^{**} [2.2174]	1.3318^{**} [15.0111]	0.6207^{**} [7.0095]
	0.0091 [0.6388]	0.8769^{**} [30.7835]	-0.2744^{**} [-3.4988]	0.4353^{**} [5.5601]
φ_2			-0.3648^{**} [-3.6117]	-0.5762^{**} [-5.8201]
			0.2893^{**} [3.2397]	0.4642^{**} [5.3020]

Table 3 : VARN(p) Factor-Based Term Structure Models. Parameter estimates of $\varphi = (\varphi_1, \varphi_2)$. (**) denotes a parameter significant at 0.05; (*) denotes a parameter significant at 0.1.

If we consider the parameter estimates of Tables 2 and 3, we observe that the joint historical dynamics of short rate and spread is not conditionally Markovian of order one, given that, in the VARN(2) specification, the parameters in the second autoregressive matrix φ_2 are significantly different from zero. Table 7 shows also that the constant term $(\nu_1, \nu_2)'$ is not significative for both AR orders.

	VARN(1)			VARN(2)		
k	$LBshort_k$	$LBspread_k$	Q_k	$LBshort_k$	$LBspread_k$	Q_k
5	10.2909 **	22.8735	58.9900	5.8051 **	12.4876 *	29.9870
10	13.5065 **	34.1952	82.6678	9.0209 **	19.8705 *	51.2640 *
15	29.2167 *	63.4161	134.2488	23.0114 **	35.8770	91.3809
20	36.2600 *	69.7833	165.8618	30.0413 **	45.5552	127.9043

Table 4 : VARN(p) Factor-Based Term Structure Models. $LBshort_k$ and $LBspread_k$ denotes, respectively, the value of the Ljung-Box test statistic, up to lag k, for short rate residuals and for spread residuals. Q_k denotes the value of the (adjusted) Portmanteau test statistic, up to lag k, for the VARN(1) and VARN(2) model residuals. (**) denotes the null hypothesis accepted at 0.05; (*) denotes the null hypothesis accepted at 0.01.

The autocorrelation analysis of model residuals, presented in Table 4, shows that the ability of the VARN(p) model to explain the serial dependence in the (univariate and joint) short rate and spread historical dynamics, improves when we move from the VARN(1) to the VARN(2) specification, even if both models are not able to pass the portmanteau test on the bivariate residual vectors. Indeed, for $p = 1$, the Ljung-Box test on the short rate residuals is passed only at 0.01 for large lags, while, the same test on spread residuals strongly fails for each lag. When we consider $p = 2$, the Ljung-Box test, when it is applied on the short rate residuals, is always passed at 0.05, while, when it is applied on the spread residuals, is passed for five and ten lags, but still fails for larger lags. This means also that the failure of the portmanteau test (stronger when $p = 1$) comes from the difficulty of the specified models to explain the serial dependence in the spread historical dynamics. We will see in Section 6.5.1 that the introduction of (2-states non-homogeneous) switching regimes leads to overcome these limits.

6.4.2 Risk-Neutral Parameter Estimates

We present the RMSE $= [\frac{S^2(\hat{\theta}_{\mathbb{Q}})}{(T-p+1)H_2^*}]^{0.5}$ and the values of the estimated vector of risk-neutral parameters $\theta_{\mathbb{Q}} = [(\nu^*)', vec(\varphi^*)']'$, for the bivariate VARN(1) and VARN(2) Factor-Based Term Structure models, in Tables 5 and 6 [the t-values are given in parenthesis].

	VARN(1)	VARN(2)
RMSE	2.96	2.75
ν_1^*	-0.65^{**} $[-7.3259]$	-0.56^{**} $[-4.7376]$
ν_2^*	0.80^{**} $[6.3172]$	0.64^{**} $[4.2546]$

Table 5 : VARN(p) Factor-Based Term Structure Models. RMSE and parameter estimates of (ν_1^*, ν_2^*). Yields to maturity observations are expressed on a monthly basis. Parameter estimates and the RMSE $= [\frac{S^2(\hat{\theta}_{\mathbb{Q}})}{(T-p+1)H_2^*}]^{0.5}$ are multiplied by 10^4. (**) denotes a parameter significant at 0.05; (*) denotes a parameter significant at 0.1.

	VARN(1)		VARN(2)	
φ_1^*	1.0142^{**} [793]	0.1160^{**} [35.1724]	1.3097^{**} [26.3124]	0.7237^{**} [11.0842]
	-0.0168^{**} [-9.2901]	0.9012^{**} [197.9250]	-0.2596^{**} [-3.7377]	0.3874^{**} [4.2302]
φ_2^*			-0.2932^{**} [-5.8972]	-0.6010^{**} [-9.4051]
			0.2410^{**} [3.4751]	0.5081^{**} [5.6745]

Table 6 : VARN(p) Factor-Based Term Structure Models. Parameter estimates of $\varphi^* = (\varphi_1^*, \varphi_2^*)$. (**) denotes a parameter significant at 0.05; (*) denotes a parameter significant at 0.1.

We find that, also in this bivariate risk-neutral (pricing) framework, the lagged values of the short rate and spread play an important role in the model specification. In particular, one may observe the significativity of all risk-neutral AR coefficients in the VARN(2) specification.

The goodness-of-fit of the VARN(2) Factor-Based Term Structure Model outperform the results of the VARN(1) specification.

6.5 Estimation Results for SVARN$^{nh}_{mv}(p)$ Factor-Based Term Structure Models

6.5.1 Historical Parameter Estimates

The maximum value of the mean Log-Likelihood function and the values of the estimated vector of parameters $[\nu(e_1)', \nu(e_2)', vec(\varphi)', vech(\mathcal{A}(e_1))', vech(\mathcal{A}(e_2))']$ of the bivariate SVARN$^{nh}_{mv}(p)$ Factor-Based Term Structure models, for an AR order $p = 1$ and $p = 2$, are presented in Tables 7 and 8 [the t-values are given in parenthesis]. The parameters $[\,\nu(e_1)', \nu(e_2)']$ are not significantly different from zero, while the parameters $vech(\mathcal{A}(z_t))'$ are significantly different from zero in both regimes. In Table 9 we present the estimation of the parameters (a_1, b_1, a_2, b_2) in the Markov chain transition matrix. One may observe that the non-homogeneity is introduced by the short rate only : we first estimated the general model specification, where the non-homogeneity was introduced by both the short rate and the spread, but the parameter associated to the latter was not significantly different from zero and the maximum value of the mean log-likelihood function did not increase. We observe that the parameters (a_1, b_1, a_2, b_2) are always significantly different from zero.

In Table 9 we present also the sample mean of the persistence probabilities $\pi_t(e_j, e_j; r_{t+1})$, with $j \in \{1, 2\}$, and the associated (mean) invariant probabilities $\bar{P}(z_t = e_1)$ and $\bar{P}(z_t = e_2)$. We report also, for $j \in \{1, 2\}$, the sample mean of the filtered $[\bar{P}(z_t = e_j |\, \underline{x_t})]$ and smoothed transition probabilities $[\bar{P}(z_t = e_j |\, \underline{x_T})]$. We observe from the invariant probability values that the factor spends in the sample much more time in regime L than in regime H; these results find confirmation in the averaged filtered and smoothed probabilities.

As far as the autocorrelation analysis of model residuals, presented in Table 10, is concerned, we observe that, thanks to the introduction of regime switching, SVARN$^{nh}_{mv}(p)$ models explain the serial dependence in the (univariate and joint) short rate and spread historical dynamics better than the VARN(p) specifications presented in Table 4 [see Section 6.4.1]. Moreover, when we move from the SVARN$^{nh}_{mv}(1)$ to the SVARN$^{nh}_{mv}(2)$ specification, we observe that the latter is able to well explain serial dependence, for each lag, in both the short rate and the spread, marginally and jointly. The Ljung-Box test, applied to short rate residuals, is always passed at 0.05, and when it is applied to spread residuals it is passed at 0.05 or 0.01. The Portmanteau test is also passed at 0.05 or 0.01. So, if we want to propose a model able to explain serial dependence in interest rates, it seems to be important to introduce, at the same time, non linearities (switching regimes) and a more general specification of the conditional mean (AR order $p > 1$) in the historical dynamics of the factor (x_t).

		$SVARN^{nh}_{mv}(1)$	$SVARN^{nh}_{mv}(2)$			$SVARN^{nh}_{mv}(1)$	$SVARN^{nh}_{mv}(2)$
$mlogL$		13.0443	13.0712				
AIC		25.9937	26.0257				
	z_t				z_t		
ν_1	e_1	0.3508 [0.4763]	0.3326 [0.4793]	$\mathcal{A}[1,1]$	e_1	2.7753** [12.4556]	2.7767** [14.6351]
	e_2	0.5951 [0.3408]	-0.3524 [-0.2141]		e_2	11.5672** [10.0798]	11.4294** [11.4109]
ν_2	e_1	0.8048 [1.0377]	0.8241 [1.1077]	$\mathcal{A}[1,2]$	e_1	-1.9889** [-8.7303]	-1.9264** [-9.4543]
	e_2	0.5531 [0.3254]	1.5402 [0.9772]		e_2	-8.4927** [-8.6055]	-8.3688** [-9.3296]
				$\mathcal{A}[2,2]$	e_1	2.3470** [15.7459]	2.4111** [19.3093]
					e_2	4.5422** [12.0497]	4.5321** [12.1126]

Table 7 : $SVARN^{nh}_{mv}(p)$ Factor-Based Term Structure Models. Maximum value of the mean Log-Likelihood function, AIC and parameter estimates of $[\nu(e_1)', \nu(e_2)']$ and of $[\mathcal{A}(e_1), \mathcal{A}(e_2)]$. The short rate and long rate observations are expressed on a monthly basis. Parameter estimates are multiplied by 10^4. We denote with $mlogL$ the mean log-Likelihood of the $SVARN^{nh}_{mv}(p)$ model: $mlogL = logL(\theta_{\mathbb{P}}|x_{p+1}, \ldots, x_T)/(T-p)$. $(^{**})$ denotes a parameter significant at 0.05; $(^{*})$ denotes a parameter significant at 0.1. The Akaike Information Criterion (AIC) is given by $2mlogL - (2k/(T-p))$, with k denoting the dimension of $\theta_{\mathbb{P}}$.

	$SVARN^{nh}_{mv}(1)$		$SVARN^{nh}_{mv}(2)$	
φ_1	0.9853** [70.6209]	0.0376* [1.6946]	1.1479** [16.3258]	0.3256** [4.7304]
	0.0021 [0.1414]	0.9188** [41.0486]	-0.1253* [-1.7749]	0.6710** [9.4722]
φ_2			-0.1543** [-2.1704]	-0.3074** [-4.5238]
			0.1187* [1.6795]	0.2686** [3.8308]

Table 8 : $SVARN^{nh}_{mv}(p)$ Factor-Based Term Structure Models. Parameter estimates of $\varphi = (\varphi_1, \varphi_2)$. $(^{**})$ denotes a parameter significant at 0.05; $(^{*})$ denotes a parameter significant at 0.1.

	$\text{SVARN}^{nh}_{mv}(1)$	$\text{SVARN}^{nh}_{mv}(2)$		$\text{SVARN}^{nh}_{mv}(1)$	$\text{SVARN}^{nh}_{mv}(2)$
a_1	5.5471** [4.1377]	4.5367** [3.5817]	$\bar{\pi}(e_1, e_1)$	0.8056	0.8750
b_1	-638.3400** [-2.4280]	-423.1984* [-1.8248]	$\bar{\pi}(e_2, e_2)$	0.3503	0.3917
a_2	-5.3091** [-2.3262]	-7.3618** [-1.9837]	$P(z_t = e_1)$	0.7697	0.8295
b_2	849.9825** [2.3627]	1283.90** [2.0091]	$P(z_t = e_2)$	0.2303	0.1705
$\bar{P}(z_t = e_1 \mid \underline{x_t})$	0.7505	0.7621	$\bar{P}(z_t = e_1 \mid \underline{x_T})$	0.7517	0.7643
$\bar{P}(z_t = e_2 \mid \underline{x_t})$	0.2495	0.2379	$\bar{P}(z_t = e_2 \mid \underline{x_T})$	0.2483	0.2357

Table 9 : $\text{SVARN}^{nh}_{mv}(p)$ Factor-Based Term Structure Models. Parameter estimates of (a_1, b_1, a_2, b_2). $(^{**})$ denotes a parameter significant at 0.05; $(^{*})$ denotes a parameter significant at 0.1. $\bar{\pi}(e_j, e_j)$ denotes the sample mean of the persistence (historical) probability of regime j, with $j \in \{1, 2\}$; $P(z_t = e_1)$ and $P(z_t = e_2)$ denote the associated invariant probabilities. $\bar{P}(z_t = e_j \mid \underline{x_t})$ and $\bar{P}(z_t = e_j \mid \underline{x_T})$ denote, respectively, the sample mean of the filtered and smoothed probability of regime j, with $j \in \{1, 2\}$.

	$\text{SVARN}^{nh}_{mv}(1)$			$\text{SVARN}^{nh}_{mv}(2)$		
k	$LBshort_k$	$LBspread_k$	Q_k	$LBshort_k$	$LBspread_k$	Q_k
5	3.5880**	21.1701	48.5946	1.8587**	8.6148**	24.7248*
10	5.5534**	34.5825	67.9541	2.8656**	18.4149*	41.5856**
15	16.2028**	42.7335	102.2871	10.8556**	27.3656*	74.0759*
20	21.3184**	43.4506	106.8500	13.5285**	28.4819**	83.7164**

Table 10 : $\text{SVARN}^{nh}_{mv}(p)$ Factor-Based Term Structure Models. $LBshort_k$ and $LBspread_k$ denotes, respectively, the value of the Ljung-Box test statistic, up to lag k, for short rate residuals and for spread residuals. Q_k denotes the value of the (adjusted) Portmanteau test statistic, up to lag k, for the $\text{SVARN}^{nh}_{mv}(1)$ and $\text{SVARN}^{nh}_{mv}(2)$ model residuals. $(^{**})$ denotes the null hypothesis accepted at 0.05; $(^{*})$ denotes the null hypothesis accepted at 0.01.

6.5.2 Risk-Neutral Parameter Estimates

In Tables 11 and 12 we present the RMSE $= [\frac{S^2(\hat{\theta}_{\mathbb{Q}})}{(T-p+1)H_2^*}]^{0.5}$ and the values of the estimated vector of risk-neutral parameters $\theta_{\mathbb{Q}} = [\nu^*(e_1)', \nu^*(e_2)', \ vec(\varphi^*)']'$, for the bivariate $\text{SVARN}^{nh}_{mv}(1)$ and $\text{SVARN}^{nh}_{mv}(2)$ Factor-Based Term Structure models [the t-values are given in parenthesis]. We find that, also in this bivariate regime-switching risk-neutral framework, we have the significativity of all risk-neutral AR coefficients in the $\text{SVARN}^{nh}_{mv}(2)$ specification.

In Table 13 we present, for each model, the parameter estimates of the risk-neutral persistence probabilities $[\pi^*(e_1, e_1), \pi^*(e_2, e_2)]$ and the associated (risk-neutral) invariant probabilities $P^*(z_t = e_1)$ and $P^*(z_t = e_2)$. We observe that, comparing the invariant historical $[\bar{P}(z_t)]$ and risk-neutral $[P^*(z_t)]$ probabilities, the short rate spends much more time in the high (H) volatility regime and much less time in the low (L) volatility regime under $\mathbb{Q}$ than under $\mathbb{P}$. Indeed, we have that $\bar{\pi}(e_1, e_1) \gg \pi^*(e_1, e_1)$ and $\bar{\pi}(e_2, e_2) \ll \pi^*(e_2, e_2)$. This is economically coherent with the meaning of the risk-neutral probability $\mathbb{Q}$: since investors are risk-averse, under the risk-neutral probability if we want to correctly price bonds, we need to treat the (risky) H state as being more likely to occur than in reality.

In sections 6.6 and 6.7 we will compare the goodness-of-fit and the ability of these models to explain the Expectation Hypothesis Puzzle with those of other competing models in the literature.

		$\text{SVARN}^{nh}_{mv}(1)$	$\text{SVARN}^{nh}_{mv}(2)$
RMSE		2.73	2.70
	z_t		
ν_1^*	e_1	4.1422** [7.1306]	-0.5919 [-1.2276]
	e_2	-0.2182 [-1.4079]	-0.9680 [-1.2032]
ν_2^*	e_1	-4.7662** [-7.0416]	0.7150 [1.2851]
	e_2	0.2514 [1.3382]	1.1767 [1.2909]

Table 11 : $\text{SVARN}^{nh}_{mv}(p)$ Factor-Based Term Structure Models. RMSE and parameter estimates of $[\,\nu^*(e_1)', \nu^*(e_2)']$. Yields to maturity observations are expressed on a monthly basis. Parameter estimates and RMSE $= [\,\frac{S^2(\hat{\theta}_{\mathbb{Q}})}{(T-p+1)H_2^*}]^{0.5}$ are multiplied by 10^4. (**) denotes a parameter significant at 0.05; (*) denotes a parameter significant at 0.1.

	$\text{SVARN}^{nh}_{mv}(1)$		$\text{SVARN}^{nh}_{mv}(2)$	
φ_1^*	1.0050** [719.3363]	0.1128** [35.2959]	1.1614** [28.6096]	0.7415** [19.3198]
	-0.0055** [-2.7358]	0.9016** [197.2227]	-0.1614** [-3.7862]	0.2772** [6.5989]
φ_2^*			-0.1413** [-2.2370]	-0.6125** [-11.0976]
			0.1385* [1.9018]	0.6125** [10.0502]

Table 12 : $\text{SVARN}^{nh}_{mv}(p)$ Factor-Based Term Structure Models. Parameter estimates of $\varphi^* = (\varphi_1^*, \varphi_2^*)$. (**) denotes a parameter significant at 0.05; (*) denotes a parameter significant at 0.1.

	$\text{SVARN}_{mv}^{nh}(1)$	$\text{SVARN}_{mv}^{nh}(2)$
$\pi^*(e_1, e_1)$	0.4119** [3.5749]	0.8201** [2.0138]
$\pi^*(e_2, e_2)$	0.9998** [42.6672]	0.9897** [2.1379]
$P^*(z_t = e_1)$	0.0003	0.0540
$P^*(z_t = e_1)$	0.9997	0.9460

Table 13 : $\text{SVARN}_{mv}^{nh}(p)$ Factor-Based Term Structure Models. Parameter estimates of risk-neutral persistence probabilities $[\pi^*(e_1, e_1), \pi^*(e_2, e_2)]$. (**) denotes a parameter significant at 0.05; (*) denotes a parameter significant at 0.1. $P^*(z_t = e_1)$ and $P^*(z_t = e_2)$ denote the associated risk-neutral invariant probabilities.

6.6 A Pricing Error Comparison with Competing Models

The purpose of this section is to study the ability of our term structure models to fit the observed yield curves. We compare (annualized absolute) yield-to-maturity errors of our VARN(p) and $\text{SVARN}_{mv}^{nh}(p)$ models ($p \in \{1, 2\}$) with those of other competing models in the literature. These models are the two-factor (approximate) discrete-time CIR model with regime-switching (2-Factor RS-CIR) proposed by Bansal and Zhou (2002), and other competing models studied also by Bansal and Zhou (2002), like the continuous-time two-factor square-root (2-Factor CIR), three-factor square-root (3-Factor CIR), and three-factor affine $\mathbb{A}_1(3)$ (3-Factor $\mathbb{A}_1(3)$) term structure models.

For each date t and for each estimated model, we compute, over the maturities used to estimate the risk-neutral parameters, the pricing error in the following way:

$$PE_t = \frac{\sum_{h \in H_2^*} |\tilde{R}(t,h) - R(t,h)|}{H_2^*}, \tag{81}$$

where $\tilde{R}(t,h)$ and $R(t,h)$ are, respectively, the (annualized) observed and model-implied yields, and where H_2^* denotes the set and the number of maturity used to estimate $\theta_{\mathbb{Q}}$. Given the time series PE_t, we calculate (for each model) the associated mean, standard deviation, minimum and maximum value [see Table 14].

Several indications stand out from this pricing error comparison. First, if we compare single-regime term structure models, we observe that the 2-Factor VARN(1) model completely dominates the 2-Factor (heteroscedastic) CIR specification (smaller mean absolute error and standard deviation), and it produces the same average pricing error as the 3-Factor (heteroscedastic) CIR and 3-Factor $\mathbb{A}_1(3)$ models, but with a smaller standard deviation. Second, if we consider the 2-Factor VARN(2) model, the pricing error performance further improves with respect to the above mentioned models, with smaller average pricing error, standard deviation and maximum error. These facts strongly indicate the important role played by the observable nature of the factor and by the introduction of a second lagged factor value, which is possible in our discrete-time setting. Moreover, our discrete-time endogenous approach is not only characterized by a more general specification of the conditional mean of the factor (with respect to the continuous time approach), but, at the same time, it is not faced to admissibility conditions or identification problems, typically induced by a latent factor [see Dai and Singleton (2000), Dai, Singleton and Yang (2005), Gourieroux and Sufana (2006)], which turn out to further reduce the theoretical flexibility (goodness-of-fit) of these models.

Third, if we consider regime-switching term structure models, the $\text{SVARN}^{nh}_{mv}(1)$ and $\text{SVARN}^{nh}_{mv}(2)$ term structure models have average pricing errors as small as the most competing 2-Factor (discretized square-root) RS-CIR model proposed by Bansal and Zhou (2002), the same standard deviations and smaller maximum values. Moreover, VARN(2) and $\text{SVARN}^{nh}_{mv}(2)$ term structure models produce, from this pricing error point of view, almost the same results indicating the role played by the observable lagged factor value as probably being more important than the one played by the introduction of switching regimes (we will see in Section 6.7 that switching regimes play a determinant role in the explanation of the long horizon expectation hypothesis puzzle).

	2-Factor CIR	2-Factor RS-CIR	3-Factor CIR	3-Factor $\mathbb{A}_1(3)$	2-Factor VARN(1)	2-Factor VARN(2)	2-Factor $\text{SVARN}^{nh}_{mv}(1)$	2-Factor $\text{SVARN}^{nh}_{mv}(2)$
Mean	30	23	25	25	25	23	23	23
Std. Dev.	18	16	21	22	17	16	16	16
Min.	3	3	1	2	3	3	3	3
Max.	121	114	133	137	108	91	100	95

Table 14 : Annualized Absolute Pricing Error (Basis Points). In the data set there are nine maturities (1, 3, 6, 9 months; 1, 2, 3, 4, 5 years) for each of 379 dates. The reported results for models 2-Factor CIR, 2-Factor RS-CIR, 3-Factor CIR and 3-Factor $\mathbb{A}_1(3)$ are taken from Table VI in Bansal and Zhou (2002), where the models was estimated by the efficient method of moments. The absolute pricing error of 2-Factor CIR, 2-Factor VARN(1), 2-Factor VARN(2), 2-Factor $\text{SVARN}^{nh}_{mv}(1)$ and 2-Factor $\text{SVARN}^{nh}_{mv}(2)$ are over 7 points; 2-Factor RS, 3-Factor CIR and 3-Factor AF over 6 points.

6.7 The Expectation Hypothesis Puzzle

In this section we consider an additional stylized fact to test the specification of our multi-lags regime-switching term structure models. In particular, we compare the ability of our VARN(p) and $\text{SVARN}^{nh}_{mv}(p)$ models with the ability of the competing models mentioned in Section 6.6, to replicate the observed violations of the Expectation Hypothesis Theory over the short and long horizon [see Campbell and Shiller (1991), Roberds and Whiteman (1999)]. More precisely, when the yield variation $R(t+m, h-m) - R(t,h)$ is regressed on to the normalized yield spread variable $[R(t,h) - R(t,m)]\frac{m}{h-m}$, where m is a positive integer and $h > m$, that is :

$$R(\,t+m, h-m) - R(t,h) = \phi_o(h,m) + \phi_1(h,m)[R(\,t,h) - R(\,t,m)]\frac{m}{h-m} + u_{t+m}(h,m)\,, \tag{82}$$

with, for each pair (h,m), $u_{t+m}(h,m)$ a zero mean white noise, we find empirically for any fixed m and increasing h that the regression coefficient $\phi_1(h,m)$ becomes more and more negative, while for the Expectation Theory this coefficient should always be equal to one. A term structure model is indicated as able to justify the violation of the Expectation Hypothesis if it replicates, for each pair (h,m), (negative) regression coefficients close the empirical ones [see Dai and Singleton (2003) for further details].

In Tables 15 and 16 we report the estimates of the regression coefficient $\phi_1(h,m)$ [the associated standard errors is between parentheses] for several h and m, in the data set and for the various models considered in the comparison. In Table 15 we study the Expectation Hypothesis Puzzle over the short horizon with $m \in \{3,6,9\}$ months and $h \in \{6,9,12\}$ months; Table 16 shows the long horizon results with $m \in \{1,2,3\}$ years and $h \in \{4,5\}$ years. As in Bansal and Zhou (2002), the reported slope coefficients for the various models are obtained with 10000 simulations of yield trajectories with length 379. For the short horizon, we find that, for each h and m, our VARN(2) and $\text{SVARN}_{mv}^{nh}(2)$ Factor-Based Term Structure Models replicate the violation of the Expectation Theory as well as or better that the most competing 2-Factor RS-CIR model in Bansal and Zhou (2002). For the long horizon the $\text{SVARN}_{mv}^{nh}(2)$ model outperforms in general the most competing 2-Factor CIR and 2-Factor RS-CIR models, and also the VARN(2) model specification.

We observe, therefore, that the $\text{SVARN}_{mv}^{nh}(2)$ Factor-Based Term Structure Model is one of the most competing models in terms of goodness-of-fit [Table 14] and the most performing one in explaining the violation of the Expectation Hypothesis Theory over the short and long horizon [Tables 15 and 16].

Short Horizon	$m = 3$ months	$m = 6$ months	$m = 9$ months
$h = 6$ months	-0.6942 (0.2533)		
2-Factor CIR	0.8253 (1.2208)		
2-Factor RS-CIR	-0.5360 (0.5243)		
3-Factor CIR	1.1676 (1.1952)		
3-Factor $\mathbb{A}_1(3)$	1.4916 (1.3874)		
2-Factor VARN(1)	0.5828 (0.3485)		
2-Factor VARN(2)	-0.3800 (0.3837)		
2-Factor $\text{SVARN}_{mv}^{nh}(1)$	-0.1339 (0.3068)		
2-Factor $\text{SVARN}_{mv}^{nh}(2)$	-0.5323 (0.3195)		
$h = 9$ months	-0.8863 (0.3238)	-0.4023 (0.2429)	
2-Factor CIR	0.7367 (1.2166)	0.6982 (1.2043)	
2-Factor RS-CIR	-0.5757 (0.5632)	-0.3021 (0.5819)	
3-Factor CIR	1.0632 (1.1903)	1.0287 (1.1809)	
3-Factor $\mathbb{A}_1(3)$	1.4348 (1.3692)	1.4316 (1.3380)	
2-Factor VARN(1)	0.4133 (0.3469)	0.4722 (0.2693)	
2-Factor VARN(2)	-0.5480 (0.3960)	-0.3890 (0.3182)	
2-Factor $\text{SVARN}_{mv}^{nh}(1)$	-0.1421 (0.3033)	0.1526 (0.2334)	
2-Factor $\text{SVARN}_{mv}^{nh}(2)$	-0.5957 (0.3286)	-0.4928 (0.2555)	
$h = 12$ months	-1.3226 (0.3530)	-0.7867 (0.2381)	-0.4371 (0.1312)
2-Factor CIR	0.6455 (1.2137)	0.6068 (1.2008)	0.5713 (1.1889)
2-Factor RS-CIR	-0.6000 (0.6030)	-0.3145 (0.6186)	-0.0679 (0.6630)
3-Factor CIR	0.9524 (1.1874)	0.9185 (1.1783)	0.8833 (1.1659)
3-Factor $\mathbb{A}_1(3)$	1.3763 (1.3514)	1.3741 (1.3210)	1.3692 (1.2959)
2-Factor VARN(1)	0.2454 (0.3486)	0.3187 (0.2710)	0.3796 (0.2430)
2-Factor VARN(2)	-0.6935 (0.4069)	-0.5272 (0.3248)	-0.3675 (0.2930)
2-Factor $\text{SVARN}_{mv}^{nh}(1)$	-0.1767 (0.3001)	0.0411 (0.2325)	0.1743 (0.2103)
2-Factor $\text{SVARN}_{mv}^{nh}(2)$	-0.6591 (0.3363)	-0.5731 (0.2574)	-0.4906 (0.2297)

Table 15 : Short Horizon Expectation Hypothesis Tests : the first row in each block reports the empirical slope coefficient and its standard error. The second to fifth rows are the same regression slope coefficients and standard errors taken from Table VIII in Bansal and Zhou (2002). The sixth to ninth rows are the slope coefficients and standard errors based on 10000 simulation from each fitted VARN(p) and $\text{SVARN}_{mv}^{nh}(p)$ Factor-Based Term Structure Model with each simulation having a sample size of 379.

Long Horizon	$m = 1$ year	$m = 2$ years	$m = 3$ years
$h = 4$ years	-1.8078 (0.2981)	-0.8380 (0.2889)	-0.0421 (0.2682)
2-Factor CIR	-0.6928 (1.1713)	-0.7324 (1.0755)	-0.7406 (0.9935)
2-Factor RS-CIR	-0.1848 (1.0852)	-0.0309 (1.1429)	-0.0631 (1.1527)
3-Factor CIR	-0.8191 (1.1752)	-0.7238 (1.0336)	-0.6292 (0.9036)
3-Factor $\mathbb{A}_1(3)$	0.5652 (1.1344)	0.5962 (1.0712)	0.6078 (1.0239)
2-Factor VARN(1)	-0.8569 (0.3536)	-0.0085 (0.3414)	0.8626 (0.3514)
2-Factor VARN(2)	-1.4088 (0.4084)	-0.2338 (0.3864)	0.9368 (0.3843)
2-Factor SVARN$^{nh}_{mv}$(1)	-0.8036 (0.2952)	-0.3602 (0.3017)	0.0865 (0.3393)
2-Factor SVARN$^{nh}_{mv}$(2)	-1.3067 (0.2857)	-0.7469 (0.2657)	-0.0306 (0.2756)
$h = 5$ years	-1.7470 (0.3291)	-0.9720 (0.3199)	-0.2378 (0.3283)
2-Factor CIR	-1.1314 (1.1825)	-1.1130 (1.0514)	-1.0652 (0.9434)
2-Factor RS-CIR	-0.4527 (1.1837)	-0.3607 (1.2056)	-0.4120 (1.1877)
3-Factor CIR	-1.4772 (1.1850)	-1.2330 (0.9957)	-1.0217 (0.8326)
3-Factor $\mathbb{A}_1(3)$	0.2358 (1.1134)	0.2958 (1.0508)	0.3304 (1.0017)
2-Factor VARN(1)	-1.1444 (0.4102)	-0.0033 (0.3953)	1.1279 (0.3970)
2-Factor VARN(2)	-1.6686 (0.4635)	-0.2112 (0.4373)	1.2060 (0.4267)
2-Factor SVARN$^{nh}_{mv}$(1)	-1.1375 (0.3463)	-0.6103 (0.3525)	-0.0748 (0.3873)
2-Factor SVARN$^{nh}_{mv}$(2)	-1.5368 (0.3208)	-0.8823 (0.2963)	-0.0472 (0.3015)

Table 16 : Long Horizon Expectation Hypothesis Tests : the first row in each block reports the empirical slope coefficient and its standard error. The second to fifth rows are the same regression slope coefficients and standard errors taken from Table VIII in Bansal and Zhou (2002). The sixth to ninth rows are the slope coefficients and standard errors based on 10000 simulation from each fitted VARN(p) and SVARN$^{nh}_{mv}(p)$ Factor-Based Term Structure Model with each simulation having a sample size of 379.

Thus, the introduction of lagged values and non-homogeneous switching regimes in the historical dynamics of (x_t), pricing the risk associated to the factor and its lagged values [$\Gamma(z_t, \tilde{X}_t) \neq 0$, with $\tilde{X}_t = (x_{1,t}, x_{1,t-1}, x_{2,t}, x_{2,t-1})'$] and pricing regime-shift risk [$\delta_j(z_t, r_{t+1}) \neq 0$ for each $j \in \{1, 2\}$] are important elements to well replicate the observed yield curves and the Campbell-Shiller regression coefficients. Moreover, we will see in the following section how pricing the risk for the second lagged factor value is determinant to explain the long horizon Expectation Hypothesis Puzzle.

6.8 Pricing Lagged Factor Value and Regime-Shift Risks

We have seen in Section 3.2 that our regime-switching essentially affine factor risk correction coefficient $\Gamma(z_t, \tilde{X}_t)$ differs from that of Dai, Singleton and Yang (2005) because of the multi-lag specification. Moreover, unlike Naik and Lee (1997), Bansal and Zhou (2002) and Ang and Bekaert (2005), we price regime-shift risk. The purpose of this section is to understand what are the effects on our SVARN$^{nh}_{mv}$(2) model's performances, presented in sections 6.5 to 6.7, if we impose restrictions on the above mentioned risk correction features. In other words, we want to study at which level pricing the risk associated to the second lagged factor value or pricing regime-shift risk is important in explaining interest rates dynamics.

We compare the absolute pricing errors and the estimates of the Campbell-Shiller regression coefficients of model SVARN$^{nh}_{mv}$(2), with those obtained if we impose to the previous specification the restriction $\varphi_2 = \varphi_2^*$ [the risk associated to the second lagged factor value is not priced] or if we consider $\pi_t(e_j, e_j; r_{t+1}) = \pi(e_j, e_j) = \pi^*(e_j, e_j)$, for each $j \in \{1, 2\}$ and for each t [regime-shift risk is not priced]. We denote the first constrained SVARN(2) model as the $\text{SVARN}^{nh}_{cm}(2)$ model [(cm) refers to the constraint on the risk-neutral mean of (x_t) : $\varphi_2 = \varphi_2^*$], while the second constrained specification is denoted as $\text{SVARN}^{ch}_{mv}(2)$ [(ch) indicated that the Markov chain (z_t) is constrained to have the same homogeneous transition matrix under both the historical and risk-neutral probability]. From Table 17 we observe, first, that the $\text{SVARN}^{nh}_{mv}(2)$ model-implied regression coefficients are closer to the empirical ones than the coefficients replicated by the $\text{SVARN}^{nh}_{cm}(2)$ and $\text{SVARN}^{ch}_{mv}(2)$ models. Second, pricing the regime-shift risk (respectively, the second lagged factor value risk) is more important over the short (respectively, over the long) horizon. Third, pricing the risk associated to the second lagged factor value is crucial to replicate the observed violation of the Expectation Theory over the long horizon.

Short Horizon	$m = 3$ months	$m = 6$ months	$m = 9$ months
$h = 6$ months	-0.6942 (0.2533)		
2-Factor $\text{SVARN}^{nh}_{mv}(2)$	-0.5323 (0.3195)		
2-Factor $\text{SVARN}^{nh}_{cm}(2)$	-0.3747 (0.3107)		
2-Factor $\text{SVARN}^{ch}_{mv}(2)$	-0.1736 (0.2963)		
$h = 9$ months	-0.8863 (0.3238)	-0.4023 (0.2429)	
2-Factor $\text{SVARN}^{nh}_{mv}(2)$	-0.5957 (0.3286)	-0.4928 (0.2555)	
2-Factor $\text{SVARN}^{nh}_{cm}(2)$	-0.4446 (0.3158)	-0.4522 (0.2439)	
2-Factor $\text{SVARN}^{ch}_{mv}(2)$	-0.2472 (0.3008)	-0.2443 (0.2429)	
$h = 12$ months	-1.3226 (0.3530)	-0.7867 (0.2381)	-0.4371 (0.1312)
2-Factor $\text{SVARN}^{nh}_{mv}(2)$	-0.6591 (0.3363)	-0.5731 (0.2574)	-0.4906 (0.2297)
2-Factor $\text{SVARN}^{nh}_{cm}(2)$	-0.5113 (0.3216)	-0.5224 (0.2467)	-0.4703 (0.2226)
2-Factor $\text{SVARN}^{ch}_{mv}(2)$	-0.3170 (0.3057)	-0.3168 (0.2443)	-0.2817 (0.2239)
Long Horizon	$m = 1$ year	$m = 2$ years	$m = 3$ years
$h = 4$ years	-1.8078 (0.2981)	-0.8380 (0.2889)	-0.0421 (0.2682)
2-Factor $\text{SVARN}^{nh}_{mv}(2)$	-1.3067 (0.2857)	-0.7469 (0.2657)	-0.0306 (0.2756)
2-Factor $\text{SVARN}^{nh}_{cm}(2)$	-1.0239 (0.2886)	-0.2757 (0.2748)	0.6756 (0.2775)
2-Factor $\text{SVARN}^{ch}_{mv}(2)$	-1.1758 (0.2799)	-0.8034 (0.2633)	-0.3788 (0.2832)
$h = 5$ years	-1.7470 (0.3291)	-0.9720 (0.3199)	-0.2378 (0.3283)
2-Factor $\text{SVARN}^{nh}_{mv}(2)$	-1.5368 (0.3208)	-0.8823 (0.2963)	-0.0472 (0.3015)
2-Factor $\text{SVARN}^{nh}_{cm}(2)$	-1.1081 (0.3196)	-0.2237 (0.2987)	0.7983 (0.2910)
2-Factor $\text{SVARN}^{ch}_{mv}(2)$	-1.5074 (0.3198)	-1.0864 (0.2965)	-0.6075 (0.3118)

Table 17 : Short and Long Horizon Expectation Hypothesis Tests : the first row in each block reports the empirical slope coefficient and its standard error. The second to fourth rows are the slope coefficients and standard errors based on 10000 simulation from each fitted 2-Factor $\text{SVARN}^{nh}_{mv}(2)$, 2-Factor $\text{SVARN}^{nh}_{cm}(2)$ and $\text{SVARN}^{ch}_{mv}(2)$ Factor-Based Term Structure Model with each simulation having a sample size of 379.

In particular, the model-implied regression coefficient $\phi_1(h, m)$, for $h \in \{4, 5\}$ and $m = 3$, moves, respectively, from -0.0306 to 0.6756 [the empirical value is -0.0421] and from -0.0472 to 0.7983 [the empirical value is -0.2378], when we impose to model $\text{SVARN}^{nh}_{mv}(2)$ the restriction

$\varphi_2 = \varphi_2^*$. As far as the absolute pricing error comparison is concerned, we find that, for both $\text{SVARN}_{cm}^{nh}(2)$ and $\text{SVARN}_{mv}^{ch}(2)$ models that the mean absolute error is equal to 24 and the standard deviation is 16 [Min. = 3 and Max. = 93 in the first case, and Min. = 3 and Max. = 95 in the second one]. Moreover, the fact to impose a homogeneous transition matrix to the regime indicator function (z_t), leads the $\text{SVARN}_{mv}^{ch}(2)$ model to perform worse than the non-constrained $\text{SVARN}_{mv}^{nh}(2)$ specification in terms of Portmanteau test for short rate and spread residuals. The values of the Portmanteau test statistic Q_k, for a number of lags $k \in \{5, 10, 15, 20\}$, are respectively given by $Q_5 = 26.1034$, $Q_{10} = 47.0264$, $Q_{15} = 85.5460$ and $Q_{20} = 95.6638$. Thus, the test is passed, for 5, 10 and 20 lags, at 0.01, while for 15 lags the null hypothesis of absence of autocorrelation is rejected. This means that, both the risk associated to the second lagged factor value and the regime-shift risk play an important role for the $\text{SVARN}_{mv}^{nh}(2)$ model's performances (interest rate serial dependence, goodness-of-fit, Campbell-Shiller regressions). Moreover, the former seems to be determinant in explaining the long horizon Expectation Hypothesis Puzzle.

7 Conclusions

This paper has developed a general discrete-time modeling of the term structure of interest rates able to take into account at the same time several important features : a) an historical dynamics of the factor (x_t) involving several lagged values and switching regimes; b) a specification of the exponential-affine stochastic discount factor (SDF) with time-varying risk correction coefficients implying stochastic risk premia, functions of the present and past values of the factor (x_t) and the regime indicator function (z_t); c) explicit or quasi explicit formulas for zero-coupon bond (the Generalized Linear Term Structure formula) and interest rate derivative prices; d) the positiveness of the yields at each maturity (in the Autoregressive Gamma framework), regardless the observable or latent nature of the factor (x_t). We have studied, in the Gaussian framework, the SARN(p) and the SVARN(p) Factor-Based Term Structure Models, providing a generalization of the recent modelisation proposed by Dai, Singleton and Yang (2005). In the Autoregressive Gamma setting, we have proposed the SARG(p) Factor-Based Term Structure Models, extending several discrete time CIR term structure models like Bansal and Zhou (2002) [see Pegoraro (2006) for a presentation of the SVARG(p) Factor-Based Term Structure Models].

The empirical analysis of bivariate VARN(p) and $\text{SVARN}_{mv}^{nh}(p)$ Factor-Based Term Structure Models, with $p \in \{1, 2\}$, shows that the introduction of multiple lags and switching regimes, in the historical and risk-neutral dynamics of the observable factor (x_t), leads to term structure models which are well specified under the historical probability (Ljung-Box and Portmanteau tests) and able to explain the Expectation Hypothesis Puzzle, over the short and long horizon. A pricing error and expectation hypothesis puzzle comparison with competing 2-Factor CIR, 3-Factor CIR, 3-Factor $\mathbb{A}_1(3)$ and 2-Factor RS-CIR term structure models further highlights the determinant role played by observable factor, lags and switching regimes in the term structure modeling. The observable nature of the factor and the introduction of a second lagged factor value lead the bivariate VARN(2) Factor-Based Term Structure Model to fit yields data better than (heteroskedastic) 2-Factor CIR, 3-Factor CIR, 3-Factor $\mathbb{A}_1(3)$ term structure models, and to perform as well as the 2-Factor (discretized) square-root regime-switching term structure model proposed by Bansal and Zhou (2002). The introduction of (2-states) switching regimes, the pricing of the risk associated to the factor and its lagged values and the pricing of the regime-shift risk, lead our $\text{SVARN}_{mv}^{nh}(2)$ Factor-Based Term Structure Model to explain the violation of the Expectation Hypothesis Theory, over both the short and long horizon, better than all the above mentioned models while maintaining, at the same time, a fitting of yields data competing with the bivariate VARN(2) and the 2-Factor RS-CIR term structure models.

Appendix 1 : Proof of Proposition 3

$$
\begin{aligned}
& E[\exp(u_1 y_{1,t+1} + u_2 y_{2,t+1}) | \underline{y_{1,t}}, \underline{y_{2,t}}] \\
= \; & E\left[\exp(u_2 y_{2,t+1}) E\left(\exp(u_1 y_{1,t+1}) | \underline{y_{1,t}}, \underline{y_{2,t+1}}\right) \; | \underline{y_{1,t}}, \underline{y_{2,t}}\right] \\
= \; & \exp\left[a_1(u_1)(\beta_{11} y_{1,t} + \beta_{12} y_{2,t}) + b_1(u_1)\right] E_t\left[(u_2 + a_1(u_1)\beta_o) y_{2,t+1} \, | \, \underline{y_{1,t}}, \underline{y_{2,t}}\right] \\
= \; & \exp\left[a_1(u_1)(\beta_{11} y_{1,t} + \beta_{12} y_{2,t}) + b_1(u_1)\right. \\
& \quad \left. + a_2(u_2 + a_1(u_1)\beta_o)(\beta_{21} y_{1,t} + \beta_{22} y_{2,t}) + b_2(u_2 + a_1(u_1)\beta_o)\right] \\
= \; & \exp\left\{[a_1(u_1)\beta_{11} + a_2(u_2 + a_1(u_1)\beta_o)\beta_{21}] y_{1,t}\right. \\
& \quad \left. + [a_1(u_1)\beta_{12} + a_2(u_2 + a_1(u_1)\beta_o)\beta_{22}] y_{2,t} + b_1(u_1) + b_2(u_2 + a_1(u_1)\beta_o)\right\} \; .
\end{aligned}
$$

Appendix 2 : Proof of Proposition 4

The Laplace transform of the one-period conditional risk-neutral probability is:

$$
\begin{aligned}
& E_t^{\mathbb{Q}}[\exp(u x_{t+1} + v' z_{t+1})] \\
= \; & E_t\{\exp[\Gamma(X_t, Z_t)\, \varepsilon_{t+1} - \tfrac{1}{2}\Gamma(X_t, Z_t)^2 - \delta'(Z_t, X_t) z_{t+1} \\
& \quad + u[\nu(Z_t) + \varphi(Z_t)' X_t + \sigma(Z_t)\varepsilon_{t+1}] + v' z_{t+1}]\} \\
= \; & \exp\left\{ u[\,\varphi'(Z_t) X_t + \Gamma(X_t, Z_t)\sigma(Z_t)] + u\nu(Z_t) + \tfrac{1}{2}u^2\sigma(Z_t)^2 \right\} \times \\
& \quad \textstyle\sum_{j=1}^{J} \pi\,(z_t, e_j\,; X_t) \exp\left[(v - \delta(Z_t, X_t))' e_j\right] \\
= \; & \exp\left\{ u[\,\varphi(Z_t) + \tilde{\gamma}(Z_t)\sigma(Z_t)]' X_t + u[\nu(Z_t) + \gamma(Z_t)\sigma(Z_t)] + \tfrac{1}{2}u^2\sigma(Z_t)^2 \right\} \times \\
& \quad \textstyle\sum_{j=1}^{J} \pi\,(z_t, e_j\,; X_t) \exp\left[(v - \delta(Z_t, X_t))' e_j\right] \; .
\end{aligned}
$$

Therefore, we get the result of Proposition 4.

Appendix 3 : Proof of Proposition 5

$$
\begin{aligned}
B(t,h) &= \exp(C_h' X_t + D_h' Z_t) \\
&= \exp\left(-r_{t+1}\right) E_t^Q \left[B(t+1,h-1)\right] \\
&= \exp\left[-c' X_t - d' Z_t\right] \, E_t^Q \left[\exp\left(C_{h-1}' X_{t+1} + D_{h-1}' Z_{t+1}\right)\right] \\
&= \exp\left[-c' X_t - d' Z_t\right] \times \\
& \qquad E_t^Q \left[\exp\left(C_{h-1}' \left[\Phi^* X_t + \left(\nu^{*'} Z_t + \sigma^{*'} Z_t \, \xi_{t+1}\right) e_1\right] + D_{1,h-1}' z_{t+1} + \tilde{D}_{h-1}' Z_t\right)\right] \\
&= \exp\left[\left(\Phi^{*'} C_{h-1} - c\right)' X_t + \left(-d + C_{1,h-1} \nu^* + \tfrac{1}{2} C_{1,h-1}^2 \sigma^{*2} + \tilde{D}_{h-1}\right)' Z_t\right] \times \\
& \qquad E_t^Q \left[\exp\left(D_{1,h-1}' z_{t+1}\right)\right] \\
&= \exp\left\{\left(\Phi^{*'} C_{h-1} - c\right)' X_t + \right. \\
& \qquad \left. \left[-d + C_{1,h-1} \nu^* + \tfrac{1}{2} C_{1,h-1}^2 \sigma^{*2} + \tilde{D}_{h-1} + F(D_{1,h-1})\right]' Z_t\right\} \, ,
\end{aligned}
$$

and the result follows by identification.

Appendix 4 : Proof of Proposition 6

Using the lag polynomials:

$$
\begin{aligned}
C_h(L) &= -\frac{1}{h} \left(C_{1,h} + C_{2,h} L + \ldots + C_{p,h} L^{p-1}\right) \\
D_h(L) &= -\frac{1}{h} \left(D_{1,h} + D_{2,h} L + \ldots + D_{p+1,h} L^p\right) \\
\Psi(L, Z_t) &= 1 - \varphi_1(Z_t) L - \ldots - \varphi_p(Z_t) L^p \, ,
\end{aligned}
$$

we get from (43):

$$
R(t,h) = C_h(L) x_t + D_h(L)' z_t \, ,
$$

and

$$
\begin{aligned}
\Psi(L, Z_t) \, R(t+1,h) &= C_h(L) \, \Psi(L, Z_t) \, x_{t+1} + D_h(L) \, \Psi(L, Z_t) \, z_{t+1}, \\
&= D_h(L) \, \Psi(L, Z_t) \, z_{t+1} + C_h(L) \, \nu(Z_t) + C_h(L) [(\sigma^{*'} Z_t) \, \varepsilon_{t+1}] \, .
\end{aligned}
$$

Appendix 5 : Proof of Proposition 8

The Laplace transform of the one-period conditional risk-neutral distribution is :

$$\begin{aligned}
& E_t^{\mathbb{Q}}[\exp(u'\tilde{x}_{t+1} + v'z_{t+1})] \\
= \;& E_t\{\exp[\Gamma(\tilde{X}_t, Z_t)'\,\varepsilon_{t+1} - \tfrac{1}{2}\Gamma(\tilde{X}_t, Z_t)'\Gamma(\tilde{X}_t, Z_t) - \delta'(Z_t, \tilde{X}_t)z_{t+1} \\
& \quad + u'[\tilde{\nu}(Z_t) + \tilde{\Phi}(Z_t)\tilde{X}_t + S(Z_t)\varepsilon_{t+1}] + v'z_{t+1}]\} \\
= \;& \exp\left\{ u'[\,\tilde{\Phi}(Z_t)\tilde{X}_t + S(Z_t)\Gamma(\tilde{X}_t, Z_t)] + u'\tilde{\nu}(Z_t) + \tfrac{1}{2}u'S(Z_t)S(Z_t)'u \right\} \times \\
& \quad \textstyle\sum_{j=1}^{J} \pi(z_t, e_j\,; \tilde{X}_t) \exp\left[(v - \delta(Z_t, \tilde{X}_t))'e_j\right] \\
= \;& \exp\left\{ u'[\,\tilde{\Phi}(Z_t) + S(Z_t)\tilde{\Gamma}(Z_t, \tilde{X}_t)]\tilde{X}_t + u'[\,\tilde{\nu}(Z_t) + S(Z_t)\gamma(Z_t)] + \tfrac{1}{2}u'S(Z_t)S(Z_t)'u \right\} \times \\
& \quad \textstyle\sum_{j=1}^{J} \pi(z_t, e_j\,; \tilde{X}_t) \exp\left[(v - \delta(Z_t, \tilde{X}_t))'e_j\right] .
\end{aligned}$$

Therefore, we get the result of Proposition 8.

Appendix 6 : Proof of Proposition 9

$$\begin{aligned}
B(t,h) \;&=\; \exp(C_h'\tilde{X}_t + D_h'Z_t) \\
&=\; \exp\left(-r_{t+1}\right) E_t^Q\left[B(t+1, h-1)\right] \\
&=\; \exp\left[-c'\tilde{X}_t - d'Z_t\right]\; E_t^Q\left[\exp\left(C_{h-1}'\tilde{X}_{t+1} + D_{h-1}'Z_{t+1}\right)\right] \\
&=\; \exp\left[-c'\tilde{X}_t - d'Z_t\right] \times \\
& \quad E_t^Q\Big[\exp\Big(C_{h-1}'\tilde{\Phi}^*\tilde{X}_t + C_{1,h-1}(\nu_1^* Z_t + S_1^*(Z_t)\xi_{t+1}) \\
& \qquad + C_{p+1,h-1}(\nu_2^* Z_t + S_2^*(Z_t)\xi_{t+1}) + D_{1,h-1}'z_{t+1} + \tilde{D}_{h-1}'Z_t\Big)\Big] \\
&=\; \exp\Big[\left(\tilde{\Phi}^{*'}C_{h-1} - c\right)' X_t + \Big[-d + C_{1,h-1}\nu_1^{*'} + C_{p+1,h-1}\nu_2^{*'} \\
& \qquad + \tfrac{1}{2}C_{1,h-1}^2(\sigma_1^{*2} + \varphi_o^{*2}\sigma_2^{*2}) + (C_{1,h-1})(C_{p+1,h-1})\varphi_o^{*2}\sigma_2^{*2} \\
& \qquad + \tfrac{1}{2}C_{p+1,h-1}^2\sigma_2^{*2} + \tilde{D}_{h-1} + F(D_{1,h-1})\Big]' Z_t\Big] ,
\end{aligned}$$

and the result follows by identification.

Appendix 7 : Proof of Lemma 1

$$\begin{aligned}
\tilde{a}(u+\alpha;\rho,\mu)-\tilde{a}(\alpha;\rho,\mu) &= \frac{\rho(u+\alpha)}{1-(u+\alpha)\mu}-\frac{\rho\alpha}{1-\alpha\mu} \\
&= \rho\frac{u}{(1-\alpha\mu)^2-u\mu(1-\alpha\mu)} \\
&= \frac{\rho}{(1-\alpha\mu)^2}\,\frac{u}{1-\frac{u\mu}{1-\alpha\mu}} \\
&= \frac{\rho^* u}{1-u\mu^*}=\tilde{a}(u;\rho^*,\mu^*)\,;
\end{aligned}$$

$$\begin{aligned}
\tilde{b}(u+\alpha;\nu,\mu)-\tilde{b}(\alpha;\nu,\mu) &= -\nu\log(1-(u+\alpha)\mu)+-\nu\log(1-\alpha\mu) \\
&= -\nu\log\left[\frac{1-(u+\alpha)\mu}{1-\alpha\mu}\right] \\
&= -\nu\log\left[1-\frac{u\mu}{1-\alpha\mu}\right] \\
&= -\nu\log(1-u\mu^*) \\
&= \tilde{b}(u;\nu,\mu^*)\,.
\end{aligned}$$

Appendix 8 : Proof of Proposition 10

$$\begin{aligned}
B(t,h) &= \exp(C_h' X_t + D_h' Z_t) \\
&= \exp\left[-c'X_t - d'Z_t\right]\, E_t^Q\left[\exp\left(C_{h-1}' X_{t+1} + D_{h-1}' Z_{t+1}\right)\right] \\
&= \exp\left(-c'X_t - d'Z_t + \tilde{C}_{h-1}' X_t + \tilde{D}_{h-1}' Z_t\right) \\
&\qquad E_t^Q\left[\exp\left(C_{1,h-1} x_{t+1} + D_{1,h-1}' z_{t+1}\right)\right] \\
&= \exp\Big[-c'X_t - d'Z_t + \tilde{C}_{h-1}' X_t + \tilde{D}_{h-1}' Z_t + A^*(C_{1,h-1})' X_t \\
&\qquad -\nu^{*'} Z_t \log(1 - C_{1,h-1}\mu^*) + F'(D_{1,h-1}) Z_t\Big]\,,
\end{aligned}$$

and the result follows by identification.

Appendix 9 : A General Proof of the Kim's Smoothing Algorithm

The proof of the Kim's Smoothing algorithm for the general model:

$$\begin{aligned} y_{t+1} &= \vartheta_y \left(\underline{y_t}, \underline{z_{t-p+1}^{t+1}}, \eta_{t+1} \right) \\ z_{t+1} &= \vartheta_z \left(\underline{y_t}, z_t, \varepsilon_{t+1} \right) , \end{aligned} \tag{A.1}$$

with (ε_t) and (η_t) independent white noise processes, (y_t) an observable process, (z_t) a non-homogeneous (latent) Markov chain, where $\underline{y_t} = (y_t, y_{t-1}, \ldots)$, $\underline{z_{t-p+1}^{t+1}} = (z_{t-p+1}, \ldots, z_{t+1})$, and where $p \in \mathbb{N}$ and $h \in \mathbb{N}_+$, is based on the following three lemmas.

Lemma 1 : Model (A.1) can be written, for each integer $h \geq 2$, in the following way :

$$\begin{aligned} y_{t+h} &= \vartheta_y^{(h)} \left(\underline{y_t}, \underline{z_{t-p+1}^{t+1}}, \eta_{t+1}, \ldots, \eta_{t+h}, \varepsilon_{t+2}, \ldots, \varepsilon_{t+h} \right) \\ z_{t+h} &= \vartheta_z^{(h)} \left(\underline{y_t}, \underline{z_{t-p+1}^{t+1}}, \eta_{t+1}, \ldots, \eta_{t+h-1}, \varepsilon_{t+2}, \ldots, \varepsilon_{t+h} \right) , \end{aligned} \tag{A.2}$$

[Proof : by recursive substitution, starting from (y_{t+1}, z_{t+1})]. In particular, for each $h \geq 1$ and replacing t by $t+p$, we have:

$$\begin{aligned} y_{t+p+h} &= \vartheta_y^{(h)} \left(\underline{y_{t+p}}, \underline{z_{t+1}^{t+p+1}}, \eta_{t+p+1}, \ldots, \eta_{t+p+h}, \varepsilon_{t+p+2}, \ldots, \varepsilon_{t+p+h} \right) \\ z_{t+p+h} &= \vartheta_z^{(h)} \left(\underline{y_{t+p}}, \underline{z_{t+1}^{t+p+1}}, \eta_{t+p+1}, \ldots, \eta_{t+p+h-1}, \varepsilon_{t+p+2}, \ldots, \varepsilon_{t+p+h} \right) , \end{aligned} \tag{A.3}$$

where, with the notation $I_\varepsilon(t+p, h) := (\varepsilon_{t+p+2}, \ldots, \varepsilon_{t+p+h})$, we assume $I_\varepsilon(t+p, 1) = \emptyset$.

Lemma 2 : If $I_1 \subset I \subset I_2$ and $\mathbb{P}[z_t|I_1] = \mathbb{P}[z_t|I_2]$, then $\mathbb{P}[z_t|I_1] = \mathbb{P}[z_t|I] = \mathbb{P}[z_t|I_2]$ [Proof : straightforward].

Lemma 3 : Given model (A.1), the following relation holds :

$$\mathbb{P}\left[z_t \,|\, \underline{z_{t+1}^{t+p+1}}, \underline{y_T}\right] = \mathbb{P}\left[z_t \,|\, \underline{z_{t+1}^{t+p+1}}, \underline{y_{t+p}}\right] . \tag{A.4}$$

Proof : Given the three sets $I_1 = \left(\underline{z_{t+1}^{t+p+1}}, \underline{y_{t+p}}\right)$, $I = \left(\underline{z_{t+1}^{t+p+1}}, \underline{y_T}\right)$ and $I_2 = \left(\underline{z_{t+1}^{t+p+1}}, \underline{y_{t+p}}, \underline{\eta_{t+p+1}^T}, \underline{\varepsilon_{t+p+2}^T}\right)$, we have that :

(i) $I_1 \subset I$,

(ii) $I \subset I_2$,

(iii) $(\underline{\eta_{t+p+1}^T}, \underline{\varepsilon_{t+p+2}^T}) \perp (\underline{z_{t+1}^{t+p+1}}, \underline{y_{t+p}}, z_t)$.

The proof of relation (i) is straightforward. Relation (ii) holds given that, from Lemma 1, we can always write, for any $s > t+p$:

$$y_s = \vartheta_y^{(s-t-p)} \left(\underline{y_{t+p}}, \underline{z_{t+1}^{t+p+1}}, \eta_{t+p+1}, \ldots, \eta_s, \varepsilon_{t+p+2}, \ldots, \varepsilon_s \right) , \tag{A.5}$$

and, therefore, $(\underline{y_T}) \subset \left(\underline{y_{t+p}}, \underline{z_{t+1}^{t+p+1}}, \underline{\eta_{t+p+1}^T}, \underline{\varepsilon_{t+p+2}^T} \right)$, that is, $I \subset I_2$.
With regard to relation (iii), from Lemma 1 applied to (z_{t+1}, y_{t+1}), we have:

$$\begin{aligned} z_{t+p+1} &= \vartheta_z^{(t+p+1)} \left(\underline{y_0}, \underline{z_{-p+1}^1}, \eta_1, \ldots, \eta_{t+p}, \varepsilon_2, \ldots, \varepsilon_{t+p+1} \right) \\ y_{t+p} &= \vartheta_y^{(t+p)} \left(\underline{y_0}, \underline{z_{-p+1}^1}, \eta_1, \ldots, \eta_{t+p}, \varepsilon_2, \ldots, \varepsilon_{t+p} \right) , \end{aligned} \tag{A.6}$$

and given that $z_t = \vartheta_z \left(\underline{y_{t-1}}, z_{t-1}, \varepsilon_t \right)$, we conclude (using the notation $\mathbb{P}$ for the p.d.f.) :

$$\mathbb{P} \left(\underline{\eta_{t+p+1}^T}, \underline{\varepsilon_{t+p+2}^T} \mid \underline{z_{t+1}^{t+p+1}}, \underline{y_{t+p}}, z_t \right) = \mathbb{P} \left(\underline{\eta_{t+p+1}^T}, \underline{\varepsilon_{t+p+2}^T} \right) , \tag{A.7}$$

and relation (iii) is proved. Now, given property (iii), we have:

$$\begin{aligned} \mathbb{P}[z_t \mid I_2] &= \frac{\mathbb{P}[z_t, I_2]}{\mathbb{P}[I_2]} \\ &= \frac{\mathbb{P}[\underline{z_{t+1}^{t+p+1}}, \underline{y_{t+p}}, z_t, \underline{\eta_{t+p+1}^T}, \underline{\varepsilon_{t+p+2}^T}]}{\mathbb{P}[\underline{z_{t+1}^{t+p+1}}, \underline{y_{t+p}}, \underline{\eta_{t+p+1}^T}, \underline{\varepsilon_{t+p+2}^T}]} \\ &= \frac{\mathbb{P}[\underline{z_{t+1}^{t+p+1}}, \underline{y_{t+p}}, z_t]\mathbb{P}[\underline{\eta_{t+p+1}^T}, \underline{\varepsilon_{t+p+2}^T}]}{\mathbb{P}[\underline{z_{t+1}^{t+p+1}}, \underline{y_{t+p}}]\mathbb{P}[\underline{\eta_{t+p+1}^T}, \underline{\varepsilon_{t+p+2}^T}]} \\ &= \frac{\mathbb{P}[z_t, I_1]}{\mathbb{P}[I_1]} = \mathbb{P}[z_t \mid I_1] , \end{aligned} \tag{A.8}$$

and applying Lemma 2, we prove (A.4).

If $p \geq 1$, the Kim's smoothing formula is :

$$\mathbb{P}[z_t, \ldots, z_{t+p} \mid \underline{y_T}] = \frac{\mathbb{P}[z_t, \ldots, z_{t+p} \mid \underline{y_{t+p}}\,]}{\mathbb{P}[z_{t+1}, \ldots, z_{t+p} \mid \underline{y_{t+p}}\,]} \sum_{z_{t+p+1}} \mathbb{P}[z_{t+1}, \ldots, z_{t+p+1} \mid \underline{y_T}\,] . \tag{A.9}$$

Proof : Applying Lemma 3, we can write

$$\begin{aligned} & \mathbb{P}[z_t, \ldots, z_{t+p+1} \mid \underline{y_T}] \\ = \; & \mathbb{P}[z_t \mid z_{t+1}, \ldots, z_{t+p+1}, \underline{y_T}]\, \mathbb{P}[z_{t+1}, \ldots, z_{t+p+1} \mid \underline{y_T}] \\ = \; & \mathbb{P}[z_t \mid z_{t+1}, \ldots, z_{t+p+1}, \underline{y_{t+p}}]\, \mathbb{P}[z_{t+1}, \ldots, z_{t+p+1} \mid \underline{y_T}] \\ = \; & \mathbb{P}[z_{t+1}, \ldots, z_{t+p+1} \mid \underline{y_T}] \frac{\mathbb{P}[z_t, \ldots, z_{t+p+1}, \underline{y_{t+p}}\,]}{\mathbb{P}[z_{t+1}, \ldots, z_{t+p+1}, \underline{y_{t+p}}\,]} \\ = \; & \mathbb{P}[z_{t+1}, \ldots, z_{t+p+1} \mid \underline{y_T}] \frac{\mathbb{P}[z_{t+p+1} \mid z_{t+p}, \ldots, z_t, \underline{y_{t+p}}\,]\mathbb{P}[z_{t+p}, \ldots, z_t \mid \underline{y_{t+p}}\,]}{\mathbb{P}[z_{t+p+1} \mid z_{t+p}, \ldots, z_{t+1}, \underline{y_{t+p}}\,]\mathbb{P}[z_{t+p}, \ldots, z_{t+1} \mid \underline{y_{t+p}}\,]} \end{aligned} \tag{A.10}$$

and, (z_t) being (conditionally) a Markov chain, relation (A.10) can be written:

$$\begin{aligned}
&\mathbb{P}[z_t, \ldots, z_{t+p+1} \,|\, \underline{y_T}] \\
= \;& \mathbb{P}[z_{t+1}, \ldots, z_{t+p+1} \,|\, \underline{y_T}] \frac{\mathbb{P}[z_t, \ldots, z_{t+p} \,|\, \underline{y_{t+p}}\,]}{\mathbb{P}[z_{t+1}, \ldots, z_{t+p} \,|\, \underline{y_{t+p}}\,]}\,;
\end{aligned} \tag{A.11}$$

now, if we integrate out z_{t+p+1} on the LHS and RHS of (A.11), we obtain (A.9). The smoothing algorithm start, at $t = T - p - 1$, from $\mathbb{P}[z_{T-p}, \ldots, z_{T-1} \,|\, \underline{y_T}] = \sum_{z_T} \mathbb{P}[z_{T-p}, \ldots, z_T \,|\, \underline{y_T}]$, with $\mathbb{P}[z_{T-p}, \ldots, z_T \,|\, \underline{y_T}]$ provided by the Kitagawa-Hamilton filter.

If $p = 0$, the Kim's smoothing formula is :

$$\mathbb{P}[z_t \,|\, \underline{y_T}] = \mathbb{P}[z_t \,|\, \underline{y_t}\,] \sum_{z_{t+1}} \frac{\mathbb{P}[z_{t+1} \,|\, z_t, \underline{y_t}\,]\mathbb{P}[z_{t+1} \,|\, \underline{y_T}\,]}{\mathbb{P}[z_{t+1} \,|\, \underline{y_t}\,]}\,. \tag{A.12}$$

Proof : Given that

$$\begin{aligned}
&\mathbb{P}[z_t, z_{t+1} \,|\, \underline{y_T}] \\
= \;& \mathbb{P}[z_t \,|\, z_{t+1}, \underline{y_T}]\, \mathbb{P}[z_{t+1} \,|\, \underline{y_T}] \\
= \;& \mathbb{P}[z_t \,|\, z_{t+1}, \underline{y_t}]\, \mathbb{P}[z_{t+1} \,|\, \underline{y_T}] \\
= \;& \mathbb{P}[z_{t+1} \,|\, \underline{y_T}] \frac{\mathbb{P}[z_t, z_{t+1} \,|\, \underline{y_t}\,]}{\mathbb{P}[z_{t+1} \,|\, \underline{y_t}\,]} \\
= \;& \mathbb{P}[z_{t+1} \,|\, \underline{y_T}] \frac{\mathbb{P}[z_{t+1} \,|\, z_t, \underline{y_t}\,]\mathbb{P}[z_t \,|\, \underline{y_t}]}{\mathbb{P}[z_{t+1} \,|\, \underline{y_t}\,]}\,,
\end{aligned} \tag{A.13}$$

if we integrate out z_{t+1} from the LHS and RHS of (A.13) we obtain (A.12).

R E F E R E N C E S

Ahn, D., Dittmar, R., and R. Gallant (2002) : "Quadratic Term Structure Models : Theory and Evidence", Review of Financial Studies, 15, 243-288.

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

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

Ang, A., and G. Bekaert, (2002b) : "Short Rate Nonlinearities and Regime Switches", Journal of Economic Dynamics and Control, 26, 1243-1274.

Ang, A., and G. Bekaert, (2005) : "Term Structure of Real Rates and Expected Inflation", Working Paper.

Ang, A., Piazzesi, M., and M. Wei (2005) : "What does the Yield Curve tell us about GDP Growth", forthcoming Journal of Econometrics.

Bansal, R., and H. Zhou, (2002) : "Term Structure of Interest Rates with Regime Shifts", Journal of Finance, 57, 1997-2043.

Beaglehole, D. R., and M. S. Tenney (1991) : "General Solutions of Some Interest Rate Contingent Claim Pricing Equations", Journal of Fixed Income, September, 69-83.

Billio, M., and A. Monfort (1998) : "Switching State-Space Models Likelihood Function, Filtering and Smoothing", Journal of Statistical Planning and Inference", 68, 65-103.

Boudoukh, J., Richardson, M., Smith, T., and R. F. Whitelaw (1999) : "Regime Shifts and Bond Returns", Working Paper.

Cai, J., (1994) : "A Markov Model of Unconditional Variance in ARCH", Journal of Business and Economic Statistics, 12, 309-316.

Campbell, J., and R. J. Shiller (1991) : "Yield Spreads and Interest Rate Movements : A Bird's Eye View", Review of Economic Studies, 58, 495-514.

Cheng, P., and O. Scaillet, (2005) : "Linear-Quadratic Jump-Diffusion modeling with Application to Stochastic Volatility", Working Paper.

Cheridito, R., Filipovic, D., and R. Kimmel (2005) : "Market Prices of Risk in Affine Models: Theory and Evidence", forthcoming Journal of Financial Economics.

Christiansen, C., (2002) : "Regime Switching in the Yield Curve", Working Paper.

Christiansen, C., and J. Lund (2005) : "Revisiting the Shape of the Yield Curve: the Effect of Interest Rate Volatility", Working Paper.

Cochrane, J., and M. Piazzesi, (2005) : "Bond Risk Premia", American Economic Review, 95/1, 138-160.

Cox, J.C., Ingersoll, J.E. and S. A. Ross (1985) : "A theory of the term structure of interest rates", Econometrica, 53/2, pp. 385-407.

Dai, Q., and K. Singleton (2000) : "Specification Analysis of Affine Term Structure Models", Journal of Finance, 55, 1943-1978.

Dai, Q., and K. Singleton (2002) : "Expectations Puzzles, Time-varying Risk Premia, and Affine Model of the Term Structure", Journal of Financial Economics, 63, 415-441.

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

Dai, Q., Singleton, K., and W. Yang (2005) : "Regime Shifts in a Dynamic Term Structure Model of U.S. Treasury Bond Yields", Working Paper.

Dai, Q., Le, A., and K. Singleton (2006) : "Discrete-time Term Structure Models with Generalized Market Prices of Risk", Working Paper.

Darolles, S., Gourieroux, C. and J. Jasiak (2006) : "Structural Laplace Transform and Compound Autoregressive Models", forthcoming Journal of Time Series Analysis.

Driffill, J., and M. Sola, (1994) : "Testing the Term Structure of Interest Rates using a Stationary Vector Autoregression with Regime Switching", Journal of Economic Dynamics and Control, 18, 601-628.

Driffill, J., T., Kenc, and M. Sola, (2003) : "An Empirical Examination of Term Structure Models with Regime Shifts", Working Paper.

Duarte, J. (2004) : "Evaluating An Alternative Risk Preference in Affine Term Structure Models", Review of Financial Studies, 17, 379-404.

Duffee, G. R., (2002) : "Term Premia and Interest Rate Forecasts in Affine Models", Journal of Finance, 57, 405-443.

Duffie, D., Filipovic, D. and W. Schachermayer (2003) : "Affine processes and applications in finance", The Annals of Applied Probability 13, 3.

Duffie, D., and R., Kan (1996) : "A yield-factor model of interest rates", Mathematical Finance, 6, 379-406.

Duffie, D., J. Pan, and K. Singleton (2000) : "Transform Analysis and Asset Pricing for Affine Jump Diffusions", Econometrica, 68, 1343-1376.

Evans, M. (2003) : "Real Risk, Inflation Risk, and the Term Structure", The Economic Journal, 113, 345-389.

Garcia, R. and P. Perron (1996) : "An Analysis of Real Interest Rate Under Regime Shifts", Reviews of Economics and Statistics, 1, 111-125.

Gourieroux, C., and J. Jasiak (2006) : "Autoregressive Gamma Processes", forthcoming Journal of Forecasting.

Gourieroux, C., Jasiak, J., and R. Sufana (2004) : "A Dynamic Model for Multivariate Stochastic Volatility : The Wishart Autoregressive Process", Working Paper.

Gourieroux, C., and A. Monfort (2006) : "Domain Restrictions on Interest Rates Implied by No Arbitrage", Crest DP.

Gourieroux, C., and A. Monfort (2006) : "Econometric Specifications of Stochastic Discount Factor Models", forthcoming Journal of Econometrics.

Gourieroux, C., Monfort, A. and V. Polimenis (2003) : "Discrete Time Affine Term Structure Models", Crest DP.

Gourieroux, C., Monfort, A. and V. Polimenis (2006) : "Affine Model for Credit Risk Analysis", Journal of Financial Econometrics, 4, 3, 494-530.

Gourieroux, C., and R. Sufana, (2003) : "Wishart Quadratic Term Structure Models", Working Paper.

Gray, S., (1996) : "Modeling the Conditional Distribution of Interest Rates as a Regime Switching Process", Journal of Financial Economics, 42, 27-62.

Hamilton, J. D., (1988) : "Rational-Expectations Econometric Analysis of Changes in Regimes - An Investigation of the Term Structure of Interest Rates", Journal of Economic Dynamics and Control, 12, 385-423.

Hamilton, J. D., (1994) : "Time Series Analysis", Princeton University Press.

Kim, C. J. (1994) : "Dynamic Linear Models with Markov Switching", Journal of Econometrics, 60, 1-22.

Monfort, A., and F. Pegoraro (2006) : "Multi-Lag Term Structure Models with Stochastic Risk Premia", CREST DP.

Pegoraro, F. (2006) : "Discrete Time Factor Models for Asset Pricing", PhD Thesis, Université Paris-Dauphine (France).

Piazzesi, M. (2003) : "Affine Term Structure Models", forthcoming Handbook of Financial Econometrics.

Polimenis, V. (2001) : "Essays in Discrete Time Asset Pricing", Ph. D. Thesis, Wharton School, University of Pennsylvania.

Roberds, W., and C. H. Whiteman (1999) : "Endogenous Term Premia and Anomalies in the Term Structure of Interest Rates : Explaining the Predictability Smile", Journal of Monetary Economics, 44, 555-580.

Stanton, R. (1997) : "A Nonparametric Model of the Term Structure Dynamics and the Market Price of Interest Rate Risk", The Journal of Finance, 52, 1973-2002.

Vasicek, O. (1977) : "An Equilibrium Characterization of the Term Structure", Journal of Financial Economics, 5, 177-188.

Wu, S., and Y. Zeng (2005) : "A General Equilibrium Model of the Term Structure of Interest Rates under Regime-Switching Risk", International Journal of Theoretical and Applied Finance, 8, 839-869.