

# Staying at Zero with Affine Processes: An Application to Term Structure Modelling

Alain MONFORT\*, Fulvio PEGORARO†

Jean-Paul RENNE‡ and Guillaume ROUSSELLET§

26 March 2015

## Abstract

We build an Affine Term Structure Model that provides non-negative yields at any maturity and that is able to accommodate a short-term rate that stays at the zero lower bound (ZLB) for extended periods of time while longer-term rates feature high volatilities. We introduce these features through a new univariate non-negative affine process called ARG-Zero, and its multivariate affine counterpart (VARG), entailing conditional distributions with a zero-point masses. The affine property of this new class of processes implies both explicit bond pricing and quasi-explicit lift-off probability formulas. We provide an empirical application to Japanese Government Bond (JGB) yields, observed weekly from June 1995 to May 2014 with maturities from six months to ten years. Our four-factor specification is able to closely match yield levels and to capture conditional yield volatilities.

**JEL Codes:** E43, G12.

**Key-words:** Zero Lower Bound, Affine Process, Term-Structure Model, Lift-Off probabilities.

---

\*Corresponding author. CREST, Banque de France, [alain.monfort@ensae.fr](mailto:alain.monfort@ensae.fr)

†Banque de France, CREST, [fulvio.pegoraro@banque-france.fr](mailto:fulvio.pegoraro@banque-france.fr)

‡Banque de France, [jeanpaul.renne@banque-france.fr](mailto:jeanpaul.renne@banque-france.fr)

§Banque de France, CREST and CEREMADE [guillaume.roussetlet@banque-france.fr](mailto:guillaume.roussetlet@banque-france.fr)

The authors are thankful to Damir FILIPOVIC, Lorian Mancini, Julien HUGONNIER, Mickael ROCKINGER, Anh LE, Glenn RUDEBUSCH, Michael BAUER, Jens CHRISTENSEN, Mike CHERNOV, Francis LONGSTAFF, Eduardo SCHWARTZ, Daniel ANDREI, Michael BRENNAN, Hanno LUSTIG, Mark GRINBLATT, Robert GESKE, Christian GOURIEROUX, Lasse PEDERSEN, Oreste TRISTANI, Peter HORDAHL, Wolfgang LEMKE, Hans DEWACHTER, Fabio TROJANI, Claudio TEBALDI, Patrick GAGLIARDINI, Jean-Sebastien FONTAINE, Bruno FEUNOU, Olivier SCAILLET, Don KIM, Olesya GRISCHCHENKO, Marcel PRIEBSCHE, Tobias ADRIAN, Emanuel MOENCH, Nina BOYARCHENKO, Richard CRUMP, Nour MEDDAHI, Bruno FEUNOU, Caio ALMEIDA, Rodrigo GUIMARAES, Iryna KAMINSKA, Andrew MELDRUM, Refet GURKAYNAK, Darrel DUFFIE, David BACKUS, Robin LUMSDAINE and Simone MANGANELLI for their useful comments. We are also grateful to seminar participants at EPFL, CREST, ECB, University of North Carolina at Chapel Hill, San Francisco Federal reserve, UCLA Anderson School of Management, Board of the U.S. Federal Reserve, New York Federal Reserve, Bank of Canada, Geneva University and GFRI, UK Debt Management Office, Bank of England, BNP-Paribas, and to participants to TSE Financial Econometrics Conference, 31<sup>st</sup> Spring French Finance Association Conference, 7<sup>th</sup> annual SoFiE Conference and 6<sup>th</sup> French Econometrics Conference. Particular thanks are addressed to Andrew SIEGEL for thought-provoking discussions that led to this paper.

The views expressed in this paper are those of the authors and do not necessarily reflect those of the Banque de France.

# 1 Introduction

Assuming that storing cash is costless, nominal interest rates cannot turn negative since cash provides a zero interest rate and is always an alternative investment to bonds (see e.g. [Black \(1995\)](#)). In other words, the sole existence of currency implies a zero lower bound (ZLB) on bond yields.<sup>1</sup> Before the outbreak of the 2008 financial crisis, the Bank of Japan was the only large central bank that had brought its policy rates – which drive the short-end of the yield curve – to zero. From 2010 on however, bringing policy rates close to the ZLB has become a common situation for the Fed, the ECB, and the BoE. In all of these currency areas, sharp decreases of short-term interest rates have led the medium- to long-term yields to drop deeply, pushing the entire yield curves to unprecedented low levels.

In this context, reproducing low but non-negative interest rates has become a great concern for the specification of term structure models, and still represents a challenging task.<sup>2</sup> Specifically, to the best of our knowledge, no existing term-structure model is able to simultaneously match the three following characteristics:<sup>3</sup>

- (i) consistency with non-negative yields;
- (ii) availability of closed-form bond pricing formulas; and
- (iii) the ability to accommodate extended periods of zero short-term rates and to evaluate associated of leaving the zero lower bound (lift-off probabilities).

In this paper, we first introduce a new affine process that opens the way to term-structure models consistent with (i), (ii), and (iii) simultaneously. This process, which we call Autoregressive Gamma-zero (ARG<sub>0</sub>), builds on the original ARG process (see [Gourieroux and Jasiak \(2006\)](#), [Dai, Le, and Singleton \(2010\)](#) or [Creal and Wu \(2015\)](#)) by extending it to a zero degree-of-freedom parameter. This process has a crucial distinctive feature: its conditional distribution given the past values encompasses a point-mass at zero.<sup>4</sup> This attractive property allows its dynamics to satisfy (i) and (iii).<sup>5</sup> We explore the properties of this univariate process, explicitly disclosing its exponential-affine conditional Laplace transform and its first two conditional and unconditional moments. This univariate affine process is then extended to a multivariate affine process which we call Vectorial Autoregressive Gamma (VARG). We adequately exploit these processes to build

<sup>1</sup>In reality, holding cash is not costless since it is subject to theft or physical destruction and is complicated to use for large transactions. These features, along with flight-to-safety phenomena or non-conventional monetary-policy measures, may result in negative interest rates. The framework we develop in the present paper is consistent with the existence of a lower bound, which can be negative.

<sup>2</sup>Typically, in the widely-used Gaussian no-arbitrage models, the yields of all maturities can take negative values with a strictly positive probability (see e.g. [Dai and Singleton \(2003\)](#), [Piazzesi \(2010\)](#), [Diebold and Rudebusch \(2013\)](#), [Duffee \(2012\)](#) or [Gurkaynak and Wright \(2012\)](#)).

<sup>3</sup>While the model proposed by [Renne \(2012\)](#) is consistent with these three points, it can only generate a discrete number of positive yield curves. That is, in Renne's framework, the support of the positive short-term (policy) rate is discrete. Here, we consider short rates whose support is the set of non-negative real numbers (denoted by  $\mathbb{R}^+$ ).

<sup>4</sup>This appealing feature is obtained by building on [Siegel \(1979\)](#), who introduces a non-central Chi-squared distribution with zero degree of freedom. This distribution has also a Dirac mass at zero.

<sup>5</sup>As noted by [Kim \(2008\)](#), coping with those two features for a short-term interest rate is of utmost importance when building a term-structure model with observed option prices.

a multi-factor term-structure model in which the yields at all maturities are non-negative, the short-term interest rate can stay at zero for extended periods of time, and the lift-off probabilities are easily computed under both measures.

We directly address the issue of point *(ii)*, making closed-form bond-pricing formulas available. Indeed, our short-term interest rate is specified as a linear combination of components that follow VARG processes under both historical and risk-neutral measures. Hence our framework boils down to an Affine Term-Structure Model (ATSM) and the zero-coupon yields for all maturities are explicit affine functions of the factors where the loadings are computable recursively (see e.g. [Duffie and Kan \(1996\)](#) or [Darolles, Gouriéroux, and Jasiak \(2006\)](#)).

The historical and risk-neutral affine property of our term-structure model allows for a great flexibility at the estimation stage. First, assuming the presence of latent factors, the estimation technique is computationally simple using Kalman filtering techniques. Indeed, transition equations of the underlying state-space model are simply given by the VAR representation of our factors' dynamics. Second, it implies that (a) forecasts and (b) conditional variances of yields are affine functions of the factors. Accordingly, this allows us to easily augment the set of measurement equations by relating linear combinations of our latent factors with observable proxies of (a) surveys of professional forecasters and (b) conditional (GARCH-based) yield variances. Including these equations respectively improves (a) the estimation of the factors' physical dynamics (see [Kim and Orphanides \(2012\)](#)) and (b) the consistency of the estimated model with sample moments of order two.<sup>6</sup>

As Japan has been confronted with extremely low interest rates since the mid-90s, the sovereign bond yields of this country constitute a relevant source of data to examine the ability of term-structure models to handle the ZLB.<sup>7</sup> Our estimated model both shows a very good fit of yields and of conditional yield volatilities across maturities. We also find differences between historical and risk-neutral lift-off probabilities. Our model's estimates imply that at the 5-year horizon, the difference between the risk-neutral and historical probabilities of exiting the ZLB can be as large as 35 percentage points.

The present article relates to the small but fast-growing literature that develops and investigates ZLB-consistent models. Three main approaches stand out: shadow-rate models, quadratic term-structure models (QTSM) and models involving square-root (CIR) processes. The shadow-rate model was introduced by [Black \(1995\)](#) and has been adopted by several recent contributions (see e.g. [Ueno, Baba, and Sakurai \(2006\)](#), [Ichiue and Ueno \(2007\)](#), [Ichiue and Ueno \(2013\)](#), [Kim and](#)

---

<sup>6</sup>Among others, [Collin-Dufresne, Goldstein, and Jones \(2002\)](#), [Adrian and Wu \(2009\)](#), [Andersen and Benzoni \(2006\)](#), [Trolle and Schwartz \(2009\)](#), [Jacobs and Karoui \(2009\)](#), [Almeida, Graveline, and Joslin \(2011\)](#), [Bikbov and Chernov \(2011\)](#), [Creal and Wu \(2014\)](#) and [Christensen, Lopez, and Rudebusch \(2014\)](#) study the ability of term-structure models to fit conditional volatilities of yields.

<sup>7</sup>See e.g. [Gorovoi and Linetsky \(2004\)](#), [Ueno, Baba, and Sakurai \(2006\)](#), [Ichiue and Ueno \(2007\)](#), [Kim and Singleton \(2012\)](#), [Christensen and Rudebusch \(2013\)](#), [Kim and Priebsch \(2013\)](#), [Krippner \(2013\)](#).

Singleton (2012), Krippner (2012, 2013), Bauer and Rudebusch (2013), Christensen and Rudebusch (2013), Kim and Priebsch (2013) and Wu and Xia (2013)). In this model, the short-term rate is defined as the maximum between zero and the so-called shadow rate and ZLB periods occur when the latter turns negative. Typically, if the shadow rate follows a Gaussian process, the model can generate prolonged periods of ZLB, making it consistent with features (i) and (iii). However, there are no closed-form formulas available for bond prices (this inadequately addresses point (ii)) and one has to resort to simulation or approximation techniques to estimate the model (see Kim and Priebsch (2013) or Wu and Xia (2013)). By contrast, QTSM and models based on square-root processes provide closed-form bond pricing formulas and positive yields (seminal contributions are those of Ahn, Dittmar, and Gallant (2002), Leippold and Wu (2002), Cox, Ingersoll, and Ross (1985), Pearson and Sun (1994) and Dai and Singleton (2000)). Nevertheless, these models treat the ZLB as a reflecting barrier. In that case, the probability of having an unchanged short-term rate for two subsequent periods is zero, which makes them inconsistent with feature (iii).<sup>8</sup>

The remainder of the paper is organized as follows. Section 2 introduces the non-negative ARG<sub>0</sub> process and highlights its ability to stay at zero. Section 3 presents the associated affine term-structure model and derives tractable lift-off probability formulas. Section 4 describes the estimation strategy and presents the empirical results. Section 5 examines the distributions of future lift-off dates. Section 6 concludes and an Appendix gathers proofs and technical results.

## 2 Non-negative affine processes with zero lower bound spells

In this section we introduce the univariate Gamma-zero distribution and extend it to the dynamic case with a new class of processes that we call Autoregressive Gamma-Zero (see Section 2.1). A multivariate generalization will be considered in Section 3. Like the continuous-time Cox, Ingersoll, and Ross (1985) process – or its discrete-time counterpart, the Autoregressive Gamma process of Gourieroux and Jasiak (2006) – it cannot take negative values. However, the Autoregressive Gamma-zero can reach the zero value with a strictly positive probability, stay at this lower bound for an extended period of time and become positive again. We present its main properties in Section 2.2 and a generalization to the Extended Autoregressive Gamma process is developed in Section 2.3.

### 2.1 The ARG<sub>0</sub> process and the zero lower bound

Let us first recall that a Gamma distribution  $\gamma_\nu(\mu)$  is a positive distribution defined by a shape (or degree of freedom) parameter  $\nu > 0$  and a scale parameter  $\mu > 0$ . Its probability density function (p.d.f.) is given by:

$$f_X(x; \nu, \mu) = \frac{\exp(-x/\mu) x^{\nu-1}}{\Gamma(\nu) \mu^\nu} \mathbb{1}_{\{x>0\}}.$$

---

<sup>8</sup>More precisely, in the case of the CIR process, zero is either a reflecting barrier or an absorbing state [see Karlin and Taylor (1981) and Francis (1989)].

Note that  $\gamma_\nu(\mu)$  converges in distribution to the Dirac distribution at zero when  $\nu$  goes to zero. A non-central Gamma distribution can be defined as an extension of the gamma distribution. Consider a Poisson random variable  $Z$  of positive parameter  $\lambda$ , the non-central Gamma distribution  $\gamma_\nu(\lambda, \mu)$  is a mixture of  $\gamma_{\nu+Z}(\mu)$  distributions ( $Z$  being the mixing variable), defined on the set of strictly positive real numbers (denoted by  $\mathbb{R}^{++}$ ), where  $\nu$ ,  $\lambda$  and  $\mu$  are strictly positive. Remarkably, although its p.d.f. is complicated, its Laplace transform is extremely simple. Indeed, if  $X \sim \gamma_\nu(\lambda, \mu)$ , we have:

$$\varphi_X(u) = \mathbb{E}[\exp(uX)] = \exp \left[ -\nu \log(1 - u\mu) + \lambda \frac{u\mu}{1 - u\mu} \right], \quad \text{for } u < \frac{1}{\mu}.$$

This distribution can be adapted to the case  $\nu = 0$  if  $\gamma_0(\mu)$  is considered as the Dirac distribution at zero. We thus obtain, by definition, a Gamma-zero distribution featuring a point mass at zero.

**Definition 2.1** *Let  $X$  be a non-negative random variable. We say that  $X$  follows a Gamma-zero distribution with parameters  $\lambda > 0$  and  $\mu > 0$ , denoted  $X \sim \gamma_0(\lambda, \mu)$ , if its conditional distribution given  $Z \sim \mathcal{P}(\lambda)$  is:*

$$X | Z \sim \gamma_Z(\mu). \tag{1}$$

The p.d.f. and the Laplace transform of  $X$ , respectively  $f_X(x; \lambda, \mu)$  and  $\varphi_X(u; \lambda, \mu)$ , are given by:

$$\begin{aligned} f_X(x; \lambda, \mu) &= \sum_{z=1}^{+\infty} \left[ \frac{\exp(-x/\mu) x^{z-1}}{(z-1)! \mu^z} \times \frac{\exp(-\lambda) \lambda^z}{z!} \right] \mathbf{1}_{\{x>0\}} + \exp(-\lambda) \mathbf{1}_{\{x=0\}} \tag{2} \\ \varphi_X(u; \lambda, \mu) &= \exp \left[ \lambda \frac{u\mu}{(1-u\mu)} \right] \quad \text{for } u < \frac{1}{\mu}. \end{aligned}$$

(Note that the p.d.f. is with respect to the sum of the Lebesgue measure on  $\mathbb{R}^{++}$  and the unit mass at zero.)

Again, despite the complexity of the density function of Equation (2), the Laplace transform of the Gamma-zero distribution is very easy to manipulate. Also, Equation (2) sheds light on a key feature of the Gamma-zero distribution: it has a point-mass located at  $x = 0$ , and  $\mathbb{P}(X = 0) = \exp(-\lambda)$ . It is extremely easy to simulate in  $\gamma_0(\lambda, \mu)$  by first simulating  $Z$  in  $\mathcal{P}(\lambda)$  and then  $X$  in  $\gamma_z(\mu)$ , where  $z$  is the result of the first simulation. As  $Z$  equals zero with a strictly positive probability,  $X$  also equals zero with a strictly positive probability.<sup>9</sup>

We now turn to the dynamic case, where  $(X_t)$  is a discrete-time random process that we call *Autoregressive Gamma-zero (ARG-Zero)* process, denoted by  $\text{ARG}_0(\alpha, \beta, \mu)$  (where  $\alpha \geq 0$ ,  $\beta \geq 0$ ,  $\mu > 0$ ).

---

<sup>9</sup>Observe also that  $\gamma_0(\lambda, \mu)$  is an infinitely divisible dsintribution given that  $\varphi_X(u; \lambda, \mu) = [\varphi_X(u; \lambda/n, \mu)]^n$  (see Filipovic and Zabczyk (2002)).

**Definition 2.2** *The random process  $(X_t)$  is a  $ARG_0(\alpha, \beta, \mu)$  process of order one if the conditional distribution of  $X_{t+1}$ , given  $\underline{X}_t = (X_t, X_{t-1}, \dots)$ , is the Gamma-zero distribution:*

$$(X_{t+1} | \underline{X}_t) \sim \gamma_0(\alpha + \beta X_t, \mu) \quad \text{for } \alpha \geq 0, \mu > 0, \beta > 0.$$

*The conditional probability density function  $f(X_{t+1} | X_t; \alpha, \beta, \mu)$  and the conditional Laplace transform  $\varphi_{X,t}(u; \alpha, \beta, \mu)$  of the  $ARG_0(\alpha, \beta, \mu)$  process are respectively given by:*

$$\begin{aligned} f(X_{t+1} | X_t; \alpha, \beta, \mu) &= \sum_{z=1}^{+\infty} \left[ \frac{\exp(-X_{t+1}/\mu) X_{t+1}^{z-1}}{(z-1)! \mu^z} \times \frac{\exp[-(\alpha + \beta X_t)] (\alpha + \beta X_t)^z}{z!} \right] \mathbb{1}_{\{X_{t+1} > 0\}} \\ &\quad + \exp(-\alpha - \beta X_t) \mathbb{1}_{\{X_{t+1} = 0\}}; \end{aligned} \quad (3)$$

$$\begin{aligned} \varphi_{X,t}(u; \alpha, \beta, \mu) &:= \mathbb{E} [\exp(u X_{t+1}) | \underline{X}_t] \\ &= \exp \left[ \frac{u\mu}{1 - u\mu} (\alpha + \beta X_t) \right], \quad \text{for } u < \frac{1}{\mu}. \end{aligned} \quad (4)$$

As for the static Gamma-zero distribution, the second element of Equation (3) emphasizes the zero-point mass of the  $ARG_0$  process. The conditional probability of the process reaching zero at date  $t + 1$  is time-varying and given by  $\exp(-\alpha - \beta X_t)$ . Note that there are two main differences between this family of processes and the ARG processes introduced in [Gourieroux and Jasiak \(2006\)](#). First, in our case we take the shape parameter equal to 0, which allows the presence of the zero-point mass. Second, we introduce a positive intercept  $\alpha$  in the Poisson intensity parameter, preventing the zero lower bound from being an absorbing state. Indeed, when  $X_t = 0$ , the value  $X_{t+1}$  equals zero with probability  $\mathbb{P}(X_{t+1} = 0 | X_t = 0) = \exp(-\alpha) < 1$ . (In the multivariate case, this probability will depend on the information available at date  $t$ ).

It is also readily seen from relation (4) that  $(X_t)$  is a discrete-time affine, or  $Car(1)$ , process (see [Darolles, Gourieroux, and Jasiak \(2006\)](#)) since  $\varphi_{X,t}(u; \alpha, \beta, \mu)$  is exponential-affine in  $X_t$ . This class of processes is particularly useful for building term structure models of interest rates, allowing for simple computation of moments and closed-form or tractable pricing formulas. In particular, we use in the next sections the fact that recursive formulas are available for the computation of multi-horizon Laplace transforms defined as:

$$\varphi_{t,h}(u_1, \dots, u_h) = \mathbb{E}_t [\exp(u_1 X_{t+1} + \dots + u_h X_{t+h})].$$

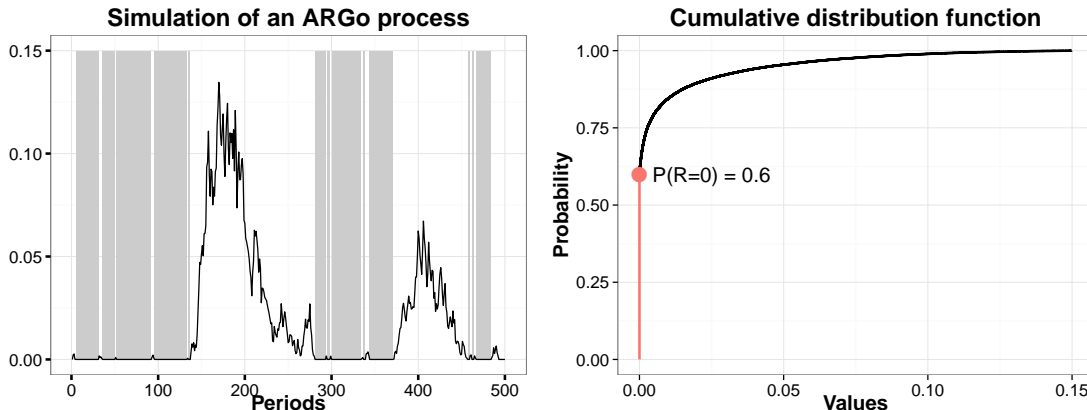
We illustrate the aforementioned properties of the  $ARG_0(\alpha, \beta, \mu)$  process and its relevance for interest rate modeling in a ZLB setting with a simple simulation exercise. Let us denote by  $r_t$  the risk-free rate between  $t$  and  $t + 1$  (known in  $t$ ) and let us assume that its dynamics is given by the following univariate  $ARG_0$  process:

$$(r_t | \underline{r}_{t-1}) \sim \gamma_0(\alpha + \beta r_{t-1}, \mu), \quad (5)$$

where  $\alpha$  and  $\beta$  are positive scalars. A model for the short-term rate dynamics described by relation (5) can accommodate both protracted periods of zero short-term rates and periods of fluctuations. We simulate this process for 500 periods with parameters calibrated as  $\alpha = 0.1$ ,  $\beta = 990$  and  $\mu = 0.001$ . These parameters are such that the marginal mean and standard deviation of process ( $r_t$ ) are about 0.01 and 0.001, respectively. For such parameters, the conditional probability of staying at the zero lower bound is around 0.9. Figure (1) presents the simulated trajectory (left panel) and the computation of the marginal cumulative distribution function (right panel).

As expected, several episodes of prolonged zero lower bound are observed among the 500 simulated values. The grey-shaded areas emphasize the large persistence of the process, which hardly takes off from zero for the first 150 periods. Over the sample, the simulated process hits zero for about 250 periods, that is half of the sample length. The right panel of Figure (1) shows that the unconditional probability of the process to be at zero is 0.6. When it is not at zero, the process experiences persistent spikes of between 100 to 150 periods. This behavior of the  $ARG_0$  process appears particularly appealing to model the dynamics of short-term interest rates in a zero lower bound environment.

Figure 1: Simulation of an  $ARG_0$  process: a short-term rate with zero lower bound spells



*Notes:* This figure displays on the left panel the simulated path of a short-term rate dynamics defined by the following conditional distribution:  $r_t|r_{t-1} \sim \gamma_0(0.1 + 990r_{t-1}, 0.001)$ . The grey zones correspond to periods where the simulated short rate hits zero. On the right panel we have the associated marginal cumulative distribution function.

## 2.2 Moments, stationarity and lift-off probabilities of $ARG_0$ processes

The exponential-affine form of the Laplace transform given in Equation (4) allows for an easy derivation of the properties of  $ARG_0(\alpha, \beta, \mu)$  processes. In this subsection, we show that  $ARG_0$  processes possess simple closed-form formulas for conditional and unconditional moments, stationarity conditions, and especially for calculating conditional probabilities of reaching zero, staying at zero or leaving zero (lift-off).

First, note that the affine property of the  $ARG_0$  process implies that all conditional cumulants are

affine functions of the lagged value of the process. Their derivation is made simple by the use of the log-Laplace transform. Proposition (2.1) and associated corollaries derive the first two conditional and unconditional moments of an  $ARG_0$  process.

**Proposition 2.1** *Let  $(X_t)$  be an  $ARG_0(\alpha, \beta, \mu)$  process. We use the notation  $\rho := \beta\mu$ . The conditional mean  $\mathbb{E}_t(X_{t+1})$  and variance  $\mathbb{V}_t(X_{t+1})$  of  $X_{t+1}$  given its past are respectively given by:*

$$\mathbb{E}_t(X_{t+1}) = \alpha\mu + \rho X_t \quad \text{and} \quad \mathbb{V}_t(X_{t+1}) = 2\mu^2\alpha + 2\mu\rho X_t = 2\mu\mathbb{E}_t(X_{t+1}). \quad (6)$$

**Corollary 2.1.1**  *$(X_t)$  has the following weak AR(1) representation:*

$$X_{t+1} = \alpha\mu + \rho X_t + \varepsilon_{t+1}, \quad (7)$$

where  $(\varepsilon_t)$  is a conditionally heteroskedastic martingale difference, whose conditional variance is  $\mathbb{V}(\varepsilon_{t+1} | \underline{\varepsilon}_t) = 2\mu^2\alpha + 2\mu\rho X_t$ .

**Corollary 2.1.2**  *$(X_t)$  is stationary if and only if  $\rho < 1$  and, in this case, its unconditional mean and variance are respectively given by:*

$$\mathbb{E}(X_t) = \frac{\alpha\mu}{1-\rho} \quad \text{and} \quad \mathbb{V}(X_t) = \frac{2\alpha\mu^2}{(1-\rho)(1-\rho^2)}. \quad (8)$$

**Proof** See Appendix A.1. ■

In particular, from the conditional moments given in Proposition 2.1, we derive simple expressions for a weak AR(1) representation that helps calculating the unconditional first-two moments of the process. Two key features of the  $ARG_0$  are worth noticing. First, the time-varying conditional variance is proportional to the conditional mean and, thus, it linearly shrinks with the level of  $X_t$ . This implies that, in a low-level environment, the  $ARG_0$  process shows low conditional volatility, a typical feature of interest-rates during zero lower bound periods (see Filipovic, Larsson, and Trolle (2014)). Note also that the conditional variance of the  $ARG_0$  process is bounded from below by  $2\mu^2\alpha$  when  $X_t$  reaches zero. Second, the closed-form availability of the first-two conditional and unconditional moments implies that simple estimation procedures can be used such as the generalized method of moments, or pseudo-maximum likelihood techniques.

We concentrate now on conditional probabilities of an  $ARG_0$  process to reach zero, to stay at zero for more than a certain number of periods, or to *lift-off* in exactly  $h$  periods. To investigate this sojourn in state zero and the associated lift-off probability, the following lemma proves useful.

**Lemma 2.1** *Let  $Z$  be a random variable valued in  $\mathbb{R}^+$  and  $\varphi_Z(u)$  is its Laplace transform. Then, we have:*

$$\mathbb{P}_Z\{0\} = \lim_{u \rightarrow -\infty} \varphi_Z(u). \quad (9)$$

**Proof** See Appendix A.2. ■



This Lemma makes the computation of the conditional probabilities of hitting zero very simple. The main formulas are presented in the following proposition.

**Proposition 2.2** *Let  $(X_t)$  be an  $ARG_0(\alpha, \beta, \mu)$  process and let us denote by  $\varphi_{t,h}(u_1, \dots, u_h) = \mathbb{E}_t[\exp(u_1 X_{t+1} + \dots + u_h X_{t+h})]$  its multi-horizon conditional Laplace transform. Then, the following properties hold:*

- (i) 
$$\begin{aligned} \mathbb{P}(X_{t+h} = 0 | X_t) &= \lim_{u \rightarrow -\infty} \varphi_{t,h}(0, \dots, 0, u) \\ &= \exp \left\{ -(1 - \rho) \left[ \frac{\rho^h}{\mu(1 - \rho^h)} X_t + \alpha \sum_{k=0}^{h-1} \frac{\rho^k}{1 - \rho^{k+1}} \right] \right\}, \end{aligned}$$
- (ii) 
$$\begin{aligned} \mathbb{P}[X_{t+1} = 0, \dots, X_{t+h} = 0 | X_t] &= \lim_{u \rightarrow -\infty} \varphi_{t,h}(u, \dots, u) \\ &= \exp(-\alpha h - \beta X_t), \end{aligned}$$
- (iii) 
$$\mathbb{P}[X_{t+1} = 0, \dots, X_{t+h} = 0, X_{t+h+1} > 0 | X_t] = \exp[-\alpha h - \beta X_t] [1 - \exp(-\alpha)].$$

**Proof** See Appendix A.2. ■

**Corollary 2.2.1** *If  $X_t = 0$ , the probability to stay in state 0 for the next  $(h - 1)$  periods **only** is  $(1 - p)p^{h-1}$  with  $p = \exp(-\alpha)$ , and the average sojourn time in zero is given by:*

$$(1 - p) \sum_{h=1}^{+\infty} h p^{h-1} = \frac{1}{1 - p} = [1 - \exp(-\alpha)]^{-1}.$$

When  $\alpha = 0$ , this average sojourn time is  $+\infty$  and the zero lower bound becomes an absorbing state.

Proposition 2.2 is key for calculating lift-off probabilities in economic applications. Corollary 2.2.1 stresses the role of the  $\alpha$  parameter: the average sojourn time in zero is entirely controlled by  $\alpha$  for univariate  $ARG_0$  processes. From an economic point of view, if the short-term interest rate is modeled by an  $ARG_0$  process,  $\alpha$  quantifies the average persistence of zero lower bound regimes.

### 2.3 The Extended $ARG_\nu(\alpha, \beta, \mu)$ process

The  $ARG_0(\alpha, \beta, \mu)$  process described in the previous section and the  $ARG_\nu(\beta, \mu)$  process of [Gourieroux and Jasiak \(2006\)](#) are nested in a general class of Extended  $ARG_\nu(\alpha, \beta, \mu)$  processes characterized by a degree of freedom parameter  $\nu \geq 0$  and a parameter  $\alpha \geq 0$ . Combining the definitions of Sections 2.1 and 2.2, we obtain the following:

**Definition 2.3** *The univariate random process  $(X_t)$  is an Extended  $ARG_\nu(\alpha, \beta, \mu)$  process of order one if the conditional distribution of  $X_{t+1}$ , given  $\underline{X}_t = (X_t, X_{t-1}, \dots)$ , is a non-centered Gamma distribution such that:*

$$(X_{t+1} | X_t) \sim \gamma_\nu(\alpha + \beta X_t, \mu), \quad \text{for } \alpha \geq 0, \nu \geq 0, \mu > 0, \beta > 0.$$

The conditional probability density function  $f(X_{t+1} | X_t; \nu, \alpha, \beta, \mu)$  and the conditional Laplace transform  $\varphi_{X,t}(u; \nu, \alpha, \beta, \mu)$  of the Extended  $ARG_\nu(\alpha, \beta, \mu)$  process are respectively given by:

$$\begin{aligned} f(X_{t+1} | X_t; \nu, \alpha, \beta, \mu) &= \sum_{z=1}^{+\infty} \left[ \frac{\exp(-X_{t+1}/\mu) X_{t+1}^{\nu+z-1}}{\Gamma(\nu+z) \mu^{\nu+z}} \times \frac{\exp[-(\alpha + \beta X_t)] (\alpha + \beta X_t)^z}{z!} \right] \mathbb{1}_{\{X_{t+1} > 0\}} \\ &\quad + \exp(-\alpha - \beta X_t) \mathbb{1}_{\{X_{t+1}=0, \nu=0\}}, \\ \varphi_{X,t}(u; \nu, \alpha, \beta, \mu) &:= \mathbb{E} [\exp(uX_{t+1}) | X_t] \\ &= \exp \left[ \frac{u\mu}{1-u\mu} \beta X_t + \alpha \frac{u\mu}{1-u\mu} - \nu \log(1-u\mu) \right], \quad \text{for } u < \frac{1}{\mu}. \end{aligned} \quad (10)$$

Note that the difference with the  $ARG_0$  process, in terms of conditional Laplace transform, is the additional term  $[-\nu \log(1-u\mu)]$  in the exponential. However, a process with Extended ARG dynamics and  $\nu > 0$  does not experience prolonged periods of zero. In line with Proposition 2.1, and following the same steps as in Appendix A.1, we derive the conditional and unconditional first two moments of an Extended ARG process.

**Proposition 2.3** *Let  $(X_t)$  be an Extended  $ARG_\nu(\alpha, \beta, \mu)$  process and  $\rho := \beta\mu$ . The conditional mean and variance of  $X_{t+1}$  are respectively given by:*

$$\mathbb{E}_t(X_{t+1}) = \mu(\nu + \alpha) + \rho X_t \quad \text{and} \quad \mathbb{V}_t(X_{t+1}) = \mu^2(\nu + 2\alpha) + 2\mu\rho X_t. \quad (11)$$

**Corollary 2.3.1**  *$(X_t)$  is stationary if and only if  $\rho < 1$  and, in this case, its unconditional mean and variance are respectively given by:*

$$\mathbb{E}(X_t) = \frac{(\alpha + \nu)\mu}{1 - \rho} \quad \text{and} \quad \mathbb{V}(X_t) = \frac{2\alpha\mu^2 + \mu^2\nu(1 + \rho)}{(1 - \rho)(1 - \rho^2)}.$$

Setting  $\nu = 0$ , we get the  $ARG_0(\alpha, \beta, \mu)$  family presented in Section 2.1 and, assuming  $\alpha = 0$  with  $\nu > 0$ , we obtain the classical  $ARG_\nu(\beta, \mu)$  family. It is also worth noting from relation (10) that, using the extension to random coefficients models, in particular regime-switching models (see [Gourieroux, Monfort, Pegoraro, and Renne \(2014\)](#)), it would be possible to make the parameters  $\alpha$  and  $\nu$  exogenously random and affine, or linearly dependent on  $X_t$ , while staying in the class of affine processes for the augmented process.

In the following sections, we use the previous univariate distributions to construct our multivariate non-negative affine term-structure model where the state vector is composed (under both the risk-neutral and historical probability) of conditionally independent factors with Gamma-zero and Extended Gamma distributions. This assumption of conditional independence characterizing the so-called Vector Autoregressive Gamma process (VARG, say) makes the zero-coupon bond pricing model specification simple while guaranteeing at the same time enough flexibility to match relevant ZLB-linked interest rates stylized facts (see Section 4).<sup>10</sup>

<sup>10</sup>A general specification of the VARG process with conditional dependence is proposed in [Monfort, Pegoraro, Renne, and Roussellet \(2014\)](#).

### 3 The Non-Negative Affine Term Structure Model

#### 3.1 The VARG risk-neutral state dynamics and the affine yield curve formula

In this section we introduce the multivariate non-negative affine term-structure model (NATSM) by directly specifying the risk-neutral ( $\mathbb{Q}$ ) dynamics of the  $n$ -dimensional latent state vector  $X_t = (X_t^{(1)'}, X_t^{(2)'})'$ , where  $\dim(X_t^{(1)}) = n_1$ ,  $\dim(X_t^{(2)}) = n_2$ , and  $n = n_1 + n_2$ . We also denote by  $r_t$  the unobservable short-term rate between  $t$  and  $t + 1$ , known at date  $t$ . More specifically, we assume that the risk-neutral dynamics of  $X_t$  is a Vector ARG (or VARG) process.

**Assumption 1** *The risk-neutral distribution of  $X_{t+1}$ , conditionally on  $\underline{X}_t$ , is given by the product of the following conditional distributions:*

$$(X_{j,t+1} | \underline{X}_t) \stackrel{\mathbb{Q}}{\sim} \gamma_{\nu_j} \left( \alpha_j^{\mathbb{Q}} + \beta_j^{\mathbb{Q}} X_t, \mu_j^{\mathbb{Q}} \right), \quad j \in \{1, \dots, n\}, \quad (12)$$

where  $\nu_j = 0$  for any  $j \in \{1, \dots, n_1\}$ , while  $\nu_j \geq 0$  if  $j \in \{n_1 + 1, \dots, n\}$ ;  $\alpha_j^{\mathbb{Q}} \geq 0$ ,  $\mu_j^{\mathbb{Q}} > 0$  and  $\beta_j^{\mathbb{Q}}$  is an  $n$ -dimensional vector of positive components.

In other words, conditionally on  $X_t$ , the  $n_1$  components of  $X_{t+1}^{(1)}$  follow independent Gamma-zero distributions, while the  $n_2$  components of  $X_{t+1}^{(2)}$  follow independent Non-central Gamma distributions.

Given the conditional (on  $\underline{X}_t$ ) independence between the scalar components in  $X_{t+1}$ , the risk-neutral conditional Laplace transform of  $X_{t+1}$  given  $X_t$  is immediately obtained:

**Proposition 3.1** *The risk-neutral Laplace transform of  $X_{t+1}$ , conditionally on  $\underline{X}_t$ , is given by:*

$$\varphi_t^{\mathbb{Q}}(u) = \mathbb{E}^{\mathbb{Q}} \left[ \exp \left( \sum_{j=1}^n X_{j,t+1} \right) \middle| \underline{X}_t \right] = \exp \left[ \sum_{j=1}^n a_j^{\mathbb{Q}}(u_j)' X_t + b_j^{\mathbb{Q}}(u_j) \right] \quad (13)$$

where, for any  $j \in \{1, \dots, n\}$ , we have:

$$a_j^{\mathbb{Q}}(u_j) = \frac{u_j \mu_j^{\mathbb{Q}}}{1 - u_j \mu_j^{\mathbb{Q}}} \beta_j^{\mathbb{Q}} \quad \text{and} \quad b_j^{\mathbb{Q}}(u_j) = \frac{u_j \mu_j^{\mathbb{Q}}}{1 - u_j \mu_j^{\mathbb{Q}}} \alpha_j^{\mathbb{Q}} - \nu_j \log(1 - u_j \mu_j^{\mathbb{Q}}). \quad (14)$$

The process  $(X_t)$  is therefore a discrete-time affine ( $Car(1)$ ) process.

**Corollary 3.1.1** *The process  $(X_t)$  is  $\mathbb{Q}$ -stationary if and only if, for all  $j \in \{1, \dots, n\}$ , we have  $\rho_j^{\mathbb{Q}} = \beta_{j,j}^{\mathbb{Q}} \mu_j^{\mathbb{Q}} < 1$ .*

**Proof** See [Monfort, Pegoraro, Renne, and Roussellet \(2014\)](#). ■

**Assumption 2** *The nominal short rate process  $(r_t)$  is given by the linear combination of the first*

$n_1$  components of  $X_t$  only, that is:

$$r_t = \sum_{j=1}^{n_1} \delta_j X_{j,t} = \delta' X_t, \quad (15)$$

where  $\delta = [(\delta_j)_{j=\{1,\dots,n_1\}}, 0_{n_2}]'$  has the first  $n_1$  entries strictly positive, the remaining ones being equal to zero.<sup>11</sup>

It is straightforward to see that the short-term interest rate still possesses the zero-point mass property given that it is a linear combination of conditionally independent variables following Gamma-zero distributions. Besides, observe that a non-zero short rate lower bound is allowed (as, for instance, in [Priebsch \(2013\)](#)) by simply adding  $r_{min} \neq 0$  (say) on the right hand side of Equation (15).

In matrix form, the conditional Laplace transform presented in [Proposition 3.1](#), can be written as:

$$\varphi_t^{\mathbb{Q}}(u) = \exp \left[ \tilde{a}^{\mathbb{Q}}(u)' X_t + \tilde{b}^{\mathbb{Q}}(u) \right],$$

where:

$$\begin{aligned} \tilde{a}^{\mathbb{Q}}(u) &= \beta^{\mathbb{Q}} \left( \frac{u \odot \mu^{\mathbb{Q}}}{1 - u \odot \mu^{\mathbb{Q}}} \right) \\ \tilde{b}^{\mathbb{Q}}(u) &= \alpha^{\mathbb{Q}'} \left( \frac{u \odot \mu^{\mathbb{Q}}}{1 - u \odot \mu^{\mathbb{Q}}} \right) - \nu' \log (1 - u \odot \mu^{\mathbb{Q}}) \\ \mu^{\mathbb{Q}} &= (\mu_1^{\mathbb{Q}}, \dots, \mu_n^{\mathbb{Q}})', \quad \beta^{\mathbb{Q}} = (\beta_1^{\mathbb{Q}}, \dots, \beta_n^{\mathbb{Q}}), \\ \alpha^{\mathbb{Q}} &= (\alpha_1^{\mathbb{Q}}, \dots, \alpha_n^{\mathbb{Q}})', \quad \nu = (0, \dots, 0, \nu_{n_1+1}, \dots, \nu_n)', \end{aligned}$$

and where  $\odot$  denotes the element-by-element product and where, with abuse of notations, the division and log operators work element-by-element when applied to vectors.

Given the exponential-affine form of the risk-neutral conditional Laplace transform of  $(X_t)$ , it is easy to obtain the following explicit zero-coupon bond pricing formula (see [Appendix A.3](#) for a proof):

**Proposition 3.2** *If the  $n$ -dimensional state vector  $(X_t)$  has a risk-neutral dynamics defined by Equation (13) and if the short-term interest rate is defined as in Assumption 2, then the price at date  $t$  of the zero-coupon bond with residual maturity  $h$ , denoted by  $P_t(h)$ , is given by:*

$$P_t(h) = \exp \left( A_h' X_t + B_h \right), \quad (16)$$

<sup>11</sup>Note that  $\delta_j$  and  $\mu_j$  cannot be both identified. In the application, we impose that  $\mu_j^{\mathbb{Q}} = 1$  for all  $j$  to ensure identification constraints.

where  $A_h$  and  $B_h$  satisfy the following recursive equations:

$$\begin{aligned} A_h &= -\delta + \tilde{a}^{\mathbb{Q}}(A_{h-1}) \\ &= -\delta + \beta^{\mathbb{Q}} \left( \frac{A_{h-1} \odot \mu^{\mathbb{Q}}}{1 - A_{h-1} \odot \mu^{\mathbb{Q}}} \right) \end{aligned} \quad (17)$$

$$\begin{aligned} B_h &= B_{h-1} + \tilde{b}^{\mathbb{Q}}(A_{h-1}) \\ &= B_{h-1} + \alpha^{\mathbb{Q}'} \left( \frac{A_{h-1} \odot \mu^{\mathbb{Q}}}{1 - A_{h-1} \odot \mu^{\mathbb{Q}}} \right) - \nu' \log(1 - A_{h-1} \odot \mu^{\mathbb{Q}}), \end{aligned} \quad (18)$$

with starting conditions  $A_0 = 0$  and  $B_0 = 0$ . The date  $t$  continuously-compounded yield associated with a zero-coupon bond maturing in  $h$  periods is therefore given by the following non-negative affine function of  $X_t$ :

$$\begin{aligned} R_t(h) &= \bar{A}'_h X_t + \bar{B}_h, \\ \bar{A}_h &= -\frac{1}{h} A_h, \quad \text{and} \quad \bar{B}_h = -\frac{1}{h} B_h, \quad h \geq 1. \end{aligned} \quad (19)$$

The non-negativeness of our NATSM can be easily established from the usual no-arbitrage formula  $R_t(h) = -\frac{1}{h} \log \mathbb{E}_t^{\mathbb{Q}} [\exp(-r_t - \dots - r_{t-h+1})]$  since the short-term rate is a positive combination of the  $X_{i,t}$ 's which are all positive.

### 3.2 The VARG historical state dynamics

We have defined the risk-neutral dynamics of  $X_t$  in Assumption 1. Let us now determine the historical ( $\mathbb{P}$ ) dynamics of the state vector ( $X_t$ ). For this, we assume that the one-period stochastic discount factor is based on an exponential-affine change of probability measure  $\frac{d\mathbb{P}_{t,t+1}}{d\mathbb{Q}_{t,t+1}} = \exp[\theta' X_{t+1} - \psi_t^{\mathbb{Q}}(\theta)]$ , where  $\psi_t^{\mathbb{Q}}(u) = \log \varphi_t^{\mathbb{Q}}(u)$  denotes the risk-neutral conditional log-Laplace transform of ( $X_t$ ), and  $\theta = (\theta_1, \dots, \theta_n)'$  denotes the  $n$ -dimensional vector of market prices of risk factors. Then, we have:

**Proposition 3.3** *The historical distribution of  $X_{t+1}$ , conditionally on  $\underline{X}_t$ , is given by the product of the conditional distributions:*

$$(X_{j,t+1} | \underline{X}_t) \stackrel{\mathbb{P}}{\sim} \gamma_{\nu_j} \left( \alpha_j^{\mathbb{P}} + \beta_j^{\mathbb{P}'} X_t, \mu_j^{\mathbb{P}} \right), \quad \text{for } j \in \{1, \dots, n\}, \quad (20)$$

where  $\alpha_j^{\mathbb{P}} \geq 0$ ,  $\mu_j^{\mathbb{P}} > 0$ , and  $\beta_j^{\mathbb{P}}$  is an  $n$ -dimensional vector of strictly positive components and the historical Laplace transform of  $X_{t+1}$ , conditionally to  $\underline{X}_t$ , is given by:

$$\varphi_t^{\mathbb{P}}(u) = \exp \left[ \sum_{j=1}^n a_j^{\mathbb{P}}(u_j)' X_t + b_j^{\mathbb{P}}(u_j) \right] \quad (21)$$

where, for any  $j \in \{1, \dots, n\}$ , we have:

$$\begin{aligned}
 a_j^{\mathbb{P}}(u_j) &= \frac{u_j \mu_j^{\mathbb{P}}}{1 - u_j \mu_j^{\mathbb{P}}} \beta_j^{\mathbb{P}} & \text{and} & & b_j^{\mathbb{P}}(u_j) &= \frac{u_j \mu_j^{\mathbb{P}}}{1 - u_j \mu_j^{\mathbb{P}}} \alpha_j^{\mathbb{P}} - \nu_j \log(1 - u_j \mu_j^{\mathbb{P}}), \\
 \text{with } \alpha_j^{\mathbb{P}} &= \frac{\alpha_j^{\mathbb{Q}}}{1 - \theta_j \mu_j^{\mathbb{Q}}}, & \beta_j^{\mathbb{P}} &= \frac{1}{1 - \theta_j \mu_j^{\mathbb{Q}}} \beta_j^{\mathbb{Q}} & \text{and} & & \mu_j^{\mathbb{P}} &= \frac{\mu_j^{\mathbb{Q}}}{1 - \theta_j \mu_j^{\mathbb{Q}}}.
 \end{aligned} \tag{22}$$

**Proof** See Appendix A.4. ■

Note that the  $\nu_j$ 's are the same in the risk-neutral and the historical worlds. In particular, if  $\nu_j = 0$  in the risk-neutral dynamics, it is also true in the historical one, as implied by the fact that the negligible sets must be the same in both conditional distributions in order to guarantee the equivalence of the associated probabilities. In line with the notation adopted in the previous section, this historical conditional Laplace transform can be represented in matrix form:

$$\varphi_t^{\mathbb{P}}(u) = \exp \left[ \tilde{a}^{\mathbb{P}}(u)' X_t + \tilde{b}^{\mathbb{P}}(u) \right]$$

where:

$$\begin{aligned}
 \tilde{a}^{\mathbb{P}}(u) &= \beta^{\mathbb{P}} \left( \frac{u \odot \mu^{\mathbb{P}}}{1 - u \odot \mu^{\mathbb{P}}} \right) \\
 \tilde{b}^{\mathbb{P}}(u) &= \alpha^{\mathbb{P}'} \left( \frac{u \odot \mu^{\mathbb{P}}}{1 - u \odot \mu^{\mathbb{P}}} \right) - \nu' \log(1 - u \odot \mu^{\mathbb{P}}) \\
 \mu^{\mathbb{P}} &= (\mu_1^{\mathbb{P}}, \dots, \mu_n^{\mathbb{P}})', \quad \beta^{\mathbb{P}} = (\beta_1^{\mathbb{P}}, \dots, \beta_n^{\mathbb{P}}), \quad \text{and} \quad \alpha^{\mathbb{P}} = (\alpha_1^{\mathbb{P}}, \dots, \alpha_n^{\mathbb{P}})'.
 \end{aligned}$$

### 3.3 Lift-off Probabilities

Let us move now to the problem of investigating the sojourn in state zero of the short rate process ( $r_t$ ), and the associated lift-off probability. As we have seen in the previous sections, our multivariate non-negative yield curve model has the convenient property of being affine under both the risk-neutral and historical dynamics. Consequently, our model allows to easily compute multi-horizon Laplace transforms in both worlds and, thus, to explicitly calculate lift-off probabilities.

Let us first remember that, given the exponential-affine nature of the conditional historical Laplace transform of ( $X_t$ ) (see relation (23)), its multi-horizon Laplace transform until  $t + k$  is given by:

$$\begin{aligned}
 \varphi_{t,k}^{\mathbb{P}}(u_1, \dots, u_k) &= \mathbb{E}^{\mathbb{P}} \left[ \exp \left( u_1' X_{t+1} + \dots + u_k' X_{t+k} \right) \middle| X_t \right] \\
 &= \exp \left[ \mathcal{A}'_k X_t + \mathcal{B}_k \right]
 \end{aligned} \tag{23}$$

where, for any  $i \in \{1, \dots, k\}$ ,  $u_i$  is an  $n$ -dimensional vector. The  $\mathcal{A}_k$  and  $\mathcal{B}_k$  loadings are obtained

as the final values  $\mathcal{A}_k = \mathcal{A}_k^{(k)}$ ,  $\mathcal{B}_k = \mathcal{B}_k^{(k)}$  of the  $k$ -step recursion:

$$\left\{ \begin{array}{l} \mathcal{A}_0^{(k)} = 0 \quad \text{and} \quad \mathcal{B}_0^{(k)} = 0, \\ \mathcal{A}_i^{(k)} = \tilde{a}^{\mathbb{P}} \left( u_{k+1-i} + \mathcal{A}_{i-1}^{(k)} \right) = \beta^{\mathbb{P}} \left( \frac{\left( u_{k+1-i} + \mathcal{A}_{i-1}^{(k)} \right) \odot \mu^{\mathbb{P}}}{1 - \left( u_{k+1-i} + \mathcal{A}_{i-1}^{(k)} \right) \odot \mu^{\mathbb{P}}} \right), \\ \mathcal{B}_i^{(k)} = \tilde{b}^{\mathbb{P}} \left( u_{k+1-i} + \mathcal{A}_{i-1}^{(k)} \right) + \mathcal{B}_{i-1}^{(k)} \\ = \alpha^{\mathbb{P}'} \left( \frac{\left( u_{k+1-i} + \mathcal{A}_{i-1}^{(k)} \right) \odot \mu^{\mathbb{P}}}{1 - \left( u_{k+1-i} + \mathcal{A}_{i-1}^{(k)} \right) \odot \mu^{\mathbb{P}}} \right) - \nu' \log \left[ 1 - \left( u_{k+1-i} + \mathcal{A}_{i-1}^{(k)} \right) \odot \mu^{\mathbb{P}} \right] + \mathcal{B}_{i-1}^{(k)}. \end{array} \right. \quad (24)$$

**Proof** See Proposition 3 in [Gourieroux, Monfort, Pegoraro, and Renne \(2014\)](#). ■

Given that the yield  $R_t(h)$  is an affine function of  $X_t$ , it is easily seen that, for any  $k$ -dimensional vector  $v$ :

$$\begin{aligned} \varphi_{R,t,k}^{(h)\mathbb{P}}(v) &:= \varphi_{R,t,k}^{(h)\mathbb{P}}(v_1, \dots, v_k) = \mathbb{E} \left[ \exp(v_1 R_{t+1}(h) + \dots + v_k R_{t+k}(h)) \mid X_t \right] \\ &= \varphi_{t,k}^{\mathbb{P}}(v_1 \bar{A}_h, \dots, v_k \bar{A}_h) \exp \left( \bar{B}_h \sum_{j=1}^k v_j \right), \end{aligned} \quad (25)$$

where  $v_1, \dots, v_k$  are the scalar entries composing  $v$ . Therefore, Equation (23) can be used to calculate the yields' multi-horizon conditional Laplace transform. Now, in order to determine lift-off probability formulas, let us introduce the following lemma, generalizing Lemma 2.1 to the multivariate framework.

**Lemma 3.1** *If  $Z$  is an  $n$ -dimensional random variable valued in  $\mathbb{R}_+^n$  and  $\varphi_Z(u_1, \dots, u_n)$  is its Laplace transform, we have:*

$$\mathbb{P}_Z\{0, \dots, 0\} = \lim_{u \rightarrow -\infty} \varphi_Z(u, \dots, u).$$

**Proof** Straightforward generalization of the proof of Lemma 2.1 using the fact that, here,  $Z = 0$  is equivalent to  $\mathbf{e}'Z = 0$  (with  $\mathbf{e} = (1, \dots, 1)'$ ). ■

Then, as far as the lift-off probabilities for the short rate are concerned, we have the following proposition:

**Proposition 3.4** *Let us consider the short rate process  $(r_t)$ . The following properties hold:*

- (i)  $\mathbb{P}[r_{t+k} = 0 \mid X_t] = \lim_{u \rightarrow -\infty} \varphi_{R,t,k}^{(1)\mathbb{P}}(0, \dots, 0, u)$ ;
- (ii)  $\mathbb{P}[r_{t+1} = 0, \dots, r_{t+k} = 0 \mid X_t] = \lim_{u \rightarrow -\infty} \varphi_{R,t,k}^{(1)\mathbb{P}}(u, \dots, u) = p_{r,t,k}$  (say);
- (iii)  $\mathbb{P}[r_{t+1} = 0, \dots, r_{t+k-1} = 0, r_{t+k} > 0 \mid X_t] = p_{r,t,k-1} - p_{r,t,k}$ ,

where  $p_{r,t,0} = 1$ .

The last relation gives the distribution of the first lift-off date. The average sojourn time in state zero is then given by:

$$\sum_{k=1}^{\infty} h (p_{r,t,k-1} - p_{r,t,k}).$$

In the previous proposition we have introduced explicit formulas concerning the probability of lift-off from the zero lower bound for the short rate process. Using the formula for truncated Laplace transform in the case of affine processes (see [Duffie, Pan, and Singleton \(2000\)](#) for details), it is possible to provide some tractable formulas if the zero lower bound is replaced by a positive floor  $\lambda > 0$  (e.g.  $\lambda = 10$  bps). Besides, such formulas are available for interest rates of any maturity. More precisely:

**Proposition 3.5** *Let us consider the yield process  $(R_t(h))$  of maturity  $h$  with the multi-horizon conditional Laplace transform given in Equation (25). The following properties hold:*

$$\begin{aligned} (i) \quad & \tilde{p}_{t,k}^{(h)}(v, \lambda) := \mathbb{P} \left[ v' R_{t+1}^{(t+k)}(h) > \lambda \mid X_t \right] = \frac{1}{2} + \frac{1}{\pi} \int_0^{+\infty} \frac{\text{Im} \left[ \varphi_{R,t,k}^{(h)\mathbb{P}}(i v x) \exp(-i \lambda x) \right]}{x} dx; \\ (ii) \quad & \mathbb{P} [R_{t+k}(h) > \lambda \mid X_t] = \tilde{p}_{t,k}^{(h)}(e_k, \lambda); \\ (iii) \quad & \mathbb{P} \left( \frac{R_{t+k-m+1}(h) + R_{t+k-m+2}(h) + \dots + R_{t+k}(h)}{m} > \lambda \mid X_t \right) = \tilde{p}_{t,k}^{(h)} \left( \frac{1}{m} e_{k-m+1}^{(k)}, \lambda \right), \end{aligned}$$

where  $R_{t+1}^{(t+k)}(h) = (R_{t+1}(h), \dots, R_{t+k}(h))'$  and  $v = (v_1, \dots, v_k)'$ ;  $e_k$  is the  $k^{\text{th}}$  column of the  $(k, k)$  identity matrix and  $e_{k-m+1}^{(k)} = (\underbrace{0, \dots, 0}_{k-m \text{ times}}, \underbrace{1, \dots, 1}_m)'$  denotes here a  $k$ -dimensional vector of zeros for the first  $k - m$  components and ones for the  $m$  others.

Observe that these formulas do not determine the probability that  $t + k$  be the first date (between  $t$  and  $t + k$ ) at which  $R_{t+k}(h) > \lambda$ . Nevertheless, this latter information can always be obtained by simulation.

## 4 Empirical analysis of NATSMs

### 4.1 Data and stylized facts

As in [Kim and Singleton \(2012\)](#) and [Christensen and Rudebusch \(2013\)](#), we concentrate on zero-coupon Japanese Government Bond (JGB) yields. The data are weekly (Fridays) and cover the period from June 16, 1995 to May 30, 2014, with residual maturities of six months and one, two, four, seven and ten years.<sup>12</sup> A graphical representation of the yields is provided on [Figure 2](#) and descriptive statistics are presented in [Table 1](#).

<sup>12</sup>The data are extracted from Bloomberg, the tickers of the time series are F10506M, F10501Y, F10502Y, F10504Y, F10507Y, F10510Y.



[ Insert Figure 2 about here. ]

During the first years of our sample, we observe a large decrease in the yields at all maturities. From 1996 to 2001, the 6-month yield stabilizes around 40bps whereas other maturities continue to decrease until 1999, and experience large fluctuations after. From 2001 to 2006, yields literally enter the zero-lower-bound phase, with the 6-month rate stable at virtually zero.<sup>13</sup> As already noted in [Kim and Singleton \(2012\)](#), during this period, the longer-term yields continued showing large variance. We examine more closely this behavior by computing three different measures of univariate conditional variances. For each yield in the data, we fit a GARCH(1,1) and a EGARCH(1,1) models and extract the associated fitted variances. We also compute a two-month rolling-window variance measure on daily data. All those measures are normalized in the same fashion, taking volatilities expressed in annualized terms. Standard descriptive statistics of those proxies are presented in [Table 1](#), and they are represented in [Figure 3](#) (for the 2-year and 10-year maturities). This figure illustrates that, for a given maturity, the three variance proxies are close to each other. We hence consider them to be coherent and credible proxies of conditional volatilities of interest rates. This proximity is confirmed by [Table 2](#), which presents the correlations between the level of interest rates and the conditional volatility proxies. The correlations between the three volatility proxies exceed 0.75 for all maturities.

[ Insert Figure 3, [Table 1](#) and [Table 2](#) about here. ]

Besides, we observe that the behavior of conditional volatility proxies differs substantially across maturities. For the 2-year maturity ([Figure 3](#), top panel), the conditional volatility proxies drop very close to zero when the 6-month rate hits the zero lower bound in 2001. For the longest maturity, the behavior of the three proxies does not show the same decreasing trend (from 1999 to 2004) as for the 2-year yield, even though they experience large spikes in 1995, 1999, and at the end of 2003.

The previous observations allow us to exhibit three important stylized facts: short-term yields can stay at zero for extended periods of time; longer-term yields show substantial variations even in a ZLB period and proxies of conditional yield variances show different profiles across the maturity spectrum. A well-specified term structure model should be able to replicate these empirical features.

## 4.2 Estimation Strategy

Since our term-structure model is affine, a natural estimation technique is to use the linear Kalman filter (as in [Duan and Simonato \(1999\)](#) and [de Jong \(2000\)](#)). The model can be easily represented in a (linear) state-space form where the measurement equations are the yield formula ([Equation](#)

<sup>13</sup>Between May 2001 and February 2006, the 6-month yield has mean and standard deviations respectively equal to 1.37bps and 1.42bps.

(19)) and the transition equations are given by the factor dynamics. Moreover, the affine nature of yields forecasts and of conditional variances naturally provided by our model opens the way to easily introduce new affine measurement equations while preserving the linear specification of the state-space model.

Our first kind of extra measurement equations relate the 2-year and the 10-year conditional variance proxies to their model-implied counterparts (specifically, we retain the EGARCH-based proxy). Recent papers in the term structure literature have highlighted that the estimations of term-structure models based only on yield in levels fail to satisfyingly replicate fluctuations in conditional volatilities (see for instance [Jacobs and Karoui \(2009\)](#) and [Cieslak and Povala \(2015\)](#)).

Second, we augment our state-space model with measurement equations relating the model-implied yields forecasts and survey-based ones. This approach, introduced by [Kim and Orphanides \(2012\)](#), is aimed at handling the persistence problem affecting the estimation of term-structure models (see [Kozicki and Tinsley \(2001a\)](#), [Kozicki and Tinsley \(2001b\)](#) and [Jardet, Monfort, and Pegoraro \(2013\)](#)). More precisely, we use three- and twelve-month-ahead forecasts of the ten-year yield coming from the Consensus Forecasts by Consensus Economics. The latter forecasts are available only from 1999 onward and at the monthly frequency. This missing-data issue is nevertheless easily handled with the Kalman Filter.

In summary, directly fitting survey-based forecasts and conditional variances of yields help to estimate historical and risk-neutral parameters of the factor  $X_t$ . This contributes to get reliable model-implied measures of long-horizon interest-rate forecasts and of lift-off probabilities under both measures ( $\mathbb{P}$  and  $\mathbb{Q}$ ).

Let us now detail the chosen specifications of the factor dynamics. Preliminary estimations have suggested that the data call for the inclusion of a single factor in the specification of the short-term rate and four factors seem necessary to get a satisfying fit for both the levels and the conditional variances of yields. Accordingly, we present in the following the estimation of the model where  $n_1 = 1$  and  $n_2 = 3$ . However, since there are causal relationships between the four factors, longer-term yields are combinations of both  $X_t^{(1)}$  and  $X_t^{(2)}$ , which allows them to vary even if  $X_t^{(1)}$  and the short rate are equal to zero. Moreover, we set  $\nu$  at zero for all components; that is, the conditional distributions of the four factors are  $\gamma_0$ .

In the following, we formally present our state-space model. Using the multivariate adaptation of Equation (11) and the historical dynamics given in Section 3.2, the transition equations can be

expressed as follows:

$$\begin{aligned} X_{t+1} &= \underbrace{\mu^{\mathbb{P}} \odot \alpha^{\mathbb{P}} + \mu^{\mathbb{P}} \odot \beta^{\mathbb{P}'} X_t}_{\mathbb{E}_t(X_{t+1})} + \underbrace{\left\{ \text{diag} \left[ \mu^{\mathbb{P}} \odot \mu^{\mathbb{P}} \odot \left( 2\alpha^{\mathbb{P}} + 2\beta^{\mathbb{P}'} X_t \right) \right] \right\}^{1/2}}_{V_t(X_{t+1})^{1/2}} \varepsilon_{t+1} \\ &= m + MX_t + \Sigma_t^{1/2} \varepsilon_{t+1}, \end{aligned} \quad (26)$$

where  $(\varepsilon_t)$  is a martingale difference with zero-mean and identity variance-covariance matrix.

The measurement equations describe the relationship between three types of observable variables and their model-implied (affine) counterparts: the JGB yields described previously, the EGARCH(1,1) conditional variance proxies for the two and ten-year maturities as well as the three and twelve months-ahead surveys of professional forecasters of the ten-year yield. Observed variables are replicated by the model up to some measurement errors, that we assume to be mutually independent and serially uncorrelated.

The vector of observed yields is denoted by  $R_t = [R_t(h)]_{h \in H}$ , where  $H = \{26, 52, 104, 208, 364, 520\}$  is the list of available maturities in weeks. Besides,  $V_t = [V_t(h)]_{h \in \{104, 520\}}$  denotes the conditional variance proxies for yield of maturity  $h$ ,  $S_t = [S_t^{(q)}(h)]_{h=520, q \in \{12, 52\}}$  denotes the survey of professional forecasters  $q$ -periods ahead for the  $h$ -maturity yield. The measurement equations for the yields and the survey variables are directly derived from Equation (19):

$$R_t(h) = \bar{B}_h + \bar{A}'_h X_t + \sigma_R \eta_{R,h,t}, \quad h \in H \quad (27)$$

$$\begin{aligned} S_t^{(q)}(h) &= \bar{B}_h + \bar{A}'_h \mathbb{E}_t^{\mathbb{P}}(X_{t+q}) + \sigma_{S,h}^{(q)} \eta_{S,h,t}^{(q)} \\ &= \bar{B}_h + \bar{A}'_h \left( \sum_{i=0}^{q-1} M^i m + M^q X_t \right) + \sigma_{S,h}^{(q)} \eta_{S,h,t}^{(q)}, \quad \text{for } \begin{cases} h = 520, \\ q \in \{13, 52\} \end{cases} \end{aligned} \quad (28)$$

where  $\sigma_R$  is the same for all maturities  $h$ , and  $\eta_{R,h,t}$  and  $\eta_{S,h,t}^{(q)}$  are i.i.d. Gaussian white noises.

Eventually, we introduce measurement equations for the volatility proxies based on the conditional variance-covariance matrix of the latent process  $X_t$ . As already emphasized, the affine property of the VARG distribution implies that the conditional variance-covariance matrix of  $X_{t+1}$  given its own past is affine in  $X_t$ . Specifically, the new measurement equations read:

$$\begin{aligned} V_t(h) &= \bar{A}'_h \left\{ \text{diag} \left[ \mu^{\mathbb{P}} \odot \mu^{\mathbb{P}} \odot \left( \nu + 2\alpha^{\mathbb{P}} + 2\beta^{\mathbb{P}'} X_t \right) \right] \right\} \bar{A}_h + \sigma_{V,h} \eta_{V,h,t} \\ &= (\bar{A}_h \odot \bar{A}_h)' \left[ \mu^{\mathbb{P}} \odot \mu^{\mathbb{P}} \odot \left( \nu + 2\alpha^{\mathbb{P}} + 2\beta^{\mathbb{P}'} X_t \right) \right] + \sigma_{V,h} \eta_{V,h,t}, \quad h \in \{104, 520\}, \end{aligned} \quad (29)$$

where  $\eta_{V,h,t}$  is a i.i.d. Gaussian white noise. We then denote the vector of observable variables by  $Y_t = [R'_t, S'_t, V'_t]'$ . Our vector of observables therefore contains 10 different variables (6 yields,

2 conditional variance proxies and 2 survey-based forecast series). Stacking the transition and measurement equations, we obtain the following state-space model representation:

$$\begin{cases} X_{t+1} &= m + MX_t + \Sigma_t^{1/2} \varepsilon_{t+1} \\ Y_t &= \Gamma_0 + \Gamma_1 X_t + \Omega \eta_t \end{cases} . \quad (30)$$

where  $\eta_t = (\eta'_{R,t}, \eta'_{S,t}, \eta'_{V,t})' \sim \mathcal{IIN}(0, I)$ , and  $\Gamma_0$  and  $\Gamma_1$  are based on Equations (27:29). To estimate the model, we use pseudo-maximum likelihood where an approximation of the likelihood function is derived from the linear Kalman filter. The latter is slightly modified to accommodate the fact that the latent factor  $X_t$  is conditionally heteroskedastic. To do so, we run the Kalman filter replacing the real – intractable – log-likelihood derived from conditional Gamma distributions by that obtained from Gaussian distributions, i.e. we approximate  $\varepsilon_{t+1}$  by a standard Gaussian white noise. The availability of a linear state-space model makes the application of such a procedure very easy. For identification purpose, we impose that  $\mu^{\mathbb{P}} = (1, \dots, 1)'$ . In addition, we take a lower-triangular  $\beta^{\mathbb{P}}$  matrix, which implies that  $X_t^{(2)}$  (that does not directly appear in the short-term interest rate specification) Granger-causes  $X_t^{(1)} \equiv X_{1,t}$  (that does appear in the short-term interest rate specification), but  $X_t^{(1)}$  does not Granger-cause  $X_t^{(2)}$ .

We estimate all risk-neutral parameters and the four market prices of risk in a single step. Historical parameters are then deduced from the estimated parameters. We also estimate the short-term interest rate loading  $\delta_1$  (we have  $r_t = \delta_1 X_t^{(1)}$ ) and the measurement-noise standard deviations of the yields. The computation of the parameters standard errors for a preliminary fully-parameterized specification pointed to the non-statistical significativity of some parameters. The latter were further constrained to zero and we eventually end up estimating 16 parameters in an embedded specification. The estimation results are presented in Table 3.

### 4.3 Cross-sectional fit

Most of the parameter estimates are highly significantly different from zero. We observe that most of the factors are highly persistent under both measures with the autocorrelation parameters  $\rho_i = \mu_i \beta_{i,i}$  being close to one. We present a graphical representation of the filtered factors on Figure 4.

[ Insert Table 3 about here. ]

[ Insert Figure 4 about here. ]

[ Insert Figure 5 about here. ]

The first three factors experience long periods at zero, notably during those periods when the 6-month interest rate is at its lowest level (2001 to 2006). By contrast, Factor 4 experiences large and persistent fluctuations during the whole sample. The interpretation of the factors is facilitated by the analysis of the so-called factor loadings, that describe the affine relationships between the levels and conditional variances of yields on the one hand and the factors on the other one. These loadings are plotted in Figure 5. Panel (a) of this figure shows for instance that the first and fourth factor are particularly important to account for the fluctuations of short-term and long-term yields, respectively. The influence of the second factor is more evenly spread across the yield curve. Moreover, Panel (b) suggests that changes in the second and third factors have more important impacts on the conditional variances of medium- to long-term yields than the first two factors.

We now turn to the empirical performances of the VARG term-structure model. First, we observe a remarkable cross-sectional fit of the JGB yields with the measurement-noise standard deviations of yields being 4bps (see Figure 6).

[ Insert Figure 6 about here. ]

The top panel of Figure 7 presents the fit of the observable conditional variance proxies obtained by our term structure model. This fit is satisfactory, the main periods of volatility spikes being captured by the model. Also, these plots demonstrate the ability of the model to accommodate different patterns in the volatilities of yields across maturities.

[ Insert Figure 7 about here. ]

The second panel of Figure 7 presents the fit obtained on the survey of professional forecasters equations. For both the 3-month and 1-year horizons, model-implied forecasts of the 10-year yield nicely reproduces the behavior of observed surveys. Note that the standard errors on the survey measurement noise are parameterized with values that are commensurate with the disagreement among forecasters, as measured by the average standard deviations of the professional forecasters declarations (10bps).

On the whole, these results show a great flexibility of our VARG term-structure model, being able not only to closely reproduce both the level and the conditional volatility behavior of yields across maturities, but also to provide expectations under the historical measure that are coherent with survey-based forecasts.

## 5 Lift-off probabilities

As described in previous sections, our specifications entail closed-form and semi closed-form formulas for calculating the distribution of future yields. As an application, we compute the model-implied probabilities that the short-term interest rate will remain low for a certain amount of time. We can compute such probabilities under both  $\mathbb{P}$  and  $\mathbb{Q}$  measures. The discrepancies existing between the  $\mathbb{P}$  and  $\mathbb{Q}$  probabilities stem from the risk aversion of investors.

We first consider the time-series behavior of such probabilities in Figure 8. Specifically, this exercise is based on the probability formula of the short-rate hitting zero in  $k$  periods ( $r_{t+k} = 0$ , see Proposition 3.4, (i)). We apply the expression for 2- and 5-year-ahead forecasts (i.e. with  $k = 104$  and  $k = 260$ ) for both the historical and risk-neutral probabilities (resp. grey and black lines of top panels of Figure 8). A second exercise exploits the Duffie, Pan, and Singleton (2000) formula (see 3.5) to calculate the probabilities of the short-term interest rate being below 25 bps 2- and 5-years ahead, also for both measures (bottom panels of Figure 8).

[ Insert Figure 8 about here. ]

Let us focus first on the top-left panel, representing both  $\mathbb{P}_t(r_{t+k} = 0)$  and  $\mathbb{Q}_t(r_{t+k} = 0)$  for a 2-year horizon. Until 1998, both probabilities are small and begin experiencing fluctuations from that date on. In 1998, coherently with the low level of yields and of with the increase in yield volatilities (see Figure 3), we obtain a dramatic increase in the probability to hit the ZLB. After that volatility spike, the probability decreases until 2001. The ZLB period of 2001-2006 corresponds to large increases in both probabilities, reaching levels highest than 75% during 2003. This peak is coherent with a flattening of the yield curve at that date: as short rates stay low and long-term rates begin to drop, agents expect a higher probability of the short rate staying at zero for 2 years on. In some sense, those probabilities are a convenient way to represent information contained in the yield curve. The probability to reach and/or stay at the ZLB increases again at the end of the sample, amid the last financial crisis (from 2009 onwards).

We turn now to the same probabilities for a 5-year horizon (top-right plot of Figure 8). First and unsurprisingly, the probabilities under both measures are on average lower than for the 2-year horizon. Second, the differences between  $\mathbb{P}$  and  $\mathbb{Q}$  probabilities under the two measures are smaller than for the previous horizon. This being said, the difference between  $\mathbb{P}$  and  $\mathbb{Q}$  probabilities are not negligible. In particular, in 2007, in a context of rising short-term rates,  $\mathbb{Q}$  probabilities are twice lower than their physical counterparts. This implies that neglecting the existence of risk premia - i.e. working in the risk-neutral world - results in a substantial underestimation of the persistence of the ZLB regime. In other words, the short-term interest rate is expected to stay at zero for a longer period under the physical measure. Third, it is interesting to note that over the last five

years of data, even if the observed short- to medium-term yields (up to 4 years) are fairly stable (see Figure 6), the probability of the short-rate being at zero in years has substantially grown. This phenomenon is consistent with the decrease in longer-term yields, which points to an increase in the perceived expected length of the low-interest rate environment.

The bottom panels of Figure 8 help confirming the previous results. Since the threshold is now different from 0 (25 bps), we observe higher probability values under both measures. For instance, the historical and risk-neutral probabilities of going below 25 bps at the 2-year maturity (bottom-left tile) are close to 1 in 2003, and fluctuates between 0.5 and 1 during the ZLB period. The divergence between  $\mathbb{P}$  and  $\mathbb{Q}$  probabilities are larger than for the upper two plots: in many instances, the physical probabilities of being in a low-rate environment are two to three times larger than the risk-neutral ones.

[ Insert Figure 9 about here. ]

Figure 9 shows conditional  $\mathbb{P}$  and  $\mathbb{Q}$  probabilities of having low short rates over a richer spectrum of horizons. We consider two dates, the first in late 2007 and the second at the end of the sample (respectively black and grey lines, Figure 9). For each date, the forecast horizon varies between 6 months and 5 years. Our term-structure model generates different profiles of low-rate probabilities with respect to the forecast horizon: for the earlier date, the horizon structure is globally increasing whereas it is hump-shaped for the latest date of the sample. This illustrates the ability of the model to generate flexible expected paths of future short-term interest rates.

## 6 Conclusion

In this paper, we introduce a new class of Affine Term Structure Models able to provide at the same time non-negative yields at any maturity and a short rate which can stay at zero for extended periods of time (the ZLB being a non-absorbing state) while the longer-term rates can still vary. These characteristics are obtained by the introduction of a new univariate non-negative affine process called Autoregressive Gamma-zero and its multivariate affine extension (VARG), involving conditional distributions with zero-point masses. The affine nature of our model allows for a great flexibility at the estimation stage. First, a Kalman-filter-based maximum likelihood approach is allowed. Second, the estimation procedure is easily enhanced by explicitly taking into account relevant information like interest rate survey-based forecasts, conditional yield variance proxies. Third, explicit and quasi-explicit formulas are easily derived for calculating the physical and risk-neutral probabilities of the short-term rate staying at –or close to– zero at different forecast horizons.

We assess the model performances with an application to Japanese government bond yields. Our four-latent-factors VARG term-structure model is able to fit both yield levels and conditional volatilities of yields. We also compute time-varying probabilities of being at the ZLB in the future under the historical and risk-neutral measures. Our results show that the differences between such physical and risk-neutral probabilities can be substantial.



## A Appendix

### A.1 Conditional moments of the $ARG_0(\alpha, \beta, \mu)$ process

The conditional cumulant-generating function is  $\psi_t(u) = \log(\varphi_t(u)) = \frac{u\mu}{1-u\mu}\beta X_t + \frac{u\mu}{1-u\mu}\alpha$ . Deriving this function with respect to  $u$  gives us the conditional expectation and variance of  $X_{t+1}$  given  $X_t$ :

$$\begin{aligned} \frac{d}{du}\psi_t(0) &= \frac{\rho(1-u\mu) + \mu(u\rho)}{(1-u\mu)^2}X_t + \frac{\mu\alpha(1-u\mu) + \mu(u\mu\alpha)}{(1-u\mu)^2} \Big|_{u=0} \\ &= \frac{\rho}{(1-u\mu)^2}X_t + \frac{\mu\alpha}{(1-u\mu)^2} \Big|_{u=0} \\ &= \alpha\mu + \rho X_t \\ \frac{d}{du^2}\psi_t(0) &= \frac{2\mu\rho}{(1-u\mu)^3}X_t + \frac{2\mu^2\alpha}{(1-u\mu)^3} \Big|_{u=0} \\ &= 2\mu^2\alpha + 2\mu\rho X_t \end{aligned}$$

Let us introduce now the following notations:  $m_{1,t} = \mathbb{E}(X_t)$  and  $m_{2,t} = \mathbb{V}(X_t)$ . It easily seen that these unconditional moments are defined by the following system of difference equations:

$$\begin{aligned} m_{1,t} &= \rho m_{1,t-1} + \alpha\mu \\ m_{2,t} &= 2\mu^2\alpha + 2\mu\rho m_{1,t-1} + \rho^2 m_{2,t-1} \end{aligned}$$

that can be represented in matrix form as:

$$\begin{pmatrix} m_{1,t} \\ m_{2,t} \end{pmatrix} = \begin{pmatrix} \rho & 0 \\ 2\mu\rho & \rho^2 \end{pmatrix} \begin{pmatrix} m_{1,t-1} \\ m_{2,t-1} \end{pmatrix} + \begin{pmatrix} \mu\alpha \\ 2\mu^2\alpha \end{pmatrix}. \quad (31)$$

This system admits a stationary solution if and only if  $\rho < 1$ , and it is given by:

$$\begin{pmatrix} m_1 \\ m_2 \end{pmatrix} = \begin{pmatrix} \frac{\alpha\mu}{1-\rho} \\ \frac{2\alpha\mu^2}{(1-\rho)(1-\rho^2)} \end{pmatrix}. \quad (32)$$

$m_1$  and  $m_2$  are therefore the marginal mean and marginal variance of the stationary  $ARG_0(\alpha, \beta, \mu)$  process.

## A.2 Sojourn time and lift-off probability of the $ARG_0(\alpha, \beta, \mu)$ process

**Proof** of Lemma 2.1

$$\varphi_X(u) = \int_{\mathbb{R}^+} \exp(ux) d\mathbb{P}_X(x) = \mathbb{P}_X\{0\} + \int_{x>0} \exp(ux) d\mathbb{P}_X(x)$$

Since  $x > 0$ ,  $\exp(ux) \rightarrow 0$  when  $u \rightarrow -\infty$ , and, using Lebesgue theorem, the integral tends towards 0. ■

**Proof** of Proposition 2.2

(i) Let us consider an  $ARG_0(\alpha, \beta, \mu)$  process  $X_t$  and let us study, first, the limit of:

$$\mathbb{E}[\exp(uX_{t+h}) | X_t] = \exp \left\{ a^{\circ h}(u) X_t + \sum_{k=0}^{h-1} b[a^{\circ k}(u)] \right\},$$

when  $u \rightarrow -\infty$ , in order to calculate  $\mathbb{P}(X_{t+h} = 0 | X_t)$ . It can be shown recursively that:

$$\begin{aligned} a^{\circ h}(u) &= \frac{\rho^h u}{1 - u\mu \left[ \frac{1 - \rho^h}{1 - \rho} \right]} \\ \sum_{k=0}^{h-1} b[a^{\circ k}(u)] &= (1 - \rho)\alpha u\mu \sum_{k=0}^{h-1} \frac{\rho^k}{1 - \rho - u\mu + u\mu\rho^{k+1}}, \end{aligned}$$

and, when  $u \rightarrow -\infty$ , we have:

$$\begin{aligned} \mathbb{P}(X_{t+h} = 0 | X_t) &= \exp \left[ -\frac{\rho^h X_t}{\mu \left( \frac{1 - \rho^h}{1 - \rho} \right)} - (1 - \rho)\alpha \sum_{k=0}^{h-1} \frac{\rho^k}{1 - \rho^{k+1}} \right] \\ &= \exp \left\{ -(1 - \rho) \left[ \frac{\rho^h X_t}{\mu (1 - \rho^h)} + \alpha \sum_{k=0}^{h-1} \frac{\rho^k}{1 - \rho^{k+1}} \right] \right\}, \end{aligned} \tag{33}$$

and the result is proved. ■

(ii) From Definition 2.2 we know that, when  $X_t$  follows an  $ARG_0(\alpha, \beta, \mu)$  process,  $\mathbb{P}(X_{t+1} = 0 | X_t) = \exp(-\alpha - \beta X_t)$ . Then, if we denote by  $f_h(X_t) = \mathbb{P}(X_{t+h} = 0, \dots, X_{t+1} = 0 | X_t)$ , we can always write:

$$\begin{aligned} f_h(X_t) &= \mathbb{P}(X_{t+h} = 0, \dots, X_{t+1} = 0 | X_t) \\ &= \mathbb{P}(X_{t+h} = 0 | X_{t+h-1} = 0, \dots, X_{t+1} = 0; X_t) f_{h-1}(X_t) \\ &= \mathbb{P}(X_{t+h} = 0 | X_{t+h-1} = 0) f_{h-1}(X_t) \end{aligned}$$

and the result is easily proved by recursion. ■

(iii)

$$\begin{aligned}
& \mathbb{P}(X_{t+h} > 0, X_{t+h-1} = 0, \dots, X_{t+1} = 0 | X_t) \\
&= \mathbb{P}(X_{t+h} > 0 | X_{t+h-1} = 0, \dots, X_{t+1} = 0; X_t) \mathbb{P}(X_{t+h-1} = 0, \dots, X_{t+1} = 0 | X_t) \\
&= [1 - \mathbb{P}(X_{t+h} = 0 | X_{t+h-1} = 0)] \exp[-\alpha(h-1) - \beta X_t] \\
&= [1 - \exp(-\alpha)] \exp[-\alpha(h-1) - \beta X_t].
\end{aligned}$$

### A.3 Risk-neutral conditional Laplace transform and yield-to-maturity formula

**Proof** of Proposition 3.2

Given that, from Assumption 2, we have  $r_t = \delta' X_t$ , where the first  $n_1$  components are different from zero and the remaining ones are equal to zero, we can write:

$$\begin{aligned}
P_t(h) &= \exp(A_h + B'_h X_t) = \mathbb{E}_t^{\mathbb{Q}} \left[ \exp(-r_t) \exp(A_{h-1} + B'_{h-1} X_{t+1}) \right] \\
&= \exp(-r_t + A_{h-1}) \mathbb{E}_t^{\mathbb{Q}} \left[ \exp(B'_{h-1} X_{t+1}) \right] \\
&= \exp \left[ A_{h-1} + \sum_{j=1}^n b_j^{\mathbb{Q}}(B_{j,h-1}) + \left( \sum_{j=1}^n a_j^{\mathbb{Q}}(B_{j,h-1}) - \delta \right)' X_t \right]
\end{aligned}$$

and the result follows by identification. ■

### A.4 Historical conditional Laplace transform of the state vector

First of all, the following result holds:

**Proposition A.1** *Let us consider a scalar Extended ARG $_{\nu}(\alpha, \beta, \mu)$  process  $(X_t)$  with conditional log-Laplace transform  $\psi_t(u) = \frac{\rho u}{1-u\mu} X_t + \frac{u\mu}{1-u\mu} \alpha - \nu \log(1-u\mu)$ , with  $\rho = \beta\mu$ . The associated conditional Esscher transform, with parameter  $\theta \in \mathbb{R}$ , generates the family of probability distributions characterized by the following conditional log-Laplace transform:*

$$\psi_t^*(u) = \frac{u\rho^*}{1-u\mu^*} X_t + \frac{u\mu^*}{1-u\mu^*} \alpha^* - \nu \log(1-u\mu^*), \quad (34)$$

which is the log-Laplace transform of an EAR $G_{\nu}(\alpha^*, \beta^*, \mu^*)$  process with

$$\begin{aligned}
\rho^* &= \frac{\rho}{(1-\theta\mu)^2}, \quad \mu^* = \frac{\mu}{1-\theta\mu}, \quad \alpha^* = \frac{\alpha}{1-\theta\mu}, \\
\beta^* &:= \frac{\rho^*}{\mu^*} = \frac{\rho}{\mu(1-\theta\mu)} = \frac{\beta}{1-\theta\mu}.
\end{aligned}$$

**Proof** of Proposition 3.3

If we consider our change of probability measure  $\frac{d\mathbb{P}_{t,t+1}}{d\mathbb{Q}_{t,t+1}} = \exp[\theta' \tilde{X}_{t+1} - \psi_t^{\mathbb{Q}}(\theta)]$ , where  $\mathbb{P}_{t,t+1}$  is

the conditional Esscher transform of  $\mathbb{Q}_{t,t+1}$  associated with  $\theta$ , we have  $\psi_{j,t}^{\mathbb{P}}(u_j) = \psi_{j,t}^{\mathbb{Q}}(u_j + \theta_j) - \psi_{j,t}^{\mathbb{Q}}(\theta_j)$  for any  $j \in \{1, \dots, n\}$ , and applying Proposition A.1, Proposition 3.3 is easily proved. ■

## References

- Adrian, T. and H. Wu (2009). The term structure of inflation expectations. Working Paper Series 2009-362, Federal Reserve Bank of New York.
- Ahn, D.-H., R. F. Dittmar, and A. R. Gallant (2002, March). Quadratic term structure models: Theory and evidence. *Review of Financial Studies* 15(1), 243–288.
- Almeida, C., J. Graveline, and S. Joslin (2011). Do interest rate options contain information about excess returns? *Journal of Econometrics* 164, 35–44.
- Andersen, T. G. and L. Benzoni (2006). Can Bonds Hedge Volatility Risk in the U.S. Treasury Market? A Specification Test for Affine Term Structure Models. Working papers, Carlson School of Management, University of Minnesota.
- Bauer, M. D. and G. D. Rudebusch (2013). Monetary policy expectations at the zero lower bound. Working Paper Series 2013-18, Federal Reserve Bank of San Francisco.
- Bikbov, R. and M. Chernov (2011). Yield curve and volatility: Lessons from eurodollar futures and options. *Journal of Financial Econometrics* 9(1), 66–105.
- Black, F. (1995, December). Interest rates as options. *Journal of Finance* 50(5), 1371–76.
- Christensen, J. H., J. A. Lopez, and G. D. Rudebusch (2014, January). Can spanned term structure factors drive stochastic yield volatility? Working paper series, Federal Reserve Bank of San Francisco.
- Christensen, J. H. and G. D. Rudebusch (2013). Modeling yields at the zero lower bound: are shadow rates the solution? Working Paper Series 2013-39, Federal Reserve Bank of San Francisco.
- Cieslak, A. and P. Povala (2015). Information in the term structure of yield curve volatility. *Journal of Finance* forthcoming.
- Collin-Dufresne, P., R. S. Goldstein, and C. S. Jones (2002, 08). Do Bonds Span the Fixed Income Markets? Theory and Evidence for Unspanned Stochastic Volatility. *Journal of Finance* 57(4), 1685–1730.
- Cox, J. C., J. Ingersoll, Jonathan E, and S. A. Ross (1985, March). A theory of the term structure of interest rates. *Econometrica* 53(2), 385–407.

- 
- Creal, D. and J. C. Wu (2015). Estimation of affine term structure models with spanned or unspanned stochastic volatility. *The Journal of Econometrics* 185(2), 60–81.
- Creal, D. D. and J. C. Wu (2014). Interest Rate Uncertainty and Economic Fluctuations. Working papers, Chicago Booth.
- Dai, Q., A. Le, and K. J. Singleton (2010). Discrete-time affine term structure models with generalized market prices of risk. *Review of Financial Studies* 23(5), 2184–2227.
- Dai, Q. and K. Singleton (2003, July). Term structure dynamics in theory and reality. *Review of Financial Studies* 16(3), 631–678.
- Dai, Q. and K. J. Singleton (2000, October). Specification analysis of affine term structure models. *Journal of Finance* 55(5), 1943–1978.
- Darolles, S., C. Gourieroux, and J. Jasiak (2006, 07). Structural laplace transform and compound autoregressive models. *Journal of Time Series Analysis* 27(4), 477–503.
- de Jong, F. (2000, July). Time series and cross-section information in affine term-structure models. *Journal of Business & Economic Statistics* 18(3), 300–314.
- Diebold, F. X. and G. D. Rudebusch (2013). *Yield Curve Modeling and Forecasting*. Introductory Chapters. Princeton University Press.
- Duan, J.-C. and J. G. Simonato (1999). Estimating exponential-affine term structure models by Kalman filter. *Review of Quantitative Finance and Accounting* 13(2), 111–135.
- Duffee, G. R. (2012). Bond pricing and the macroeconomy. *Hanbook of Economics and Finance* (forthcoming).
- Duffie, D. and R. Kan (1996). A yield-factor model of interest rates. *Mathematical Finance* 6(4), 379–406.
- Duffie, D., J. Pan, and K. Singleton (2000). Transform analysis and asset pricing for affine jump-diffusions. *Econometrica* 68(6), 1343–1376.
- Filipovic, D., M. Larsson, and A. Trolle (2014). Linear-rational term structure models. Working paper, Swiss Finance Institute at EPFL.
- Filipovic, D. and J. Zabczyk (2002). Markovian Term-Structure Models in Discrete Time. *Annals of Applied Probability* 12(2), 710–729.
- Francis, L. (1989). A nonlinear general equilibrium model of the term structure of interest rates. *Journal of Financial Economics* 23, 195–224.
- Gorovoi, V. and V. Linetsky (2004). Black’s model of interest rates as options, eigenfunction expansions and japanese interest rates. *Mathematical Finance* 14(1), 49–78.

- 
- Gourieroux, C. and J. Jasiak (2006). Autoregressive gamma processes. *Journal of Forecasting* 25(2), 129–152.
- Gourieroux, C., A. Monfort, F. Pegoraro, and J.-P. Renne (2014). Regime switching and bond pricing. *Journal of Financial Econometrics* 12(2), 237–277.
- Gurkaynak, R. S. and J. H. Wright (2012, June). Macroeconomics and the Term Structure. *Journal of Economic Literature* 50(2), 331–67.
- Ichiue, H. and Y. Ueno (2007). Equilibrium yield curve and the yield curve in low interest rate environment. Working paper series, Bank of Japan.
- Ichiue, H. and Y. Ueno (2013). Estimating term premia at the zero bound: An analysis of Japanese, u.s., and u.k. yields. Working paper series, Bank of Japan.
- Jacobs, K. and L. Karoui (2009). Conditional volatility in affine term-structure models: Evidence from treasury and swap markets. *Journal of Financial Economics* 91, 288–318.
- Jardet, C., A. Monfort, and F. Pegoraro (2013). No-arbitrage Near-Cointegrated VAR(p) term structure models, term premia and GDP growth. *Journal of Banking & Finance* 37(2), 389–402.
- Karlin, S. and H. M. Taylor (1981). *A Second Course in Stochastic Processes*. Academic Press.
- Kim, D. H. (2008). Zero bound, option-implied pdfs, and term structure models. Finance and Economics Discussion Series 2008-31, Board of Governors of the Federal Reserve System (U.S.).
- Kim, D. H. and A. Orphanides (2012, February). Term structure estimation with survey data on interest rate forecasts. *Journal of Financial and Quantitative Analysis* 47(01), 241–272.
- Kim, D. H. and M. Priebsch (2013). Estimation of multi-factor shadow-rate term structure models. Federal reserve board discussion paper series, Federal Reserve Board.
- Kim, D. H. and K. J. Singleton (2012). Term structure models and the zero bound: An empirical investigation of Japanese yields. *Journal of Econometrics* 170(1), 32–49.
- Kozicki, S. and P. Tinsley (2001a). Shifting Endpoints in the Term Structure of Interest Rates. *Journal of Monetary Economics* 47, 613–652.
- Kozicki, S. and P. Tinsley (2001b). Term Structure Views of Monetary Policy Under Alternative Models of Agent Expectations. *Journal of Economic Dynamics and Control* 25, 149–184.
- Krippner, L. (2012, March). Modifying gaussian term structure models when interest rates are near the zero lower bound. Reserve Bank of New Zealand Discussion Paper Series DP2012/02, Reserve Bank of New Zealand.
- Krippner, L. (2013). Measuring the stance of monetary policy in zero lower bound environments. *Economics Letters* 118(1), 135–138.

- 
- Leippold, M. and L. Wu (2002, June). Asset pricing under the quadratic class. *Journal of Financial and Quantitative Analysis* 37(02), 271–295.
- Monfort, A., F. Pegoraro, J.-P. Renne, and G. Roussellet (2014). Recursive discrete-time affine processes and asset pricing. Technical report, mimeo.
- Pearson, N. D. and T.-S. Sun (1994, September). Exploiting the conditional density in estimating the term structure: An application to the cox, ingersoll, and ross model. *Journal of Finance* 49(4), 1279–1304.
- Piazzesi, M. (2010). Affine term structure models. In *Handbook of Financial Econometrics, Volume 1* (Yacine Ait-Sahalia and Lars Peter Hansen North Holland ed.), Chapter 12, pp. 389–472.
- Pribsch, M. (2013). Computing arbitrage-free yields in multi-factor gaussian shadow-rate term structure models. Federal reserve board discussion paper series, Federal Reserve Board.
- Renne, J.-P. (2012). A model of the euro-area yield curve with discrete policy rates. Working papers, Banque de France.
- Siegel, A. F. (1979, August). The noncentral chi-squared distribution with zero degrees of freedom and testing for uniformity. *Biometrika* 66(2), 381–386.
- Trolle, A. B. and E. S. Schwartz (2009, May). A General Stochastic Volatility Model for the Pricing of Interest Rate Derivatives. *Review of Financial Studies* 22(5), 2007–2057.
- Ueno, Y., N. Baba, and Y. Sakurai (2006). The use of the black model of interest rates as an option for monitoring the jgb market expectations. Working paper series, Bank of Japan.
- Wu, J. C. and F. D. Xia (2013). Measuring the macroeconomic impact of monetary policy at the zero lower bound. Working paper series, Chicago Booth.

Table 1: Mean and standard deviations of yields and volatility proxies

	Maturity	6m	1y	2y	4y	7y	10y
Mean	Yields	0.2142	0.2581	0.3811	0.7091	1.2025	1.5850
	GARCH(1,1)	0.0019	0.0020	0.0029	0.0040	0.0052	0.0047
	EGARCH(1,1)	0.0020	0.0021	0.0030	0.0042	0.0052	0.0046
	rolling-window	0.0022	0.0023	0.0028	0.0040	0.0055	0.0051
Std.	Yields	0.2082	0.2440	0.3347	0.5025	0.6799	0.6576
	GARCH(1,1)	0.0017	0.0019	0.0024	0.0025	0.0024	0.0022
	EGARCH(1,1)	0.0017	0.0019	0.0024	0.0025	0.0021	0.0020
	rolling-window	0.0021	0.0021	0.0023	0.0027	0.0029	0.0026

*Notes:* Yields are expressed in annualized percentage points. GARCH and EGARCH models are computed on weekly data whereas the rolling-window volatility is computed on a 60-day window of daily data and converted to the weekly frequency keeping only Fridays data. Our volatility proxies are the square roots of the estimated conditional variance proxies; they are normalized to make them homogeneous to annualized yields. 'Mean' and 'Std.' respectively present sample means and standard deviations of our proxies.

Table 2: Correlation between rates and volatility proxies

Maturity	6m			1y		
	yield	volatility		yield	volatility	
		GARCH	EGARCH		GARCH	EGARCH
GARCH	0.63	1		0.68	1	
EGARCH	0.68	0.96	1	0.72	0.98	1
Rw	0.58	0.75	0.76	0.65	0.88	0.89
Maturity	2y			4y		
GARCH	0.74	1		0.72	1	
EGARCH	0.78	0.97	1	0.74	0.95	1
Rw	0.70	0.92	0.92	0.67	0.90	0.90
Maturity	7y			10y		
GARCH	0.54	1		0.31	1	
EGARCH	0.56	0.95	1	0.35	0.93	1
Rw	0.60	0.87	0.86	0.54	0.80	0.80

*Notes:* Yields are expressed in annualized percentage points. GARCH and EGARCH models are computed on weekly data whereas the rolling-window volatility is computed on a 60-day window of weekly data.

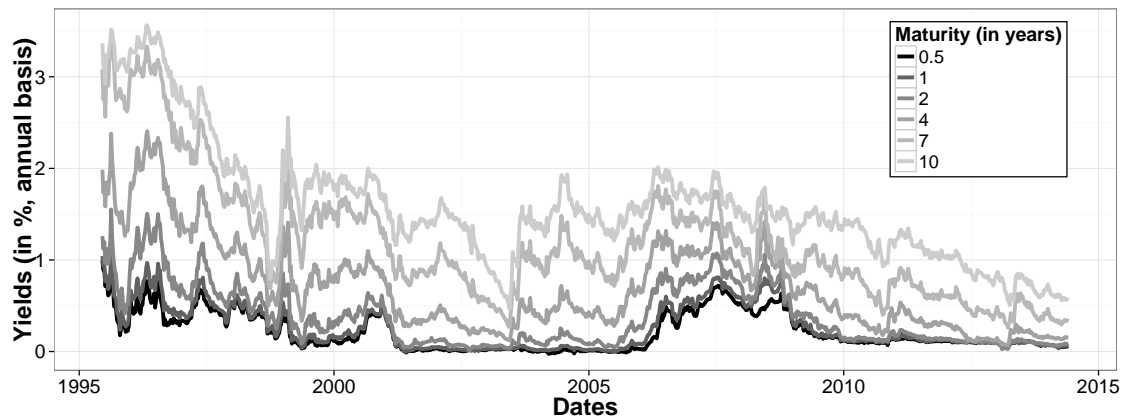


Table 3: Parameter estimates

	P-parameters		Q-parameters		
	Estimates	Std.	Estimates	Std.	
$\alpha_4$	3.2455	0.1118	3.2347	0.1113	
$\beta_{1,1}$	0.9663	0.0078	0.9794	0.0042	
$\beta_{2,2}$	0.9978	0.0005	0.9957	0.0006	
$\beta_{3,3}$	0.9486	0.0044	0.9705	0.0023	
$\beta_{4,4}$	0.9967	0.0005	0.9933	0.0003	
$\beta_{2,1}$	0.0308	0.0041	0.0308	0.0041	
$\beta_{3,2}$	0.1094	0.0059	0.1120	0.0061	
$\beta_{4,3}$	$3.88 \cdot 10^{-4}$	$2.28 \cdot 10^{-5}$	$3.87 \cdot 10^{-4}$	$2.27 \cdot 10^{-5}$	
$\mu_1$	1	–	1.0135	0.0040	
$\mu_2$	1	–	0.9980	0.0005	
$\mu_3$	1	–	1.0231	0.0023	
$\mu_4$	1	–	0.9967	0.0003	
Other Parameters					
$\delta_1$	0.0030	0.0003			
$\theta_1$	-0.0133	0.0039	$\theta_2$	0.0020	0.0005
$\theta_3$	-0.0226	0.0022	$\theta_4$	0.0033	0.0003
$\sigma_R$	0.0407	0.0003			
$\sigma_V$	$3 \cdot 10^{-3}$	–	$\sigma_S$	0.15	–

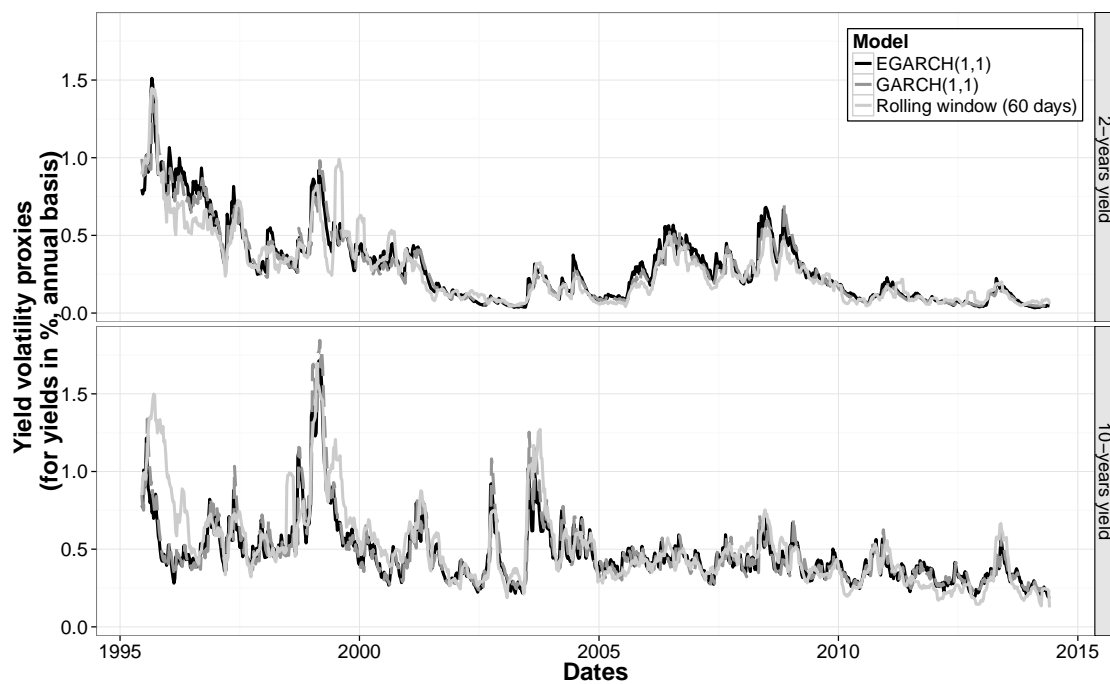
*Note:* This table reports the estimated parameters of a four-factor model where  $r_t = \delta_1 X_{1,t}$ . Standard deviations (Std.) are calculated from the outer product of the log-likelihood gradient at the estimated parameter values. The symbol '–' in the standard-deviation column indicates that the parameter has been calibrated. The  $\sigma_V$ 's are set to twice the standard deviations of the differences between the GARCH, EGARCH and rolling window variance proxies. The  $\sigma_S$ 's are set at the in-sample mean of standard-deviations of forecasts among the professional forecasters.  $\sigma_S$  and  $\sigma_R$  are expressed in percentage points. Last, we impose that the unconditional mean of the short-term interest rate is equal to 100bps.

Figure 2: Japanese yields data



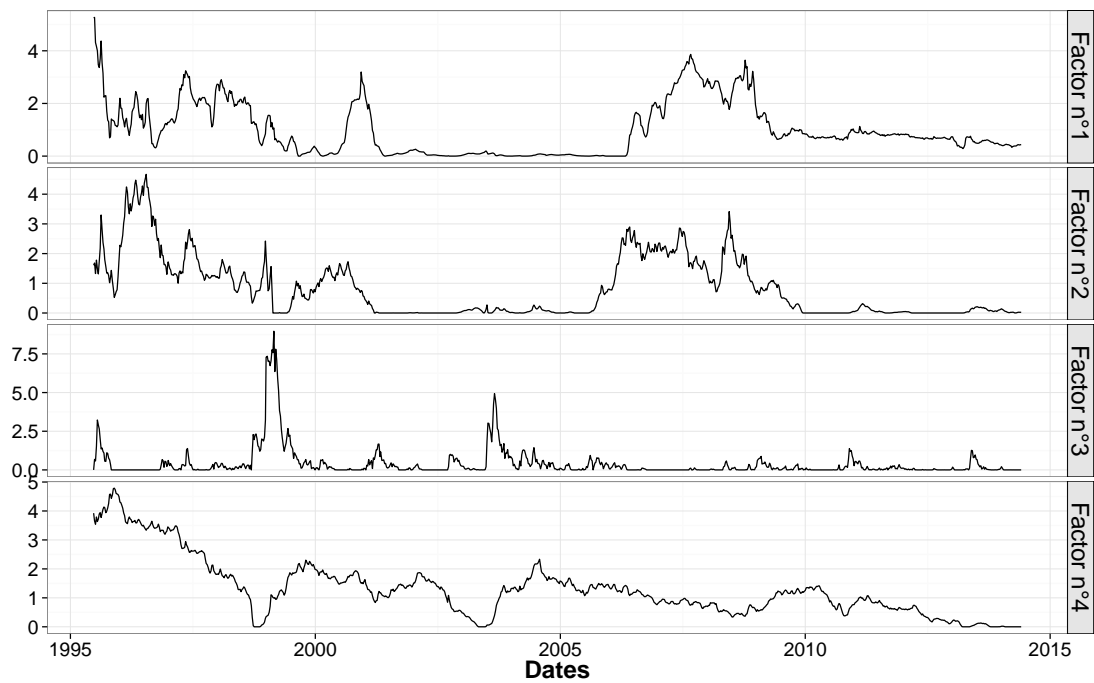
Notes: Yields are weekly data from June 16, 1995 to May 30, 2014. Yields are expressed in annualized percentage points figures, with maturity from 6 months (darkest line) to 10 years (lightest line).

Figure 3: Conditional volatility proxies



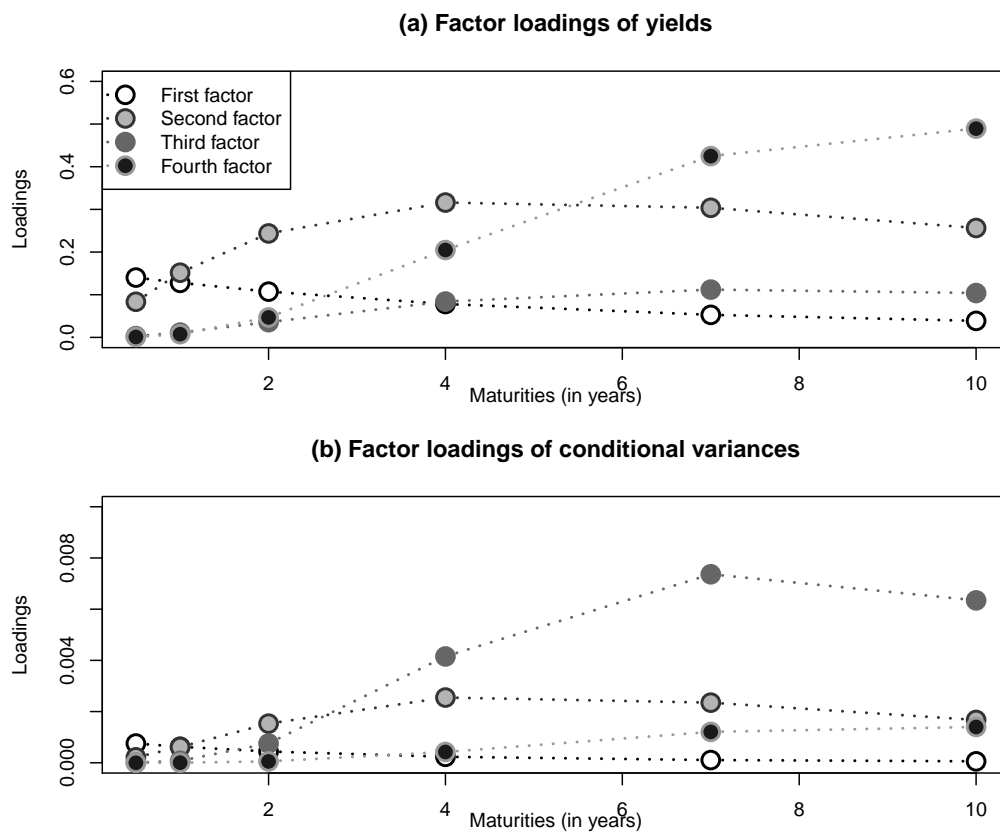
Notes: Top and bottom panels respectively present the volatility proxies for the 2-year and the 10-year yields. GARCH and EGARCH conditional volatility models are computed on weekly yield changes whereas the rolling-window volatility is computed on a 2-month window of daily data. We take the square-root of estimated proxies of conditional variance and obtain our conditional volatility proxies. We normalize them to be comparable to annualized yields. We take estimated proxies and normalize them to be comparable to annualized yields.

Figure 4: Estimated Factors



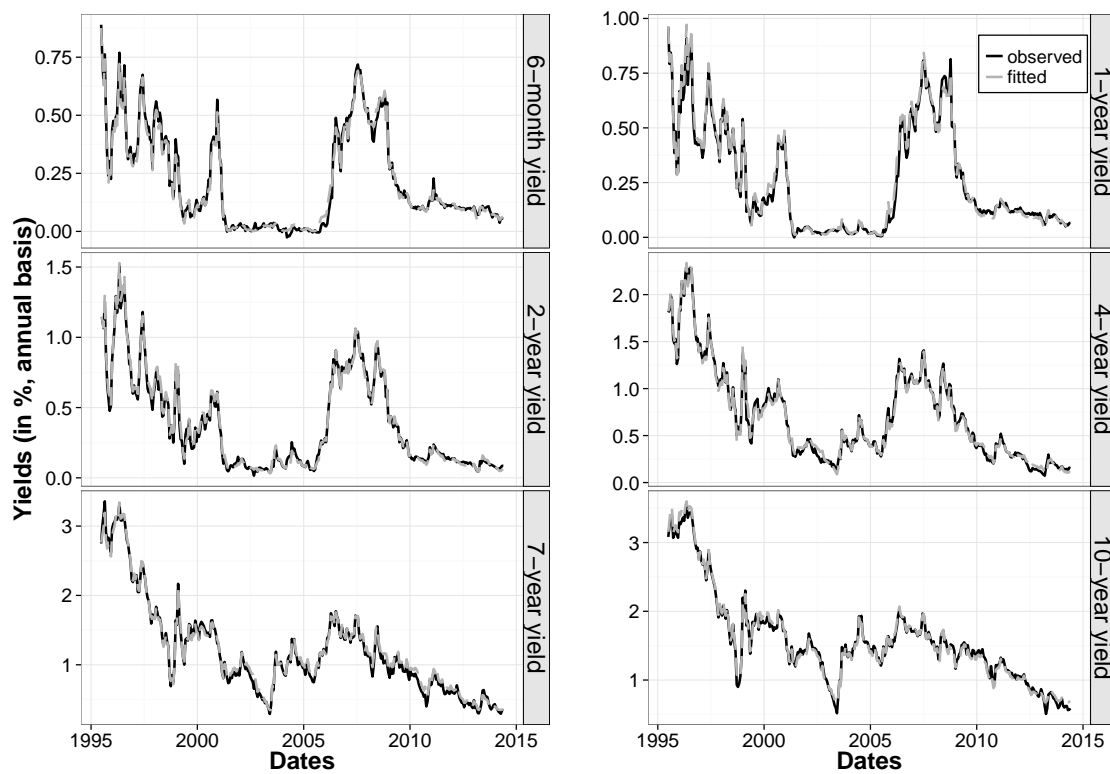
*Notes:* Factors are filtered estimates from the linear Kalman filter on the full sample (June 1995 to May 2014). The short-term rate  $r_t$  is equal to  $\delta_1 X_{1,t}$ . For all  $j > i$  factor  $X_{j,t}$  Granger-causes factor  $X_{i,t}$ .

Figure 5: Estimated Factors



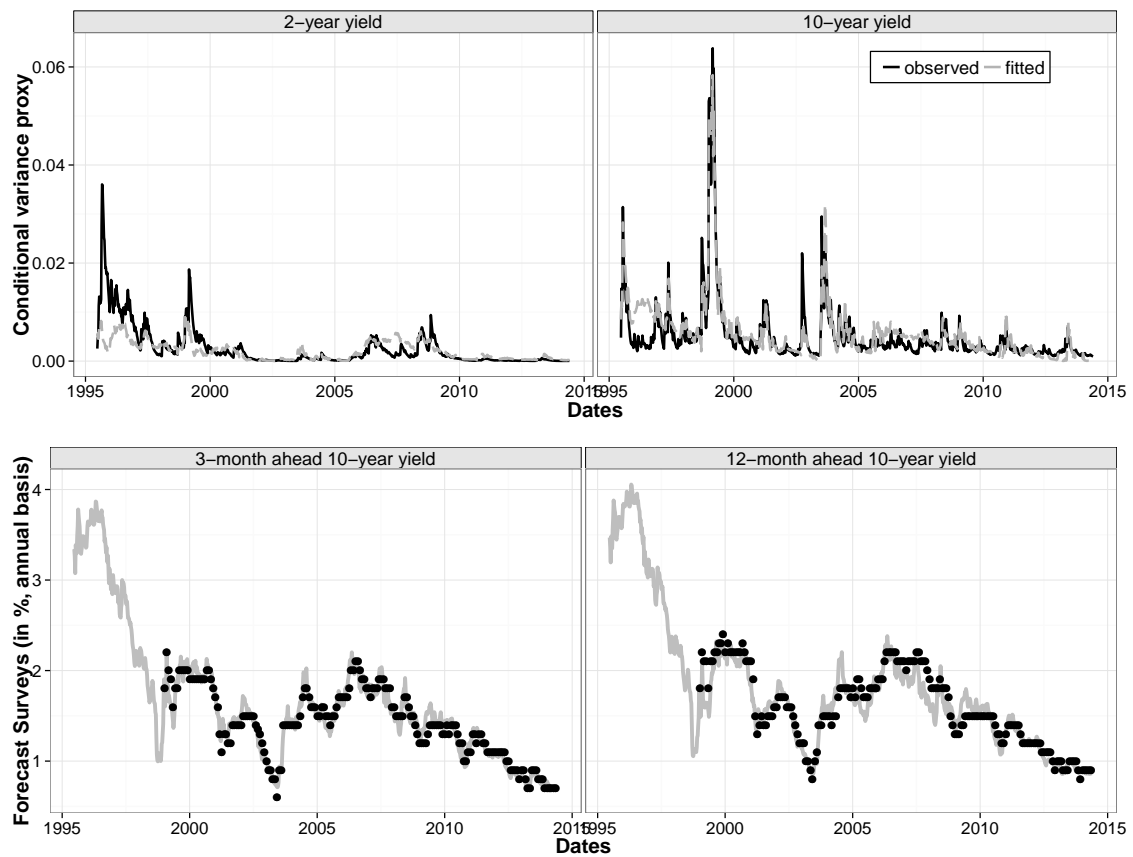
*Notes:* This figure displays the factor loadings of the levels (Panel (a)) and the conditional variances (Panel (b)) of yields. These levels and variances are affine in factors  $X_t$ . Panel (a) relate to Equation (27) and Panel (b) to Equation (29). The loadings are divided by the sample standard deviations of the estimated factors, that is, using the notations of Equation (30), the plotted loadings are of the form  $\Gamma_{1,i,j}/\sqrt{\text{Var}(X_{j,t})}$ , where  $\Gamma_{1,i,j}$  is the entry  $\{i, j\}$  of matrix  $\Gamma_1$ .

Figure 6: Observed and model-implied yields

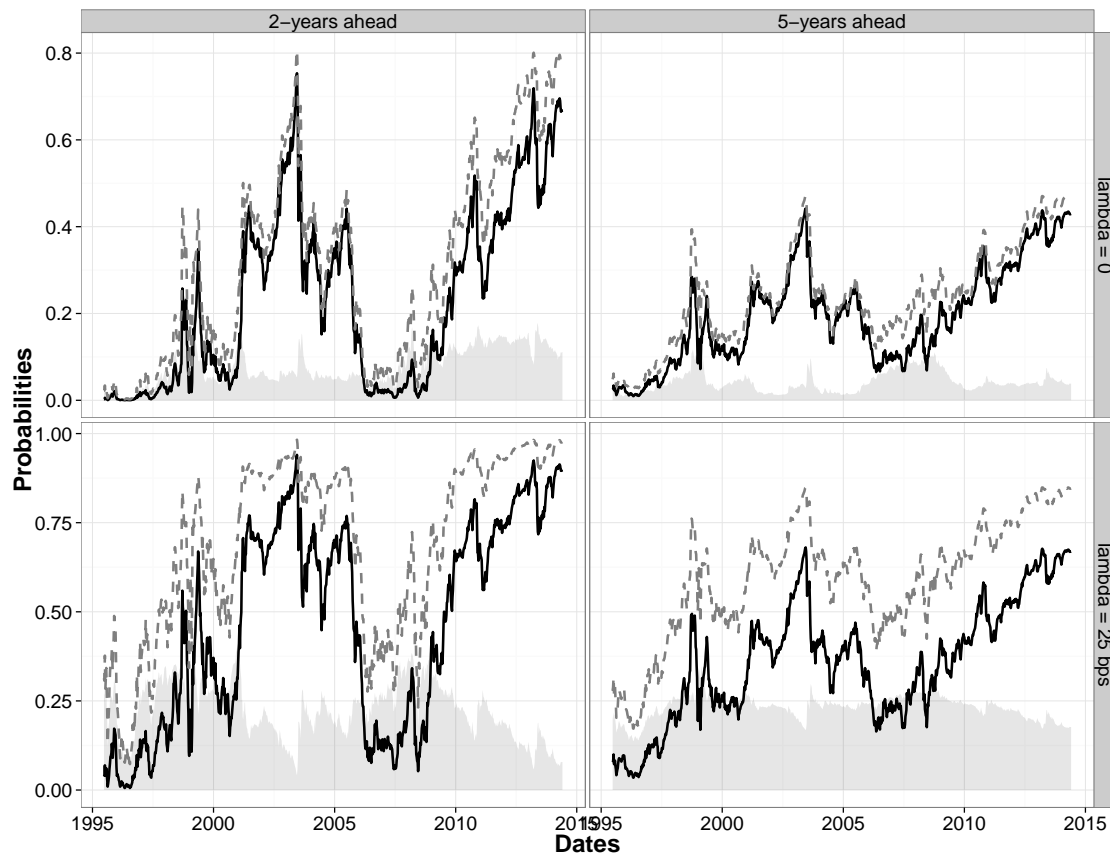


*Notes:* Yields are observed at the weekly frequency from June 16, 1995 to May 30, 2014. Yields are expressed in annualized percentage points, with maturities from 6 months to 10 years. The black solid lines are the observed yields and the grey dashed lines are the model-implied (or fitted) yields using the term structure framework of Section 3 with 4 factors ( $n_1 = 1$  and  $n_2 = 3$ ).

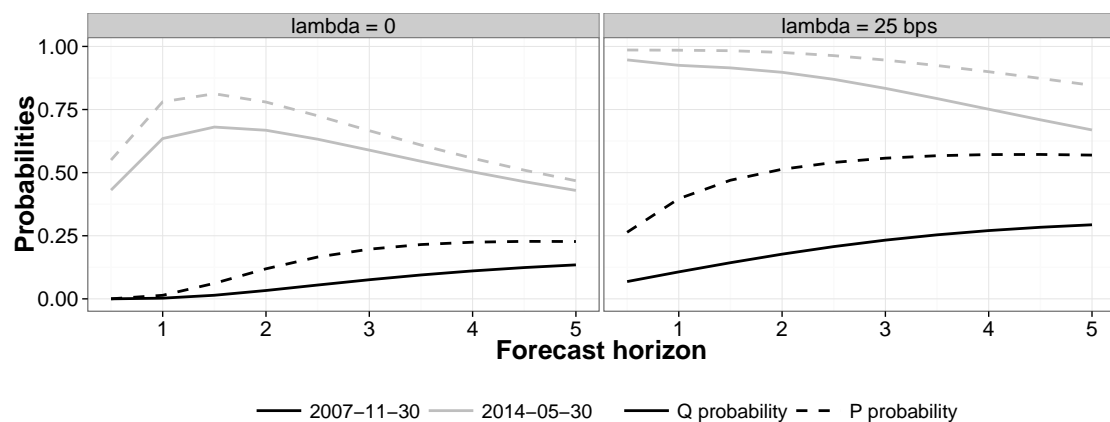
Figure 7: Fitted conditional variance proxies and surveys



*Notes:* The top panel presents the two conditional variance proxies  $V_t(h)$  estimated with an EGARCH(1,1) model on 2- and 10-year yields (left and right tiles) of weekly data from June 30, 1995 to May 30, 2014. The black solid lines are the observed variance proxies and the grey dotted lines are the model-implied (or fitted) equivalent. The bottom panel presents the survey of professional forecasters for the 10-year yield, 3- and 12-months ahead. Survey-based data are available at the monthly frequency from 1999 to the end of the sample. The black dots correspond to the observed data, and the grey solid lines are the fitted equivalent.

Figure 8: Time-series of ZLB probabilities:  $\mathbb{P}_t(r_{t+k} \leq \lambda)$  and  $\mathbb{Q}_t(r_{t+k} \leq \lambda)$ 

Notes: Probabilities are computed with weekly data from June 16, 1995 to May 30, 2014. The top panels present the probabilities of the short-rate hitting zero in two years (top-left panel) and 5 years (top-right panel). On bottom panels, we represent the probabilities of the short-rate being below 25 bps in 2 years (bottom-left panel) and 5 years (bottom-right panel). Black solid lines are the risk-neutral probabilities whereas grey dashed lines are the historical ones; grey-shaded areas are the difference between the two probabilities.

Figure 9: Horizon structure of ZLB probabilities:  $\mathbb{P}_t(r_{t+k} \leq \lambda)$  and  $\mathbb{Q}_t(r_{t+k} \leq \lambda)$ .

Notes: X axis is the horizon  $k$  of the short rate  $r_{t+k}$  being exactly 0 (left tile), or below 25 bps (right tile). Black and grey curves distinguish the date at which these probabilities are evaluated, and respectively correspond to November 30, 2007 and May 30, 2014. Solid and dashed lines represent respectively to  $\mathbb{Q}$  and  $\mathbb{P}$ -probabilities.