Staying at Zero with Affine Processes:
A New Dynamic Term Structure Model

Alain Monfort[*], Fulvio Pegoraro[†],
Jean-Paul Renne[‡], and Guillaume Roussellet[§]

March, 2014

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. These features are allowed for by the introduction of a new univariate non-negative affine process called ARG-Zero, and its multivariate affine counterpart (VARG), entailing conditional distributions with 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 January 1995 to March 2008 with maturities from six months to ten years. Our multivariate (four-dimensional) specification is able to closely match yield levels *and* to capture conditional yield volatilities at any maturity.

JEL Codes: E43, G12.

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

[*]Crest, Banque de France, alain.monfort@ensae.fr
[†]Banque de France, Crest, fulvio.pegoraro@banque-france.fr
[‡]Banque de France, jeanpaul.renne@banque-france.fr
[§]Corresponding author. Banque de France, Crest and Ceremade guillaume.roussellet@banque-france.fr

Particular thanks are addressed to Andrew Siegel for thought-provoking discussions that led to this paper.

1 Introduction

Assuming that storing cash is costless, nominal interest rates cannot turn negative since cash, as an alternative investment to bonds, provides a zero interest rate (see e.g. Cecchetti (1988) or Black (1995)). In other words, the sole existence of currency implies a zero lower bound (ZLB) on bond yields.[1] 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 the short-term 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 term-structure modeling and still represents a challenging task.[2] More generally, to the best of our knowledge, no existing term-structure model is able to simultaneously match the three following characteristics:[3]

(i) consistency with non-negative yields;

(ii) availability of closed-form bond pricing formulas; and

(iii) the ability to accommodate extended periods of zero or close-to-zero short-term rates and to evaluate 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)* at the same time. This process, which we call Autoregressive Gamma Zero, builds on the original ARG process (see Gourieroux and Jasiak (2006), Dai, Le, and Singleton (2010) or Creal and Wu (2013)) by extending it to a zero degree-of-freedom parameter. This process, which we denote ARG_0, possesses a crucial distinctive feature: its conditional distribution given the past encompasses a point-mass at zero.[4] This attractive property allows the non-negative ARG_0 dynamics to satisfy *(i)* and *(iii)*.[5] 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

[1] Note that real interest rates are not constrained by such a lower bound since inflation is not bounded. However, building term-structure models of nominal interest rates is essential given the overwhelming importance of nominal interest-rates instruments (versus inflation-indexed ones) in financial markets. As an illustration, only an average of 9% of government bonds is indexed to inflation in G7 countries (OECD (2012)).

[2] 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)).

[3] 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 $\mathbb{R}^+$.

[4] 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.

[5] 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.

exploit these processes to build a multi-factor term-structure model in which the yields at all maturities are non-negative and 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, Gourieroux, 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 of yields, (b) conditional variances of yields and (c) excess bond returns 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, (b) conditional (GARCH-based) yield variances and (c) expected excess bond-returns predictors *à la* Cochrane and Piazzesi (2005). Including these equations respectively improves (a) the estimation of the factors' physical dynamics (see Kim and Orphanides (2012)), (b) the consistency of the estimated model with sample moments of order two,[6] and (c) the ability of the model to predict excess bond returns.

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.[7] For the sake of comparison, we use the same yield data as in Kim and Singleton (2012). Our estimated model both shows a very good fit and strongly outperforms alternative ZLB-consistent models in capturing conditional yield volatilities across maturities, especially for the medium and long end.

Since we have a historical and risk-neutral modeling, we exploit our estimated model to investigate the behavior of risk premia in a low-yield environment, thereby extending the literature that uses term-structure models to measure the influence of agents' risk aversion on bond prices. Over recent decades, term-structure models have significantly improved our understanding of risk premia.[8] By

[6] Among others, Collin-Dufresne and Goldstein (2002), Trolle and Schwartz (2009), Jacobs and Karoui (2009), Almeida, Graveline, and Joslin (2011), Andersen and Benzoni (2006), 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.

[7] 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).

[8] In particular, these models have contributed to build the compelling evidence gathered against the expectation

definition, these premia are the differences between observed yields and those that would prevail if agents were risk-neutral. Through the specification of a stochastic discount factor, bridging the historical and the risk-neutral dynamics, no-arbitrage term-structure models conveniently reproduce the time variation in risk premia.[9] While numerous papers have studied such premia in a context of far-from-zero yields, a limited number of contributions consider them in a ZLB environment. Arguably, this comes from the lack of tractable models that satisfyingly capture the dynamics of rates in this specific (but now common) context. Our results suggest that, in a ZLB context, risk premia remain substantial and explain an important share of yields' fluctuations. Interestingly, this is not only the case for long-term yields: even for short horizons, risk-neutral distributions of future short-term yields differ from their historical counterparts. In particular, it means that forward rates are biased measures of market expectations of future short-term (policy) rates, even at short maturities. We show that this translates into sizeable differences between historical and risk-neutral distributions of the lift-off date, that is the date at which the central bank exits the ZLB by increasing its policy rate (see e.g. Bauer and Rudebusch (2013) or Swanson and Williams (2013)). For instance we calculate that, at the 2-year horizon, the difference between the risk-neutral and historical probabilities of exiting the ZLB can be as large as 75 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 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 adresses 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)*.[10]

hypothesis. A typical implication of departing from this hypothesis is that forward rates differ from expected future interest rates (see e.g. Fama and Bliss (1987), Campbell and Shiller (1991), Backus, Foresi, Mozumdar, and Wu (2001), Cochrane and Piazzesi (2005), or Faust and Wright (2011)).

[9] See among others, Kim and Wright (2005), Cochrane and Piazzesi (2009), Kim and Orphanides (2012) and Jardet, Monfort, and Pegoraro (2013).

[10] More precisely, in the case of the CIR process, zero is either a reflecting barrier or an absorbing state.

The remainder of the paper is organized as follows. Section 2 introduces the non-negative ARG_0 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 the 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 is a non-negative process. However, the Autoregressive Gamma-Zero can reach the zero value with a strictly positive probability and stay at this lower bound for an extended period of time. 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_0 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:

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

Note that $\gamma_\nu(\mu)$ converges in distribution to the Dirac distribution at zero when ν 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 λ. Then, 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 $\mathbb{R}^+$, where ν, λ and μ are strictly positive. Remarkably, although its p.d.f. is complicated, its Laplace transform is extremely simple. Let $X \sim \gamma_\nu(\lambda,\mu)$, we have:

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

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 positive 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:

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

(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.

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 $\mathrm{ARG}_0(\alpha,\beta,\mu)$ (where $\alpha\geq 0$, $\beta\geq 0$, $\mu>0$).

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 $\underline{X_t}=(X_t,X_{t-1},\ldots)$, is the Gamma-zero distribution:*

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

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:

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

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

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 α in the Poisson 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 $\mathbb{P}\left(X_{t+1}=0 \,|\, X_t=0\right)=\exp(-\alpha)<1$.

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:

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

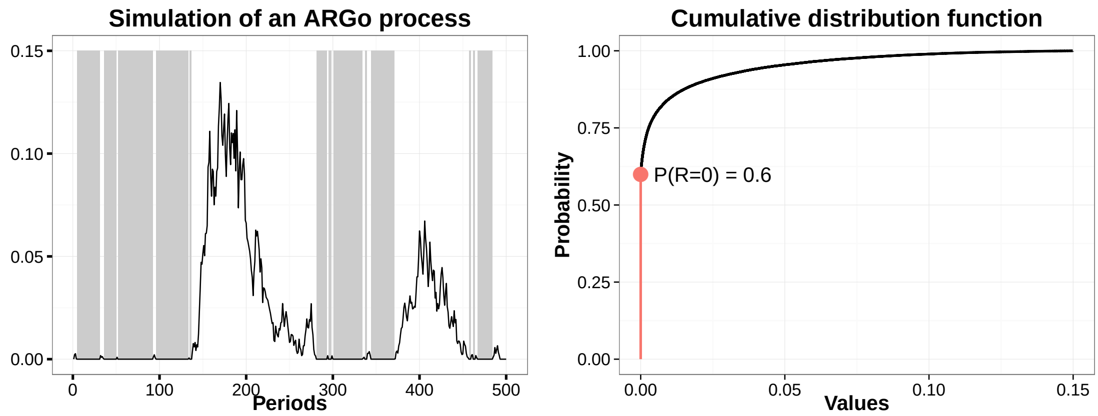
We illustrate the aforementioned properties of the $\mathrm{ARG}_0(\alpha,\beta,\mu)$ process and its relevance for interest rate modeling in a ZLB setting with a simple simulation exercise. Let us denote by r_t the risk-free rate between t and $t+1$ (known in t) and let us assume that its dynamics is given by the following univariate ARG_0 process:

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

where α and β 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 probablities 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 $\text{ARG}_0(\alpha, \beta, \mu)$ 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:*

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

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

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

where (ε_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:*

$$\mathbb{E}(X_t) = \frac{\alpha\mu}{1-\rho} \qquad \textit{and} \qquad \mathbb{V}(X_t) = \frac{2\alpha\mu^2}{(1-\rho)(1-\rho^2)}\,. \tag{8}$$

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 (2013)). Note also that the ARG_0 process' conditional variance is bounded from below by $2\mu^2\alpha$ when X_t reaches zero. Second, the closed-from 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:*

$$\mathbb{P}_Z\{0\} = \lim_{u\to-\infty} \varphi_Z(u)\,. \tag{9}$$

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 $\varphi_{t,h}(u_1,\ldots,u_h) = \mathbb{E}_t[\exp(u_1X_{t+1}+\ldots+u_hX_{t+h})]$ its multi-horizon conditional Laplace transform. Then, the following properties hold:*

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

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:*

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

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 α parameter: the average sojourn time in zero is entirely controlled by α for univariate ARG_0 processes. From an economic point of view, if the short-term interest rate is modeled by an ARG_0 process, α quantifies the average persistence of zero lower bound regimes.

2.3 The Extended $\text{ARG}_\nu(\alpha, \beta, \mu)$ process

The $\text{ARG}_0(\alpha, \beta, \mu)$ process described in the previous section and the $\text{ARG}_\nu(\beta, \mu)$ process of Gourieroux and Jasiak (2006) are nested in a general class of Extended $\text{ARG}_\nu(\alpha, \beta, \mu)$ 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 iif the conditional distribution of X_{t+1}, given $\underline{X_t} = (X_t, X_{t-1}, \ldots)$, is a non-centered Gamma distribution such that:*

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

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:

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

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:*

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

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:*

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

Setting $\nu = 0$, we get the $\mathrm{ARG}_0(\alpha, \beta, \mu)$ family presented in Section 2.1 and, assuming $\alpha = 0$ with $\nu > 0$, we obtain the classical $\mathrm{ARG}_\nu(\beta, \mu)$ 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 (2013)), it would be possible to make the parameters α and ν 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). For the sake of completeness, we also present a general specification of the VARG process in Appendix A.5, where conditional dependence is introduced, while preserving the risk-neutral and historical affine property of the multivariate process.

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 $X_t = \left(X_t^{(1)'}, X_t^{(2)'}\right)'$, where $\dim\left(X_t^{(1)}\right) = n_1$, $\dim\left(X_t^{(2)}\right) = n_2$, 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 *Assuming conditional independence given the past, the risk-neutral distribution of*

X_{t+1}, conditionally on $\underline{X_t}$, is given by the product of the following conditional distributions:

$$(X_{j,t+1} \,|\, \underline{X_t}) \overset{\mathbb{Q}}{\sim} \gamma_{\nu_j}\left(\alpha_j^{\mathbb{Q}} + \beta_j^{\mathbb{Q}'} X_t, \mu_j^{\mathbb{Q}}\right)\,, \;\; j \in \{1,\ldots,n\}\,, \tag{12}$$

where $\nu_j = 0$ for any $j \in \{1,\ldots,n_1\}$, while $\nu_j \geq 0$ if $j \in \{n_1+1,\ldots,n\}$; $\alpha_j^{\mathbb{Q}} \geq 0$, $\mu_j^{\mathbb{Q}} > 0$ 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:*

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

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

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

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

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

Proof See Appendix A.5. ∎

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:*

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

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

These assumptions are aimed at replicating relevant stylized facts that characterize interest rates during ZLB periods.[12] First, it is straightforward to see that the short-term interest rate still

[11] Note that δ_j and μ_j cannot be both identified. In the application, we impose that $\mu_j^{\mathbb{Q}} = 1$ for all j to ensure identification constraints.

[12] 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 relation (15).

possesses the zero-point mass property given that it is a linear combination of conditionally independent variables following Gamma-zero distributions. Second, we introduce more than one factor in the short-term interest rate equation (15) in order to breakdown the exact proportionality between conditional mean and variance of univariate ARG_0 processes (see Proposition 2.1). Indeed, as we consider $n_1 = 2$ short rate factors in our model, it easily seen that the conditional short rate variance-expectation ratio is no more constant but given by the following time-varying convex combination:

$$\frac{\mathbb{V}_t(r_{t+1})}{\mathbb{E}_t(r_{t+1})} = 2\,\mu_1\,\delta_1\,\frac{\delta_1\,\mathbb{E}_t(X_{1,t+1})}{\delta_1\,\mathbb{E}_t(X_{1,t+1}) + \delta_2\,\mathbb{E}_t(X_{2,t+1})} + 2\,\mu_2\,\delta_2\,\frac{\delta_2\,\mathbb{E}_t(X_{2,t+1})}{\delta_1\,\mathbb{E}_t(X_{1,t+1}) + \delta_2\,\mathbb{E}_t(X_{2,t+1})}\,.$$

Assuming $\mu_1\delta_1 < \mu_2\delta_2$, it is easily seen that this ratio fluctuates between $2\mu_1\delta_1$ and $2\mu_2\delta_2$.

In matrix form, the conditional Laplace transform presented in Proposition 3.1, can be written as:

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

where:

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

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

Now, 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 risk-neutral dynamics defined by Equation (13) and 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:*

$$P_t(h) = \exp\left(A_h^{'}\,X_t + B_h\right)\,, \tag{16}$$

where A_h and B_h satisfy the following recursive equations:

$$
\begin{aligned}
A_h &= -\delta + \widetilde{a}^{\mathbb{Q}}(A_{h-1}) \\
&= -\delta + \beta^{\mathbb{Q}} \left(\frac{A_{h-1} \odot \mu^{\mathbb{Q}}}{1 - A_{h-1} \odot \mu^{\mathbb{Q}}} \right) \qquad (17) \\
B_h &= B_{h-1} + \widetilde{b}^{\mathbb{Q}}(A_{h-1}) \\
&= B_{h-1} + \alpha^{\mathbb{Q}\prime} \left(\frac{A_{h-1} \odot \mu^{\mathbb{Q}}}{1 - A_{h-1} \odot \mu^{\mathbb{Q}}} \right) - \nu' \log \left(1 - A_{h-1} \odot \mu^{\mathbb{Q}} \right) , \qquad (18)
\end{aligned}
$$

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:

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

The non-negativeness of our NATSM can be easily established from the usual no-arbitrage formula $R_t(h) = -\frac{1}{h} \log \mathbb{E}_t^{\mathbb{Q}} \left[\exp\left(-r_t - \ldots - r_{t-h+1} \right) \right]$ 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 $\frac{d\mathbb{P}_{t,t+1}}{d\mathbb{Q}_{t,t+1}} = \exp\left[\theta' X_{t+1} - \psi_t^{\mathbb{Q}}(\theta) \right]$, where $\psi_t^{\mathbb{Q}}(u) = \log \varphi_t^{\mathbb{Q}}(u)$ denotes the risk-neutral conditional log-Laplace transform of (X_t), and $\theta = (\theta_1, \ldots, \theta_n)'$ 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:*

$$
(X_{j,t+1} \,|\, \underline{X_t}) \overset{\mathbb{P}}{\sim} \gamma_{\nu_j} \left(\alpha_j^{\mathbb{P}} + \beta_j^{\mathbb{P}\prime} \, X_t, \mu_j^{\mathbb{P}} \right) , \quad \text{for} \quad j \in \{1, \ldots, n\} \, , \qquad (20)
$$

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:

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

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

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

Proof See Appendix A.4. ■

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

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

where:

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

3.3 Lift-off Probabilities

Let us move now to the problem of investigating the sojourn period 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:

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

where, for any $i \in \{1, \ldots, k\}$, 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:

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

Proof See Proposition 3 in Gourieroux, Monfort, Pegoraro, and Renne (2013). ■

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:

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

where $v_1, \ldots, v_k$ 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 $\varphi_Z(u_1, \ldots, u_n)$ is its Laplace transform, we have:*

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

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 $\mathbf{e} = (1, \ldots, 1)'$). ■

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). Then, the following properties hold:*

$$\begin{array}{ll} (i) & \mathbb{P}\left[r_{t+k} = 0 \mid X_t\right] = \lim\limits_{u \to -\infty} \varphi_{R,t,k}^{(1)\mathbb{P}}(0, \ldots, 0, u)\,; \\ (ii) & \mathbb{P}\left[r_{t+1} = 0, \ldots, r_{t+k} = 0 \mid X_t\right] = \lim\limits_{u \to -\infty} \varphi_{R,t,k}^{(1)\mathbb{P}}(u, \ldots, u) = p_{r,t,k} \quad \textit{(say)}\,; \\ (iii) & \mathbb{P}\left[r_{t+1} = 0, \ldots, r_{t+k-1} = 0, r_{t+k} > 0 \mid X_t\right] = p_{r,t,k-1} - p_{r,t,k}\,. \end{array}$$

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:

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

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). Then, the following properties hold:*

$$(i) \qquad \widetilde{p}_{t,k}^{(h)}(v,\lambda) := \mathbb{P}\left[v' R_{t+1}^{(t+k)}(h) > \lambda \,|\, X_t\right] = \frac{1}{2} + \frac{1}{\pi}\int_0^{+\infty} \frac{Im\left[\varphi_{R,t,k}^{(h)\mathbb{P}}(i\,v\,x)\exp(-i\,\lambda\,x)\right]}{x}\,dx\,;$$

$$(ii) \qquad \mathbb{P}\left[R_{t+k}(h) > \lambda \,|\, X_t\right] = \widetilde{p}_{t,k}^{(h)}(e_k,\lambda)\,;$$

$$(iii) \qquad \mathbb{P}\left(\frac{R_{t+k-m+1}(h) + R_{t+k-m+2}(h) + \ldots + R_{t+k}(h)}{m} > \lambda \,|\, X_t\right) = \widetilde{p}_{t,k}^{(h)}\left(\frac{1}{m}e_{k-m+1}^{(k)},\lambda\right),$$

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

Our empirical exercise is based on the same data set as in Kim and Singleton (2012). We concentrate on Japanese Government Bond (JGB) yields, sampled weekly (Fridays) from January 1995 to March 2008, with residual maturities of six months and one, two, four, seven and ten years (see Kim and Singleton (2012)).[13] Taking exactly the same data as Kim and Singleton (2012) makes it possible to compare the performances of our model with the Black (1995)-type shadow-rate models of the zero lower bound. A graphical representation of the yields is provided on Figure 2 and descriptive statistics are presented in Table 1.

[13] We are particularly thankful to Don H. Kim for providing the data.

[Insert Figure 2 about here.]

During the first year in the data, we observe a large drop of 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.[14] As already noted in Kim and Singleton (2012), during this period, the longer-term yields continued showing larger variance. We examine more closely this behavior by computing three different measures of univariate conditional variances. For each maturity in the data, we fit a GARCH(1,1) and a EGARCH(1,1) models and extract the fitted variances. We also compute a sixty-day rolling-window standard-deviation 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).

[Insert Figure 3 and Table 1 about here.]

First, we observe that the behavior of conditional volatility proxies are very distinct across maturities. For instance, 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 any trend, even though they experience large spikes in 1995, 1999, and at the end of 2003. Second, for a given maturity, the three volatility proxies are very 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 presenting the correlations between the level of interest rates and the conditional volatility proxies. The correlations between the three volatility proxies exceed 0.9 for maturities up to four years, and lie between 0.8 and 0.95 for maturities of seven and ten years.

[Insert Table 2 about here.]

Table 2 also presents an important typical stylized fact of ZLB period, concerning interest rates levels and associated conditional volatilities. For maturities up to four years, the correlation between a given conditional yield volatility proxy and the underlying interest rate level is between 0.5 and 0.65 (see column *total* of Table 2). As already noted in Filipovic, Larsson, and Trolle (2013), this linear dependence increases slightly for low yield levels (we calculate the same correlations for yield levels below the median value, see column *low half* of Table 2). Nevertheless, it is also important to highlight that this increasing level-dependence in the conditional volatility, when the associated yield moves towards the zero lower bound, seems not to concern the 7 and 10-year maturities. In other words, as can be seen from Figure 4, the magnitude of the above-mentioned

[14] Between May 2001 and February 2006, the 6-month yield has mean and standard deviations respectively equal to 1.37bps and 1.42bps.

correlations seem to diminish at the long end of the yield curve, even in a ZLB environment.

[Insert Figure 4 about here.]

We take those stylized facts to be a benchmark of key features that should be reproduced by a term structure model. As detailed in the next subsection, to be in line with those stylized facts, we directly incorporate the information given by EGARCH conditional variance proxies for the 2-year and the 10-year maturities in the estimation procedure. Directly fitting the conditional variances of yields of a short and a long maturity will help to estimate historical and risk-neutral dynamics coherent with this behavior. The volatility humps observed on Figure 4 can be generated by our term structure model as long as more than one ARG_0 factor enters the short-term interest rate specification with different loadings (or equivalently with different scaling parameters), thus we impose $n_1 = 2$.

4.2 Estimation Strategy

Since our term-structure model is affine, a natural estimation technique is to use the linear Kalman filter (see e.g. Duan and Simonato (1999) and de Jong (2000)). To that purpose, we formulate the model in state-space form. We take $n_1 = 2$ and $n_2 = 2$, which implies that the short-term interest rate r_t is a linear combination of two factors only. However, since there are causal relationships between the four factors, longer-term yields are combinations of both $X_t^{(1)}$ and $X_t^{(2)}$. Using the multivariate adaptation of Equation (11) and the historical dynamics given in Section 3.2, the transition equations can be expressed as follows:

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

where $\nu = [0'_{n_1}, \nu'_2]'$ is such that the factors entering the short-term interest rate are conditionally independent with a γ_0 distribution, and (ε_t) is a i.i.d. martingale difference with zero-mean and unit variance-covariance matrix.

For the measurement equations, we consider four types of observable variables that are directly used in the estimation procedure: the JGB yields described previously, the EGARCH(1,1) conditional variance proxies for the two and ten-year maturities, the three and twelve months-ahead surveys of professional forecasters of the ten-year yield, and two proxies for the 6-month holding period expected excess-return for the two and ten year yields.[15] We build these two proxies in three

[15]We do not directly incorporate observed excess returns since they are not coherent with the available set of information at date t. Using realized excess returns would introduce autocorrelations in the measurement errors, thus altering the Kalman filter estimation. Indeed, the realized k-holding-period excess return between t and $t+k$ is observable at date $t+k$. We therefore consider only proxies of expected excess returns that are functions of observable variables at time t.

steps, using the same methodology as Cochrane and Piazzesi (2005). First, we construct 6-month holding period observed excess-returns for a total of 19 maturities (from 1 to 10 years) and compute their average across maturities to obtain a single time-series. Second, we regress this average series on both the 2-year spot rates, the 2-year-in-4-year forward rate, and the 2-year-in-8-year forward rate. The fitted values of this regression explain nearly 60% of the excess return average and represent a nice excess return forecasting factor. Last, we regress the 6-month holding period excess return for the two and ten year maturities on this factor (6-month-lagged). The fitted values of these regressions are kept as observable variables as proxies for the expected excess returns.[16]

Considering additional variables to yields in the observation vector greatly increases the economic interpretation of our results, and is allowed by our affine structure. In fact, the excess return proxies and the survey of professional forecasters help to identify more efficiently the parameters under the historical measure (see for instance Kim and Orphanides (2012)).[17] 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 h-maturity yield, and $K_t = [K_t^{(k)}(h)]_{k=26, h\in\{104,520\}}$ denotes the proxies of expected excess returns for the two and ten-year maturities. We authorize all observable variables to be measured with errors. The measurement equations for the yields and the survey variables are directly derived from Equation (19):

$$
\begin{aligned}
R_t(h) &= \overline{B}_h + \overline{A}_h' X_t + \sigma_R \eta_{R,h,t}, \quad h \in H && (27) \\
S_t^{(q)}(h) &= \overline{B}_h + \overline{A}_h' \mathbb{E}_t^{\mathbb{P}}(X_{t+q}) + \sigma_{S,h}^{(q)} \eta_{S,h,t}^{(q)} \\
&= \overline{B}_h + \overline{A}_h' \left(\sum_{i=0}^{q-1} M^i m + M^q X_t \right) + \sigma_{S,h}^{(q)} \eta_{S,h,t}^{(q)}, \quad \text{for} \begin{cases} h = 520, \\ q \in \{13, 52\} \end{cases} && (28)
\end{aligned}
$$

where σ_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. We can derive model-implied excess return predictions in the same fashion:

$$
\begin{aligned}
K_t^{(k)}(h) &= \frac{1}{k}\mathbb{E}_t^{\mathbb{P}}\left[\log\left(\frac{P_{t+k}(h-k)}{P_t(h)}\right)\right] - R_t(k) + \sigma_{K,h}^{(k)}\eta_{K,h,t}^{(k)}, \qquad \text{for} \begin{cases} h \in \{104, 520\} \\ k = 26 \end{cases} \\
&= \frac{1}{k}\left[B_{h-k} + B_k - B_h + (A_k - A_h)' X_t + A_{h-k}' \mathbb{E}_t^{\mathbb{P}}(X_{t+k})\right] + \sigma_{K,h}^{(k)}\eta_{K,h,t}^{(k)} \\
&= \frac{1}{k}\left[B_{h-k} + B_k - B_h + A_{h-k}' \sum_{i=0}^{k-1} M^i m + (A_k + M^{k'} A_{h-k} - A_h)' X_t\right] + \sigma_{K,h}^{(k)}\eta_{K,h,t}^{(k)} && (29)
\end{aligned}
$$

[16] R-squared from these regressions are satisfyingly high, with respective values of 0.69 and 0.48.

[17] Survey-based forecasts are available only from 2003 onward and at the monthly frequency. Therefore, we face a missing-data problem. This issue is nevertheless easily handled with the Kalman Filter.

where $\eta^{(k)}_{K,h,t}$ are i.i.d. Gaussian white noises.

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

$$\begin{aligned} V_t(h) &= \overline{A}'_h \left\{\text{diag}\left[\mu^{\mathbb{P}} \odot \mu^{\mathbb{P}} \odot \left(\nu + 2\alpha^{\mathbb{P}} + 2\beta^{\mathbb{P}'} X_t\right)\right]\right\} \overline{A}_h + \sigma_{V,h}\,\eta_{V,h,t} \\ &= (\overline{A}_h \odot \overline{A}_h)' \left[\mu^{\mathbb{P}} \odot \mu^{\mathbb{P}} \odot \left(\nu + 2\alpha^{\mathbb{P}} + 2\beta^{\mathbb{P}'} X_t\right)\right] + \sigma_{V,h}\,\eta_{V,h,t}\,, \qquad h \in \{104, 520\}\,, \end{aligned} \tag{30}$$

where $\eta_{V,h,t}$ is a i.i.d. Gaussian white noise. We then denote the vector of observable variables by $Y_t = [R'_t, S'_t, V'_t, K_t]'$. Our vector of observables therefore contains 12 different variables (6 yields, 2 conditional variance proxies, 2 survey-based forecast series, and 2 expected excess returns proxies). Stacking the transition and measurement equations, we obtain the following state-space model representation:

$$\left\{ \begin{aligned} X_{t+1} &= m + MX_t + \Sigma_t^{1/2}\varepsilon_{t+1} \\ Y_t &= \Gamma_0 + \Gamma_1 X_t + \Omega\,\eta_t \end{aligned} \right. , \tag{31}$$

with:

$$m = \mu^{\mathbb{P}} \odot (\nu + \alpha^{\mathbb{P}})\,, \quad MX_t = \mu^{\mathbb{P}} \odot \beta^{\mathbb{P}'} X_t \quad \text{and} \quad \Sigma_t = \text{diag}\left[\mu^{\mathbb{P}} \odot \mu^{\mathbb{P}} \odot \left(\nu + 2\alpha^{\mathbb{P}} + 2\beta^{\mathbb{P}'} X_t\right)\right],$$

where $\eta_t \sim \mathcal{IIN}(0, I)$, and Γ_0 and Γ_1 are constructed based on Equations (27:30). To estimate the model, we use pseudo-maximum likelihood of order two with the linear Kalman filter in a slightly modified version since 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 ε_{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 purposes, we impose that $\mu^{\mathbb{Q}} = (1, \ldots, 1)'$. Further, we constrain the latent processes to be stationary under both measures. In addition, we take a triangular superior β' matrix, which implies that $X_t^{(2)}$ (that does not enter directly the short-term interest rate specification) Granger-causes $X_t^{(1)}$ (that does enter directly the short-term interest rate), but $X_t^{(1)}$ does not Granger-cause $X_t^{(2)}$. Also, for parsimony reasons, we impose that $(\alpha_1, \alpha_2) = (0, 0)$.[18]

4.3 Cross-sectional fit

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 two

[18]Preliminary estimations without that constraint suggested that those estimated α's are not statistically different from zero. This does not imply that 0 is an absorbing state as (α_3, α_4) are different from 0.

short-term interest rate loadings δ_1 and δ_2, and the measurement-noise standard deviations of the yields. We end up estimating 21 parameters. The estimation results are presented in Table 3.

[Insert Table 3 about here.]

Most of the estimates are highly significantly different from zero. We observe that most of the factors are highly persistent under both measures with the $\rho_i = \mu_i \beta_{i,i}$ parameters reaching values above 0.96. We present a graphical representation of the filtered factors on Figure 5.

[Insert Figure 5 about here.]

Factors 1 and 2 experience very long periods at zero, during those periods when the 6-month interest rate is at its lowest level (2001 to 2006). Whereas Factor 3 experiences large and persistent fluctuations during the whole sample, Factor 4 possesses large spikes in 1999 and at the end of 2003. During the rest of the sample, this fourth factor is alternating between staying at zero and low spikes.

We now turn to the empirical performances of the VARG term-structure model. First, we observe a very nice cross-sectional fit of the JGB yields with the measurement-noise standard deviations of yields being 9bps (see also Figure 6). This nice performance is confirmed on the first row of Table 4, this indicates that the in-sample RMSEs on yield levels are between 5bps for the one-year maturity, and 12bps for the four-year maturity. This empirical performance is comparable to that obtained by the Quadratic shadow-rate model of Kim and Singleton (2012).

[Insert Figure 6 about here.]

Our model is therefore able to reproduce the behavior of interest-rates with affine processes that are positive and consistent with the zero-lower bound.

4.4 Model-implied yields and volatilities

Looking again at Table 3, we see that the orders of magnitude of the short-term interest rate loadings (δ_1 and δ_2) are very different (as stated previously, such a difference is key to generate the volatility humps that we observe in the data). The top panel of Figure 7 presents the fit obtained by our term structure model for the observable conditional variance proxies V_t.

[Insert Figure 7 about here.]

Again, the fit on both proxies is very good: the fluctuations of the 2-year conditional variance proxy are very closely reproduced, whereas errors on the 10-year conditional variance proxy are slightly larger. This difference is not surprising as the two-year proxy experience fluctuations similar to the yield level, whereas the ten-year proxy tells a somewhat different story. To confirm that our model is able to capture the fluctuations of conditional yield volatilities for all maturities, we follow Kim and Singleton (2012) and run regressions of conditional volatility proxies on the model-implied ones. The regressors are therefore defined by the square-root of the right-hand side of Equation (30) (excluding measurement errors). The results are reported in Table 4.

[Insert Table 4 about here.]

For all maturities, the R^2 of the regressions are above 0.6. The very short- and long-end of the yield curve conditional volatilities shows the lowest R^2 of our regressions, between 0.64 and 0.76, and medium-term maturities (1-y, 2-y, 4-y, 7-y) R^2 are higher than 0.8, respectively 0.8, 0.92, 0.87, and 0.8. We also provide the intercept and slope of the regressions, very close respectively to 0 and 1 for all maturities, to ensure that this good fit does not suffer from a scaling effect.

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 broadly reproduces the behavior of observed surveys. Note that the standard errors on the measurement noise are parameterized with values equal to the disagreement among forecasters (to the average standard deviations of the professional forecasters declarations), and are accordingly set to 13 and 21bps, respectively. These high values are consistent with the slightly poorer capacity of the model to fit the last year of the sample for 12-month ahead forecast of the 10-year yield.

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 for all maturities, but also to provide expectations under the historical measure that are coherent with survey-based forecasts.

5 Bond risk-premia and lift-off probabilities

In this section, we exploit our model to examine risk premia deriving from our estimation. These risk premia stem from the fact that we allow for deviations between the historical and the risk-neutral measures. Their existence implies that excess returns are partially predictable, which we consider in a first part of this section. Second, we show that these risk premia translate into substantial differences between lift-off probabilities under the two measures. In other words, investors appear to be averse to the risk of exiting the ZLB regime.

We regress observed excess bond returns for 6 months holding period on their model-implied predicted equivalent. The intercepts, slopes and R-squared are presented on the last rows of Table

4. As we directly incorporate our excess return proxies in the measurement equations, the model nicely shows R-squared between 0.37 for the ten-year to 0.75 for the one-year excess return. This performance is in line with existing models of the ZLB (see Kim and Singleton (2012)). Also, Table 4 shows that this performance does not hide a scaling effect with intercepts and slopes close to respectively 0 and 1.

This good performance regarding the ability to predict excess bond returns and to fit the surveys of professional forecasters ensures that we can rely on the historical VARG dynamics of the factors with confidence. This is crucial if one wants to use the model to extract agents expectations (under $\mathbb{P}$) regarding future short rates. As described in previous sections, our term-structure model allows for both closed-form and semi closed-form formulas for calculating distribution of future short-rates. As an application, we estimate the probabilities that the short-term interest rate will remain low for a certain amount of time under both measures. Intuitively, the risk-neutral probability can be directly backed out from observable market prices of bonds. As long as the representative investor is risk-averse, this risk-neutral probabilities implicitly incorporates the evolution of the future short-rate (under $\mathbb{P}$) and the term premia. Conversely, the historical probabilities represents the *real-world* future distribution of the short-rate, leaving aside the evolution of term premia. We first consider the time-series behavior of such probabilities in Figure 8.

Our first exercise exploits 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 $k = 26$ and $k = 104$ weeks ahead, for both the historical and risk-neutral probabilities (resp. red and black lines of top panels of Figure 8). A second exercise exploits the Duffie, Pan, and Singleton (2000) formula to calculate the probabilities of the short-term interest rate being below 10bps for $k = 26$ and $k = 104$ weeks 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 $k = 26$ weeks ahead. Until 1999, both probabilities are virtually 0 and begin experiencing fluctuations from that date on. The ZLB period of 2001-2006 corresponds to large increases in both probabilities, reaching a highest of nearly 70% 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 6 months on. In some sense, those probabilities are a convenient way to represent information contained in the yield curve. Note also that probabilities under both measures are not very different from each other in that case, with a difference being below 5 percentage points during the whole period. This implies that for relatively short horizons, the probability of hitting the ZLB is not very different across the two measures, and term premia do not play an important role. At the end date of the sample, agents do not anticipate a period of short-rates hitting zero in the next 6 months as the interest rates began to increase since 2007.

We turn now to the same probabilities for a two-year horizon (top-right plot of Figure 8). Unsurprisingly, the probabilities under both measures are on average lower than at the 6-month horizon, to a highest of 40% in 2003 in the historical world. It is worth noticing two fundamental differences with the previous case. First the differences between the probabilities under the two measures can reach nearly 20 percentage points. Second, at the end of the sample, we observe that the $\mathbb{P}$-probability increases to 20% whereas the $\mathbb{Q}$ probability stays close to zero. Consequently, even if the observed yields are fairly stable during this period the probability of the short-rate hitting zero can grow subsequently. This phenomenon is also coherent with the out-of-sample evolution of yields from 2008 on: the 6-month maturity rate continues to grow until October 2008 to 60bps, i.e. 7 months after the end of our estimation sample, then begins to drop to stabilize around 10bps in January 2010 for a prolonged period, that is nearly 2 years after the end of our sample.

The bottom panels of Figure 8 help confirming the previous results. Since the threshold is now different from 0 (10bps), we observe higher probability values under both measures. For instance, the historical and risk-neutral probabilities of going below 10bps at the 6-month maturity (bottom-left tile) is nearly equal to 1 in 2003, and whipsaws between 0.75 and 1 during the ZLB period. We also observe the same divergence pattern between probabilities under the two measures at the end of the sample, even for the 6-month horizon. The model consequently predicts the short-rate being very low with 25% chance 6-month after the end of the sample, but not 0. This information complements the first calculation stating that even if monetary policy has escaped zero, there is still a large possibility to remain in a very low yield environment. This observation is even more obvious at the two-year horizon (bottom-right panel). Though the $\mathbb{Q}$-probability possesses roughly the same pattern as its counterpart for $\lambda = 0$, the $\mathbb{P}$-probability is globally increasing during the whole sample to reach 0.8 at the last date. Again, this evaluation is coherent with the out-of-sample observations. Strikingly, the divergence between $\mathbb{P}$ and $\mathbb{Q}$ probabilities goes up to 75 percentage points at the 2-year maturity. This spread is of utmost importance since considering only observed yields – i.e. the $\mathbb{Q}$-dynamics – for calculating lift-off dates would result in a large underestimation. Under the physical measure, the short-term interest rate is expected to stay at zero for a longer period. It seems that those differences grow with the threshold value λ and with the forecast horizon k. We reinforce this statement considering the probabilities with respect to the forecast horizon k on Figure 9).

[Insert Figure 9 about here.]

Again, we calculate $\mathbb{P}$ and $\mathbb{Q}$ probabilities for the short rate, for a λ threshold of 0 and 10bps. We evaluate those probabilities at mid-June 2003 and at the end of the sample (respectively black and red lines, Figure 9). The forecast horizon varies between 6 months to 5 years. Our term-structure model is able to generate different profiles of probabilities with respect to the forecast horizon: for the first date, the horizon structure is globally increasing whereas it is hump-shaped at the end of the sample. This particularly implies that the model possesses a significant flexibility in generating future paths of the short-term interest rate. As noted on the time series of probabilities in Figure 8, the differences between probabilities under the two measures tend to increase with the forecast horizon, although it is nearly negligible for low k. This pattern is particularly remarkable on the right plot of Figure 9, where the threshold is equal to 10bps. For mid 2003, the probabilities are both very close to one under the two measures. However, at the two and four years horizon, the $\mathbb{P}$-probabilities exceed $\mathbb{Q}$-probabilities of respectively around 15p.p. and 30p.p.. These differences tend however to be lower for the zero-threshold case, being comprised between 0 and 25 percentage points.

6 Conclusion

In this paper, we introduce the first Affine Term Structure Model able to provide at the same time non-negative yields at any maturity and a short rate staying at zero for extended periods of time (the ZLB being a non-absorbing state). 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 and expected excess return 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 zero at different forecast horizon.

We assess our model performances with an application to Japanese government bond yields, using the same data as Kim and Singleton (2012). Our four-latent-factors VARG term-structure model either outperforms or equals the shadow-rate, QTSM and CIR models with respect to fitting yield levels, conditional volatilities of yields, and expected 6-month holding period excess bond returns. We also compute probabilities of staying at the ZLB for a prolonged period under the historical and risk-neutral measures. Our results show that even though the difference between the probabilities under the two measures tend to be small for short horizons, it can reach 75 percentage points at the two year horizon. This suggests in particular that using only the risk-neutral measure to back out lift-off dates without a no-arbitrage framework results in a systematic underestimation.

The affine framework we develop in this paper can also be used to easily price fixed-income derivatives such as options, or to include observable macro variables in the analysis. We left those different directions for further future research.

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:

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

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:

$$
\begin{pmatrix} m_{1,t} \\ m_{2,t} \end{pmatrix} = \begin{pmatrix} \rho & 0 \\ 2\mu\rho & \rho^2 \end{pmatrix} \begin{pmatrix} m_{1,t-1} \\ m_{2,t-1} \end{pmatrix} + \begin{pmatrix} \mu\alpha \\ 2\mu^2\alpha \end{pmatrix}. \tag{32}
$$

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

$$
\begin{pmatrix} m_1 \\ m_2 \end{pmatrix} = \begin{pmatrix} \dfrac{\alpha\mu}{1-\rho} \\ \dfrac{2\alpha\mu^2}{(1-\rho)(1-\rho^2)} \end{pmatrix}. \tag{33}
$$

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

$$
\begin{aligned}
\varphi_X(u) &= \int_{\mathbb{R}^+} \exp(ux)\,\mathrm{d}\mathbb{P}_X(x) \\
&= \mathbb{P}_X\{0\} + \int_{x>0} \exp(ux)\,\mathrm{d}\mathbb{P}_X(x)
\end{aligned}
$$

Since $x > 0$, $\exp(ux) \to 0$ when $u \to -\infty$, 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:

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

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

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

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

$$
\begin{aligned}
\mathbb{P}(X_{t+h} = 0 \,|\, X_t) &= \exp\left[-\frac{\rho^h\, X_t}{\mu \left(\dfrac{1-\rho^h}{1-\rho}\right)} - (1-\rho)\,\alpha \sum_{k=0}^{h-1} \frac{\rho^k}{1-\rho^{k+1}} \right] \\
&= \exp\left\{ -(1-\rho) \left[\frac{\rho^h\, X_t}{\mu\,(1-\rho^h)} + \alpha \sum_{k=0}^{h-1} \frac{\rho^k}{1-\rho^{k+1}} \right] \right\},
\end{aligned}
\tag{34}
$$

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)

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

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

Proof of Proposition 3.2

Given that, from Proposition 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:

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

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 $\psi_t(u) = \dfrac{\rho u}{1 - u\mu} X_t + \dfrac{u\mu}{1 - u\mu}\, \alpha - \nu \log(1 - u\mu)$, 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:*

$$
\psi_t^*(u) = \frac{u\rho^*}{1 - u\mu^*} X_t + \frac{u\mu^*}{1 - u\mu^*}\, \alpha^* - \nu \log(1 - u\mu^*) \,, \tag{35}
$$

which is the log-Laplace transform of an $EARG_\nu(\alpha^*, \beta^*, \mu^*)$ 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 $\dfrac{d\mathbb{P}_{t,t+1}}{d\mathbb{Q}_{t,t+1}} = \exp\left[\theta' \widetilde{X}_{t+1} - \psi_t^{\mathbb{Q}}(\theta)\right]$, where $\mathbb{P}_{t,t+1}$ is the Esscher transform of $\mathbb{Q}_{t,t+1}$ associated with (θ), we have $\psi^{\mathbb{P}}_{j,t}(u_j) = \psi^{\mathbb{Q}}_{j,t}(u_j + \theta_j) - \psi^{\mathbb{Q}}_{j,t}(\theta_j)$ for any $j \in \{1, \dots, n\}$, and applying Proposition A.1, Proposition 3.3 is easily proved. ■

A.5 Multivariate non-negative affine processes with ZLB spells: the VARG$_\nu(\alpha, \beta, \mu)$ class of processes

A.5.1 Risk-neutral dynamics of Vector Autoregressive Gamma ($VARG$) processes

The purpose of this section is to specify, under the risk-neutral ($\mathbb{Q}$) probability, a family of multivariate processes which possess simultaneously several important features: positivity, zero lower bound spells, contemporaneous and lagged correlations between the scalar components, conditional heteroscedasticity and tractability induced by an exponential-affine conditional Laplace transform. The Vector $ARG_\nu(\alpha^{\mathbb{Q}}, \beta^{\mathbb{Q}}, \mu^{\mathbb{Q}})$ class of processes, denoted $VARG_\nu(\alpha^{\mathbb{Q}}, \beta^{\mathbb{Q}}, \mu^{\mathbb{Q}})$, satisfies all these requirements.

Definition A.1 *The n-dimensional process (X_t) is a risk-neutral $VARG_\nu(\alpha^{\mathbb{Q}}, \beta^{\mathbb{Q}}, \mu^{\mathbb{Q}})$ process if, for any $j \in \{1, \dots, n\}$, the risk-neutral distribution of $X_{j,t+1}$, conditionally on $(X_{1,t+1}, \dots, X_{j-1,t+1}, \underline{X_t})$ (on $\underline{X_t}$, if $j = 1$), is:*

$$\gamma_{\nu_j}\left(\alpha^{\mathbb{Q}}_j + \sum_{k=1}^{j-1} \omega^{\mathbb{Q}}_{j,k}\, X_{k,t+1} + \sum_{k=1}^{n} \beta^{\mathbb{Q}}_{j,k}\, X_{k,t}, \mu^{\mathbb{Q}}_j\right), \tag{36}$$

where $\nu_j \geq 0$, $\alpha^{\mathbb{Q}}_j \geq 0$, $\mu^{\mathbb{Q}}_j > 0$; $\omega^{\mathbb{Q}}_{j,k} \geq 0$ ($\omega^{\mathbb{Q}}_{j,k} = 0$, if $j = 1$), $\beta^{\mathbb{Q}}_{j,k} \geq 0$. Then, the Laplace transform of $X_{j,t+1}$, conditionally on $(X_{1,t+1}, \dots, X_{j-1,t+1}, \underline{X_t})$, is given by:

$$\varphi^{\mathbb{Q}}_{t,j}(u_j) \quad = \quad \exp\left(\sum_{k=1}^{j-1} c^{\mathbb{Q}}_{j,k}(u_j)\, X_{k,t+1} + \sum_{k=1}^{n} a^{\mathbb{Q}}_{j,k}(u_j)\, X_{k,t} + b^{\mathbb{Q}}_j(u_j)\right) \tag{37}$$

where

$$c^{\mathbb{Q}}_{j,k}(u_j) = \frac{u_j\, \mu^{\mathbb{Q}}_j\, \omega^{\mathbb{Q}}_{j,k}}{1 - u_j\, \mu^{\mathbb{Q}}_j}\,, \quad a^{\mathbb{Q}}_{j,k}(u_j) = \frac{u_j\, \mu^{\mathbb{Q}}_j\, \beta^{\mathbb{Q}}_{j,k}}{1 - u_j\, \mu^{\mathbb{Q}}_j}\,, \quad b^{\mathbb{Q}}_j(u_j) = \frac{u_j\, \mu^{\mathbb{Q}}_j\, \alpha^{\mathbb{Q}}_j}{1 - u_j\, \mu^{\mathbb{Q}}_j} - \nu_j\, \log(1 - u_j\, \mu^{\mathbb{Q}}_j)\,. \tag{38}$$

From relation (36) we see that all the components of (X_t) take non-negative values and the vector of parameters $\nu = (\nu_1, \dots, \nu_n)'$ controls for zero lower bound spells; indeed, any component $X_{j,t}$ featuring $\nu_j = 0$ will provide $\mathbb{P}\left(X_{j,t+1} = 0 \,|\, X_{1,t+1}, \dots, X_{j-1,t+1}, X_t\right) = \exp\left[-\alpha^{\mathbb{Q}}_j - \sum_{k=1}^{j-1} \omega^{\mathbb{Q}}_{j,k}\, X_{k,t+1} - \sum_{k=1}^{n} \beta^{\mathbb{Q}}_{j,k}\, X_{k,t}\right] > 0$. Another key property of a $VARG_\nu(\alpha^{\mathbb{Q}}, \beta^{\mathbb{Q}}, \mu^{\mathbb{Q}})$ process is that it is a discrete-time affine process. Indeed, we have:

Proposition A.2 *The n-dimensional $VARG_\nu(\alpha^{\mathbb{Q}}, \beta^{\mathbb{Q}}, \mu^{\mathbb{Q}})$ process (X_t) is a risk-neutral discrete-time affine (Car(1)) process characterized by the following exponential-affine Laplace transform:*

$$\varphi^{\mathbb{Q}}_t(u) \quad = \quad \exp\left(\sum_{k=1}^{n} \widetilde{a}^{\mathbb{Q}}_{n,k}(u_1, \dots, u_n) X_{k,t-1} + \widetilde{b}^{\mathbb{Q}}_n(u_1, \dots, u_n)\right) \tag{39}$$

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

$$
\left\{
\begin{array}{l}
\widetilde{a}^{\mathbb{Q}}_{j,k}(u_1, \dots, u_j) = \widetilde{a}^{\mathbb{Q}}_{j-1,k}\left[u_1 + c^{\mathbb{Q}}_{j,1}(u_j), \dots, u_{j-1} + c^{\mathbb{Q}}_{j,j-1}(u_j)\right] + a^{\mathbb{Q}}_{j,k}(u_j) \\
\widetilde{b}^{\mathbb{Q}}_{j}(u_1, \dots, u_j) = \widetilde{b}^{\mathbb{Q}}_{j-1}\left[u_1 + c^{\mathbb{Q}}_{j,1}(u_j), \dots, u_{j-1} + c^{\mathbb{Q}}_{j,j-1}(u_j)\right] + b^{\mathbb{Q}}_{j}(u_j)\,, \\
j \in \{2, \dots, n\}\,,
\end{array}
\right. \tag{40}
$$

with starting conditions $\widetilde{a}^{\mathbb{Q}}_{1,k}(u_1) = a^{\mathbb{Q}}_{1,k}(u_1)$ and $\widetilde{b}^{\mathbb{Q}}_{1}(u_1) = b^{\mathbb{Q}}_{1}(u_1)$.

Proof See Monfort, Pegoraro, Renne, and Roussellet (2014).

A.5.2 Historical dynamics of the $VARG$ process

Let us consider now our one-period change of probability measure $\frac{d\mathbb{P}_{t,t+1}}{d\mathbb{Q}_{t,t+1}} = \exp\left[\theta' X_{t+1} - \psi^{\mathbb{Q}}_t(\theta)\right]$. Then, the associated historical dynamics of (X_t) remains $VARG$, that is discrete-time (recursive) affine. Indeed, we have:

Proposition A.3 *Given the n-dimensional risk-neutral $VARG_{\nu}(\alpha^{\mathbb{Q}}, \beta^{\mathbb{Q}}, \mu^{\mathbb{Q}})$ process $X_t = (X_{1,t}, \dots, X_{n,t})'$ introduced in Definition A.1 and given the one-period change of probability measure $\frac{d\mathbb{P}_{t,t+1}}{d\mathbb{Q}_{t,t+1}} = \exp\left[\theta' X_{t+1} - \psi^{\mathbb{Q}}_t(\theta)\right]$, then the historical Laplace transform of $X_{j,t+1}$, conditionally on $(X_{1,t+1}, \dots, X_{j-1,t+1}, \underline{X}_t)$, is given by:*

$$
\varphi^{\mathbb{P}}_{t,j}(u_j) \quad = \quad \exp\left(\sum_{k=1}^{j-1} c^{\mathbb{P}}_{j,k}(u_j)\, X_{k,t+1} + \sum_{k=1}^{n} a^{\mathbb{P}}_{j,k}(u_j)\, X_{k,t} + b^{\mathbb{P}}_{j}(u_j)\right) \tag{41}
$$

where

$$
c^{\mathbb{P}}_{j,k}(u_j) = \frac{u_j\, \mu^{\mathbb{P}}_j\, \omega^{\mathbb{P}}_{j,k}}{1 - u_j\, \mu^{\mathbb{P}}_j}\,, \quad a^{\mathbb{P}}_{j,k}(u_j) = \frac{u_j\, \mu^{\mathbb{P}}_j\, \beta^{\mathbb{P}}_{j,k}}{1 - u_j\, \mu^{\mathbb{P}}_j}\,, \quad b^{\mathbb{P}}_{j}(u_j) = \frac{u_j\, \mu^{\mathbb{P}}_j\, \alpha^{\mathbb{P}}_j}{1 - u_j\, \mu^{\mathbb{P}}_j} - \nu_j\, \log(1 - u_j\, \mu^{\mathbb{P}}_j)\,, \tag{42}
$$

and where

$$
\begin{array}{rcl}
\alpha^{\mathbb{P}}_j & = & \dfrac{\alpha^{\mathbb{P}}_j}{1 - \widetilde{\theta}_j\, \mu^{\mathbb{Q}}_j}\,, \quad \omega^{\mathbb{P}}_{j,k} = \dfrac{\omega^{\mathbb{Q}}_{j,k}}{1 - \widetilde{\theta}_j\, \mu^{\mathbb{Q}}_j}\,, \\
\beta^{\mathbb{P}}_{j,k} & = & \dfrac{\beta^{\mathbb{Q}}_{j,k}}{1 - \widetilde{\theta}_j\, \mu^{\mathbb{Q}}_j}\,, \quad \mu^{\mathbb{P}}_j = \dfrac{\mu^{\mathbb{Q}}_j}{1 - \widetilde{\theta}_j\, \mu^{\mathbb{Q}}_j}\,,
\end{array} \tag{43}
$$

with

$$
\left\{
\begin{array}{rcl}
\widetilde{\theta}_n & = & \theta_n \\
\widetilde{\theta}_j & = & \theta_j + \displaystyle\sum_{k=1}^{n-j} c^{\mathbb{Q}}_{n-k+1,j}(\widetilde{\theta}_{n-k+1})\,, \quad j \in \{n-1, \dots, 1\}\,.
\end{array}
\right. \tag{44}
$$

Hence, the historical Laplace transform of X_{t+1}, conditionally on $\underline{X}_t$, is:

$$
\varphi^{\mathbb{P}}_t(u) \quad = \quad \exp\left(\sum_{k=1}^{n} \widetilde{a}^{\mathbb{P}}_{n,k}(u_1, \dots, u_n) X_{k,t-1} + \widetilde{b}^{\mathbb{P}}_n(u_1, \dots, u_n)\right) \tag{45}
$$

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

$$\begin{cases} \widetilde{a}^{\mathbb{P}}_{j,k}(u_1, \dots, u_j) = \widetilde{a}^{\mathbb{P}}_{j-1,k}\left[u_1 + c^{\mathbb{P}}_{j,1}(u_j), \dots, u_{j-1} + c^{\mathbb{P}}_{j,j-1}(u_j)\right] + a^{\mathbb{P}}_{j,k}(u_j) \\ \widetilde{b}^{\mathbb{P}}_{j}(u_1, \dots, u_j) = \widetilde{b}^{\mathbb{P}}_{j-1}\left[u_1 + c^{\mathbb{P}}_{j,1}(u_j), \dots, u_{j-1} + c^{\mathbb{P}}_{j,j-1}(u_j)\right] + b^{\mathbb{P}}_{j}(u_j)\,, \\ j \in \{2, \dots, n\}\,, \end{cases} \tag{46}$$

with starting conditions $\widetilde{a}^{\mathbb{P}}_{1,k}(u_1) = a^{\mathbb{P}}_{1,k}(u_1)$ and $\widetilde{b}^{\mathbb{P}}_{1}(u_1) = b^{\mathbb{P}}_{1}(u_1)$.

Proof See Monfort, Pegoraro, Renne, and Roussellet (2014).

A.5.3 Moments, VAR representations, stationarity and lift-off

Given the results already presented in the previous sections, it is straightforward to verify that the conditional mean and variance of the component $X_{j,t+1}$ of a $VARG_\nu(\alpha, \beta, \mu)$ process, given $(X_{1,t+1}, \dots, X_{j-1,t+1}, X_t)$, are respectively given by:

$$\begin{aligned} \overline{m}_{j,t+1} &= \mu_j \left(\alpha_j + \sum_{k=1}^{j-1} \omega_{j,k}\, X_{k,t+1} + \sum_{k=1}^{n} \beta_{j,k}\, X_{k,t}\right) + \mu_j\, \nu \\ \overline{\sigma}^2_{j,t+1} &= 2\,\mu_j \left(\alpha_j + \sum_{k=1}^{j-1} \omega_{j,k}\, X_{k,t+1} + \sum_{k=1}^{n} \beta_{j,k}\, X_{k,t}\right) + \mu_j^2\, \nu\,. \end{aligned}$$

It is also easy to check that the process (X_t) has the following *Recursive VAR* representation:

$$X_{j,t+1} = \overline{m}_{j,t+1} + \overline{\sigma}^2_{j,t+1}\, \varepsilon_{j,t+1}\,,$$

where $\varepsilon_t = (\varepsilon_{1,t}, \dots, \varepsilon_{n,t})'$ is a martingale difference sequence with $\mathbb{E}_{t-1}(\varepsilon_t) = 0$ and $\mathbb{V}_{t-1}(\varepsilon_t) = I$. The conditional mean and variance-covariance matrix of X_{t+1}, given X_t, are respectively given by:

$$\begin{aligned} m_{t+1} &= (I - \Delta_m)^{-1}\,(A_m\, X_t + b_m) \\ \Sigma_{t+1} &= (I - \Delta_m)^{-1}\, diag(\sigma^2_{t+1})\,(I - \Delta_m)^{-1}\,, \end{aligned}$$

where $\Delta_m = diag(\mu_j)\,\Delta$, Δ being the lower triangular matrix with elements $\Delta_{i,j} = 0$, if $j \geq i$, and $\Delta_{i,j} = \omega_{i,j}$ for $j < i$; $A_m = diag(\mu_j)\,B$, with $B_{ij} = \beta_{ij}$ and $b_m = diag(\mu_j)\,(\alpha + \nu)$; $\sigma^2_{t+1} = \Delta_\vartheta\,(I - \Delta_m)^{-1}\,(A_m\, X_t + b_m) + A_\vartheta X_t + b_\vartheta = A_\sigma\, X_t + b_\sigma$ (say), where $\Delta_\vartheta = 2\,diag(\mu_j^2)\Delta$, $A_\vartheta = 2\,diag(\mu_j^2)B$ and $b_\vartheta = diag(\mu_j^2)(2\,\alpha + \nu)$. We get therefore the following canonical VAR representation:

$$X_{t+1} = (I - \Delta_m)^{-1}\, b_m + (I - \Delta_m)^{-1}\, A_m\, X_t + \Omega_{t+1}\eta_{t+1}\,, \tag{47}$$

with $\Omega_{t+1} = (I - \Delta_m)^{-1}\, diag(\sigma_{t+1})$ and where η_t is a martingale difference sequence with $\mathbb{E}_{t-1}(\eta_t) = 0$ and $\mathbb{V}_{t-1}(\eta_t) = I$. If we introduce now the notation $\widetilde{A}_m = (I - \Delta_m)^{-1}\, A_m$ for the autoregressive matrix in (47), it is easily seen that the unconditional mean and variance-covariance matrix of (X_t), respectively denoted by $\widetilde{m}_t$ and $\widetilde{\Sigma}_t$, satisfy a recursive system of the form:

$$\begin{pmatrix} \widetilde{m}_{t+1} \\ vec(\widetilde{\Sigma}_{t+1}) \end{pmatrix} = \begin{bmatrix} \widetilde{A}_m & 0 \\ (I-\Delta_m)^{-1} \otimes (I-\Delta_m)^{-1}\, C & \widetilde{A}_m \otimes \widetilde{A}_m \end{bmatrix} \begin{pmatrix} \widetilde{m}_t \\ vec(\widetilde{\Sigma}_t) \end{pmatrix} + b$$

where

$$b = \begin{pmatrix} (I-\Delta_m)^{-1}\, b_m \\ (I-\Delta_m)^{-1} \otimes (I-\Delta_m)^{-1}\, c \end{pmatrix},$$

and where $C\widetilde{m}_t + c = vec\,[diag(A_\sigma\, \widetilde{m}_t + b_\sigma)]$. This system admits a stationary solution if and only if the moduli of the eigenvalues of $\widetilde{A}_m$ are all smaller than one (since the eigenvalues of $\widetilde{A}_m \otimes \widetilde{A}_m$ are all the possible products of the eigenvalues of $\widetilde{A}_m$). This solution is given by:

$$\begin{aligned} \widetilde{m} &= (I-\widetilde{A}_m)^{-1}\, (I-\Delta_m)^{-1}\, b_m \\ vec(\widetilde{\Sigma}) &= (I-\widetilde{A}_m \otimes \widetilde{A}_m)^{-1}\, (I-\Delta_m)^{-1} \otimes (I-\Delta_m)^{-1} vec(diag(A_\sigma\, \widetilde{m} + b_\sigma))\,. \end{aligned}$$

Moreover, iterating (47), we easily find $Cov(X_{t+h}, X_t) = \widetilde{A}_m^h\, \widetilde{\Sigma}_t$ and we thus have the following property:

Proposition A.4 *The $VARG_\nu(\alpha,\, \beta, \mu)$ process is (asymptotically) second order stationary if the moduli of the eigenvalues of $\widetilde{A}_m$ are all smaller than one and, in this case, $\mathbb{E}(X_t) \to \widetilde{m}$, $\mathbb{V}(X_t) \to \widetilde{\Sigma}$ and $Cov(X_{t+h}, X_t) \to \widetilde{A}_m^h\, \widetilde{\Sigma}$ when $t \to +\infty$.*

Corollary A.4.1 *If B is lower triangular, that is to say if $X_{k,t}$ does not cause $X_{j,t}$ when $k > j$, then the matrix $\widetilde{A}_m = (I-\Delta_m)^{-1}\, A_m = (I-\Delta_m)^{-1}\, diag(\mu_j)B$ is lower triangular, with diagonal entries equal to $\mu_j\, \beta_{jj}$ and, therefore, the stationary condition is $\mu_j\, \beta_{jj} < 1$ for all $j \in \{1, \ldots, n\}$.*

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

	Maturity	6m	1y	2y	4y	7y	10y
Mean	Yields	0.2930	0.3451	0.5144	0.9369	1.5327	1.8988
	GARCH(1,1)	0.0023	0.0025	0.0038	0.0050	0.0060	0.0055
	EGARCH(1,1)	0.0024	0.0026	0.0038	0.0050	0.0060	0.0054
	rolling-window	0.0020	0.0022	0.0033	0.0049	0.0061	0.0054
Std.	Yields	0.3842	0.4064	0.4891	0.6517	0.8207	0.7782
	GARCH(1,1)	0.0019	0.0021	0.0025	0.0025	0.0022	0.0022
	EGARCH(1,1)	0.0020	0.0022	0.0024	0.0023	0.0019	0.0019
	rolling-window	0.0017	0.0019	0.0023	0.0024	0.0025	0.0022

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. We normalize them 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	
	total	low half	Garch	Egarch	total	low half	Garch	Egarch
Garch	0.54	0.53	1		0.58	0.60	1	
Egarch	0.56	0.63	0.96	1	0.59	0.70	0.97	1
Rw	0.54	0.62	0.95	0.94	0.60	0.72	0.96	0.96
Maturity	2y				4y			
Garch	0.60	0.65	1		0.58	0.54	1	
Egarch	0.64	0.75	0.97	1	0.61	0.61	0.96	1
Rw	0.65	0.73	0.95	0.96	0.58	0.61	0.91	0.91
Maturity	7y				10y			
Garch	0.41	0.36	1		0.15	−0.05	1	
Egarch	0.44	0.45	0.93	1	0.19	0.01	0.90	1
Rw	0.43	0.40	0.88	0.86	0.14	0.04	0.85	0.83

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. We take estimated proxies and normalize them to be comparable to annualized yields.

Table 3: Parameter estimates

	$\mathbb{Q}$-parameters			$\mathbb{P}$-parameters	
	Estimates	Std.		Estimates	Std.
α_1	0	–		0	–
α_2	0	–		0	–
α_3	0.090	0.006		0.090	0.006
α_4	0.043	0.004		0.048	0.004
ρ_1	0.999	$9.011 \cdot 10^{-5}$		0.959	$9.055 \cdot 10^{-4}$
ρ_2	0.999	$2.076 \cdot 10^{-13}$		0.316	$1.560 \cdot 10^{-2}$
ρ_3	0.997	$9.489 \cdot 10^{-5}$		0.996	$3.484 \cdot 10^{-4}$
ρ_4	0.996	$3.481 \cdot 10^{-4}$		0.999	$2.893 \cdot 10^{-4}$
$\beta_{2,1}$	1.649	0.027		0.926	0.132
$\beta_{3,1}$	$6.429 \cdot 10^{-4}$	$1.211 \cdot 10^{-4}$		$6.425 \cdot 10^{-4}$	$1.209 \cdot 10^{-4}$
$\beta_{3,2}$	$1.174 \cdot 10^{-5}$	$.896 \cdot 10^{-6}$		$1.174 \cdot 10^{-5}$	$1.896 \cdot 10^{-6}$
$\beta_{4,1}$	0.075	$3.592 \cdot 10^{-3}$		0.075	0.004
$\beta_{4,2}$	$2.061 \cdot 10^{-9}$	$2.381 \cdot 10^{-13}$		$2.064 \cdot 10^{-9}$	$4.115 \cdot 10^{-13}$
$\beta_{4,3}$	0.029	0.014		0.029	0.014
μ_1	1	–		0.980	$4.312 \cdot 10^{-4}$
μ_2	1	–		0.562	0.014
μ_3	1	–		0.999	$1.871 \cdot 10^{-4}$
μ_4	1	–		1.001	$1.609 \cdot 10^{-4}$
		Other Parameters			
δ_1	$5.322 \cdot 10^{-7}$	$2.302 \cdot 10^{-8}$	δ_2	$1.248 \cdot 10^{-5}$	$2.158 \cdot 10^{-6}$
θ_1	-0.021	$4.491 \cdot 10^{-4}$	θ_2	-0.780	0.044
θ_3	$-6.531 \cdot 10^{-4}$	$1.873 \cdot 10^{-4}$	θ_4	$1.340 \cdot 10^{-3}$	$1.605 \cdot 10^{-4}$
ν_1	0	–	ν_2	0	–
ν_3	3.965	0.602	ν_4	0.017	0.053
$\sigma_{V,1}$	$1 \cdot 10^{-3}$	–	$\sigma_{V,2}$	$2 \cdot 10^{-3}$	–
$\sigma_{S,1}$	0.130	–	$\sigma_{S,2}$	0.210	–
$\sigma_{K,1}$	0.228	–	$\sigma_{K,2}$	1.419	–
σ_R	0.090	0.0009			

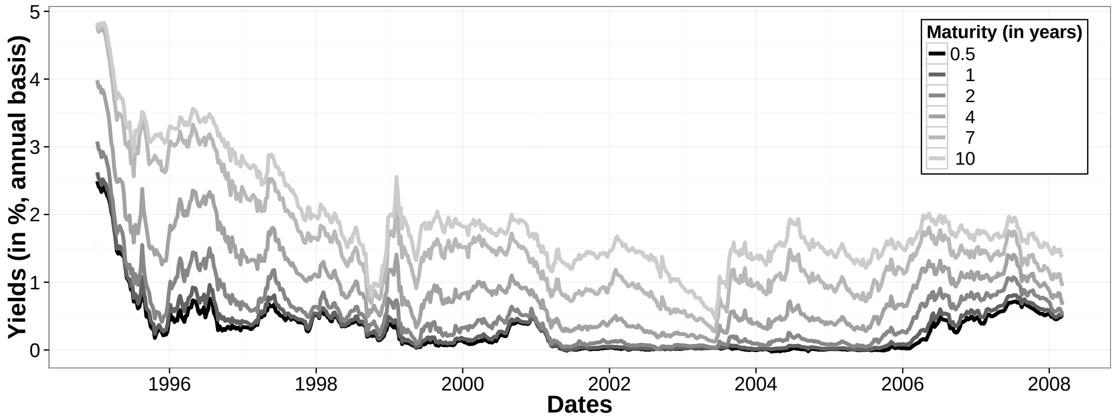
Note: The symbol '–' in the standard-deviation column indicates that the parameter has been calibrated. The σ_V's are set to be coherent with the standard deviations of the differences between the GARCH, EGARCH, and rolling window variance proxies. The σ_S's are set at the in-sample mean of standard-deviations of forecasts among the professional forecasters. The σ_K's are set at a fourth of the in-sample standard deviations of the proxies of expected excess bond-returns. Last, we impose that the unconditional mean of the short-term interest rate is equal to its in-sample mean.

Table 4: Cross-sectional fit of yield levels, conditional volatility proxies, and expected excess bond returns for 6-months holding period

Maturity		6m	1y	2y	4y	7y	10y
level	RMSE (in bps)	6.25	5.34	9.59	10.24	9.60	11.93
Volatility regressions	intercept	0.002	−0.008	−0.007	0.004	0.02	0.01
	slope	0.93	1.06	1.04	0.91	0.78	0.79
	R^2	0.64	0.80	0.92	0.87	0.80	0.76
Excess returns regressions	intercept	—	−0.10	−0.25	−0.38	0.14	0.23
	slope	—	1.02	1.06	1.15	1.08	0.99
	R^2	—	0.75	0.68	0.63	0.49	0.37

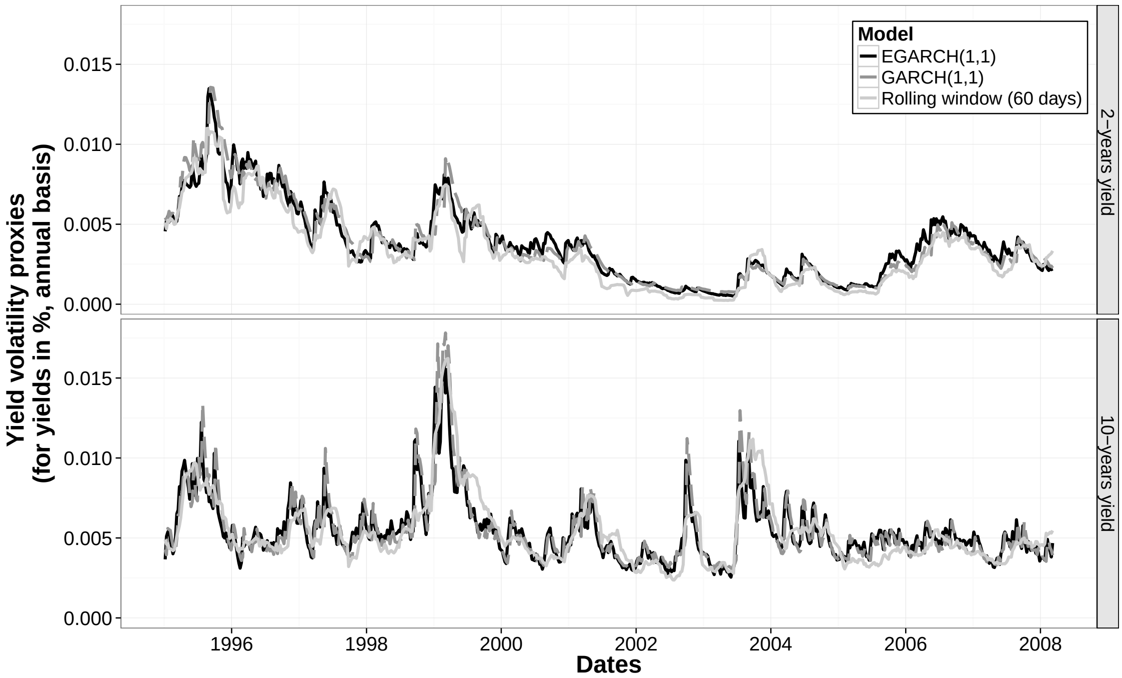
Note: The volatility regressions are the regressions of Egarch(1,1) volatility proxies on the model-implied conditional volatilities. Proxies of expected excess returns are derived based on the methodology of Cochrane and Piazzesi (2005).

Figure 2: Japanese yields data



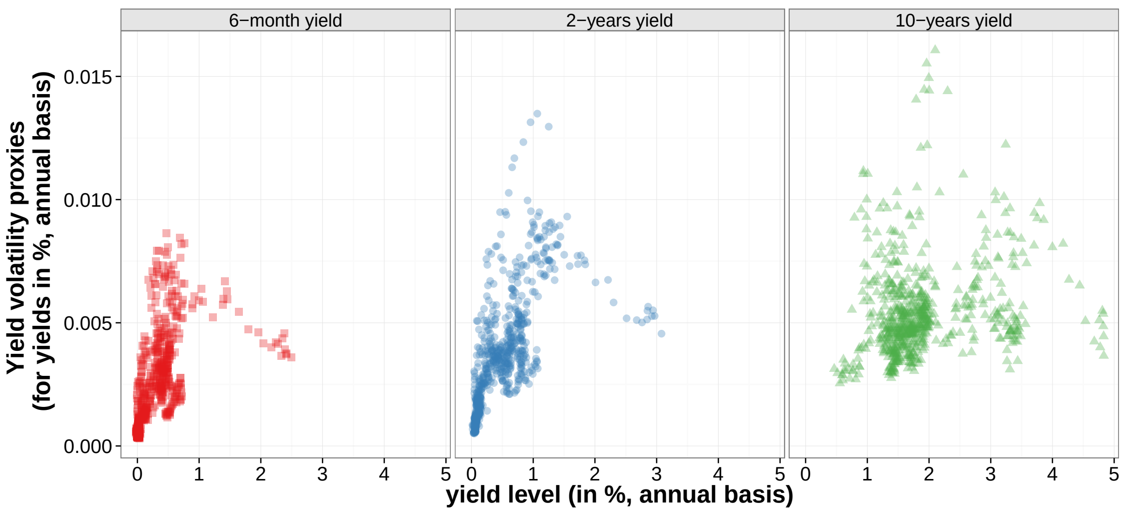
Notes: Yields are weekly data from January, 6th 1995 to March, 8th 2008. Yields are expressed in annualized percentage points figures, with maturity from 6 months (darkest line) to 10 years (lightest line).

Figure 3: Conditional volatility proxies



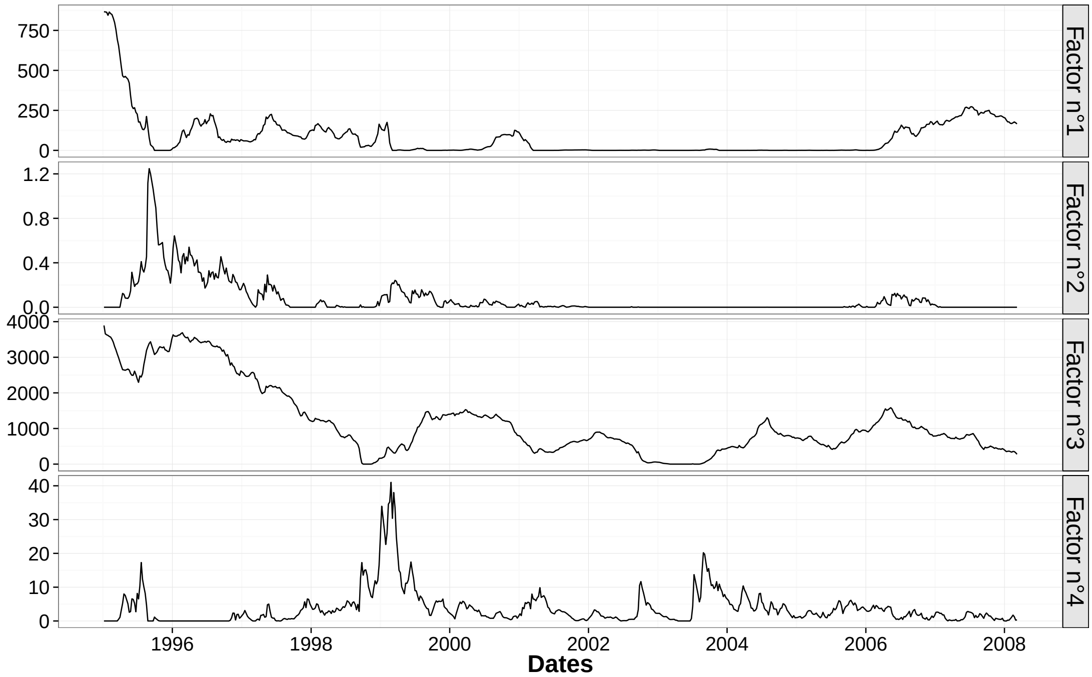
Notes: Top and bottom panels respectively present the volatility proxies for the 2-year and the 10-year yields. Garch and Egarch conditional volatility models are computed on weekly yield changes whereas the rolling-window volatility is computed on a 60-day window of weekly data. We take the square-root 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: Dependence between yield conditional volatilities and levels



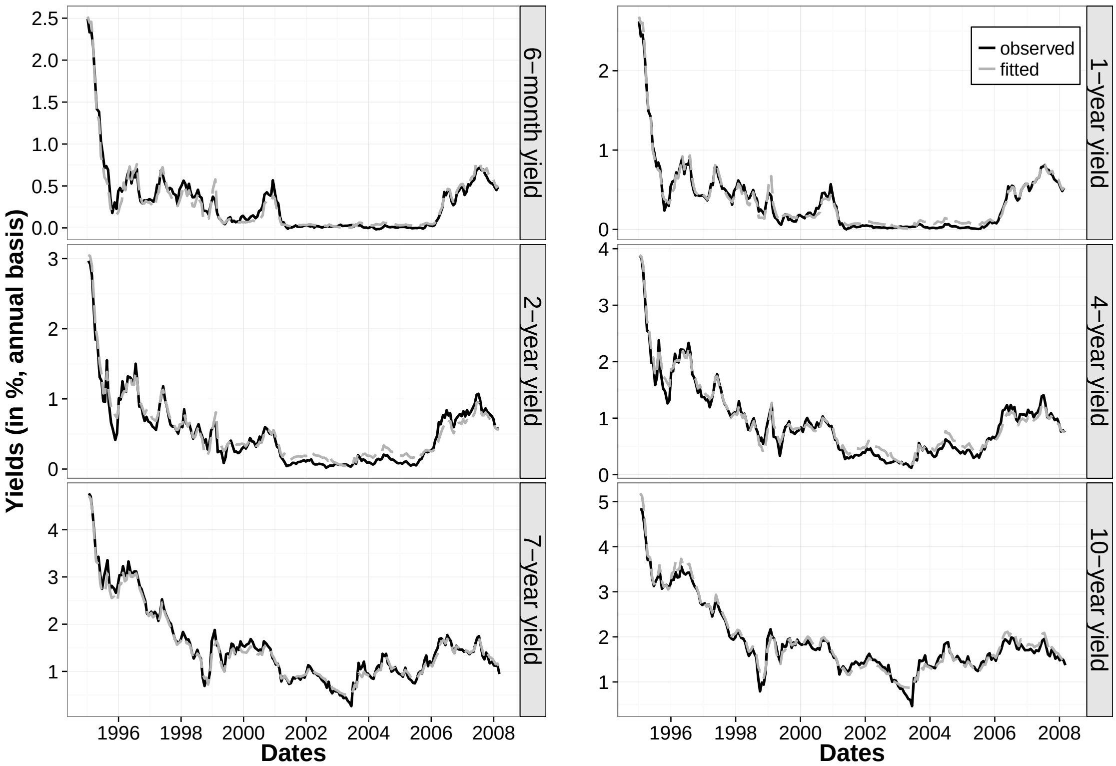
Note: The volatility proxy is extracted from an Egarch model computed on weekly yield changes. We take the square-root of estimated proxy of conditional variance and obtain our conditional volatility proxy. We take the estimated proxy and normalize it to make it homogeneous to annualized yields.

Figure 5: Estimated Factors



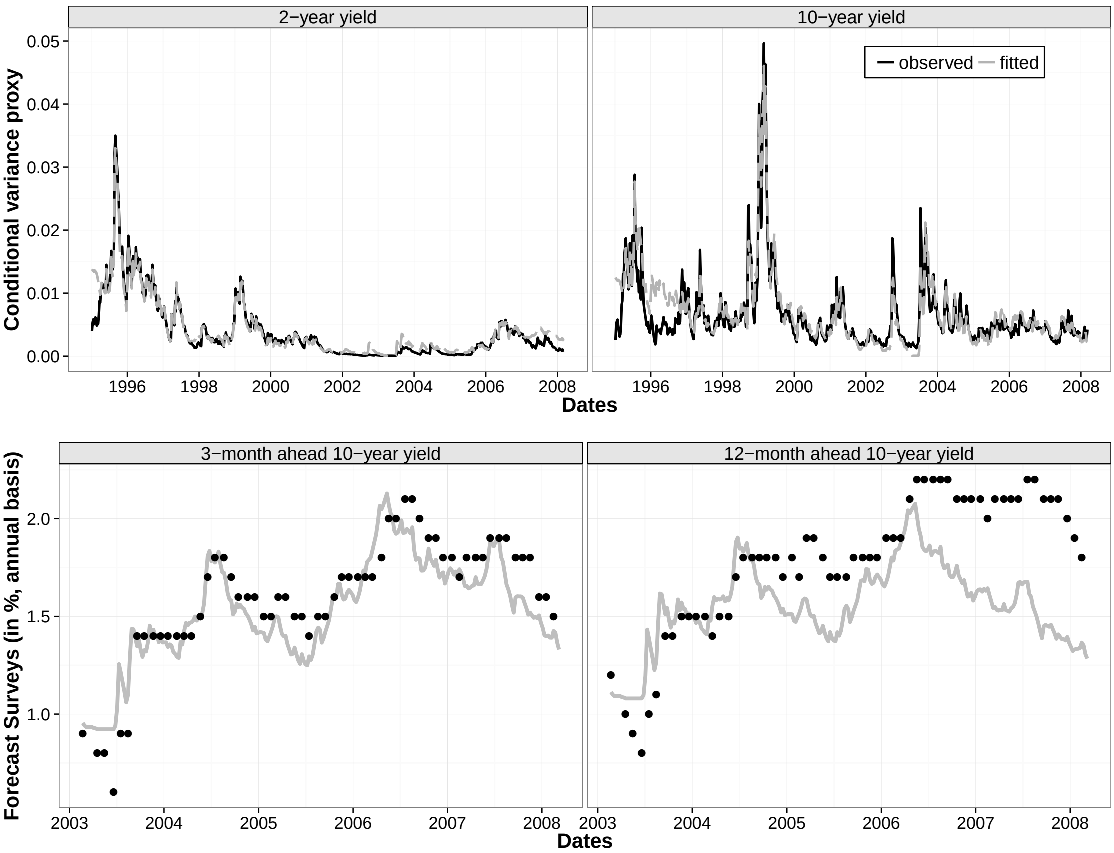
Notes: Factors are filtered estimates from the linear Kalman filter on the full sample (January 1995 to March 2008). Factors 1 and 2 correspond to $X_t^{(1)}$ which directly enters the short-term interest rate specification with loadings δ_1 and δ_2, and Factors 3 and 4 correspond to $X_t^{(2)}$.

Figure 6: Observed and model-implied yields



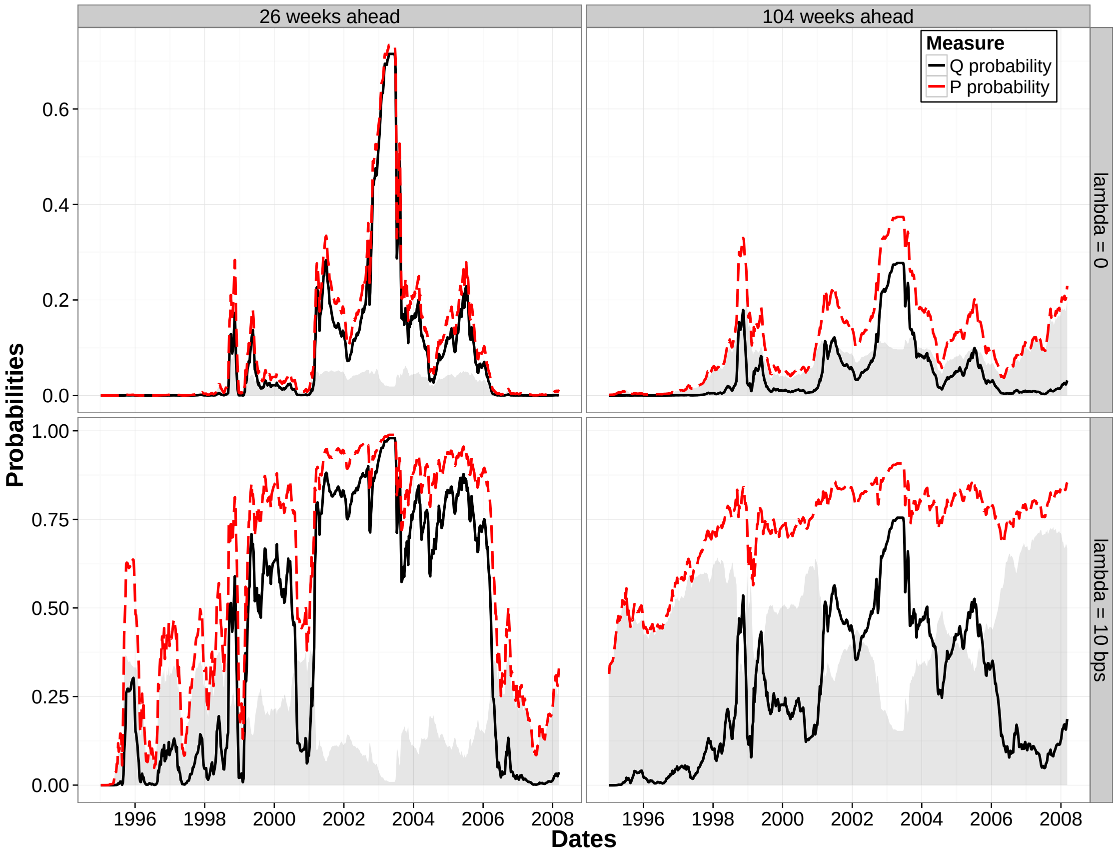
Notes: Yields are observed at the weekly frequency from January, 6th 1995 to March, 8th 2008. Yields are expressed in annualized percentage points, with maturities from 6 months to 10 years. The black solid lines are the observed yields and the grey dotted lines are the model-implied (or fitted) yields using the term structure framework of Section 3 with 4 factors ($n_1 = 2$ and $n_2 = 2$).

Figure 7: Fitted conditional variance proxies and surveys



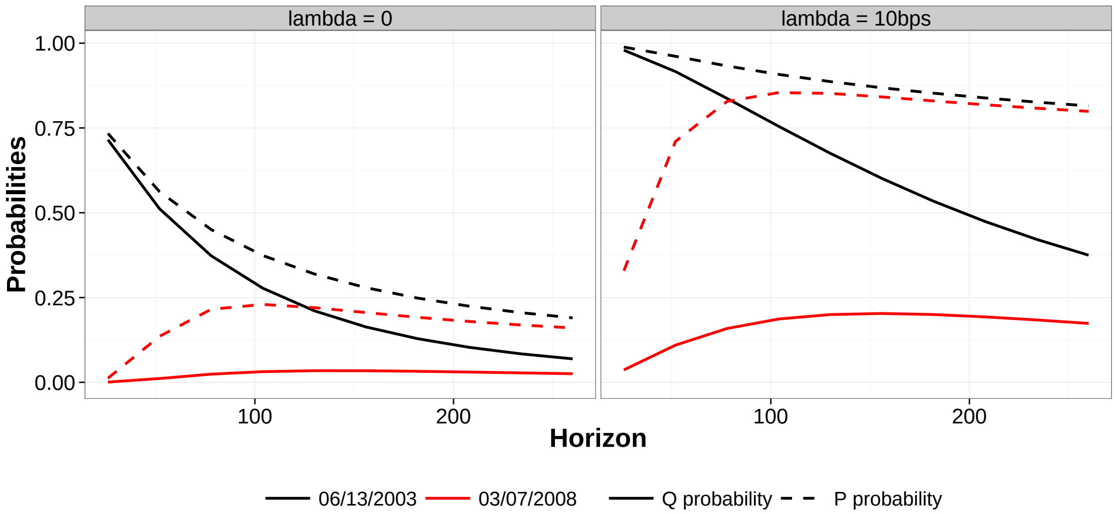
Notes: The top panel presents the two conditional variance proxies $V_t(h)$ estimated with an EGARCH(1,1) model on 2- and 10-year yields (left and right tiles) of weekly data from January, 6th 1995 to March, 8th 2008. The black solid lines are the observed variance proxies and the grey dotted lines are 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. Data is available quarterly from 2003 to 2008. The black dots correspond to the observed data, and the grey solid lines are the fitted equivalent.

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



Notes: Probabilities are computed with weekly data from January, 6th 1995 to March, 8th 2008. They represent the probabilities of the short-rate hitting zero in $k = 26$ weeks (top-left panel) and $k = 104$ weeks (top-right panel). On bottom panels, we represent the probabilities of the short-rate being below 10 bps in $k = 26$ weeks (bottom-left panel) and $k = 104$ weeks (bottom-right panel). Black solid lines are the risk-neutral probabilities whereas red dashed lines are the historical ones. Grey-shaded areas are the difference between the two.

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



Notes: X axis is the horizon k of the short rate r_{t+k} being exactly 0 (left tile), or below 10bps (right tile). Black and red curves correspond distinguish the date at which these probabilities are evaluated, and respectively correspond to the 13^{th} of June 2003 and the 7^{th} of March 2008. Solid and dashed lines represent respectively to $\mathbb{Q}$ and $\mathbb{P}$-probabilities.

References

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 retunrs? *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.

Backus, D., S. Foresi, A. Mozumdar, and L. Wu (2001, March). Predictable changes in yields and forward rates. *Journal of Financial Economics 59*(3), 281–311.

Bauer, M. D. and G. D. Rudebusch (2013). Monetary policy expectations at the zero lower bound. Technical report.

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.

Campbell, J. Y. and R. J. Shiller (1991, May). Yield spreads and interest rate movements: A bird's eye view. *Review of Economic Studies 58*(3), 495–514.

Cecchetti, S. G. (1988, December). The case of the negative nominal interest rates: New estimates of the term structure of interest rates during the great depression. *Journal of Political Economy 96*(6), 1111–41.

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? Technical report.

Cochrane, J. and M. Piazzesi (2009). Decomposing the yield curve. Technical report.

Cochrane, J. H. and M. Piazzesi (2005, March). Bond risk premia. *American Economic Review 95*(1), 138–160.

Collin-Dufresne, P. and R. S. Goldstein (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 C. Wu (2013). Estimation of Non-Gaussian Affine Term Structure Models. Technical report, University of Chicago Booth School of Business, Working paper.

Creal, D. D. and J. C. Wu (2014). Term Structure of Interest Rate Volatility and Macroeconomic Uncertainty. Working papers, Chicago Booth.

Dai, Q., A. Le, and K. J. Singleton (2010). Discrete-time affineq 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.

Fama, E. F. and R. R. Bliss (1987, September). The information in long-maturity forward rates. *American Economic Review 77*(4), 680–92.

Faust, J. and J. H. Wright (2011). Efficient Prediction of Excess Returns. *Review of Economics and Statistics 59*(93), 647–659.

Filipovic, D., M. Larsson, and A. Trolle (2013). Linear-rational term structure models. Technical report.

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 (2013). Regime switching and bond pricing. *Journal of Financial Econometrics (forthcoming).*

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. Technical report.

Ichiue, H. and Y. Ueno (2013). Estimating term premia at the zero bound: An analysis of japanese, u.s., and u.k. yields. Technical report.

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.

Kim, D. H. (2008). Zero bound, option-implied pdfs, and term structure models. Technical report.

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.

Kim, D. H. and J. H. Wright (2005). An arbitrage-free three-factor term structure model and the recent behavior of long-term yields and distant-horizon forward rates. Technical report.

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.

OECD (2012). *OECD Central Government Debt Statistics 2012* (OECD Publishing ed.).

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.

Priebsch, 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. Technical report.

Siegel, A. F. (1979, August). The noncentral chi-squared distribution with zero degrees of freedom and testing for uniformity. *Biometrika 66*(2), 381–386.

Swanson, E. T. and J. C. Williams (2013). Measuring the effect of the zero lower bound on yields and exchange rates. *Journal of International Economics forthcoming*.

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. Technical report.

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.