AESTIMATIO, THE IEB INTERNATIONAL JOURNAL OF FINANCE, 2014. **8**: 38-67

Fitting the Pareto-Lévy distribution on the yield curve: An application to forecasting

Rostan, Pierre
Rostan, Alexandra

RECEIVED: 14 JANUARY 2013
ACCEPTED: 9 JUNE 2013

Abstract

Finding the optimal distribution of innovation terms to forecast the yield curve or to price derivatives on fixed-income securities with Monte Carlo simulation is a challenge that not so many authors have taken up. We investigate the Pareto-Lévy distribution that fits the U.S. yield curve when the latter experiences different shapes: normal, inverse, flat and humped and experiences a volatile environment or not. We show that the Pareto-Lévy distribution does not improve significantly yield curve forecasting with Monte Carlo simulation when benchmarked to the Normal distribution but we discovered interesting outcomes concerning the Normal distribution such as its higher performance for fitting the yield curve and its consistency whatever the shape of the yield curve and whether the interest rate environment is volatile or not. We base our findings on 2,707 U.S. Treasury yield curves over the 2001-2012 period. Market participants who use Monte Carlo simulation, in need of a methodological framework to identify an optimal random number generator that fits the yield curve or in need of an accurate short term forecast of the yield curve, will find our paper appealing.

Keywords:

Yield curve forecasting, Distribution, Interest rate, Cox-Ingersoll-Ross model, Pareto-Lévy distribution, Normal distribution, Monte Carlo simulation, Innovation terms.

JEL classification:

C53; C63; E43; E47.

Rostan, P. American University in Cairo, Egypt
Rostan, P., Ph.D., Associate Professor AUC Avenue, P.O. Box 74 New Cairo 11835, Egypt. ✆ +20.2.2615.1000 Fax: +20.2.27957565
E-mail: prostan@aucegypt.edu

Rostan, A. American University in Cairo, Egypt
Rostan, A., M.Sc., Lecturer AUC Avenue, P.O. Box 74 New Cairo 11835, Egypt. ✆ +20.2.2615.1000 Fax: +20.2.27957565
E-mail: arostan@aucegypt.edu

AESTIMATIO, THE IEB INTERNATIONAL JOURNAL OF FINANCE, 2014. **8**: 38-67

Ajuste de la distribución Pareto-Lévy a la curva de tipos de interés: **Una aplicación a la predicción**

Rostan, Pierre
Rostan, Alexandra

Resumen

Encontrar la distribución óptima de los términos de innovación a la hora de predecir la curva de tipos de interés o valorar derivados sobre títulos de renta fija mediante simulación Monte Carlo constituye un reto que ha sido aceptado por un escaso número de investigadores. En este artículo se investiga la distribución de Pareto-Lévy que ajusta la curva de tipos del Tesoro americano bajo diferentes formas de esta última: inversa, plana, encorvada, y tanto en ambientes de volatilidad como en entornos de tipos no volátiles. Se muestra que la distribución de Pareto-Lévy no mejora significativamente la predicción de la curva de tipos con simulación Monte Carlo respecto a la distribución Normal; sin embargo, se han descubierto algunos resultados ciertamente interesantes en lo que se refiere a la distribución Normal, tales como su mejor funcionamiento a la hora de ajustar la curva de tipos y su consistencia, cualquiera que sea la forma de la curva y el entorno (de volatilidad o no) de los tipos de interés. Los resultados que se exponen en este artículo están basados en 2.707 curvas de tipos del Tesoro americano, en el periodo 2001-2012. Los participantes en el mercado que utilicen simulación Monte Carlo, ya sea por la necesidad de un marco metodológico para identificar un generador de números aleatorios óptimos que ajuste la curva de tipos o bien para obtener predicciones precisas a corto plazo, encontrarán este artículo atractivo.

Palabras clave:

Predicción de la curva de tipos de interés, distribución, tipo de interés, modelo Cox-Ingersoll-Rossl, distribución Pareto-Lévy, distribución normal, simulación Monte Carlo, términos de innovación.

1. Introduction

Forecasting the yield curve is crucial for central banks to implement adequate monetary policy and for institutional investors to actively manage portfolios because the term structure is commonly considered as a leading gauge of the economic activity. Our paper presents a methodological framework to test random numbers generators involved in Monte Carlo simulation for yield curve fitting and forecasting. As an example, we test the fitting of the Pareto-Lévy distribution to the distribution of innovation terms of the observed yield curve. To find the distribution of innovation terms, we interpolate the yield curve with five hundred points using the cubic spline methodology. We compute the differences dr. We assume that dr may be modeled with the Cox-Ingersoll-Ross (CIR, 1985) model. We deduce the innovation terms from the CIR model (refer to equation 3 below). More than an academic exercise, our aim is to help practitioners identifying an optimal distribution from which random numbers will be drawn to feed Monte Carlo simulation of the yield curve. Applications are numerous in the fields of pricing fixed income derivatives and forecasting the yield curve based on Monte Carlo simulation. Our paper will present an application to yield curve forecasting. Section 2 reviews the literature concerning the shape of the yield curve, curve fitting and the Pareto-Lévy distribution that may fit the distribution of innovation terms of the yield curve. In section 3, we present the methodology in five steps. We present the results in section 4 and we wrap up our findings in section 5.

2. Literature review

2.1. Capturing the shape of the yield curve: curve fitting

In order to explore the distribution of the yield curve, we assume that the shape of the yield curve may have an impact on the distribution. Economists classify the shape in four categories: normal, flat, humped and inverted. A "normal" yield curve means that yields increase as maturity of bills and bonds increases: the yield curve is upward sloping and reflects expectations that the current economy will grow in the future, i.e. expectations of a greater inflation in the future. When all maturities have similar yields, we observe a flat yield curve. When short-term and long-term yields are equal and medium-term yields are higher than those of the short-term and long-term, we observe a humped yield curve. Humped and flat curves indicate uncertainty in the economy. When long-term yields are lower than short-term yields, we observe an inverted yield curve, signal of recession. From the above definition, a humped yield curve should be centered on the 5-year medium maturity where it displays the highest yield. However, if we stick to this definition, none of the curves in our sample of 2,707 U.S. yield curves is humped. Table 1 gathers the definitions that we adopt in this paper for normal, flat, humped and inverted yield curves in order to adjust to the real word.

Table 1. Classification of the U.S. yield curve in four occurrences

Type of curve:	Inverted	Else: Flat	Else: Humped	Else: Normal
Criteria:	1-month rate is higher than 20-year rate	All rates remain in a range of 50 basis points	6-month rate is higher than 5-year rate	

We modify the definition of a humped curve to make it observable: our subsample of humped curves will comprise curves that are centered on the 6-month yield (the highest point of the curve), which represent 4% of the sample. Again, to make it tractable, our definition of a flat curve will comprise yield curves bounded in a range 50 basis points between the highest and the lowest yields. 6% of our sample falls in this definition. In addition, the inverted yield curve will be the one with a 1-month yield higher than the 20-year yield whatever the shape of the curve in the midst. 5% of our sample includes inverted curves. Finally, the normal yield curve will be whichever curve remaining, i.e. 85% of the curves with a positive slope.

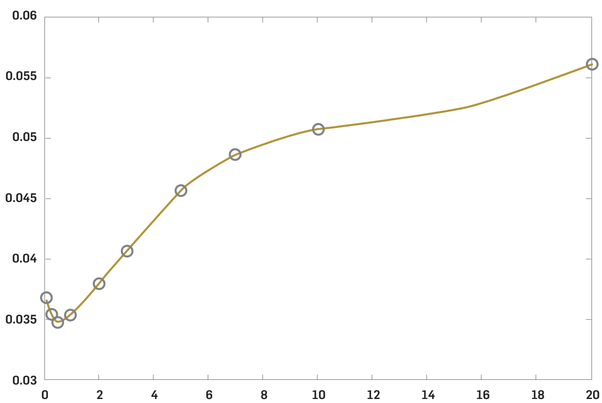
Table 2 gathers the statistics for each type of curve in our sample.

Table 2. Counting occurrences among 2,707 observed U.S. yield curves from July 31, 2001 to May 24, 2012

Type of curve:	Normal	Humped	Flat	Inverted	Total
Number of occurrences:	2,300	104	168	135	2,707
% of occurrences:	85	4	6	5	100

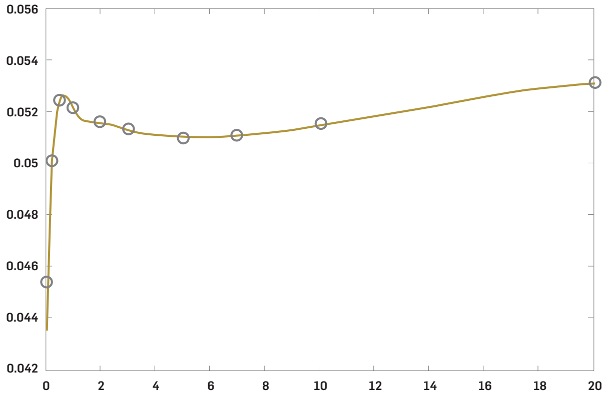
Figures 1, 2, 3, 4 illustrate examples of normal, humped, flat and inverted yield curves observed in our sample and based on the definition provided in Table 1. We will analyze the distribution of innovation terms of the yield curve in the context of its shape.

Figure 1. A "normal" U.S. yield curve on July, 31 2001- 500-points cubic spline interpolation



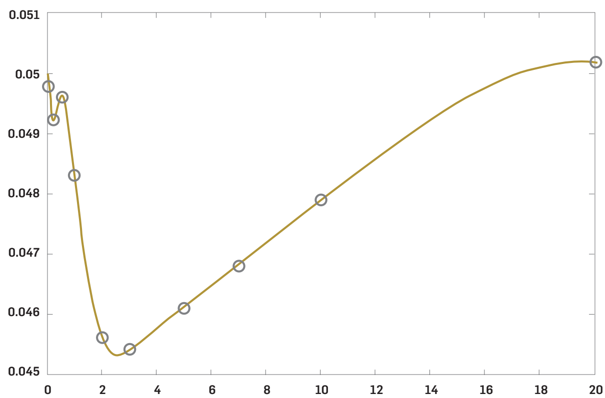
SOURCE: THE AUTHORS

Figure 2. A humped U.S. yield curve on June 30, 2006 - 500-points cubic spline interpolation. Humped means that 6-month rate is higher than 5-year rate



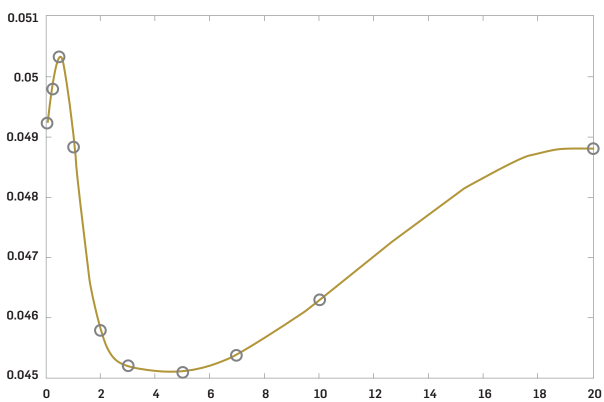
SOURCE: THE AUTHORS

Figure 3. A "flat" U.S. yield curve on July 26, 2007 - 500-points cubic spline interpolation. "Flat" means that yields are bounded in a range 50 basis points



SOURCE: THE AUTHORS

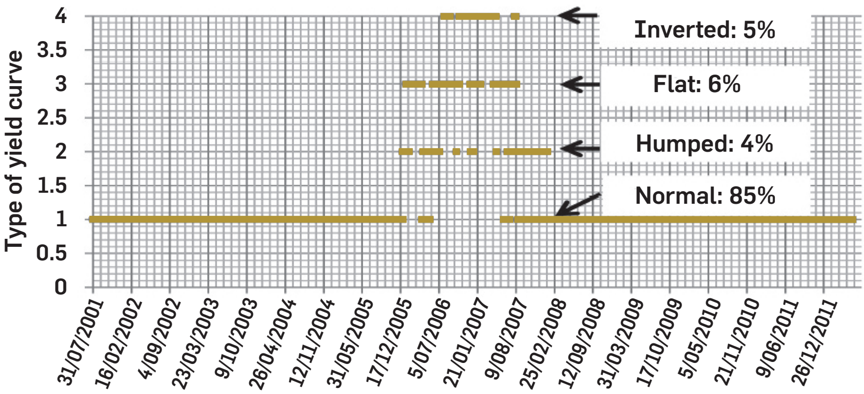
Figure 4. An "inverted" U.S. yield curve on April 24, 2007- 500-points cubic spline interpolation. "Inverted" means that 1-month yield is higher than 20-year yield whatever the shape of the curve in the midst



SOURCE: THE AUTHORS

Figure 5 illustrates how the types of yield curve are distributed in our sample. Non-"normal" curves concentrate between December 27, 2005 and January 17, 2008.

Figure 5. **Plotting the type of yield curve versus time on a sample ranging between July, 31 2001 and May 24, 2012: non normal curves concentrate between December 27, 2005 and January 17, 2008**



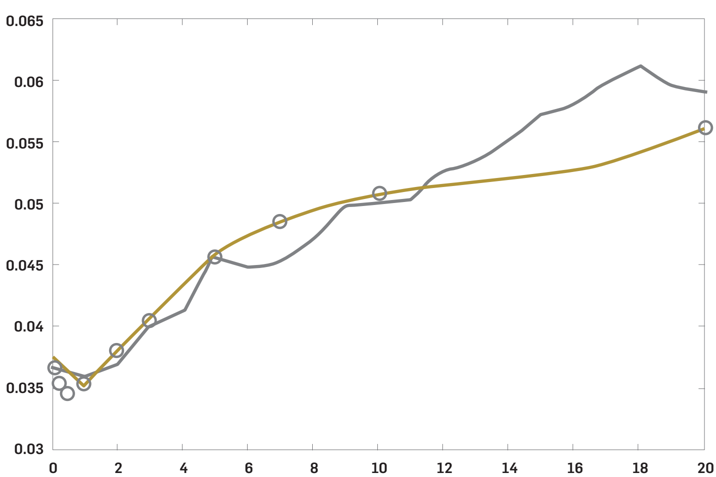
SOURCE: THE AUTHORS

Concerning term structure estimation methods, Zangari (1977) then Lin (2002) in the footsteps of Zangari (1977) classifies term structure estimation methods into two groups: theoretical (see section 2.2) and empirical (see section 2.3). Examples of the first group that models the short interest rate are Vasicek (1977) and Cox, Ingersoll and Ross (CIR, 1985). The latter model is at the epicenter of our paper. Although a one-factor model – the short-term interest rate –, we believe that a methodology implying a simplistic model is easier to reproduce by market participants and requires fewer assumptions about market factors such as market risk or correlation between factors. In this paper, we simulate the yield curve by Monte Carlo simulation using the CIR model. The second group of empirical methods is independent of any model or theory of the term structure and just tries to find a close representation of the term structure at any point in time, given some observed interest rate data. Interpolation and bootstrapping belong to this group. Finally, a mix of the two groups can be seen in Hull and White (1990) and Heath *et al.* (1990) that use an empirically determined yield curve in a theoretical model. Hagan and West (2006) survey interpolation algorithms that market participants apply for construction of curves such as forward curves, basis curves, and yield curves. They review the issue of bootstrapping yield curves. In addition, they introduce two new interpolation methods: the monotone convex method and the minimal method, which may solve problems inherent to past methods. In our paper, we apply the cubic spline interpolation to the observed yield curve. In our sample, daily yield curves are built from ten observable market yields of U.S. Treasury

Fitting the Pareto-Lévy distribution on the yield curve: an application to forecasting. *Rostan, P. and Rostan A.*
AESTIMATIO, THE IEB INTERNATIONAL JOURNAL OF FINANCE, 2014. 8: 38-67

securities (bills and notes) at 1-, 3-, 6-month, 1-,2-,3-,5-,7-,10- and 20-year constant maturity. Our assumption is that cubic spline of the ten interpolated points should stick more to reality than linear interpolation. Keller-Ressel and Steiner (2008) show that the two classical Cox-Ingersoll-Ross (1985) and Vasicek (1977) models are not flexible enough to accommodate shapes of yield curves that are different from normal, inverse or humped: 'curves with a dip (a local minimum), curves with a dip and a hump' are examples of shapes not captured by these models. They conclude that with two-factor models, yield curves with more complex shapes, including a dip for instance, may be obtained. Dealing with two-factor models such as Fong and Vasicek (1992) or Longstaff and Schwartz (1992) is obviously not as tractable as dealing with one-factor models. For example, the CIR one-factor model is parsimonious and widely-accepted by market participants since it is a good compromise between presenting some degree of complexity and being easy to use. In addition, the CIR model has a decisive advantage at step 2 of the methodology: the innovation terms of five hundred interpolated points of the yield curve are deduced from the CIR equation (refer to equation 3). A 2-factor model would make this step problematic, especially when two innovation terms are involved and correlated with some degree. However, we must acknowledge the limitation of the CIR model: the dip or local minimum, as underlined by Keller-Ressel and Steiner (2008), is observed 25% of the time in our sample of 2,707 yield curves where 685 curves display this feature. For example, Figure 6 shows the observed yield curve with a 6-month yield as local minimum on July 31, 2001. As expected the simulated yield curve obtained with the CIR model smoothes the dip, which epitomizes the clear drawback of the model.

Figure 6. **20-point cubic spline interpolation of the observed yield curve on July 31, 2001 and one simulated yield curve with the CIR model when α = 0.0583; μ = 0.0643, σ = 0.0056**



SOURCE: THE AUTHORS

2.2. Theoretical methods

'Theoretical term structure methods typically posit an explicit structure for a variable known as the short rate of interest, whose value depends on a set of parameters that might be determined using statistical analysis of market variables' (Hagan and West, 2006). Finding the optimal distribution of innovation of the yield curve depends specifically on the theoretical model that will be employed to simulate the yield curve. For example, by using the CIR model, we deduce the innovation terms from equation 3. Thus, the values of the three parameters α, μ and σ have a direct impact on the value of the innovations terms. The theory about modeling interest-rate term structure suggests that the evolution of the yield curve shape is affected by the level of interest rates, the slope of the term structure, the curvature and the volatility of the changes: for example read Litterman and Scheinkman (1991), Chen and Scott (1993), Dai and Singleton (2000), and De Jong (2000). The term structure relates to the relationship between the interest rates that shape the yield curve. Vasicek (1977) proposed a mean reverting process of the short term interest rate:

$$dr = \alpha\,(\mu - r)dt + \sigma\, dz_t \tag{1}$$

In equation 1, α is the speed of mean reversion, μ is the long-term average to which the short rate is reverting, and σ is the instantaneous volatility of the short rate. All parameters are assumed constant overtime. Vasicek and Fong (1982) proposed to model the term structure using exponential splines. For example, the widely used Cox, Ingersoll and Ross model (CIR, 1985) involves the short rate and its variance to be proportional to the level of the short rate:

$$dr = \alpha\,(\mu - r)dt + \sigma\,\sqrt{r}\, dz_t \tag{2}$$

From equation 2, the innovation term is deduced by rearranging the terms of the equation:

$$\varepsilon = \frac{dr - \alpha\,(\mu - r)dt}{\sigma\,\sqrt{r}\,\sqrt{dz_t}} \tag{3}$$

Therefore the choice of the model has a clear impact on the distribution of innovation terms. A more complex model could be for example Audrino and De Giorgi (2007) model which is 'an affine term structure model that accommodates nonlinearities in the drift and volatility function of the short-term interest rate. They derive iterative closed-form formula for the whole yield curve dynamics that can be estimated using a linearized Kalman filter.' Another example is presented by Audrino and Trojani (2007) who proposed 'a Functional Gradient Descent (FGD) estimation of the conditional mean vector and covariance matrix of a multivariate interest rate series. They apply filtered historical simulation to compute reliable out-of-sample yield curve scenarios and confidence intervals. They back-test their methodology on daily USD bond data for forecasting horizons from 1 to 10 days.'

2.3. Empirical methods

McCulloch (1971) pioneered the estimation of the term structure where coupon payments were included explicitly in a formal way. He first used quadratic splines which could be estimated by linear regression, then used cubic splines (1975). A problem with this approach, as Shea (1984, 1985) noted, is that the forward rate can become negative. In addition, Shea showed that the resulting yield function 'tends to bend sharply towards the end of the maturity range observed in the sample'. Mansi *et al.* (2001) proposed an exponential function to model the term structure that 'depends on the estimation of four parameters fitted by nonlinear least squares. In comparing the proposed model with current yield-curve-smoothing models, they found that the proposed model does best overall in terms of pricing accuracy both in sample and out of sample'. Other 'empirical studies have suggested that the evolution of the term structure of interest rates would be driven by the dynamics of several factors which can be represented by macroeconomic shocks or be related to the level, slope, and curvature' (Hong, 2001). To meet these empirical evidences, authors developed multifactor models such as Bliss (1997), Andersen *et al*. (1997), Dai and Singleton (2000) and Duffee (2002). Nelson and Siegel (1987) fitted the observed yield curve with a function of the time to maturity of the fixed income securities.

The model has later been modified by Svensson (1995) who estimated the forward rates mainly using the original Nelson and Siegel model but, in some cases using an extended version. Dolan (1999) argued that 'the curvature parameter of the yield curve, estimated using the Nelson and Siegel (1987) model, could be predicted using simple parsimonious models'. Fabozzi *et al.* (2005) tested for statistical significance in the predictive power of the Nelson and Siegel model when forecasting yield curve. Bernadell *et al.* (2005) revisited an early version of Diebold and Li paper (2003) by adding a regime-switching expansion. Diebold and Li (2006) applied the Nelson and Siegel model to forecasting by predicting the three factors which determine the shape of the yield curve with autoregressive models. Their model was encouraging for forecasting horizons longer than 24 months. In our paper, we choose the Diebold and Li (2006) model as benchmark at the chapter of forecasting the yield curve.

2.4. Exploring the optimal distribution that fits the yield curve: the Pareto-Lévy distribution

The Pareto-Lévy (1925) distribution belongs to the family of alpha-stable distribution (Veillette, 2012) that includes the Gaussian, the Cauchy and the Pareto-Lévy distributions. It is a four-parameter distribution: 'the first parameter $\alpha \in [0,2]$ called the characteristic exponent describes the tail of the distribution. The second parameter is the skewness. It specifies if the distribution is right skewed when $\beta > 0$ or left skewed when $\beta < 0$. The last two parameters are the scale (related to the variance) $\gamma > 0$, and the location $\delta \in \mathbb{R}$ (the mean).' This distribution is increasingly popular in finance (Wilmott, 2009) because it fits most of the financial data with fat tails. It has also the advantage to be a stable distribution, i.e. the sum of independent random numbers drawn from the distribution follows a Pareto-Lévy distribution itself. This is a useful property for the distribution of returns. The normal distribution is a special case of the Pareto-Lévy distribution when $\alpha=2$ and $\beta=0$, and with the parameter γ equal to half of the variance. In this paper, our objective is to identify the optimal Pareto-Lévy distribution with the four parameters that allow the simulated yield curve to fit the observed yield curve. Our intuition lays on Figure 7 representing the distribution of the innovation terms obtained from the interpolated observed yield curve on July 31, 2001. The distribution is negatively skew with fat tails. The Pareto-Lévy should be a good candidate to capture these features of the distribution. Mittnika and Rachevb (1993) discuss the fact that in economics and finance literature, stable distributions are virtually exclusively associated with stable Paretian distributions; in their paper, 'they adopt a more fundamental view and extend the concept of stability to a variety of probabilistic schemes. These schemes give rise to alternative stable distributions, which they compare empirically using S&P 500 stock return data. In this comparison, the Weibull distribution, associated with both the nonrandom-minimum and geometric-random summation schemes dominates the other stable distributions considered-including the stable Paretian model'. Jin (2007) proposes 'to fit the stable distribution to corn cash and futures prices. The stable distribution has been used as a generalized distributional model that can explain the distributions of asset returns significantly better than the conventional normal distribution. Using the stability-under-addition test, Cornew *et al.* (1984) and So (1987) confirm that a better correspondence for futures price changes is usually obtained when using the stable distribution; that is, the distribution more adequately describes futures price changes (in particular, heavy tails) than does the normal distribution'. Finally, Coronel-Brizio and Hernández-Montoya (2005) have successfully used the Pareto-Lévy distribution to describe probabilities associated to extreme variations of stock markets indexes such as the New York Stock Exchange (DJIA) and the Mexican Stock Market (IPC) indices.

3. Methodology

We work on a sample of 2,707 U.S. yield curves from July 31, 2001 to May 24, 2012. To summarize, the methodology has five steps. At step 1, we apply Kladivko's (2007) methodology to the calibration of the CIR model (equation 2). We apply a cubic spline interpolation to the observed yield curve to obtain dr. At step 2, we obtain the innovation terms from equation 3 using r and dr of the interpolated yield curve and α, μ and σ from step 1. We can thus plot a distribution of innovation terms for every daily observed yield curve of the sample. At step 3, we fit the distribution of innovation terms to Normal and Pareto-Lévy distributions. At step 4, we test the Normal and the Pareto-Lévy distributions with Monte Carlo simulation. The distributions are tested: 1) on the whole sample; 2) on each of the four subsamples (normal, humped, flat, inverted); 3) on a volatile versus non-volatile environment. Finally at step 5, we forecast the 20-day forward yield curve by simulating with Monte Carlo the CIR model (equation 2) a hundred times.

Step 1

We calibrate equation 2 with the daily observed yield curve, i.e. to find parameters α, μ and σ corresponding to the maximization of the log-likelihood function (equation 4). We apply Kladivko's (2007) methodology. The log-likelihood function of the CIR process is:

$$\ln L(\theta) = (N-1)\ln c + \sum_{i=1}^{N-1} \left\{ u_{t_i} + v_{t_{i+1}} + 0.5q \ln\left(\frac{v_{t_{i+1}}}{u_{t_i}}\right) + \ln\{I_q(2\sqrt{u_{t_i} + v_{t_{i+1}}})\} \right\} \tag{4}$$

Where $u_{t_i} = cr_{t_i} e^{-\alpha \Delta t}$ and $v_{t_{i+1}} = cr_{t_{i+1}}$. We find maximum likelihood estimates $\hat{\theta}$ of parameter vector θ by maximizing the log-likelihood function 4 over its parameter space:

$$\theta \equiv (\hat{\alpha}, \hat{\mu}, \hat{\sigma}) = \arg\max_{\theta} \ln L(\theta) \tag{5}$$

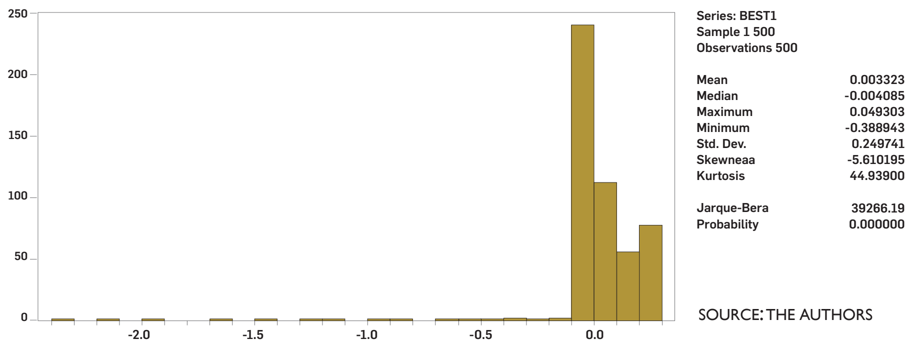
Since the logarithmic function is monotonically increasing, maximizing the log-likelihood function also maximizes the likelihood function. Refer to Kladivko's (2007) methodology, for the practical implementation of the calibration. For technical implementation, we interpolate the yield curve of ten observations with cubic spline (de Boor, 1978) to accommodate the Maximum Likelihood Estimate (MLE) that requires data with regular intervals. The choice of twenty points is purely empirical. If we use 10 points, we cannot fit curves with abnormal shapes. For example with 10 points, the dip observed on July, 31 2001 will not be captured by the interpolation. If we choose to interpolate with 500 points, the MLE returns odd estimates of α, μ and σ. The choice of 20 points offers tangible values and captures pretty well the dip. It is a good compromise. Besides, since 20 years / 20 points = 1 year, α, μ are annual rates and σ is the annual volatility of the interest-rate. Figure 6 illustrates the 20-point cubic spline interpolation and one Monte Carlo simulation of the interest rate when α = 0.0583, μ = 0.0643 and σ = 0.0056.

Step 2

We interpolate the observed yield curve with a 500-point cubic spline interpolation. We then extract a distribution of innovation terms from equation 3. In equation 3, r and dr are obtained from the interpolated yield curve of 500 points, α, μ and σ are computed at step 1; dt = 20 years / 500 = 0.04 year. We obtain a distribution of 500 innovations terms.

For example, Figure 7 represents the distribution of innovation terms of the interpolated yield curve observed on July 31, 2001. Based on the test of Jarque-Bera, since the probability is equal to 0%, the null hypothesis of a Normal distribution is rejected. The mean is 0.003323, the standard deviation equal to 0.249741, the skewness –5.61 and the kurtosis 44.939.

Figure 7. **Distribution of innovation terms of the 500-point interpolated yield curve on July 31, 2001 with α = 0.0583; μ = 0.0643, σ = 0.0056**



SOURCE: THE AUTHORS

The non-normality of the innovation terms pushes us to search an alternative distribution to the standard Normal distribution. The Pareto-Lévy distribution is a good candidate to capture skewness and heavy tails as explained in section 2.4 of the literature review.

Step 3

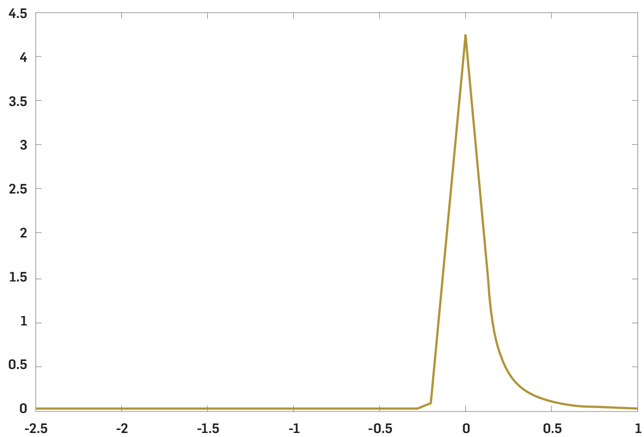
We fit the following theoretical distributions to the observed distribution obtained at step 2:

- Normal(μ ,σ) with μ the mean and σ the standard deviation.

- Pareto-Lévy(α, β, γ, δ) with α the characteristic exponent (the tail of the distribution); β the skewness; γ the scale (related to the variance); and δ the location (the mean).

"Fitting" means that we find the parameters of the theoretical distributions. The methodology of fitting differs with the distribution. We fit the observed distribution to the Normal distribution using the *normfit* function of Matlab. This function uses the iterative process of Maximum-Likelihood Estimation. With the Pareto-Lévy distribution, we apply the methodology and the algorithm of Veillette (2012) which fit the theoretical distribution to the observed distribution. The optimization is based on Koutrouvelis's (1980, 1981) method. For example, Figure 8 represents the PDF of the Pareto-Lévy (1.36858, 0.9033, 0.06372, 0.07981) fitted to the distribution of innovation terms of the interpolated yield curve on July 31, 2001. As we observe, Figures 7 and 8 share strong similarities.

Figure 8. **PDF of the Pareto-Lévy distribution (1.36858, 0.90332, 0.0637272, 0.0798125) fitted to the distribution of innovation terms of the interpolated yield curve of July 31, 2001**



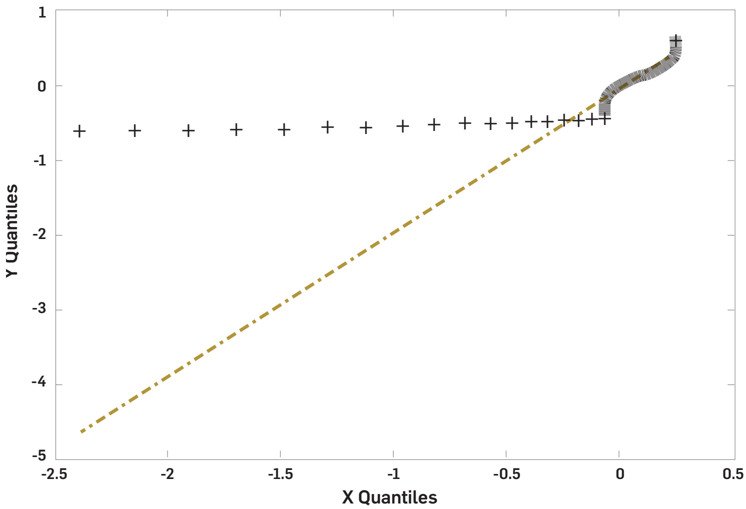
SOURCE: THE AUTHORS

We apply the Q-Q plot test of the Normal and Pareto-Lévy distributions to the observed distribution. 'The quantile-quantile plot[1] is a visual test of goodness of fit for determining whether two samples come from the same distribution (whether normally distributed or not): the plot will be linear if they come from the same distribution. The Q-Q plot has three graphical elements. The pluses are the quantile of each sample. By default the number of pluses is the number of data values in the smaller sample. The solid line joints the 25th and 75th percentiles of the samples. The dashed line extends the solid line to the extent of the sample. It is incorrect to interpret a linear plot as a guarantee that the two samples come from the same distribution. But, for assessing the validity of a statistical procedure that depends on the two samples coming from the same distribution (e.g., ANOVA), a linear quantile-quantile plot should be sufficient'. Figures 9 and 10 show that the left tail

[1] *http://www.mathworks.com/help/toolbox/stats/qqplot.html Accessed on May 16, 2013.*

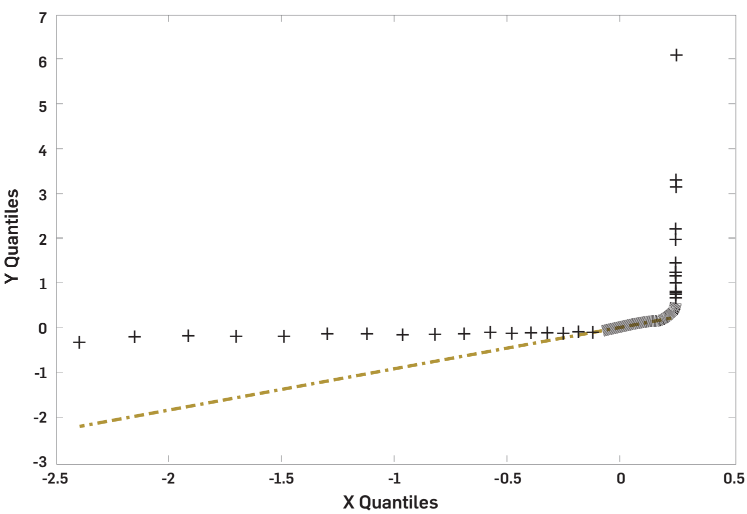
is better captured by the Pareto-Lévy distribution, whereas the Normal distribution captures well the right tail contrarily to the Pareto-Lévy distribution.

Figure 9. July 31, 2001: Quantile-Quantile (QQ) plot of the observed distribution (X-quantiles) versus the theoretical Normal distribution (0.0033228, 0.249740) (Y-quantiles)



SOURCE: THE AUTHORS

Figure 10. July 31, 2001: Quantile-Quantile (QQ) plot of the observed distribution (X-quantiles) versus the theoretical Pareto-Lévy distribution (1.3685, 0.9033, 0.06372, 0.07981) (Y-quantiles)



SOURCE: THE AUTHORS

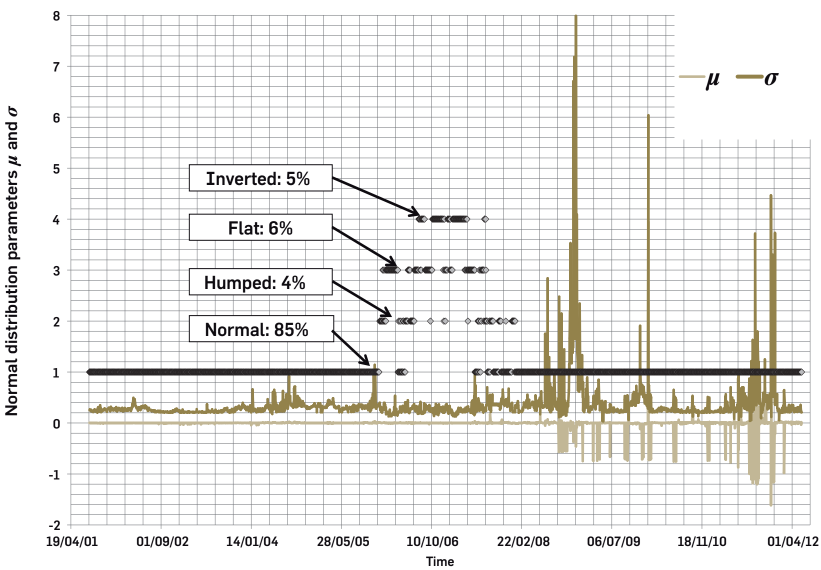
The fitting of the observed distribution of innovations terms give the following results in Table 3.

Table 3. Fitting the theoretical distributions Normal and Pareto-Lévy to the observed distribution on July 31, 2001

Theoretical distribution:	Normal	Pareto-Lévy
	$\mu = 0.0033228$	$\alpha = 1.36858$
	$\sigma = 0.249740$	$\beta = 0.9033$
		$\gamma = 0.06372$
		$\delta = 0.07981$

Figures 11 and 12 illustrate the variability of the distribution parameters over the sample of 2,707 yield curves when the distribution of observed innovation terms has been fitted to the Normal and Pareto-Lévy distributions.

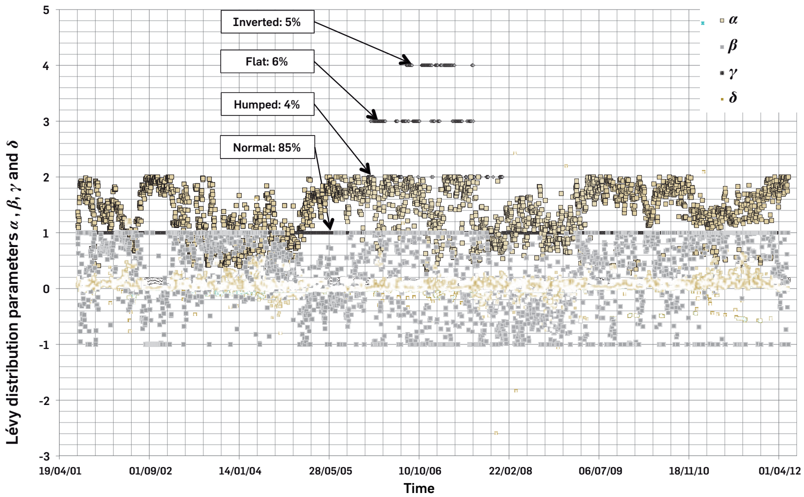
Figure 11. Variability of the Normal distribution parameters over the sample of 2,707 yield curves when the distribution of observed innovation terms has been fitted to the Normal distribution Long-term average of μ = –0.0043702; long-term average of σ = 0.34738



SOURCE: THE AUTHORS

From Figure 11, we observe that the parameters of the Normal distribution μ and σ are rather stable overtime with a long-term average of –0.0043702 for μ and 0.34738 for σ. Clearly, the type of yield curve (normal, humped, flat or inverted) has no impact on μ and σ. On the other side, the volatility of interest rates produced by the credit crisis has a clear impact on the variability of σ, starting to increase by July 2008 and reaching a peak at 7.98 in December 2008. μ is quiet stable overtime and centered on zero until September 2008 when it starts to enter in negative territory, steadily declining up to –1.61 in December 2011 before coming back to zero in February 2012. From Figure 11, we can conclude that the standard normal distribution may be a good candidate by default but on average the random numbers should be drawn from a Normal (–0.0043702, 0.34738).

Figure 12. Variability of the Pareto-Lévy distribution parameters over the sample of 2,707 yield curves when the distribution of observed innovation terms has been fitted to the Pareto-Lévy distribution. Long-term averages: α = 1.382977; β = 0.237807; γ = 0.094706; δ = 0.018099



SOURCE: THE AUTHORS

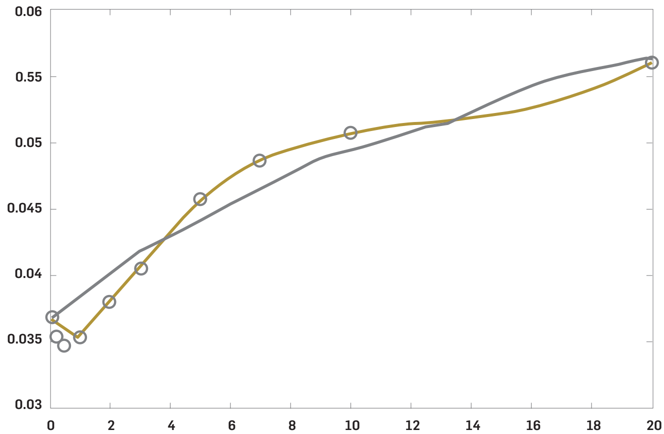
Finally from Figure 12, we observe the following long-term averages: α = 1.382977; β = 0.237807; γ = 0.094706; δ = 0.018099. γ is centered on zero all across the sample and does not vary much. δ is less stable but still centered on its long-term average that is close to zero. 99% of δ's are in the interval [–1,+1]. α and β are much more volatile. α represents the tails and oscillates in the interval [0,+2], β represents the asymmetry and spreads out in the interval [–1,+1]. One very interesting characteristic of α is the sinusoidal pattern across the sample that attests a degree of persistence in the time series that could be detected by the Hurst coefficient: high levels of α are followed by high levels, and conversely for low levels. During the credit crisis (2008-2009): 1) α is not at its highest – highest is equal to 2 – but is centred on 1; 2) β is more often in negative territory; 3) γ – the variance – sticks to zero, inversely to the Normal distribution which records higher variances (refer to Figure 14); 4) δ – the mean – records negative values and outliers. The period of 'non-normal' yield curves that we have mentioned between December 2005 and January 2008 is in a cycle of high α, widespread β and close-to-zero-values of γ and δ. Again, the type of yield curve (normal, humped, flat or inverted) does not seem to have an impact on the parameters.

Step 4

Using Monte Carlo simulation of the CIR model calibrated in step 1, we simulate a 100 times the yield curve with the two theoretical distributions that have been calibrated in step 3 to fit the observed distribution.

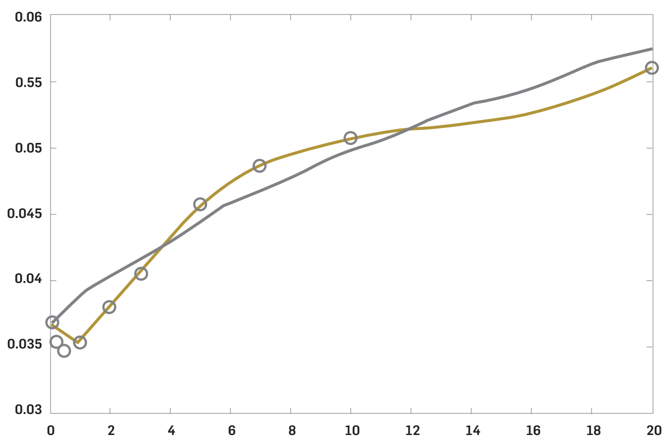
Figures 13 and 14 illustrate the best simulated yield curve out of one hundred that minimizes the *RMSE* using random numbers generators from the Normal and Pareto-Lévy distributions calibrated on July 31, 2001 as explained in step 3.

Figure 13. **July 31, 2001: One simulated yield curve with the CIR model with α = 0.0583; μ = 0.0643, σ = 0.0056, 500 steps, that minimizes *RMSE* = 0.0012 out of 100 simulations, Random numbers generated with calibrated N(0.00332,0.2497)**



SOURCE: THE AUTHORS

Figure 14. **July 31, 2001: One simulated yield curve with the CIR model with α = 0.0583; μ = 0.0643, σ = 0.0056, 500 steps, that minimizes *RMSE* = 0.0014 out of 100 simulations, Random numbers generated with calibrated Pareto-Lévy (1.3685, 0.9033, 0.0637, 0.07981)**



SOURCE: THE AUTHORS

In step 4, we answer the following questions:

- Between Normal and Pareto-Lévy, which distribution works best overall?
- Between Normal and Pareto-Lévy, which distribution works best for a given type of yield curve: normal, humped, flat and inverted?
- Between Normal and Pareto-Lévy, which distribution works best for a volatile/non-volatile interest rates environment?
- Can we generalize the results with a Normal or Pareto-Lévy distribution?

Step 5

We forecast the 20-day forward yield curve by simulating the CIR model (equation 2) a hundred times using a variance reduction technique called the stock dog technique (see section 3.5). The random numbers generators are based on the five distributions presented at step 4. We produce yield forecasts based on an underlying univariate AR(1) specification, as:

$$dr_{t+h/t} = \alpha_{t+h/t}\,(\mu_{t+h/t} - r)dt + \sigma_{t+h/t}\sqrt{r}\,dz_1 \quad (6)$$

where:

$$\alpha_{t+h/t} = C_1 + \omega_1\alpha_t \quad (7)$$

$$\mu_{t+h/t} = C_2 + \omega_2\mu_t \quad (8)$$

$$\sigma_{t+h/t} = C_3 + \omega_3\sigma_t \quad (9)$$

C_i and ω_i are obtained by regressing α_t on an intercept and α_{t-h}, μ_t on an intercept and μ_{t-h}, and finally σ_t on an intercept and σ_{t-h}. The forecasting horizon is $h = 20$ days. We regress the first set of α, μ, σ from 1 to 250 days with the set of 250 α, μ, σ obtained between 20 and 270 days, then moving forward one-day at a time. The resulting α, μ, σ are then plugged in equation 10 in order to compute the forecasted yield curve in 20 days. During the simulation of equation 6, we have 10 variable time steps dt with $dt = [0.0833, 0.1667, 0.25, 0.5, 1, 1, 2, 2, 3, 10]$. We simulate equation 6 a hundred times and compute the average simulated yield curve that becomes the 20-day forecast.

We apply the *RMSE* criteria (equation 10) to the out-of-sample of 2,416 20-day forecasted yield curves from September 27, 2002 to May 24, 2012:

$$RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(Forecasted\ yield_i - Observed\ yield_i)^2} \quad (10)$$

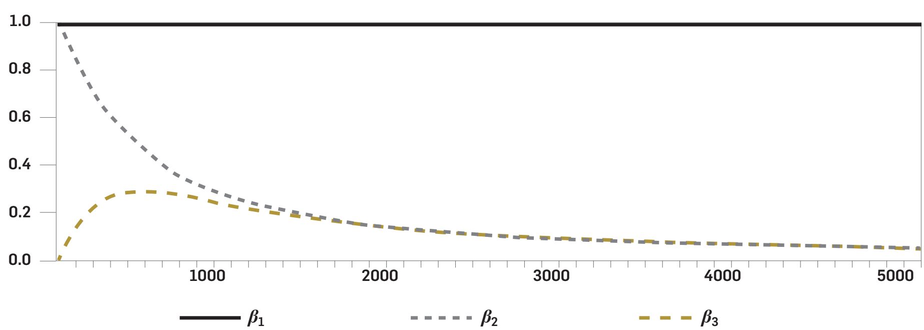
3.5. The stock dog technique

Our paper uses a variance reduction technique called the stock dog technique, developed by Rostan and Rostan (2012) to improve the Monte Carlo simulation results. This technique consists in building upper and lower bands during the simulation, using the information embedded in the shapes of the most 20 recent yield curves. These bands are built to reflect the intrinsic dynamic forces of the interest rate market that is responsible of the future shape of the term structure. The choice of the bands is based on the assumption that the current dynamic forces of the interest rate market are captured by the level β_{1t}, the slope β_{2t} and the curvature β_{3t} provided by the Nelson and Siegel model from the 20 most recent daily fitted U.S. Treasury yield curves. Therefore, the future yield curve shape in 20 days depends partly on the shapes of the most recent yield curves. Market news occurring in the next 20 days will bring the final touch to the shape. The choice of the 20 past days is conditional to the forecasting horizon of 20 days and should be adjusted to the forecasting horizon. We pick β_{1t}, β_{2t}, β_{3t}, independently among the 20 most recent days such as that, based on equation 11, their combination maximizes $y_t(\tau)$ to obtain the upper band and minimizes $y_t(\tau)$ to obtain the lower band. Nelson and Siegel model (1987) fit the yield curve using a three-factor model:

$$y_t(\tau) = \beta_{1t} + \beta_{2t}\,(1-e^{-\lambda_t\tau})/(\lambda_t\tau) + \beta_{3t}\,\{(1-e^{-\lambda_t\tau})/(\lambda_t\tau) - e^{-\lambda_t\tau}\} \quad (11)$$

The loadings (i.e. coefficients) of β_{1t}, β_{2t}, β_{3t}, are a function of time and are graphed in Figure 15.

Figure 15. **Factor loadings of Nelson and Siegel (1987) model for λ = 0.91**



SOURCE: THE AUTHORS

From equation 11, the loading of β_{1t} is equal to the constant 1, the loading of β_{2t} is $(1-e^{-\lambda_t\tau})/(\lambda_t\tau)$ and the loading of β_{3t} is $\{(1-e^{-\lambda_t\tau})/(\lambda_t\tau) - e^{-\lambda_t\tau}\}$. The corresponding function of β_{2t} loading starting at 1 decreases gradually overtime. The corresponding function of β_{3t} loading starts at time zero, then increases to reach a maximum at 1.97 year (when λ is set at 0.91), and finally decreases steadily. We obtain λ over a period of 250 days from the first 250 U.S. Treasury yield curves of our database (7/31/2001 to 7/31/2002) by minimizing the average *RMSE* over the period. We find estimates $\hat{\theta}$ of parameter vector θ by optimizing equations 12 and 13 over their parameter spaces:

$$\theta \equiv (\widehat{\beta_{1t}}, \widehat{\beta_{2t}}, \widehat{\beta_{3t}}) = \arg\max_{\theta} \textstyle\sum_{i=1}^{10} y_i(\tau)(\theta) \quad (12)$$

Equation 12 is used to obtain the upper band.

$$\theta \equiv (\widehat{\beta_{1t}}, \widehat{\beta_{2t}}, \widehat{\beta_{3t}}) = \arg\min_{\theta} \textstyle\sum_{i=1}^{10} y_i(\tau)(\theta) \quad (13)$$

Equation 13 is used to obtain the lower band.

With $\text{Min}(\beta_{i,t-20:t}) < (\widehat{\beta_{1t}}) < \text{Max}(\beta_{i,t-20:t})$ for $i = 1, 2, 3$.

And $y_i(\tau)$ obtained from equation 11.

The intuition behind the bands is that, in the last 20 days, the more volatile the interest rate market was, the wider the range of β_i values was, the wider the bands will be. With a volatile market, the shape of the yield curve in 20 days is more likely to be different from the ones of the most recent yield curves, since the number of possible combinations of β_{1t}, β_{2t}, β_{3t} increases. Inversely, with a low-volatility market, the narrower the range of β_i values was, the tighter the bands will be. With a limited number of combinations of β_i, the shape of the yield curve in 20 days is more likely to look the same as the most recent shapes. Therefore, the bands provided by the "stock dog" technique behave very much like Bollinger bands to reflect the volatility of the market (refer to Théoret *et al.*, 2005) but capture additional information related to the shape of the future yield curve. In conclusion, the upper and lower bands are reasonable limits constructed on intrinsic information obtained from the shapes of the 20 most recent yield curves, beyond which the current market forces make the 20-day forecasted yield curve unlikely to be.

We benchmark the CIR model coupled with the stock dog technique to the Diebold and Li (2003, 2006) forecasting technique. Diebold and Li (2006) forecast the Nelson and Siegel factors as univariate AR(1) process. They produce yield forecasts based on an underlying univariate AR(1) specification, as:

$$y_{t+h/t}(\tau) = \beta_{1,t+h/t} + \beta_{2,t+h/t}(1-e^{-\lambda\tau})/(\lambda\tau) + \beta_{3,t+h/t}((1-e^{-\lambda\tau})/(\lambda\tau) - e^{-\lambda\tau}) \quad (14)$$

where:

$$\beta_{i,t+h/t} = C_i + \omega_i \beta_{it} \quad (15)$$

C_i and ω_i are obtained by regressing β_{it} on an intercept and $\beta_{i,t-h}$.

In our paper, the forecasting horizon is $h = 20$ days. We regress the first set of β_{1t}, β_{2t}, β_{3t} from 1 to 250 days with the set of 250 $\{\beta_{1t}, \beta_{2t}, \beta_{3t}\}$ obtained between 20 and 270 days, then moving forward one-day at a time. The resulting $\{\beta_{1t}, \beta_{2t}, \beta_{3t}\}$ are then plugged in equation 11 in order to compute the forecasted interest-rate in 20 days corresponding to each maturity of the term structure.

3.6. Database

The database includes market yields of U.S. Treasury securities (bills and notes) at 1-, 3-, 6-month, 1-,2-,3-,5-,7-,10- and 20-year constant maturity, quoted on investment basis yields on actively traded non-inflation-indexed issues adjusted to constant maturities. The U.S. yield curves extend from July 31, 2001 to May 24, 2012 or 2,707 days and are obtained from the Federal Reserve website[2]. We forecast 2,418 U.S. yield curves from Sept. 27, 2002 to May 24, 2012. Since the 30-year Treasury constant maturity series was discontinued on February 18, 2002, and reintroduced on February 9, 2006, we discard the 30-year maturity.

We divide the database in four sub samples. We classify the four occurrences of the U.S. yield curves, using the criteria presented in Table 1. We count the occurrences of our sample of 2,707 daily yield curves in Table 2.

4. Results

As explained at step 4, we compute the *RMSE* between the simulated yield curve and the observed yield curve. When simulating the curve, we use either the Normal distribution or the Levy distribution that fit the observed distribution of the daily yield curve. Calibrating the Pareto-Lévy distribution at step 3 of the methodology leads to abnormal results in 2.4% of the time (64 out of 2,707 yield curves) when at step 4 we simulate 100 simulations and keep the simulation with the lowest *RMSE*. Obviously a high *RMSE* is explained by a poor model calibration. In our paper, we consider a high *RMSE* when its value is higher than the highest *RMSE* recorded in the Normal distribution sample, i.e. equal to 3.54. When calibrating the Pareto-Lévy distribution, the optimization is based on Koutrouvelis's (1980, 1981) method which may not accommodate all types of yield curve (normal, humped, flat and inverted). Indeed, out of 64 outliers, 62 are normal curve (97%) and 2 inverted curves (3%). We may deduce that humped and flat curve have no problem with Koutrouvelis's calibration process.

Removing the 64 outliers, the results are gathered in Table 4.

[2] *http://www.federalreserve.gov/releases/h15/data.htm Accessed on May 16, 2013.*

Table 4. Computing the average *RMSE* of the best simulated yield curve out of 100 (observed yield curve is the benchmark) over 2,643 days (we remove 61 outliers from our initial sample) for the Normal and the Pareto-Lévy distributions. Impact on *RMSE* of the type of curve (normal, humped, flat, inverted) and the volatile environment on the average *RMSE*

Type of Distribution used as Random Numbers Generator (1 to 5)	1. Normal (parameters calibrated daily)	2. Normal (Long-term Average) μ = -0.004370 σ = 0.34738	3. Pareto-Lévy (Long-term Average) α = 1.382977 β = 0.237807 γ = 0.094706 δ = 0.018099	4. *N*(0,1)	5. Pareto-Lévy (parameters calibrated daily)	Test for Equality of Means Between Series: **Type of curve** (Standard error) Anova *F*-statistic *p*-value
Average *RMSE* (2,646 obs.) (Standard error of Mean)	0.00516 (0.000492)	0.00768 (0.000937)	0.00797 (0.000971)	0.00889 (0.001052)	0.01574 (0.001853)	(0.000515) 11.89687 0.0000
Average *RMSE* Normal curve (2,237 obs.) (Standard error)	0.00591 (0.000580)	0.00887 (0.001105)	0.00920 (0.001145)	0.01019 (0.001241)	0.01714 (0.002107)	(0.000596) 9.807182 0.0000
Average *RMSE* - Humped curve (104 obs.) (Standard error)	0.00147 (7.84 e-0.5)	0.00155 (9.44 e-0.5)	0.00145 (8.49 e-0.5)	0.00216 (0.000129)	0.01932 (0.010609)	0.002137 2.773098 0.0266
Average *RMSE* - Flat curve (168 obs.) (Standard error)	0.00088 (2.15 e-0.5)	0.0010 (2.05 e-0.5)	0.0011 (7.75 e-0.5)	0.0015 (3.09 e-0.5)	0.0015 (0.000210)	(4.64 e-0.5) 8.771787 0.0000
Average *RMSE* - Inverted curve (133 obs.) (Standard error)	0.00094 (2.94 e-0.5)	0.0011 (2.92 e-0.5)	0.0011 (4.82 e-0.5)	0.0016 (4.54 e-0.5)	0.0072 (0.005337)	(0.001069) 1.307836 0.2656
Type of Distribution used as Random Numbers Generator (1 to 5)	**1. Normal (parameters calibrated daily)**	**2. Normal (Long-term Average) μ = -0.004370 σ = 0.34738**	**3. Pareto-Lévy (Long-term Average) α = 1.382977 β = 0.237807 γ = 0.094706 δ = 0.018099**	**4. *N*(0,1)**	**5. Pareto-Lévy (parameters calibrated daily)**	Test for Equality of Means Between Series: **Volatile versus non-volatile environment** - (Standard error) Anova *F*-statistic *p*-value
Average *RMSE* - Non-volatile environment (2,260 obs.) (Standard error)	0.00362 (0.000421)	0.00481 (0.000726)	0.00499 (0.000740)	0.00583 (0.000839)	0.01455 (0.002010)	(0.000491) 16.3375 0.0000
Average *RMSE* - volatile environment (382 obs.) (Standard error)	0.01431 (0.002264)	0.02469 (0.004762)	0.02562 (0.005006)	0.02697 (0.005230)	0.02274 (0.004766)	(0.002030) 1.226155 0.2976

Overall, based on a test for equality of means, values are significantly different and the ranking is as follows:

1. Normal distribution with parameters calibrated daily;
2. Normal distribution with long-term averages of parameters given μ = -0.004370; σ = 0.34738;
3. Pareto-Lévy distribution with long-term averages of parameters given α = 1.382977; β = 0.237807; γ = 0.094706; δ = 0.018099;
4. Standard Normal distribution *N*(0,1);
5. Pareto-Lévy distribution with parameters calibrated daily.

Fitting the Pareto-Lévy distribution on the yield curve: an application to forecasting. *Rostan, P. and Rostan A.*
AESTIMATIO, THE IEB INTERNATIONAL JOURNAL OF FINANCE, 2014. 8: 38-67

We conclude that, in order to use Monte Carlo simulation to fit the yield curve, we should fit the observed distribution of innovation terms with the Normal distribution that has parameters calibrated daily. Using the Normal distribution with $\mu = -0.004370$ and $\sigma = 0.34738$ (long-term averages of parameters) offers a good alternative. Obviously, if we seek fairly accurate results, the standard Normal distribution makes the deal. We note that the Pareto-Lévy with long-term averages of parameters is beating the standard normal distribution. In addition, we analyze the results based on the type of yield curve and the volatility of the environment. We observe from Table 4 that the type of yield curve has no impact on the performance of the distributions. The Normal distribution calibrated daily consistently outperforms the others (except for the humped curve where the Pareto-Lévy with long-term averages of parameters is best). Concerning the volatility of the environment, we identify volatile periods from Figure 11 where periods with high standard deviations computed from the fitting of the Normal distribution indicates volatile periods: the interest rate environment is highly volatile between June 5, 2008 to April 23, 2009, November 13, 2009 to January 21, 2010, May 27, 2011 to December, 19 2011. We have 408 volatile days out of 2,643 (14% of the sample). The remaining periods displays a low-volatile environment. Again, whatever the volatility of the period, the Normal distribution calibrated daily consistently outperforms the others.

We may conclude that whatever the shape of the yield curve or the degree of volatility of the interest-rate environment, the Normal distribution calibrated daily is best to fit the yield curve using Monte Carlo simulation.

We test the Pareto-Lévy and the Normal distributions in a second test (step 5 of the methodology) that measures their ability to forecast the yield curve in 20-day. Table 5 illustrates the average *RMSE* of the 20-day forecasted yield curves versus the observed yield curve over 2,416 days with random numbers drawn from the Normal and the Pareto-Lévy distributions using Monte Carlo simulation coupled to the stock dog technique. In addition, Table 5 illustrates the impact of 1) the **type of curve** (normal, humped, flat, inverted) and 2) the **volatile environment** on the average *RMSE*.

Overall, Table 5 shows that the standard Normal distribution N(0,1) used as random number generator (column 4) offers the lowest *RMSE* and therefore the highest precision in forecasting the 20-day forward yield curve, whatever the shape of the yield curve -except for flat curves where the Diebold and Li (2006) model is best- and whether the interest-rate environment is volatile or not.

Based on the Test for Equality of Means between Series, line 1 of Table 5 suggests that over 2,416 days, the average *RMSE* is not statistically different between the forecasts of the 5 random numbers generators and the Diebold and Li (2006) forecasts. Although, based on standard errors, the standard Normal distribution

$N(0,1)$ is not statistically a better random number generator than the Normal distribution with parameters calibrated daily, the Normal distribution with long-term averages of parameters and the Pareto-Lévy distribution with parameters calibrated daily, it statistically offers better yield curve forecasts than the Diebold and Li model (Table 6).

Table 5. Computing the average *RMSE* of 20-day forecasted yield curves versus the observed yield curve over 2,416 days using Monte Carlo simulation coupled to the stock-dog technique and random numbers drawn from the Normal and the Pareto-Lévy distributions. Impact of the type of curve (normal, humped, flat, inverted) and the volatile environment on the average *RMSE*

Type of Distribution used as Random Numbers Generator (1 to 5)	1. Normal (parameters calibrated daily)	2. Normal (Long-term Average computed over 250 days) $\mu = 0.0008162$ $\sigma = 0.273843$	3. Pareto-Lévy (Long-term Average computed over 250 days) $\alpha = 1.3707$ $\beta = 0.4504$ $\gamma = 0.08983$ $\delta = 0.06591$	4. $N(0,1)$	5. Pareto-Lévy (parameters calibrated daily)	6. Diebold and Li forecasts	Test for Equality of Means Between Series: **Type of curve** (Standard error) Anova *F*-statistic *p*-value
Line 1: Average *RMSE* - (2,416 obs.) (Standard error)	0.003010 (3.44 e-05)	0.003031 (3.50 e-05)	0.003038 (3.51 e-05)	0.002939 (3.45 e-05)	0.003025 (3.47 e-05)	0.003086 (4.19 e-05)	(1.47 e-05) 1.767885 0.1157
L 2: Average *RMSE* - Normal curve (2009 obs.) (Standard error)	0.003163 (3.90 e-05)	0.003188 (3.98 e-05)	0.003195 (3.99 e-05)	0.003085 (3.94 e-05)	0.003180 (3.94 e-05)	0.003255 (4.83 e-05)	(1.68 e-05) 1.799374 0.1093
L 3: Average *RMSE* - Humped curve (104 obs.) (Standard error)	0.002934 (0.000160)	0.002941 (0.000162)	0.002940 (0.000163)	0.002903 (0.000155)	0.002934 (0.000161)	0.003083 (0.000156)	(6.49 e-05) 0.161437 0.9765
L 4: Average *RMSE* - Flat curve (168 obs.) (Standard error)	0.002171 (4.67 e-05)	0.002168 (4.66 e-05)	0.002165 (4.60 e-05)	0.002143 (4.70 e-05)	0.002177 (4.67 e-05)	0.001989 (4.56 e-05)	(1.90 e-05) 2.447397 0.0324
L 5: Average *RMSE* - Inverted curve (135 obs.) (Standard error)	0.001836 (5.40 e-05)	0.001832 (5.43 e-05)	0.001849 (5.47 e-05)	0.001783 (5.39 e-05)	0.001849 (5.41 e-05)	0.001940 (5.91 e-05)	(2.25 e-05) 0.856461 0.5100
Type of Distribution used as Random Numbers Generator (1 to 5)	1. Normal (parameters calibrated daily)	2. Normal (Long-term Average computed over 250 days) $\mu = 0.0008162$ $\sigma = 0.273843$	3. Pareto-Lévy (Long-term Average computed over 250 days) $\alpha = 1.3707$ $\beta = 0.4504$ $\gamma = 0.08983$ $\delta = 0.06591$	4. $N(0,1)$	5. Pareto-Lévy (parameters calibrated daily)	6. Diebold and Li forecasts	Test for Equality of Means Between Series: **Volatile versus non-volatile environment** (Standard error) Anova *F*-statistic *p*-value
L 6: Average *RMSE* – Non-volatile environment (2,008 obs.)	0.002900 (3.54 e-05)	0.002902 (3.54 e-05)	0.002908 (3.54 e-05)	0.002818 (3.52 e-05)	0.002904 (3.52 e-05)	0.002836 (3.38 e-05)	(1.43 e-05) 1.312181 0.2555
L 7: Average *RMSE* - Volatile environment (408 obs.)	0.003547 (0.000101)	0.003665 (0.000107)	0.003674 (0.000108)	0.003535 (0.000104)	0.003624 (0.000106)	0.004316 (0.000171)	(4.88 e-05) 6.122095 0.0000

Table 6. **Test for Equality of Means of *RMSE*s between forecasting the yield curve with 1) Monte Carlo simulation coupled to the stock-dog technique and random numbers drawn from *N*(0,1) and 2) Diebold and Li**

Test for Equality of Means Between forecasting *RMSE* of *N*(0,1) and Diebold and Li	Random Numbers Generator from the *N*(0,1) distribution	Diebold and Li forecasts	Test for Equality of Means Between Series: **Type of curve** - (Standard error) Anova *F*-statistic *p*-value
Line 1: Average *RMSE* - (2,416 obs.) (Standard error)	0.002939 (3.45 e-05)	0.003086 (4.19 e-05)	(2.72 e-05) 7.329329 0.0068
L 2: Average *RMSE* - Normal curve (2009 obs.) (Standard error)	0.003085 (3.94 e-05)	0.003255 (4.83 e-05)	(3.12 e-05) 7.417389 0.0065
L 3: Average *RMSE* - Humped curve (104 obs.) (Standard error)	0.002903 (0.000155)	0.003083 (0.000156)	(0.000110) 0.675739 0.4120
L 4: Average *RMSE* - Flat curve (168 obs.) (Standard error)	0.002143 (4.70 e-05)	0.001989 (4.56 e-05)	(3.30 e-05) 5.527253 0.0193
L 5: Average *RMSE* - Inverted curve (135 obs.) (Standard error)	0.001783 (5.39 e-05)	0.001940 (5.91 e-05)	(4.02 e-05) 3.808815 0.0520
Test for Equality of Means Between forecasting *RMSE* of *N*(0,1) and Diebold and Li	**Random Numbers Generator from the *N*(0,1) distribution**	**Diebold and Li forecasts**	**Test for Equality of Means Between Series: Volatile versus non-volatile environment - (Standard error) Anova *F*-statistic *p*-value**
L 6: Average *RMSE* – Non-volatile environment (2,008 obs.)	0.002818 (3.52 e-05)	0.002836 (3.38 e-05)	(2.44 e-05) 1.139280 0.7090
L 7: Average *RMSE* - Volatile environment (408 obs.)	0.003535 (0.000104)	0.004316 (0.000171)	(0.000101) 15.14013 0.0001

Since overall, the standard Normal distribution N(0,1) offers the lowest average *RMSE* than the four other distributions (even if most of the time its value is not statistically different than 3 of them), we may compare the performance of the standard Normal distribution N(0,1) and the Diebold and Li model (Table 6). From this table, we conclude that the standard Normal distribution is best for all types of yield curve shape, except for humped and inverted curves where the two models offers the same forecast accuracy and during non-volatile environments where again the two models have equivalent forecasting power. In volatile environments, the standard Normal distribution N(0,1) returns better forecasts than the Diebold and Li model.

Finally, we can answer the following questions presented at step 4. We based our answers on Table 5:

- Between Normal and Pareto-Lévy, which distribution works best overall?
  Overall, the Pareto-Lévy (with parameters calibrated daily) and the Normal

distribution (Standard Normal, Normal with parameters calibrated daily and Normal with long-term averages of parameters) offer the same degree of accuracy in forecasting the 20-day forward yield curve.

- Between Normal and Pareto-Lévy, which distribution works best for a given type of yield curve: normal, humped, flat and inverted?
  Statistically, Table 5 shows that the Normal distribution (standard) and the Pareto-Lévy (calibrated daily) offer the same accuracy in forecasting the 20-day forward yield curve whatever the shape.
- Between Normal and Pareto-Lévy, which distribution works best for a volatile/non-volatile interest rates environment?
  Table 5 shows that the Normal distribution (standard) and the Pareto-Lévy (calibrated daily) offer statistically the same accuracy in forecasting the 20-day forward yield curve whether the interest rates environment is volatile or not.
- Can we generalize the results with a Normal or Pareto-Lévy distribution?
  Since the Pareto-Lévy and the Normal distributions offer the same degree of accuracy in forecasting the 20-day forward yield curve whatever the yield curve and whether the environment is volatile or not, it is preferable to use the random numbers drawn from the standard Normal distribution N(0,1) because it is readily available in all programming languages and statistical packages. Besides, the standard Normal distribution offers the highest precision in forecasting the yield curve (the lowest average *RMSE*).

5. Conclusion

Normal and Pareto-Lévy distributions offer statistically the same precision of yield curve forecasts (Table 5) but the Normal distribution has a higher accuracy in fitting the yield curve (Table 4).

The Monte Carlo simulation coupled with the stock dog technique and the standard Normal distribution used as random numbers generator offers statistically better 20-day forecasts of the yield curve than the Diebold and Li (2006) model (Table 6).

The type of yield curve (normal, humped, flat or inverted) or the volatile or non-volatile environment have no impact on the forecasting power of the Normal or Pareto-Lévy distributions since they offer statistically the same degree of accuracy in forecasting the yield curve. However, the Normal distribution is superior in fitting the yield curve whatever the type of curve of whether the environment is volatile or not.

In future researches, we may apply the methodological framework to other distributions, aside Student's t-distribution since our early investigation of this distribution has not been conclusive for fitting the yield curve. Alternative stable distributions, such as the Weibull distribution, associated with both the nonrandom-minimum and geometric-random summation schemes may be tested (Mittnika and Rachevb, 1993). In addition, we may assess our methodological framework to the pricing of fixed-income derivatives. Finally, we may investigate other types of interest rate models to produce a distribution of innovation terms and longer-term forecasting horizon may be tested.

Acknowledgements

The authors express gratitude to the participants of the 2012 Global Development Finance Conference, held in Cape Town, South Africa, and to the participants of the 2012 GReFA conference (*Groupe de Recherche en Finance Appliquée*) held in Sherbrooke, Canada, for their observations and comments regarding this paper.

References

- Andersen, A., Lundb, J. and Torben, G. (1997). Estimating continuous-time stochastic volatility models of the short-term interest rate, *Journal of Econometrics*, **77**, pp. 343-377.
- Audrino, F. and De Giorgi, E. (2007). Beta Regimes for the Yield Curve, *Journal of Financial Econometrics*, **5**(3), pp. 456-490.
- Audrino, F. and Trojani, F. (2007). Accurate Short-Term Yield Curve Forecasting using Functional Gradient Descent, *Journal of Financial Econometrics*, **5**(4), pp. 591-623.
- Bernadell, C., Coche, J. and Nyholm, K. (2005). Yield curve prediction for the strategic investor, Working Paper No. 472, European Central Bank.
- Bliss, R. (1997). Movements in the term structure of interest rates, *Economic Review*, Federal Reserve Bank of Atlanta, **4**, pp. 16-33.
- Chen, R. and Scott, L. (1993). Maximum Likelihood Estimation For A Multifactor Equilibrium Model Of The Term Structure Of Interest Rates, *Journal of Fixed Income*, **4**, pp. 14-31.
- Cornew, R.W., Town, D.E. and Crowson, L.D. (1984). Stable Distributions, Futures Prices, and the Measurement of Trading Performance, *The Journal of Futures Markets*, **4**(4), pp. 531-557.
- Coronel-Brizio, H.F. and Hernández-Montoya, A.R. (2005). On fitting the Pareto-Levy distribution to stock market index data: Selecting a suitable cutoff value, *Physica A: Statistical Mechanics and its Applications*, **354**, pp. 437-449.
- Cox, J.C., Ingersoll, J.E. and Ross, S.A. (1985). The theory of the term structure of interest rates, *Econometrica*, **539**(2), pp. 363-384.

- Dai, Q. and Singleton, K. (2000). Specification Analysis of Affine Term Structure Models, *Journal of Finance*, **55**, pp. 1943-1978.
- De Boor, C. (1978). *A practical guide to splines*, Springer-Verlag, New York.
- De Jong, F. (2000). Time Series and Cross-Section Information in Affine Term-Structure Models, *Journal of Business & Economic Statistics*, **18**(3), pp. 300-314.
- Diebold, F. and Li, C. (2003). Forecasting the Term Structure of Government Bond Yields, Working Paper No. 10048, National Bureau of Economic Research.
- Diebold, F. and Li, C. (2006). Forecasting the Term Structure of Government Bond Yields, *Journal of Econometrics*, **130**, pp. 337-364.
- Dolan, C. (1999). Forecasting the yield curve shape: evidence from global markets, *The Journal of Fixed Income*, **10**, pp. 40-53.
- Duffee, G. (2002). Term Premia and Interest Rate Forecasts in Affine Models, *The Journal of Finance*, **57**, pp. 405-443.
- Fabozzi, J., Martellini, L. and Priaulet, P. (2005). Predictability in the shape of the term structure of interest rates, *The Journal of Fixed Income*, **15**, pp. 40-53.
- Fong, H.G. and Vasicek, O.A. (1992). Interest Rate Volatility as a Stochastic Factor, working paper, Gifford Fong Associates.
- Hagan, P.S. and West, G. (2006). Interpolation Methods for Curve Construction, Applied *Mathematical Finance*, **13**(2), pp. 89-129.
- Heath, D., Jarrow, R. and Morton, A. (1990). Bond pricing and the term structure of interest rates: a discrete time approximation, *Journal of Financial and Quantitative Analysis*, **25**, pp. 419-440.
- Hong, Y. (2001). Dynamic Models of the Term Structure, *Financial Analysts Journal*, **57**, pp. 60-76.
- Hull, J. and White, A. (1990). Pricing interest rate derivative securities, *Review of Financial Studies*, **3**(4), pp. 573-592.
- Jin, H.J. (2007). Heavy-Tailed Behavior of Commodity Price Distribution and Optimal Hedging Demand, *Journal of Risk and Insurance*, **74**(4), pp. 863-881.
- Joy, C., Boyle, P.P. and Tan, K. S. (1996). Quasi-Monte Carlo Methods in Numerical Finance, *Management Science*, **42**(6), pp. 926-938.
- Keller-Ressel, M. and Steiner, T. (2008). Yield curve shapes and the asymptotic short rate distribution in affine one-factor models, *Finance and Stochastics*, **12**(2), pp. 149-172.
- Klaviko, K. (2007). Maximum likelihood estimation of the Cox-Ingersoll-Ross process: the Matlab implementation, Working Paper, Department of Statistics and Probability Calculus, University of Economics, Prague and Debt Management Department, Ministry of Finance of the Czech Republic.
- Koutrouvelis, I.A. (1980). Regression-Type Estimation of the Parameters of Stable Laws, *Journal of the American Statistical Association*, **72**(372), pp. 918-928.
- Koutrouvelis, I.A. (1981). An Iterative Procedure for the estimation of the Parameters of Stable Laws, *Communication in Statistics*, **10**(1), pp. 17-28.
- Pareto-Lévy, P. (1925). *Calcul des Probabilités*, Gauthier Villars, Paris.

■ Lin, B.-H. (2002). Fitting term structure of interest rates using B-Splines: the case of Taiwanese government bonds, *Applied Financial Economics*, **12**(1), pp. 57-75.

■ Litterman, R. and Scheinkman, J. (1991). Common Factors Affecting Bond Returns, *Journal of Fixed Income*, **1**, pp. 54-61.

■ Longstaff, F.A. and Schwartz, E.S. (1992). Interest rate volatility and the term structure: a two-factor general equilibrium model, *The Journal of Finance*, **47**, pp. 1259-1282.

■ McCulloch, J.H. (1971). Measuring the Term Structure of Interest Rates, *Journal of Business*, **44**, pp. 19-31.

■ McCulloch, J.H. (1975). The tax adjusted yield curve, *Journal of Finance*, **30**, pp. 811-830.

■ Mansi, S.A., Phillips, J. H. and Green, M.D. (2001). Modeling the Term Structure from the On-the-Run Treasury Yield Curve, *Journal of Financial Research*, **24**(4), pp. 545-64.

■ Mittnika, S. and Rachevb, S.T. (1993). Modeling asset returns with alternative stable distributions, *Econometric Reviews*, **12**(3), pp. 261-330.

■ Nelson, C.R. and Siegel, A.F. (1987). Parsimonious Modeling of Yield Curves, *Journal of Business*, **60**, pp. 473-489.

■ Rostan, P. and Rostan A. (2012). Forecasting the yield curve with the “stock dog” technique”, *The International Journal of Business and Management*, **7**(16), pp. 31-43.

■ Shea, G. (1984). Pitfalls in smoothing interest rate term structure data: Equilibrium models and spline approximations, *Journal of Financial and Quantitative Analysis*, **19**, pp. 253-269.

■ Shea, G. (1985). Interest rate term structure estimation with exponential splines: A note, *Journal of Finance*, **11**, pp. 319-325.

■ So, J.C. (1987). The Sub-Gaussian Distribution of Currency Futures: Stable Paretian or Nonstationary? , *The Review of Economics and Statistics*, **69**(1), pp. 100-107.

■ Svensson, L.E.O. (1995). Estimating forward interest rates with the extended Nelson and Siegel method, *Sveriges Riskbank Quarterly Review*, **3**, pp. 13-26.

■ Théoret, R., Rostan, P. and El Moussadek, A. (2005). Forecasting the Interest-rate Term Structure: Using the Model of Fong and Vasicek, the Extended Kalman Filter and the Bollinger Bands, Working Paper, SSRN.

■ Vasicek, O.A. (1977). An equilibrium characterization of the term structure, *Journal of Financial Economics*, **5**, pp. 177-188.

■ Vasicek, O.A., and Fong, H.G. (1982). Term Structure Modeling Using Exponential Splines, *Journal of Finance*, **37**(2), pp. 339-348.

■ Veillette, M. (2013). http://math.bu.edu/people/mveillet/research.html Accessed on May 16, 2013.

■ Wilmott, P. (2009). *Frequently Asked questions in Quantitative Finance*, John Wiley & Sons Ltd., Chichester, West Sussex, England.

■ Zangari, P. (1997). An investigation into term structure estimation methods for RiskMetrics, *RiskMetrics Monitor*, Third Quarter, pp. 3-48.