Sunshine-Factor Model with Treshold GARCH for Predicting Temperature of Weather Contracts

Hélène Hamisultane

Sunshine-Factor Model with Treshold GARCH for Predicting Temperature of Weather Contracts

Hélène Hamisultane [†]

August 28, 2008

ABSTRACT

Climate changes have sparked growing interest for the weather derivatives which are financial contracts relied on a meteorological index and allowing companies to hedge against climate risk. These contracts present the particularity of providing compensation to the buyer when the meteorological index crossed a limit agreed in advance with the seller. In order to evaluate these products and to manage at best the risks associated with their exchange, it is important to be able to accurately predict the evolution of the climate variable. Several processes have been proposed in the literature to model the behaviour of the temperature which is the basis of most of the traded weather instruments. These processes relate mainly to the univariate time series modelling which is founded on the study of the autocorrelation of the stationary variable. But we know that the behaviour of the temperature can be influenced by climatic factors such as rain, wind or sunshine. In our paper, we propose to take into account the impact of sunshine on the temperature as well as the asymmetric effect of the shocks on the volatility by estimating a structural model with a periodic threshold GARCH. We show that this model provides better out-sample forecasts for 30 and 60 days ahead than those obtained by the univariate autoregressive-conditional heteroskedasticity process.

Keywords : weather derivatives, structural model, Markov chain, threshold GARCH, Monte-Carlo simulations, pricing, Value-at-Risk.

JEL classification : C2, C5, G13

[†] EconomiX , 200 avenue de la République, 92000 Nanterre, France. E-mail : helene.hamisultane@free.fr

1. Introduction

The weather derivatives market which began in 1997 has gained momentum in recent years because of the increased awareness of enterprises in the vulnerability of their turnover to climatic variations (temperature, rain, snow, ...). In 2003, it was estimated that the market had grown to a $3.5 billion notional market value with approximately 4,517 contracts traded. A weather derivative is a financial contract whose payments are triggered when the underlying climate index crosses, upwards or downwards, a barrier agreed in advance between two counterparties. It allows different actors (heating company, farmer, amusement park,…) to cope with climate risk. Unlike insurance contracts, weather derivatives require no finding of actual losses to receive compensation. The meteorological index can be calculated, for instance, from measurements of temperature, precipitation or humidity. At the moment, the most actively traded contracts are on the temperature because they meet the coverage needs of the energy companies which were the first initiators of weather derivatives and which currently represent the main actors on the climate market. The use of weather contracts requires companies to know with precision the evolution of the climate variable. Regarding the temperature and typically, the daily average temperature which is the average of the maximum and minimum temperature of the day, the estimation of an ARMA(Auto Regressive Moving Average) process has often been studied. It is a model connecting the dependent variable to its past values and past error terms. The construction of the model is based on the analysis of the autocorrelation function. The major drawback of ARMA processes is that they do not consider other variables outside delayed values of the endogenous variable and shocks. But it is well-known that precipitation, wind and sunshine can affect the behaviour of the temperature. The question we ask is whether the incorporation of any of these variables in the modelling can improve or not the forecats of the daily average temperature. To find out, we propose to study the impact of hours of sunshine on the temperature through the estimation of a structural model composed of stationary variables. Moreover, it was shown that the volatility of the daily average temperature exhibited a seasonal behaviour which could be captured by means of a periodic GARCH (Generalized AutoRegressive Conditional Heteroskedasticity) process suggested by Campbell and Diebold (2005). The drawback of the GARCH process is that only symmetric effects of the shocks on the volatility are considered. To take into account also asymmetric impacts of the error terms on the volatility, we propose to extend the model of Campbell and Diebold by using the threshold GARCH process instead of the GARCH formulation. Our results demonstrate that considering the sunshine effect and the asymmetric impacts on the volatility of the temperature may give forecasts outside the sampling period (for 30 and 60 days ahead) that are better than those provided by the univariate time series model with a periodic GARCH.

The outline of the paper is as follows : section 2 explores the relationship between the temperature and the sunshine as well as the behaviour of the two variables. It will be suggested to model the sunshine dynamics by a two-state first order Markov chain to capture its discontinuous feature. Section 3 deals with the estimation of the parameters of the Markov chain and those of the structural model equipped with the periodic threshold GARCH to reproduce the movement of the daily average temperature. Section 4 applies the estimated models for the evaluation and the risk management (computation of the Value-at-Risk) of the weather derivatives. Section 5 provides the conclusions.

2. Causal relationship modelling of temperature with sunshine component and time-varying volatility

A key feature of sunshine is that it is a discontinuous variable (it may happen that some days are not sunny). As we see in Figures 1 and 2 below, the hours of sunshine are longer in summer than in winter and the number of days when it is sunny is more important in summer than in winter. From this information, we can assume that the sunshine shows a seasonal behaviour marked by a probability of occurrence and a magnitude greater in summer than in winter. Figures 3 and 4 reveal that the probability of having a sunny day in t depends on the situation in (t-1) : the chances of having a sunny day in t are more important when it is sunny in (t-1) than when it is not beautiful in (t-1). But the occurrence of sunshine in t seems to be dominant whatever the situation in (t-1) during the summer period (this is particularly true for regions with mediterranean climate like Toulouse in France).

Fig.1 Daily average temperature and sunshine hours of Toulouse

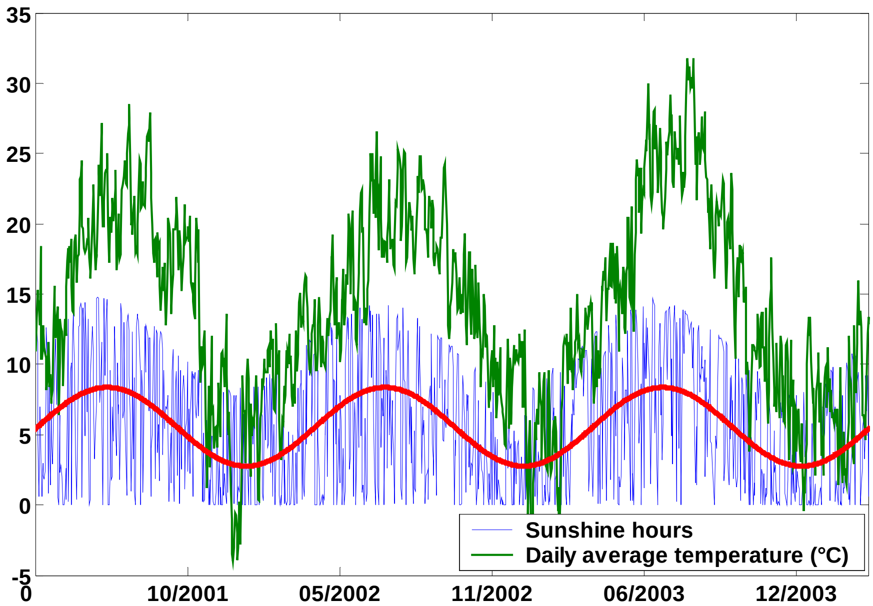


Fig.2 Number of days of sunshine in Toulouse from 01/01/1949 to 05/31/2004 (no leap year)

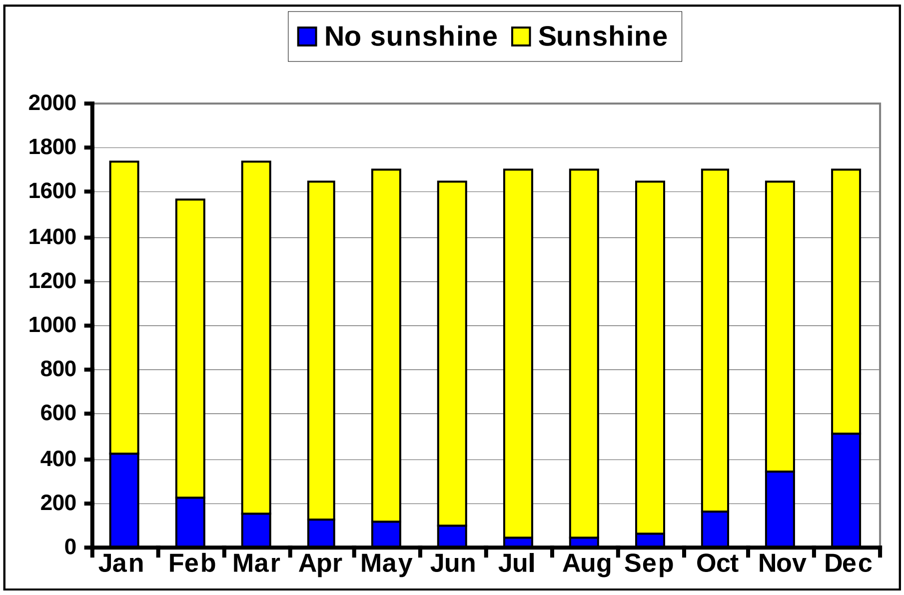


Fig.3 Probability of sunshine following sunshine in Toulouse

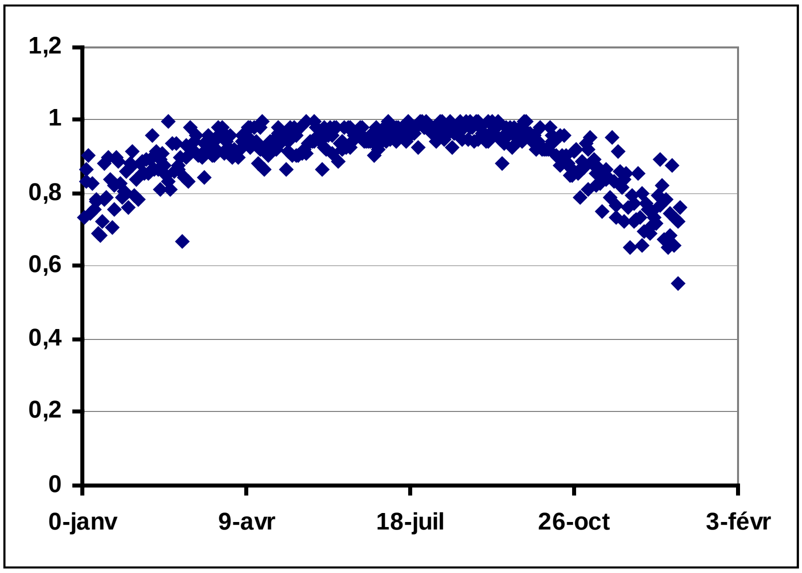
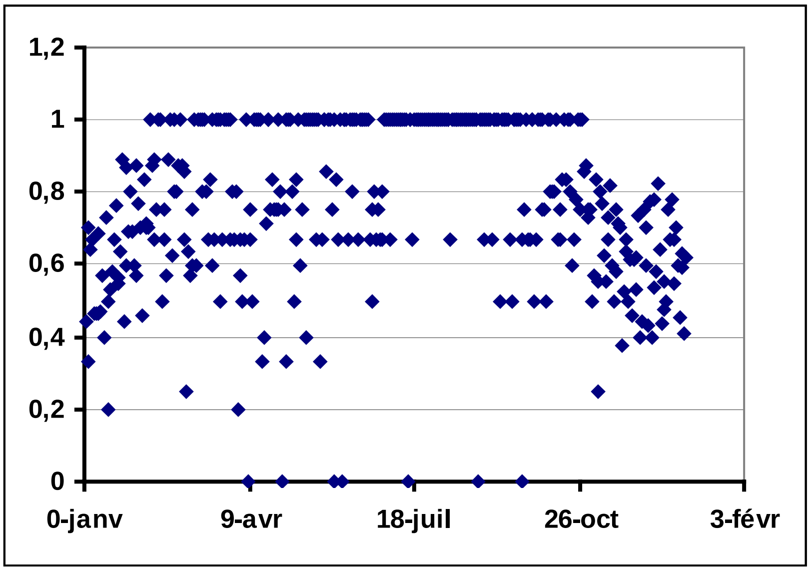


Fig.4 Probability of sunshine following no sunshine in Toulouse



About the daily average temperature, Figure 1 displays a relationship between the latter and the sunshine. It is reasonable to consider a positive correlation between these two variables. Indeed, the more the hours of sunshine are important and the higher the temperature should be. Figure 1 also puts in evidence the following characteristics of the daily average temperature : a cyclical behaviour, a mean-reversion (the variations in the temperature do not deviate from a long equilibrium value) and an autoregressive feature (a hot day will likely be followed by another hot day, same for a cold day).

Based on these observations, we propose the following model for the daily average temperature[(1)] :

$$T_t = T_t^m - \sum_{i=1}^{p} \rho_i \,(T_{t-i}^m - T_{t-i}) + \sum_{j=0}^{q} \varphi_j \, f_{t-j} \,(S_{t-j} - \overline{S}_{t-j}) + \varepsilon_t \qquad (1)$$

with

$$T_t^m = a + bt + \sum_{k=1}^{K} \delta_{c,k} \cos(k\omega t) + \delta_{s,k} \sin(k\omega t) \quad \text{and} \quad \omega = \frac{2\pi}{365} \qquad (2)$$

where f_t is a binary variable which accounts for the occurrence of sunshine at time t, S_t represents the hours of sunshine, $\overline{S}_t$ corresponds to the seasonal mean of hours of sunshine, $\rho_1, \ldots, \rho_p$ are p autoregressive parameters, T_t^m denotes the mean-reverting value of the daily average temperature with $\rho_1 < 0$ reflecting the speed of the mean-reversion, bt stands for the trend and ε_t is an uncorrelated random variable with zero mean and variance σ_t^2.

---

(1) The model could have been written in level as follows: $T_t = T_t^m + \sum_{i=1}^{p} \rho_i \, T_{t-i} + \sum_{j=0}^{q} \varphi_j \, f_{t-j} \, S_{t-j} + \varepsilon_t$.
The problem with this formulation is that it does not take into account the mean-reversion of the temperature.

The model (1) without the lagged values of sunshine corresponds to an autoregressive process which is a particular case of the ARMA processes. The ARMA processes connect the stationay endogenous variable to its past values and error terms. It was shown in numerous articles that the daily average temperature exhibited an autoregressive structure rather than a mixture of autoregressive and moving average parts. The major problem of the ARMA processes is that they do not allow to take into account other dependent variables except the lagged values of the endogenous variable and of the error terms. The formulation given by Eq.(1) has the advantage to consider external factors like here the effect of sunshine on the daily average temperature. The model represents, in fact, a multivariate linear regression model composed of the stationary lagged variables $(T^m_{t-i} - T_{t-i})$ and $(S_{t-j} - \overline{S}_{t-j})$. It cannot be regarded as a VAR (Vector Auto Regressive) process since $(S_{t-j} - \overline{S}_{t-j})$ does not appear also as a dependent variable. The coefficients of the model are to be estimated by the Maximum Likelihood Method and not by the Ordinary Least Squares because the variance σ^2_t is assumed to vary with time (Engle (1982)). Indeed, it is well-known that the estimators of the Ordinary Least Squares are no more efficient in the presence of heteroskedasticity.

In the case of the daily average temperature, many authors have pointed out the seasonality of the variance σ^2_t. To capture this characteristic, Cao and Wei (1998, 2004) and Roustant (2002) have used a sine function. Campbell and Diebold (2005) have put forward a more general model combining a k-th order Fourier series and a GARCH component. A non-parametric volatility was proposed by Benth and Šaltytė-Benth (2005). We want to extend the model of Campbell and Diebold by allowing the volatility of the temperature to respond in asymmetric way to the shocks. Indeed, concerning the GARCH formulation, a negative or positive impact on the volatility has for only effect an increase of this one and in the same amplitude because of the square form of the variance. To allow the volatility to react differently depending on the sign of the shock, we use the threshold GARCH process in place of the GARCH representation in the model of Campbell and Diebold. The process threshold GARCH of Glosten, Jaganathan, and Runkle (1993), noted GJR-GARCH (p, q) is written as [(2)] :

$$\sigma^2_t = c + \sum_{i=1}^{q} \left(\alpha_i \varepsilon^2_{t-i} + \gamma_i \varepsilon^2_{t-i} d_{(\varepsilon_{t-i}<0)} \right) + \sum_{j=1}^{p} \beta_j \sigma^2_{t-j} \qquad (3)$$

where $d_{(\varepsilon_{t-i}<0)} = 1$ if $\varepsilon_{t-i} < 0$ and $d_{(\varepsilon_{t-i}<0)} = 0$ otherwise, with $c > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$ $\forall$ i, j.

---

(2) The threshold GARCH process of Zakoian (1994) called TGARCH (p,q) is quite close to the process of Glosten, Jaganathan, and Runkle (1993) :

$$\sigma_t = c + \sum_{i=1}^{q} \left(\alpha^+_i \varepsilon^+_{t-i} - \alpha^-_i \varepsilon^-_{t-i} \right) + \sum_{j=1}^{p} \beta_j \sigma_{t-j}$$

where $\varepsilon^+_t = \max(\varepsilon_t, 0)$ and $\varepsilon^-_t = \min(\varepsilon_t, 0)$.

We can notice that this process concerns the standard deviation and not the variance. The coefficients must be non-negative.

When $\varepsilon_{t-i} \geq 0$, the variance increases of $\alpha_i \varepsilon^2_{t-i}$ whereas for $\varepsilon_{t-i} < 0$, the variance increases of $(\alpha_i + \gamma_i)\varepsilon^2_{t-i}$. If $\gamma_i > 0$, the rising of the volatility is stronger after a negative shock than after a positive shock.. If $\gamma_i < 0$, the increase of the volatility is much lower following a negative shock. In short, if $\gamma_i \neq 0$, the error terms have an asymmetric effect on the volatility.

Regarding the volatility of the daily average temperature, it is then expressed as follows :

$$\sigma^2_t = c + \sum_{\ell=1}^{L} \left(\lambda_{c,\ell} \cos(\ell\omega t) + \lambda_{s,\ell} \sin(\ell\omega t)\right) + \sum_{i=1}^{q} \left(\alpha_i \varepsilon^2_{t-i} + \gamma_i \varepsilon^2_{t-i} d_{(\varepsilon_{t-i}<0)}\right) + \sum_{j=1}^{p} \beta_j \sigma^2_{t-j} \tag{4}$$

with $\omega = \dfrac{2\pi}{365}$.

The EGARCH (Exponential GARCH) process of Nelson (1991) whose coefficients are not restricted to positive values may also be considered to take into account the asymmetric effects on volatility. The EGARCH(p,q) process is defined as (3) :

$$\ln\sigma^2_t = c + \sum_{i=1}^{q} \psi_i \left(\phi\, z_{t-i} + \gamma \left[\left|z_{t-i}\right| - E\left|z_{t-i}\right|\right]\right) + \sum_{j=1}^{p} \beta_j \ln\sigma^2_{t-j} \tag{5}$$

where $z_{t-i} = \dfrac{\varepsilon_{t-i}}{\sigma_{t-i}}$ represents the standardized innovation (with expectation 0 and variance 1) and ε_t ~>iid$(0, \sigma^2_t)$.

The logarithmic expression of the conditional variance allows the removal of the positivity constraints on the coefficients. In contrast to the CJR-GARCH process, the impacts of z_{t-i} are considered here instead of the impacts of ε_{t-i}. Indeed, if $z_{t-i} \geq 0$, the variance has the slope (γ + ø) while for $z_{t-i} < 0$, it has the slope (γ - ø). We will see in Section 3 on the estimation of the parameters of the processes that, for the temperature, impacts of ε_{t-i} are more significant than those of z_{t-i} on the volatility.

To simulate the paths of the temperature, it is necessary to model the behaviour of the sunshine. To capture the discontinuous nature of the variable, we use a two-state (sunny day and no sunny day) first-order Markov chain (the state of the variable at t depends only on its state at the moment (t-1)) :

:

$$f_t = \begin{cases} 1 & \text{if sunshine occurs} \\ 0 & \text{otherwise} \end{cases}$$

$$f_t \mid f_{t-1} \sim> \text{Markov}(P)$$

(3) The EGARCH(1,1) process is stated as : $\ln\sigma^2_t = c + \alpha_1 \dfrac{\varepsilon_{t-1}}{\sigma_{t-1}} + \alpha_2 \left(\left|\dfrac{\varepsilon_{t-1}}{\sigma_{t-1}}\right| - E\left|\dfrac{\varepsilon_{t-1}}{\sigma_{t-1}}\right|\right) + \beta_1 \ln\sigma^2_{t-1}$.

$E\left|\dfrac{\varepsilon_{t-1}}{\sigma_{t-1}}\right| = \sqrt{\dfrac{2}{\pi}}$ when the distribution is normal.

where P is the state transition probability matrix defined by :

$$P = [\, p_{ij} \,] \,, \; p_{ij} = \Pr(f_t = j \mid f_{t-1} = i), \; i, j = 0, 1. \tag{6}$$

where p_{00}, p_{01}, p_{11}, p_{10} are calculated as follows:

$$p_{00} = \frac{n_{00}}{n_{00} + n_{01}} \;, \; p_{01} = 1 - p_{00} \;, \; p_{11} = \frac{n_{11}}{n_{11} + n_{10}} \;, \; p_{10} = 1 - p_{11} \tag{7}$$

with n_{i1} representing the number of years when day t is in state 1 and day (t-1) is in state i.The probability p_{11} is plotted in Figure 3 while p_{01} is depicted in Figure 4.

To reproduce the variation with time of the transition probabilities, we fit a truncated Fourier series to the data, i.e. :

$$p_{ij} = \alpha_0 + \sum_{r=1}^{R} \alpha_{c,r} \cos(r\omega t) + \alpha_{s,r} \sin(r\omega t) \,. \tag{8}$$

These probabilities are useful to simulate the occurrence of sunshine. To do this, a number between 0 and 1 is first genererated from the uniform distribution. It is then compared to p_{11} if the previous day was sunny or to p_{01} otherwise. The current day is sunny when p_{11} or p_{01} is superior or equal to the generated number (Wilks (1999)).

When the current day is sunny ($f_t = 1$), we determine the amount S_t of sunshine by generating a value with the exponential or gamma distribution which produces only strictly positive numbers. The two distributions are given by :

$$f_{exp}(x) = \frac{e^{-x/\theta}}{\theta} \tag{9}$$

and

$$f_{gam}(x) = \frac{(x/\beta)^{\alpha-1} \exp(-x/\beta)}{\beta\, \Gamma(\alpha)} \,. \tag{10}$$

To consider the performance of the model proposed for the daily average temperature, we estimate the coefficients in the following section. The results will be compared to those given by the univariate autoregressive process with a periodic GARCH.

3. Estimation and simulation

We use data from the European Climate Assessment. They regroup the daily average temperature and hours of sunshine of Toulouse in France. They cover the period from January 1st, 1949 to July 31st, 2004 which represents 20,226 observations (February 29 is removed for leap years). We estimate the parameters of the models by using data up to the end of May, 2004. The remainder of the data is used for out-sample forecasts.

The component $\overline{S}_t$ is estimated by fitting a Fourier series to the observations of sunshine since a periodic pattern is noticeable in Figure 1. On the basis of the information criteria (Akaike (AIC) and Schwarz (SC) criteria), we have selected three orders for the function as shown in Table 1. This number of orders minimizes the information criteria. The estimation results of the temperature processes are presented in Table 2.

All the dependent variables of the models with and without sunshine appear significant (the t-statistics in absolute value are above 1.96 at 5% level) which validates our representation choice made on the analysis of the information measures. Moreover, the coefficient ρ_1 of mean-reversion has the correct sign (it appears here positive because we have replaced $-\sum_{i=1}^{p} \rho_i \, (T^m_{t-i} - T_{t-i})$ by $\sum_{i=1}^{p} \rho_i \, (T_{t-i} - T^m_{t-i})$ to make the estimation easier). Both models have no autocorrelation since the p-values associated with the Q-statistic of Ljung-Box in Table 3 are all superior to 5% (the null hypothesis of the test is that there is no autocorrelation up to order k). By comparing the information criteria in Table 2 of the model with sunshine and CJR-GARCH component and the model without sunshine equipped with the GARCH effect, we remark that considering the sunshine impacts and the asymmetric effects on the volatility helps to reduce the values of the AIC and SC measures indicating that the first model fits better the data than the second model. Moreover, the parameter γ_1 is negative and significant which means that there is indeed a phenomenon of asymmetry in the variance (the increase of the volatility is much lower after a negative shock). Table 2 reveals that the asymmetric effects of the error terms on the volatility are better considered through the estimation of the periodic CJR-GARCH process than through the estimation of the periodic EGARCH process because the first process shows the weakest values of AIC and SC. To analyze the performance of the model with sunshine and the periodic CJR-GARCH process outside the sample, we need to simulate the paths of sunshine. To do this, we need to simulate both the occurrence of sunshine events and the quantity of sunshine. As we stated earlier, the simulation of the occurrence is conducted by comparing the value of the transition probability which is given by the Fourier series estimated in Table 4 to the random uniform number. When the current day is sunny, the amount of sunhine is generated by means of the exponential or gamma distribution. The estimated coefficients of the two distributions are listed in the Table 5. When plotting the quantities from the two distributions, we note that those from the exponential distribution substantially underestimate the actual quantities (see Figure 5) while the amounts from the gamma distribution over-estimate widely the actual quantities (see Figure 6). Another problem is also noticed, the simulated quantities are not seasonal when this feature is apparent in the observations. To reproduce this characteristic, we decide to use the data of sunshine to create a series of seasonal amount. This series is determined by first combining the observations of hours of sunshine at time t per year and then by taking the average of these observations over the years for each date t giving rise to 365 averages which constitute the seasonal amount series for any given year. The performance of the model with sunshine and the periodic CJR-GARCH process in terms of forecasts for the horizons of 7 days, 30 days and 60 days are recorded in Table 6. It is measured with the RMSE (Root Mean Squared Error) and MAE (Mean Absolute Error) criteria. We remark that the model with sunshine and periodic CJR-GARCH process provides predictions which are less good than those of the model without sunshine and periodic GARCH for the horizon of 7 days. On the contrary, the first model is more effective for longer horizons of 30 and 60 days. It has a more important predictive power in the longer term.

Table 1 : Estimation of the Fourier series parameters for the seasonal mean of sunshine

	$\overline{S}$	
	Estimation	**t-statistic**
η_0	5.56	207.38
$\eta_{c,1}$	-2.81	-73.98
$\eta_{c,2}$	-0.33	-8.62
$\eta_{c,3}$	-0.18	-4.65
$\eta_{s,1}$	-0.07	-1.84
$\eta_{s,2}$	0.53	14.07
$\eta_{s,3}$	-0.05	-1.26

$$\overline{S}_t = \eta_0 + \sum_{d=1}^{D} \eta_{c,d} \cos(d\omega t) + \eta_{s,d} \sin(d\omega t)$$

Table 2 : Estimation of the models of the daily average temperature of Toulouse

	Model with no sunshine (*)		Model with sunshine (**)			
	Mean equation		**Mean equation**		**Mean equation**	
	Estimation	**t-statistic**	**Estimation**	**t-statistic**	**Estimation**	**t-statistic**
a	12.34	265.25	12.34	265.25	12.34	265.25
b	8.17×10^{-5}	20.52	8.17×10^{-5}	20.52	8.17×10^{-5}	20.52
$\delta_{c,1}$	-7.42	-225.42	-7.42	-225.42	-7.42	-225.42
$\delta_{c,2}$	-0.25	-7.48	-0.25	-7.48	-0.25	-7.48
$\delta_{c,3}$	-0.18	-5.54	-0.18	-5.54	-0.18	-5.54
$\delta_{s,1}$	-3.05	-92.70	-3.05	-92.70	-3.05	-92.70
$\delta_{s,2}$	0.99	30.03	0.99	30.03	0.99	30.03
$\delta_{s,3}$	0.03	0.92	0.03	0.92	0.03	0.92
ρ_1	0.83	114.48	0.80	105.52	0.80	106.95
ρ_2	-0.12	-12.91	-0.10	-11.05	-0.10	-11.10
ρ_3	0.05	6.40	0.05	6.69	0.05	6.80
φ_0			0.12	32.94	0.12	32.80
φ_1			0.03	8.54	0.03	8.66
φ_2			0.03	7.62	0.03	7.58
φ_3			-0.02	-5.88	-0.02	-5.97
	Periodic GARCH(1,1)		**Periodic CJR-GARCH(1,1)**		**Periodic EGARCH(1,1)**	
c	0.31	3.34	0.95	5.58	0.25	4.55
$\lambda_{c,1}$	0.09	3.22	0.43	5.41	0.10	5.26
$\lambda_{c,2}$	0.03	2.50	0.09	4.02	0.01	2.54
$\lambda_{s,1}$	-0.03	-5.42	-0.03	-2.23	-0.005	-1.54
$\lambda_{s,2}$	-0.03	-4.46	-0.05	-3.19	-0.01	-3.12
α_1	0.02	5.01	0.07	7.70	0.03	5.17
α_2					0.09	7.42
γ_1			-0.05	-5.38		
β_1	0.91	39.77	0.74	17.22	0.77	18.17
AIC (***)	4.3138		4.2375		4.2380	
SC	4.3177		4.2433		4.2439	

(*) Model with no sunshine and periodic GARCH(1,1) :

$$T_t = T_t^m + \sum_{i=1}^{p} \rho_i \,(T_{t-i} - T_{t-i}^m) + \varepsilon_t \;,\quad \varepsilon_t = \sigma_t \tilde{\varepsilon}_t \;,\quad \tilde{\varepsilon}_t \sim\!> \text{iid}(0,1)$$

$$T_t^m = a + bt + \sum_{k=1}^{K} \delta_{c,k} \cos(k\omega t) + \delta_{s,k} \sin(k\omega t) \quad \text{with } \omega = \frac{2\pi}{365}$$

and $$\sigma_t^2 = c + \sum_{\ell=1}^{L} \left(\lambda_{c,\ell} \cos(\ell\omega t) + \lambda_{s,\ell} \sin(\ell\omega t)\right) + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 .$$

(**) Model with sunshine :

$$T_t = T_t^m + \sum_{i=1}^{p} \rho_i \,(T_{t-i} - T_{t-i}^m) + \sum_{j=0}^{q} \varphi_j \, f_{t-j} \,(S_{t-j} - \overline{S}_{t-j}) + \varepsilon_t \;,\quad \varepsilon_t = \sigma_t \tilde{\varepsilon}_t \;,\quad \tilde{\varepsilon}_t \sim\!> \text{iid}(0,1)$$

$$T_t^m = a + bt + \sum_{k=1}^{K} \delta_{c,k} \cos(k\omega t) + \delta_{s,k} \sin(k\omega t) \quad \text{with } \omega = \frac{2\pi}{365}$$

$$\sigma_t^2 = c + \sum_{\ell=1}^{L} \left(\lambda_{c,\ell} \cos(\ell\omega t) + \lambda_{s,\ell} \sin(\ell\omega t)\right) + \alpha_1 \varepsilon_{t-1}^2 + \gamma_1 \varepsilon_{t-1}^2 d_{(\varepsilon_{t-1}<0)} + \beta_1 \sigma_{t-1}^2 \quad \text{for the periodic CJR-GARCH(1,1)}$$

and $$\ln\sigma_t^2 = c + \sum_{\ell=1}^{L} \left(\lambda_{c,\ell} \cos(\ell\omega t) + \lambda_{s,\ell} \sin(\ell\omega t)\right) + \alpha_1 \frac{\varepsilon_{t-1}}{\sigma_{t-1}} + \alpha_2 \left|\frac{\varepsilon_{t-1}}{\sigma_{t-1}}\right| + \beta_1 \ln\sigma_{t-1}^2 \quad \text{for the periodic EGARCH(1,1)}$$

(***) AIC = $-2\left(\frac{LL}{n}\right) + \frac{2k}{n}$ et SC = $-2\left(\frac{LL}{n}\right) + k\,\frac{\ln(n)}{n}$ where LL represents the estimated log-likelihood, n is the number of observations and k is the number of parameters of the model.

Table 3 : Test of Ljung-Box for detecting autocorrelation up to order k

	Model with no sunshine		Model with sunshine	
Order	**Q-Stat**	**p-value**	**Q-Stat**	**p-value**
1	0.0148	0.903	0.0036	0.952
2	0.0487	0.976	0.0089	0.996
3	4.1796	0.243	1.2607	0.738
4	4.2441	0.374	1.4312	0.839
5	4.2482	0.514	1.8797	0.866
6	6.7277	0.347	2.4385	0.875
7	7.3500	0.393	2.8417	0.899
8	11.268	0.187	6.6103	0.579
9	11.301	0.256	6.6663	0.672
10	11.740	0.303	7.1169	0.714
11	12.292	0.342	7.5018	0.757
12	13.225	0.353	8.1343	0.775
13	16.316	0.232	11.320	0.584
14	16.612	0.277	11.382	0.656
15	16.653	0.340	11.570	0.711
16	18.073	0.320	12.127	0.735
17	18.126	0.381	12.163	0.790
18	18.285	0.437	12.167	0.839
19	19.224	0.443	13.600	0.806
20	19.501	0.490	14.128	0.824

Table 4 : Estimation of the Fourier series for the transition probabilities

	p_{11}		p_{01}	
	Estimation	**t-statistic**	**Estimation**	**t-statistic**
α_0	0.90	386.59	0.78	73.43
$\alpha_{c,1}$	-0.09	-26.90	-0.14	-9.13
$\alpha_{c,2}$	-0.04	-12.28	-0.04	-2.44
$\alpha_{c,3}$	-0.01	-3.77		
$\alpha_{s,1}$	-0.002	-0.69	-0.03	-1.94
$\alpha_{s,2}$	0.02	6.07	0.06	4.27
$\alpha_{s,3}$	0.01	2.64		

Table 5 : Estimation of the distributions for the non-zero amount of sunshine

	Exponential distribution			**Gamma distribution**	
	Estimation	**t-statistic**		**Estimation**	**t-statistic**
θ	6.30	169.70	α	1.48	101.60
			β	4.26	82.46

Fig.5 Simulated hours of sunshine from the exponential distribution

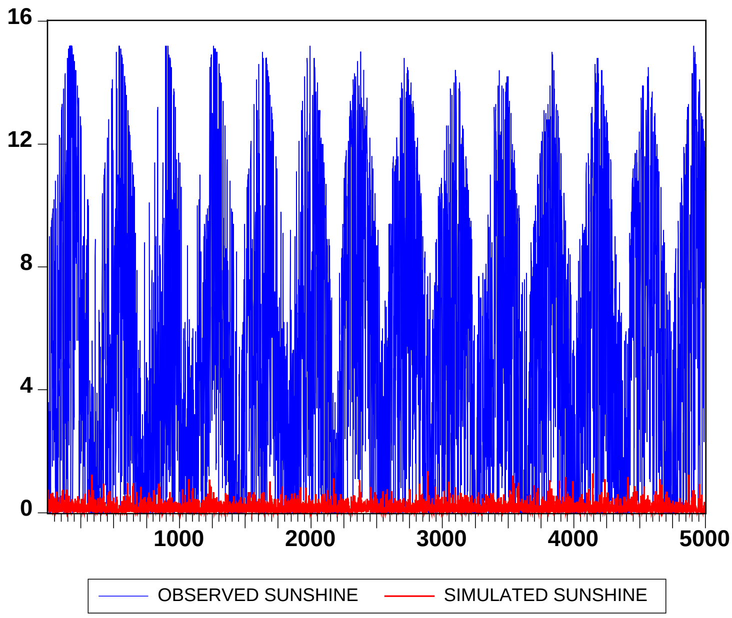


Fig.6 Simulated hours of sunshine from the gamma distribution

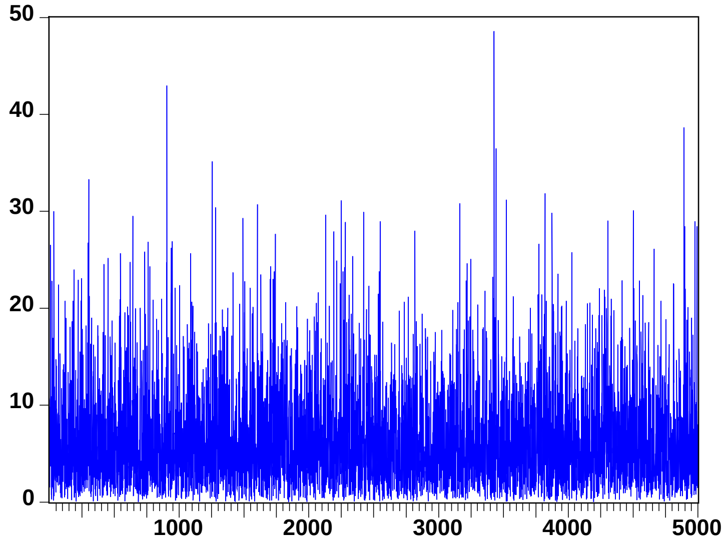


Table 6 : Comparison criteria (*) for out-sample forecats of the daily average temperature of Toulouse

	Model with no sunshine and periodic GARCH(1,1)	**Model with sunshine and periodic CJR-GARCH(1,1)**
	7 days	
RMSE (**)	1.990	2.114
MAE	1.784	1.896
	30 days	
RMSE	3.069	3.021
MAE	2.579	2.570
	60 days	
RMSE	3.330	3.020
MAE	2.697	2.455

(*) For each of the two models, we run 10,000 paths of the daily average temperature with the same noise for the 2 models. We then made the average of these 10,000 simulations and we compared it to the observed temperature in order to compute the RMSE and MAE criteria.

(**) RMSE = $\sqrt{\frac{1}{n}\sum_{t=1}^{n}(T_t - \hat{T}_t)^2}$ and MAE = $\frac{1}{n}\sum_{t=1}^{n}\left|T_t - \hat{T}_t\right|$ where $\hat{T}_t$ represents the estimated temperature and n corresponds to the number of observations.

4. Some applications to weather derivatives

A weather derivative is a financial contract whose payoff depends on the evolution of an underlying meteorological index. At the moment, the most actively traded contracts are on the temperature and more specifically on cooling degree day (CDD) which counts the daily average temperature above 65 ° F when it comes to the summer and on heating degree day (HDD) which counts the daily average temperature being below 65° F during the winter period. Currently, two major challenges of the weather derivatives are the valuation and the risk management of these instruments. For these two tasks, it is essential to accurately predict the behaviour of the weather variable. We will judge the results provided by the model with sunshine and the periodic CJR-GARCH process by using it to value the weather forward on the CDDs of Toulouse and to calculate the Value-at-Risk of this contract.

4.1 Pricing

In practice, weather derivatives are evaluated by using the actuarial method which is an alternative to the arbitrage-free method. This latter is commonly used to price derivatives. Regarding the weather derivatives, the arbitrage-free method encounters a number of difficulties in implementation. Firstly, the meteorological index is not traded on the financial market, so we can not use the Black and Scholes (1973) formula to value the weather options. Secondly, it is difficult to get round this obstacle by susbtituting the underlying for a linked exchanged security since the weather index is weakly correlated with prices of other financial assets. Thirdly, the market is incomplete for the weather derivatives, a market price of risk must be considered. However, a great part of the weather contracts are either non quoted or non liquid which does not allow the estimation of this parameter and so, the calibration of the arbitrage-free model. Besides, the actuarial method also has the advantage of being very simple to implement which attracts the practitioners. It determines the price at time t of the weather derivatives expiring in t_m in the following way :

$$C(t,T_t, I_t^C) = \delta\, e^{-r(t_m-t)} \left(E[\,\text{payoff} \mid \mathcal{F}_t\,] + \kappa\, \sigma_{\text{payoff}}\right) \tag{11}$$

for the weather call option on the CDD index,

and

$$F(t,T_t, I_t^C) = \delta \left(E[\, I_{t_m}^C \mid \mathcal{F}_t\,] + \kappa\, \sigma_{I^C}\right) \tag{12}$$

for the weather forward on the CDD index

where payoff = $\max(I_{t_m}^C - K,0)$, $I_{t_m}^C = \sum_{j=1}^{n} \max(T_j - 18C°,0)$ represents the CDD[(4)] index on n days of the contract period, $T_j = \dfrac{T_j^{min} + T_j^{max}}{2}$ is the daily average temperature, δ corresponds to the value attributed to one degree-day, K refers to the strike price, r corresponds to the risk-free interest rate, $E[\ldots \mid \mathcal{F}_t]$ designates the conditional expectation operator under the real

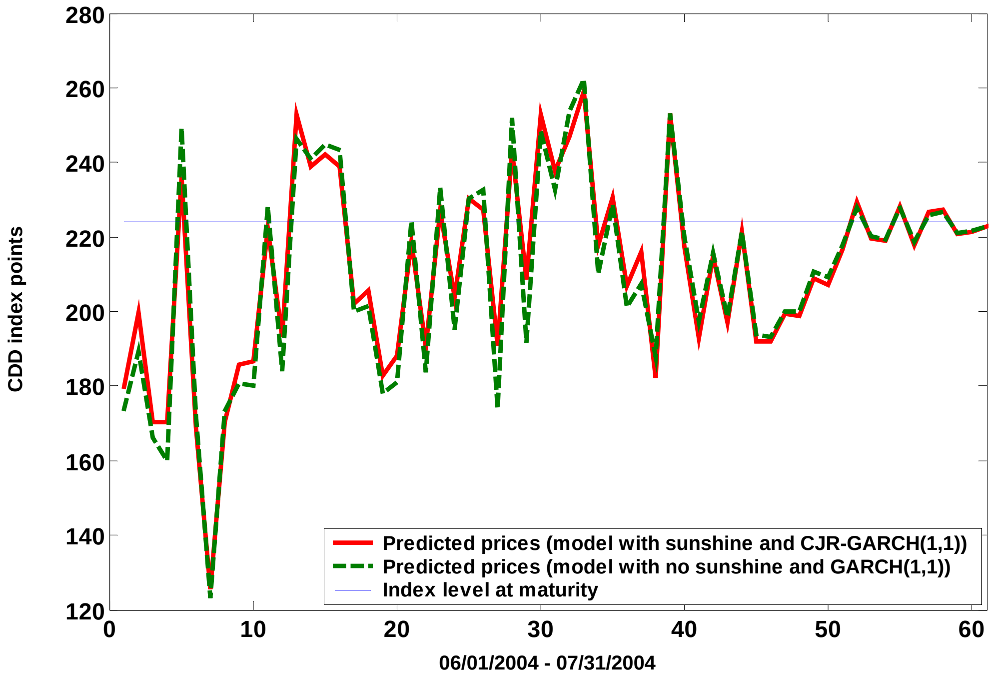
---

(4) During the cold period (November to April), it is the HDD index that is used : $I_{t_m}^H = \sum_{j=1}^{n} \max(18°C - T_j, 0)$.

probability, $\kappa\ \sigma_{payoff}$ and $\kappa\ \sigma_{I^c}$ stand for the risk premiums and σ_{payoff} and σ_{I^c} denote respectively the volatility of the payoff and the volatility of the CDD index.

The risk premium is often assumed to be equal to zero or to an arbitrary value because most of the time there is no quotation to calibrate the pricing model. As we do not have price data to estimate the risk premium of the weather forward of Toulouse, we do not keep account of this term when evaluating the contract. The conditional expectation, for which it is difficult if not impossible to be reduced to a closed-form expression, is most often calculated by using the Monte-Carlo simulations. This method consists in generating a set of paths for the daily average temperature. For each of these paths, the CDD index is constructed. The conditional expectation of the weather forward price is then equivalent to the average of the indexes from all the generated paths. We try here to calculate the prices of the weather forward on the CDDs of Toulouse over a period of 2 months (06/01/2004-07/31/2004). The paths of the temperature will be generated on the one hand, by using the model with sunshine and the periodic CJR-GARCH process and on the other hand, by resorting to the model without sunshine and the periodic GARCH for comparison. For prices at a time $t > 0$, we construct the CDD index by simulating the paths of the temperature for the moments t to t_m and by taking observations of the temperature for dates prior to t (0 to t-1). The obtained predictions of price for the 2-month horizon appear in Figure 7. We can notice that prices from the model with sunshine and asymmetric effects and from the model without sunshine and asymmetric shocks are relatively close, especially for dates near the maturity date. However, we note that for these latter dates, prices provided by the model without sunshine are closer to the level reached by the CDD index at expiration as those given by the model with sunshine. This result is not surprising since we had shown that the model without sunshine was more efficient for forecasts of temperature in the short term. On the contrary, the price differences from the two models are more pronounced in the long term with prices much closer to the effective level of the index at maturity, in general, for prices from the model with sunshine. Moreover, we note that prices from the actuarial method becomes very inaccurate beyond 7 days.

Fig.7 : Predicted prices (expressed in CDD index points) of the weather forward of Toulouse from the actuarial pricing method with 10,000 simulations for each of both models of the daily average temperature.



4.2 Evaluation of the Value-at-Risk

The Value-at-Risk (VaR) is a measure of the maximum loss that can be expected for a position or a portfolio in normal market conditions and for a probability set *a priori*. More specifically, the VAR is the loss to be exceeded with a probability of only α% over the holding period of the asset or portfolio (α representing the percentage of abnormal fluctuations in the market. Generally α is fixed to 5 %). Let p_T be the value of an asset at the moment T and p_0, the value of the asset when estimating VaR, the change in the price of the asset over the period [0; T] is given by:

$$\Delta p = p_T - p_0. \quad (13)$$

This variation of the price is also called the "PnL" (Profit and Loss). From this definition, the VAR is defined then as follows :

$$\Pr\{ \Delta p > VaR(\alpha)\} = \alpha \quad \text{ou} \quad \Pr\{ \Delta p \leq VaR(\alpha)\} = 1-\alpha . \quad (14)$$

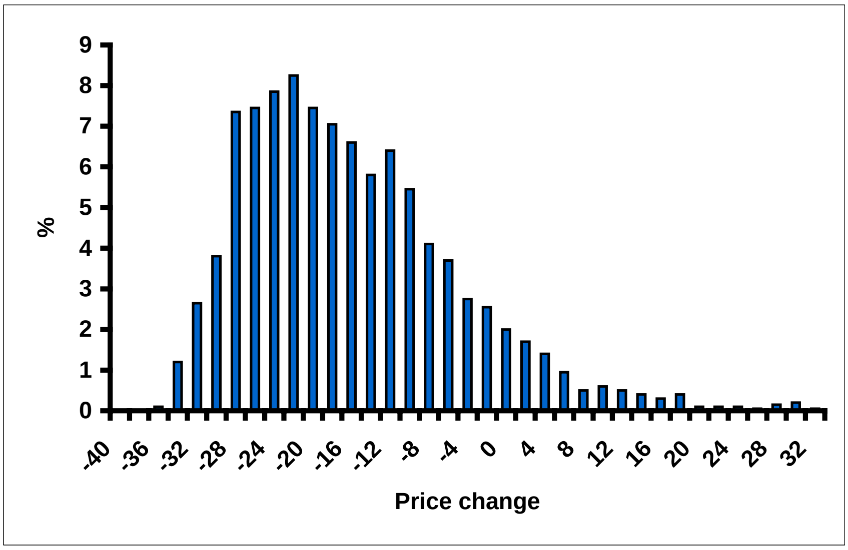
In other words,

$$VaR(\alpha) = F^{-1}(1- \alpha) \quad (15)$$

where F refers to the cumulative distribution function of the random variable Δp and F^{-1} is the inverse function of F.

Three methods are commonly used to estimate the VaR : variance-covariance method, historical simulation approach and Monte-Carlo simulation technique. In the case of the Monte-Carlo simulation method, n price changes are simulated. After that, they are arranged in ascending order. The VaR is then equal to the absolute value of $(n \times \alpha)$ i-th smallest value. For example, we assume here that an electricity company in Toulouse wants to hedge against a too strong rise in temperature during the first half of June 2004, which would increase above normal the electricity consumption for air conditioning. It signs then with a bank a forward contract on the CDD index with a perception of a compensation when the index exceeds the level 40 at the expiration date (the coverage period is of 15 days). However, the company will have to pay the bank if it is wrong in its anticipations (it will pay the positive difference between 40 and the level reached by the index at the maturity date). The level of 40 was chosen after simulation of the CDD index level at maturity (the obtained level was around 26) and comparison of this level with the average of the historical CDD indexes for the month of June (it was equal to 17 for the first half of June). The average of the historical CDD indexes was calculated from observations over the period 01/01/1949 - 05/31/ 2004. It was observed a magnitude of 91 for the CDD index in 2003 which is the highest record because of a long heatwave. To determine the VaR of the weather forward on the CDDs of Toulouse for the period 06/01/2004 - 06/15/2004, we calculate, as a first step, n = 2,000 prices for the forward contract on the date 06/15/2004. Each price are computed with 10,000 simulations of the daily average temperature. In a second step, we calculate n price variations over the considered period by setting for initial price of the contract, the level 40 of the CDD index. Finally, in last time, the n variations are ordered in ascending way to determine the VaR (see Figure 8).

Fig.8 : Frequency (in %) of the simulated PnL from the model with sunshine and periodic GJR-GARCH(1,1) for the daily average temperature



For α = 5% and a sample of 2,000 simulations, the Value-at-Risk is equal to 31.35 in CDD index points (or 31.35 € if one assumes that 1 degree-day is equal to 1 €). This value corresponds to the 101st value of the list. It indicates the potential loss on a horizon of 15 days and in 5% of the worst scenarios. To validate the model chosen for the temperature to calculate the VaR, we must verify that the actual losses would exceed the estimated VaR only in 5% of the cases during the holding period. To verify it, it would be necessary to be able to compare the daily actual losses with the daily VaR. In our case, it is difficult to carry out this task because we do not have quotations for the forward contract of Toulouse. It is therefore not possible to calculate the daily actual losses. We only have, for information, the level of the index at the maturity date (here the effective level of the index is 38.2). The electricity company has therefore lost 1.8 points. To overcome the lack of price data, practitioners determine the actual losses by using "marked-to-model" prices. In our case, there would be a strong risk of error given the inaccuracy of the estimated prices from the actuarial method beyond 7 days. However, it is possible to judge the results of a model for the temperature in the calculation of the VaR by applying this model to the temperature of a city for which there is a weather contract with quotations. If the model provides correct results in the calculation of the VaR of a quoted and liquid weather derivative, there are strong chances that this model also fits in the computation of the VaR of a non quoted weather contract.

5. Conclusion

By taking into account the impacts of sunshine and the asymmetric effects of the shocks on the volatility through the estimation of a structural model with a periodic CJR-GARCH (1,1) process, we have improved the in-sample forecasts of the daily average temperature with regard to those stemming from the univariate autoregressive process with a periodic GARCH(1,1) representation. Concerning the out-sample forecasts, the results from the structural model appeared better than those from the univariate model for the horizons of 30 and 60 days. Simulations of the daily average temperature were run by using a two-state first order Markov chain for the sunshine dynamics in order to capture the discontinuous pattern of the latter variable.Two examples of the use of the structural model were proposed for the weather risk management as the pricing of the weather forward of Toulouse and the evaluation of the Value-at-Risk of the contract. The structural model presents no difficulty of implementation and can easily be extended to other climate variables to improve the understanding of the behaviour of the temperature and also to simulate extreme and possible scenarios of the temperature.

References

Benth F.E. and Šaltytė-Benth J. (2005), “Stochastic Modelling of Temperature Variations with a View Towards Weather Derivatives”, *Applied Mathematical Finance*, Vol. 12, No. 1, 53-85.

Black F. and Scholes M. (1973), “The Pricing of Options and Corporate Liabilities”, *Journal of Political Economy*, Vol. 81, 637-659.

Campbell S.D. and Diebold F.X. (2005), “Weather Forecasting for Weather Derivatives”, *Journal of the American Statistical Association*, Vol. 100, 6-16.

Cao M. and Wei J. (1998), “Pricing Weather Derivatives : an Equilibrium Approach”, Working Paper, old version of the article titled “Weather Derivatives Valuation and Market Price of Weather Risk”.

Cao M. and Wei J. (2004), “Weather Derivatives Valuation and Market Price of Weather Risk”, *Journal of Futures Markets*, Vol. 24, No. 11, 1065-1089.

Engle R.F. (1982), “Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation”, *Econometrica*, Vol. 50, No.4, 987-1008.

Glosten L.R., Jagannathan R. and Runkle D. (1993), “On the Relation between the Expected Value and the Volatility of the Nominal Excess Returns on Stocks”, *Journal of Finance*, Vol. 48, No. 5, 1779-1791.

Nelson D. B. (1991), “Conditional heteroskedasticity in asset returns: A new approach”, *Econometrica*, Vol. 59, 347-370.

Roustant O. (2002), “Une application de deux modèles économétriques de température à la gestion des risques climatiques (1st part)”, *Banque & Marchés*, Vol. 58, 22-29.

Roustant O. (2002), “Une application de deux modèles économétriques de température à la gestion des risques climatiques (2nde part)”, *Banque & Marchés*, Vol. 59, 36-44.

Wilks D.S. (1999), “Multisite Downscaling of Daily Precipitation with a Stochastic Weather Generator”, *Climate Research*, Vol. 11, 125-136.

Zakoian J.M. (1994), “Threshold Heteroskedastic Models”, *Journal of Economic Dynamics and Control*, Vol. 18, No.5, 931-955.