Croatian Review of Economic, Business and Social Statistics (CREBSS)
UDK: 33;519,2; DOI: 10.1515/crebss; ISSN 1849-8531 (Print); ISSN 2459-5616 (Online)
Vol. 6, No. 1, 2020, pp. 27-42

Ornstein-Uhlenbeck process and GARCH model for temperature forecasting in weather derivatives valuation

Berislav Žmuk
University of Zagreb, Faculty of Economics and Business, Department of Statistics, Zagreb, Croatia
bzmuk@efzg.hr

Matej Kovač
Addiko bank, Croatia
kovacmatej95@gmail.com

Abstract
An accurate weather forecast is the basis for the valuation of weather derivatives, securities that partially compensate for financial losses to holders in case of, from their perspective, adverse outside temperature. The paper analyses precision of two forecast models of average daily temperature, the Ornstein-Uhlenbeck process (O-U process) and the generalized autoregressive conditional heteroskedastic model (GARCH model) and presumes for the GARCH model to be the more accurate one. Temperature data for the period 2000-2017 were taken from the DHMZ database for the Maksimir station and used as the basis for the 2018 forecast. Forecasted values were compared to the available actual data for 2018 using MAPE and RMSE methods. The GARCH model provides more accurate forecasts than the O-U process by both methods. RMSE stands at 3.75 °C versus 4.53 °C for the O-U process and MAPE is 140.66 % versus 144.55 %. Artificial intelligence and supercomputers can be used for possible improvements in forecasting accuracy to allow for additional data to be included in the forecasting process, such as up-to-date temperatures and more complex calculations.

Keywords: GARCH model, MAPE, Ornstein-Uhlenbeck process, RMSE, temperature forecasting, weather derivatives.

JEL classification: C22, C53.
DOI: 10.2478/crebss-2020-0003

Received: April 03, 2020
Accepted: May 10, 2020

Acknowledgements: Authors would like to thank ms. Vidić from DHMZ for her support in data collection

Introduction
Paper discusses the accuracy of two statistical prognostic models applied to the forecast of daily average air temperature in the City of Zagreb. Variability of daily air

temperature is a major factor in determining the price of weather derivatives, securities that represent sort of a bet on future meteorological conditions in certain area.

Main purpose of the paper is to determine the more accurate prognostic model of the two observed: Ornstein-Uhlenbeck process (O-U process) and the Generalized autoregressive conditional heteroskedastic model (GARCH) by discovering the one whose deviation of forecasted to actual values is the smallest. Successful forecast of temperature trends in the future is the basis for accurately constructed derivative that ensures the highest expected revenue for the derivative provider. Calculation and comparison of root mean square error (RMSE) and mean absolute percentage error (MAPE) will draw conclusions which of two mentioned models is more accurate one, with GARCH model being the more likely one before the results and analysis of the data because of better ability to capture the daily temperature jumps.

Paper is divided into five sections. After the introduction, section two provides literature review on weather derivatives forecasting and past examples of such analysis. Section three describes the basic principles and assumptions of two forecasting models. Section four discusses the accuracy of air temperature forecasting with selected models, finds the most accurate one and gives insight on possible improvements in air temperature forecasting. Section five concludes.

Literature review

Individuals can buy and sell call and put options with weather derivatives as underlying assets. Call option entitles its owner to buy the underlying asset from the provider at a predetermined price for a contractual period in exchange for a premium (Orsag, 2011). If the buyer has optimistic presumptions on the development of adverse weather which can make the price of derivatives go high, they opt to buy call options because of the capital gain caused by the fixed price the provider has to offer to the buyer. Put option, on the other hand, is again a right but not an obligation of the buyer of an option to sell the underlying asset at a fixed price (Orsag, 2011). It is usually contracted if the buyer has negative expectations on the development of adverse temperature which then causes the reduction of the derivative price.

Buyers also contract future contracts on weather derivatives, which means that they arrange to buy or sell the option for a predetermined price on a specified future date. Those kinds of contracts are interesting for the parties whose businesses usually depend on seasonality, such as construction business or tourism companies.

Weather derivatives are used more and more often since their aim is to diminish the potential loss for a company on a similar but much more accessible way compared to insurance contract. By assuming that even a small change of certain event can affect the business of the company, weather derivatives can be contracted for less significant changes in temperature or precipitation compared to insurance policies. Weather derivatives therefore cover low-risk/high-probability events, while insurance covers high-risk/low-probability events (Lazibat Županić, 2010). If, for example, there was an above average temperature during the heating season, companies would be able to obtain the amount contracted with the insurance policy in the case of extremely high temperatures, when energy companies become significantly affected by such weather. Weather derivatives, on the other hand, assume that even a small change in daily air temperature can affect revenues of the company so the contracted amount can be paid even when the air temperature deviates only a few degrees (three to five degrees Celsius) to the expected value.

Weather derivatives and their valuation became popular in modern literature at the end of the 20th century when the first such instrument was introduced and contracted between two USA based energy companies, as mentioned and explained in Considine (2000).

Since its inceptions at the start of the century, weather derivatives have developed into a liquid market paper with different versions of derivative contracts, such as swap contracts, futures/forwards or options on forward contracts with the two latter being the most described and observed ones in Lazibat and Županić (2010), Baković et al. (2011) and Till (2015) among others. All these contract types must define the weather index which is the starting point in weather derivatives valuation. Alaton et al. (2002), Cambell and Diebold (2005) and Cao and Li (2003) focus on defining heating degree day (HDD) and cooling degree day (CDD) indices which presents a starting point in weather derivatives valuation. Based on the value of the index, monetary value of each degree Celsius is calculated and the monetary value of weather derivative evaluated.

Temperature forecasting is not an easy task. Models chosen for the forecasting process have a documented history of application in many areas of stochastic modelling. Literature knows many forms of O-U processes. Models used in temperature forecasting vary from more simple equations in Alaton et al. (2002) or Benth and Benth (2005) to more complex ones in Alexandridis and Zapranis (2007). Combination of all available models lead to the more simplified notation of the model in this work. GARCH model can be found in many scientific articles, most of which cover the topics of stochastic modelling with moving averages or reverting processes. Most relevant were adapted based on the equations from Cambell and Diebold (2005), Buizza and Taylor (2004) and Gilks et al. (1996).

Some authors suggest using different methods than those selected, such as Gilks (1996) using Markov Chains or Alexandridis and Zapranis (2007) with neural works together with Berliner (2001) which uses Ensemble forecasting. All mentioned try to explain some other weather phenomena, even more stochastic than temperature, and are therefore omitted from this work.

Methodology

Characteristics of weather derivatives

Solid weather conditions, along with adverse ones, affect almost every economic activity and can cause significant financial loss to the economy. Hail or dry periods damage the agriculture, warm winters hurt energy sector, while cold or rainy summers strike tourism. Due to the high unpredictability of the weather and air temperature forecasting, it is necessary to explain weather derivatives as an instrument of protection against the variability of the weather, specifically air temperature as one of the key characteristics of the weather.

Weather derivatives are financial derivatives (forward contracts and options on forward contracts) whose payment depends on future weather conditions. They insure industries against the adverse impact of weather with help of weather index calculated as the deviation of current air temperature from the selected reference point (Baković et al., 2011).

Weather index refers to the difference between the daily average air temperature from a certain temperature reference value set by the derivative provider. Value of the reference temperature is usually set at 65 Fahrenheit (Schiller et al., 2012), corresponding to 18 °C. Derivative buyer bets that an unfavourable development of air temperature will occur to his business for a contractual period. If the buyer's forecast turns out to be correct, the provider will pay him a predetermined fee.

Meteorological Station Maksimir measures the mean daily temperature in Zagreb as an arithmetic mean of hourly temperatures, same as in Dall'Amico and Hornsteiner (2006). Arithmetic mean of the daily temperature ($\bar{T}$) is thus given by the expression:

$$\bar{T} = \frac{1}{24\,h}\int T\,dt, \tag{1}$$

where T represents the air temperature at the beginning or end of each hour in one day, while $24\,h$ stands for the hours in one day (24 hours).

Daily average temperature (T_t) is a vital part in constructing weather index. Many authors, including Alaton et al. (2002) and Considine (2000), mention two types of weather indices related to air temperature. These are: heating degree day (HDD) and cooling degree day (CDD). Heating degree day index is defined as:

$$HDD = \sum_{T1}^{T2} \max\left(T^{ref} - T_t; 0\right), \tag{2}$$

where T^{ref} refers to the reference temperature which is usually set at 18°C, and T_t is the daily average temperature according to the data collected from the meteorological station. HDD index is associated with colder weather (usually winter) that requires the use of heating in homes (H in HDD stands for heating) (Till, 2015).

Cooling degree day on the other hand, is defined as:

$$CDD = \sum_{T1}^{T2} \max\left(T_t - T^{ref}; 0\right), \tag{3}$$

For the CDD index, which is used during the summer months, it is assumed that the daily average temperature will be higher than the reference temperature, which switches the factors in the equation. It was named CDD index because of cooling (C in CDD stands for cooling) - most people use cooling during warm days, or those whose average daily temperature is higher than 18 °C with the help of cooling machines (Till, 2015).

During the contract period, probability of profit for a weather derivative provider depends on the precision of the forecasting model. Providers of weather derivatives should forecast the temperature accurately in order to plan possible payouts based on these values. Air temperature forecasting has some similarities with the forecasting of stock prices, mostly in dealing with the stochasticity of forecasting parameters with stochasticity being the variability of daily average air temperature regardless of its movement in the past.

Despite extremely unpredictable daily temperature movements and various jumps, there are some general characteristics of the daily air temperature which should be covered by the statistical models used in forecasting:

1. Autoregression - Daily average temperature depends on the temperature on the same day in some of the periods before the forecast period or few days before the forecast day in the forecast period,
2. Return to the long-term average - winter and summer temperatures represent seasonal fluctuations from the average of the daily temperature which reaches between 15 and 20 °C in a regular year,
3. Sinusoidal form of temperature distribution - during the year temperatures move like a sine function - they rise, reach a maximum, fall, reach a minimum and then rise again over a period of 365 days a year.

Google scholar, Hrčak Srce database and the databases of the Documentation Center of the Faculty of Economics and Business in Zagreb were sources used during the research, along with academic workbooks. Data on daily air temperature for the period 1.1.2000–31.12.2018 were taken from the database of average daily temperature in Maksimir station in Zagreb, provided by Croatian Meteorological and

Hydrological Service. Data processing and analysis presented in the paper were carried out through Microsoft Excel, R-Studio and E-Views.

In the sense of comparison of the two forecasting models, it is necessary to calculate the deviations of the forecasted values from the actual values with indicators applicable to both models. First used will be root mean square error (RMSE). Chai and Draxler (2014) find this indicator, which is defined by following expression, to be often used to evaluate the performance of forecasting models in meteorology:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(Y_t - \hat{Y}_t)^2}{n}}, \quad (4)$$

where Y_t represents the true value of the variable in period t, $\hat{Y}_t$ the forecasted value of the variable, while n refers to the number of periods for which the forecast is made.

Mean absolute percentage error (MAPE) is another indicator that will be used in model performance evaluation. According to Hanke and Witchern (2014) the expression of this indicator is:

$$MAPE = \frac{1}{n} * \sum_{t=1}^{n} \frac{|Y_t - \hat{Y}_t|}{|Y_t|} \times 100, \quad (5)$$

where n is the number of forecast periods.

Ornstein – Uhlenbeck process

Ornstein-Uhlenbeck process (O-U process) is a mathematical model that was developed in the 20th century, when French physicist Paul Langevin created a formula explaining stochastic motion of particles in liquids (Gillespie, 1990). However, actual creators of the model were Ornstein and Uhlenbeck (1930), extending the basic motion formula with Einstein's explanation of the free movement of particles and atoms in space. Due to the complexity of the noise analysis and calculation, the part of the equation that observes the stochastic component of temperature has been omitted from this analysis, which concentrated more on accurately capturing the three previously mentioned temperature characteristics.

Model was developed based on the continental climate (original paper observed air temperature data Bromma airport near Stockholm, Sweden) which has a seasonal characteristic (Alaton et al., 2002). Average daily air temperature, with the previous assumption of seasonality, varies from extremely cold in winter, pleasant in spring and autumn, to extremely warm in summer.

Due to the summer and winter season, the movement of daily temperature, T^s, can be simplified by a sine function that depends on the time t, where t is day of the year, as follows:

$$T^s = \sin(\omega t + \rho), \quad (6)$$

where t denotes the time in days for the current year (1-365), ω represents the term 2π / 365, which defines the period of the function that corresponds to the period of 365 days, while ρ calculates the shift from the January average temperature to the actual yearly average temperature (which mostly occurs in spring or autumn months). In the period 2000 to 2017 which is a base period for the 2018 forecast, there were five leap years (2000, 2004, 2008, 2012 and 2016) with 29 days in February which were omitted to ensure that all years have comparable data.

Basic approximation, expressed by the sine function solely with the parameter ωt, would describe the temperature distribution in such a way that the yearly average temperature would be equal to the average temperature on January 1, while the maximum temperature would be predicted on March 1 and the minimum on September 1. However, such an assumption is unrealistic, so the record includes a shift of the sine curve ρ that ensures lowest value of the function (T_{min}) in the middle of winter, and consequently the highest value (T_{max}) in the middle of summer (Alaton et al., 2002).

As a result of global warming and other climatological factors, daily temperatures have a slight upward trend as the years go by, so average daily temperatures today are somewhat higher than in the past. By combining the seasonal (summer and winter), trend components (global warming) and the expression of a shift of sine function, all factors for analysing the O-U process are ready. Equation of the prognostic O-U process that will be used in part four to forecast daily temperatures in 2018 is given by the expression:

$$T_t = A + Bt + Csin(\omega t + \rho), \quad (7)$$

where T_t represents the forecasted value of the temperature for a t day in 2018 and t is the number of days in the year (1-365). Variable A represents the average of daily air temperature for the period 1.1.2000 – 31.12.2017, variable B defines the influence of an increasing temperature trend on a yearly basis, while variable C determines the seasonality of temperatures throughout the year or how strongly winter and summer temperatures deviate from the yearly air temperature mean. Variables A, B and C, together with the shift ρ will be calculated based on historical temperature data.

GARCH model

Generalized autoregressive conditional heteroskedastic model (GARCH) predicts the variability of the future and solves the problem of heteroskedasticity - the non-constant variability of time series (Bahovec, Erjavec, 2009). GARCH model is often referred to as the GARCH (1,1) model. Factors (1,1) mean that the present variance depends on the forecast error in one period before and on the variance in one period before the present. Expression which describes the possible existence of heteroskedasticity and then corrects it, based on the expression in Frances and Dijk (1996), is presented as:

$$\sigma_t^2 = \omega + \alpha * \varepsilon_{t-1}^2 + \beta * \sigma_{t-1,}^2 \quad (8)$$

where the factors ω, α, β are calculated using the statistical software E-Views once a complete model to forecast the daily average temperature is constructed. GARCH model comes after the initial forecasting model because it works on fixing residuals which can occur only if there are actual and forecasted values that can be compared. Expression ε_{t-1}^2 represents the residual in one period before the forecast period and σ_{t-1}^2 standard deviation in that same period.

Daily average temperatures depend on their corresponding month. To capture this effect of the month on the daily air temperature, 12 dummy variables (βn, n ranges from 1 to 12) will be introduced into the forecasting model. Each indicator variable corresponds to each month of the year, January to December. Values of indicator variable are usually 0 or 1, with 0 meaning the exclusion of some feature and 1 presents the opposite. This will be used to calculate the average temperature of the month in which some day is calculated and will represent a constant.

In addition to the dummy variable as the first part of the overall GARCH model, a part that connects today's forecast temperature to the actual temperature one and two days before the forecast day with a multiple regression model should be added. In statistics, this term is known as the AR (2) process (Kölbl, 2006).

Expression (8) that was shown to explain the problem of heteroskedasticity is the notation of an AR (1) process where the dependent variable y depends on only one independent variable x. AR(2) process adds another independent variable x_2 with its corresponding coefficient β_2. When this extension of expression (8) is translated into the terms of the forecasting model for daily average temperature, AR (2) process by which the predicted value depends on the actual in the two days before can be written as:

$$T_t = \alpha + \gamma * T_{t-1} + \delta * T_{t-2} + \epsilon, \quad (9)$$

where T_{t-1} is the actual daily air temperature of the day before the forecast and T_{t-2} is the actual daily air temperature two days before the forecast. Factors γ, δ and a are the coefficients and ϵ is the residual for which the possible existence of the heteroskedasticity problem will be tested. Instead of residual ϵ, the final expression of the GARCH model will have expression (9). In expression (9) there is a free coefficient a, but in the final expression of the model it will not be visible as a but as a previously explained βn indicator variable for each month of the year.

Temperature forecasting requires AR(2) process because it makes the forecasted temperature much more accurate than only coefficients β_n as indicator variables. If the daily forecasted temperature would be just the constant β_n, forecasted temperature would be represented by a horizontal line that shows the average temperature of a corresponding month. Forecast values would show substantial variability to actual ones since it is highly unlikely that the daily temperature remains constant for one month. It is therefore necessary to involve actual values of two days before the forecasted day to capture upward or downward trend of the temperature.

Average monthly temperature with the existence of an indicator of variables β_n + AR (2) process of linking the actual daily temperature of the past two days with forecast of today + GARCH (1,1) model for residual and standard deviation analysis gives a complete expression of the model that will forecast the temperature in 2018 . GARCH model expression is therefore presented as:

$$T_t = \beta t * t + \gamma * T_{t-1} + \delta * T_{t-2} + \omega + \alpha * \varepsilon_{t-1}^2 + \beta * \sigma_{t-1}^2, \quad (10)$$

where: βt – dummy variable with t being month of the year, γ, δ – regression coefficients in AR (2) process for T_{t-1} and T_{t-2} actual air temperature values, respectively, ω, α, β – coefficients in GARCH (1,1) model.

Expression (10) simultaneously calculates the daily average temperature forecast for 2018 with the help of the AR (2) process as well as analyses the deviations of the actual values from the forecasted ones and the potential existence of heteroskedasticity with the help of the GARCH (1,1) process.

Air temperature forecasts, precision of forecasting models and possible future development

Descriptive statistics

Daily average temperatures in each of the eighteen base years (2000-2017) should show expected movements, which means that in the winter months (October to February) temperatures should be low and reach extreme low values during this period, while in the middle of the year, during the summer months (June to August), they should be highest. Period between winter and summer should show a constantly increasing trend of temperatures in the spring months, as well as a constantly decreasing trend of temperatures in the fall as the daytime temperature drops and approaches the cold winter. Picture 5 shows the actual daily average temperature at the Maksimir station for each of 365 days in a year for a base period.

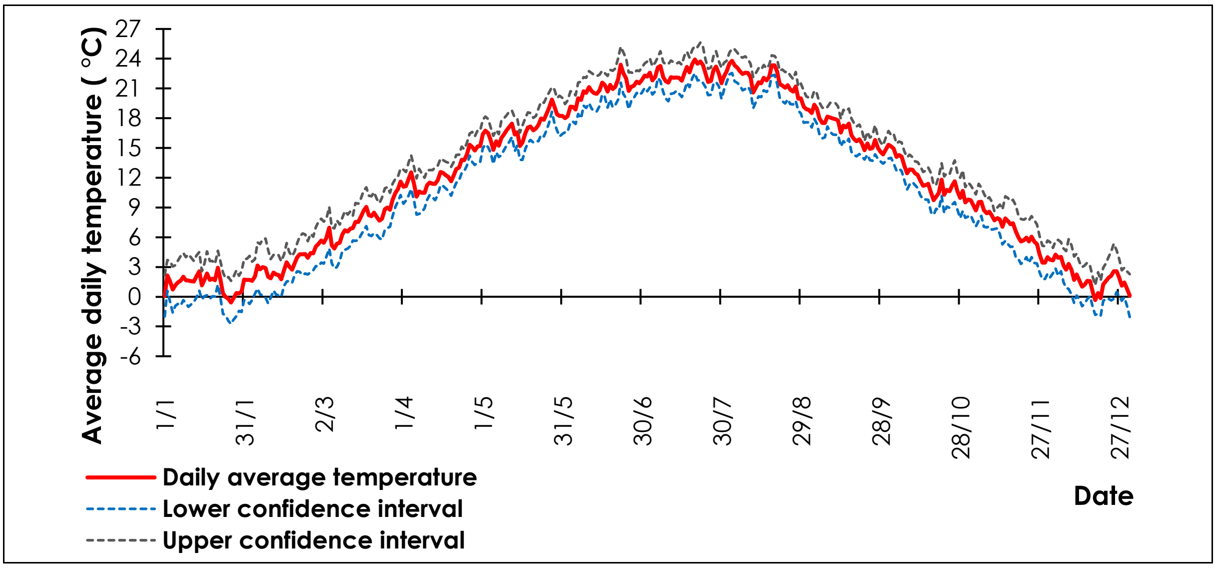

Figure 1 Daily average air temperature Zagreb (Maksimir), in °C, included period from 2000 to 2017

Source: authors based on Vidić (2019).

Daily average air temperatures in Figure 1 show that highest values correspond to summer months and the lowest values at the end of one or the beginning of the second year. Graphic, as well as temperature ranges, are similar to those in the of Buizza and Taylor (2004) and Campbell and Diebold (2005). Besides the average daily temperatures, upper and lower confidence interval limits with confidence level 95 % are shown on the graph.

Descriptive statistics of data for all 6,570 days (18 base years – without data for 29.02. of leap years) are presented in Table 1, while Table 2 shows the same data for every third year in base period. All indicators except the coefficient of variation, skewness, and kurtosis as relative indicators are expressed in degrees Celsius (°C).

Table 1 Selected indicators of descriptive-statistical analysis of daily average air temperature 2000.-2017.

Indicator	Value
No. of days observed	6,570
Mean	12.12
Standard deviation	8.47
Coefficient of variation	69.88%
Median	12.70
Mode	17.40
Skewness	-0.81
Kurtosis	-0.17
Maximum temperature	31.70
Minimum temperature	-12.40
Range	44.10

Source: authors.

Descriptive analysis of the data shows that the average value of the average daily temperatures is 12.12 °C, while the modal temperature (the one most frequently repeated in the series) is 17.40 °C. Interestingly, the mode is very similar to the temperature of 18 °C, used in time derivatives as a reference temperature for HDD and CDD indices. Coefficient of variation, which shows the variability of the data, that is, how representative the calculated average of 12.12 °C is, is just over 70% and shows the high variability of the data – consequently the calculated average has poor representativeness.

In Gaussian or normal distribution, the mean of the data is equal to the mode (the most common value) and the median (the value that divides the distribution into two equal parts). Observed distribution is slightly asymmetric to the left, showing an asymmetry coefficient (skewness) of less than 0 (-0.17). Mode of data of 17.4 °C is higher than the median of 12.7 °C and the mean of 12.12 °C. Left-sided asymmetry whose basic condition is that the mode is greater than the median greater than the arithmetic middle is thus confirmed (Mandikandan, 2011).

Data and small-scale distributions have a positive Kurtosis, pointed tip and low tails because there are few extreme values. Such distributions have very few serious unusual values (or outliers) and most of the data are close to the average, which is then more representative and accurate. Opposite case is where the distribution is slightly rounded and has thicker tails, as is the case with the observed data. Kurtosis of the observed data has a coefficient of less than 0 (-0.81), so the top of the distribution is slightly rounded and the distribution tails relatively high due to the higher representation of extreme values. This shows a widespread distribution and a lower representativeness of the calculated average of daily average temperatures.

Table 2 outlies the value of same coefficients as Table 1 but now for specific years in the base period (each 3rd year starting with 2000). This table might give a better insight in the assumption of global warming and better understanding of the movement of temperature. Question on capturing the variability of the data and stochasticity in modelling weather still plays a pivotal role in efficient forecasting.

Table 2 Indicators of descriptive statistical analysis of daily average air temperature of selected years

Indicator	Year					
	2000	2003	2006	2009	2012	2015
Number of days	365	365	365	365	365	365
Mean	12.73	11.91	11.85	12.45	12.57	12.70
Standard deviation	8.25	9.80	8.62	8.64	9.34	8.19
Coefficient of variation	64.80%	82.28%	72.74%	69.40%	74.30%	64.48%
Median	13.85	11.95	13.10	13.60	12.60	12.50
Mode	8.60	21.20	16.40	6.20	6.60	14.40
Excess	-0.49	-1.12	-0.82	-0.68	-0.58	-1.05
Kurtosis	-0.29	-0.06	-0.21	-0.36	-0.30	0.14
Maximum temperature	29.60	29.60	28.30	27.20	30.10	29.40
Minimum temperature	-11.60	-10.50	-10.00	-12.40	-10.90	-3.60
Range	41.20	40.10	38.30	39.60	41.00	33.00

Source: authors.
Note: Data for February 29 in 2000 and 2012 are excluded for the comparability of the data.

Descriptive statistics indicators show that, for example, average of daily average temperature rises slightly over the years and shows a value higher by almost 1 °C in the 10-year period (2006-2015). In 2006 the average daily average temperature was 11.85 °C, while in 2015 it was 12.70 °C. This phenomenon is included in the O-U process via parameter B in the record (6), while in GARCH model it is covered by the AR (2).

Analysis of the base period shows that the coefficient of variation (an average relative deviation from the average) is larger than 60%. Such value shows that the calculated averages of the average daily temperature are not representative and that the daily average temperature is extremely dispersed over the years. Variation of about 40 °C between the highest and lowest average daily temperature per year in almost all observed years confirms this huge dispersiveness making the base period not so stable for further forecasting.

Ornstein-Uhlenbeck process forecasting

In order to forecast the daily average temperatures of the O-U process for 2018, it is necessary to determine the parameters A, B, C and ρ from expression (7) so that only the independent variable t, representing the days of the year, can be changed and forecast calculated.

Parameter A represents arithmetic mean of all average daily temperatures for the base period 2000-2017. Value of parameter A will be a constant in the O-U expression. Parameter B shows expected increase in temperature in 2018 compared to an average of 12.12 °C will also be a constant. Parameter C represents approximately half of the maximum and minimum daily average temperature range in 2018 (2*C indicates the amount of the total forecast maximum range and minimum temperatures). Finally, the shift ρ will determine the expected position of yearly maximum and minimum temperature.

By analysing the daily temperatures for the base period, the O-U process given by expression (7), with parameters A, B, C, and ρ, takes the following form:

$$T_t = 12.12 + 3.48 + 10.654 * sin\left(\frac{2\pi}{365} * t + 1.499335\right), \quad (11)$$

where A is the coefficient calculated as the average of daily average temperatures between 1.1.2000 - 31.12.2017 and stands at 12.12 °C. Coefficient B represents the expected increase in the annual average temperature in 2018 compared to the average of the average temperatures in the base period (2000-2017) and stands at 3.48 °C.

Factor C stands at 10.65 °C which is logical and expected, as it assumes that the highest value of the average daily temperature in 2018 will be around 23 °C (12.12 °C + 10.65 °C), while the lowest will be about 1.5 °C (12.12 °C - 10.65 °C).

Sine function shift, ρ, stands at 1.49 in terms of sine function, which corresponds to 2.85 months, or nearly three months of a year, which means that the average daily temperature, previously calculated 12.12 °C, can be expected at the end of March (2.85 months behind start of the year), and that all average daily values starting from 1st of January should be lower than 12.12 °C with an upward tendency. Shift of average daily temperature to March corresponds to the movement of temperature in an average year unexposed to temperature shocks.

By adding 365 days of the year instead of t, the forecast values for 2018 were obtained through the O-U process and plotted along with the actual 2018 daily temperatures in Figure 2.

Figure 2 shows how the O-U process, without analysing stochastic jumps, is a pretty good prognostic model of daily average temperature movement. When stochastic jump analysis is excluded, the O-U process is a process that describes the sinusoidal function.

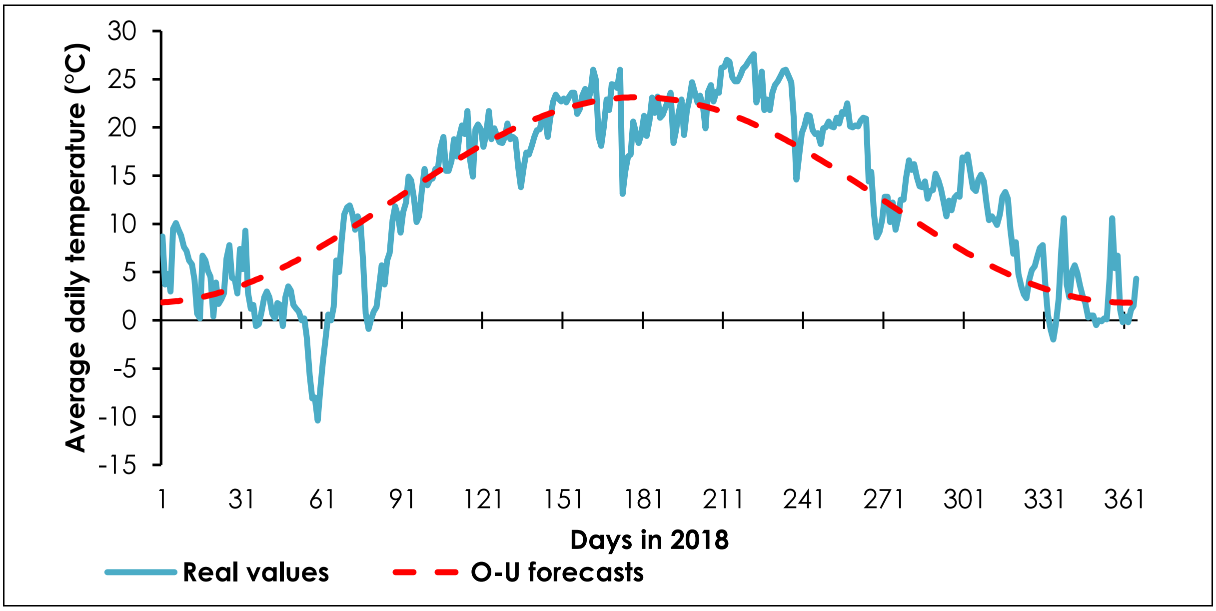

Figure 2 Actual daily average temperatures for 2018 and values forecasted by the Ornstein-Uhlenbeck process

Source: authors.

MAPE shows what is the average deviation of forecasted to actual value, relatively. Forecasted O-U values deviate 144.55 % from its actual values for 2018, on average.

According to Lewis (1982) such value can be observed as inaccurate forecasting since measures say 50% is the maximum MAPE value which can be perceived as good enough. Although RMSE indicator shows average deviation of 4.53 °C which compared presents a reasonably accurate forecast (compared to the average daily temperature in the year of 12 °C), problems with the representativeness of the average and inability to explain the stochastic jumps lead to the conclusion of somewhat inaccurate model based on MAPE indicator.

Because of simplicity, this model is incapable of accurately explaining the stochasticity that exists in each day of the year. O-U process in this form is smooth, so another representation of the actual data for 2018, perhaps in monthly rather than daily form, would help in better understanding of the fitness of the model. Monthly data are not pointed and could better go with flat sine curve presented in O-U process. Potential forecasting of daily jumps could be done as in Alexandris and Zapranis (2007) with so-called wavelet functions, which are derivatives of neural networks. Wavelet functions seek to break down each individual effect that affects the air temperature into individual parts based on a large amount of data on past temperature values, and then reassemble them in a prognostic period into one effect that will explain the temperature value.

GARCH model forecasting

Forecasting by the GARCH model for 2018, requires determining the parameters $\beta t, \gamma, \delta, \omega, \alpha, \beta$ from expression (10). All these parameters will be calculated in the E-Views program, based on daily temperatures from the base years 2000-2017.

Parameter βt, a constant in the final expression of the GARCH model, will be calculated as the average of daily average temperatures for each day in January for 18 base years, and will represent a $\beta 1$ constant for all 31 forecasted January days in 2018. Same goes for February ($\beta 2$), March ($\beta 3$) and other months.

UDK: 33;519,2; DOI: 10.1515/crebss; ISSN 1849-8531 (Print); ISSN 2459-5616 (Online)
Vol. 6, No. 1, 2020, pp. 27-42

Regression coefficients γ and δ that connect today's forecasts to the actual values for one and two days before it will be calculated using the least squares method. After that, the program will consider the differences between the actual and forecast values and, if necessary, correct the existence of heteroskedasticity using the parameters ω, α, β and the associated values of residuals and variances.

Expression (10) after the calculation of the coefficients takes the following form:

$$T_t = \beta_n + 0.8713 * T_{t-1} - 0.0932 * T_{t-2} + 1.6266 + 0.1003 * \varepsilon_{t-1}^2 + 0.6271 * \sigma_{t-1}^2, \quad (12)$$

where 0.8713 is the regression coefficient with the actual average daily temperature of the day before the forecasted day, -0.0932 is the regression coefficient with the actual average daily temperature of two days before the forecasted day. Constant factor ω from the expression 11 is 1.6266. Regression coefficient with differences in residuals in two consecutive prognostic periods is 0.1003 while the coefficient with difference of standard deviations in two consecutive periods is 0.6271. Value of coefficients β_n with the indicator variable for each month of 2018 shown in the attachments.

Regression coefficient with the variable T_{t-2}, -0.0932, shows an interesting value. Such value indicates that today's forecast of average daily temperature depends negatively on the actual average daily temperature of two days before. Another interesting feature is how, when calculating regression coefficients, program calculates and assigns greater weight to independent variables that are closer in time to the forecast period and lesser to those that are further from the forecast period. Thus, the coefficient with T_{t-1} (actual temperature of the day before the forecasted day) is 0.8, while the coefficient with T_{t-2}, i.e. with the actual temperature two days before the forecast is only -0.09.

Expression (12) with actual daily average temperatures from the base period inputted in E-Views program gives forecasted average daily temperature values for 2018, shown in Figure 3, together with actual values.

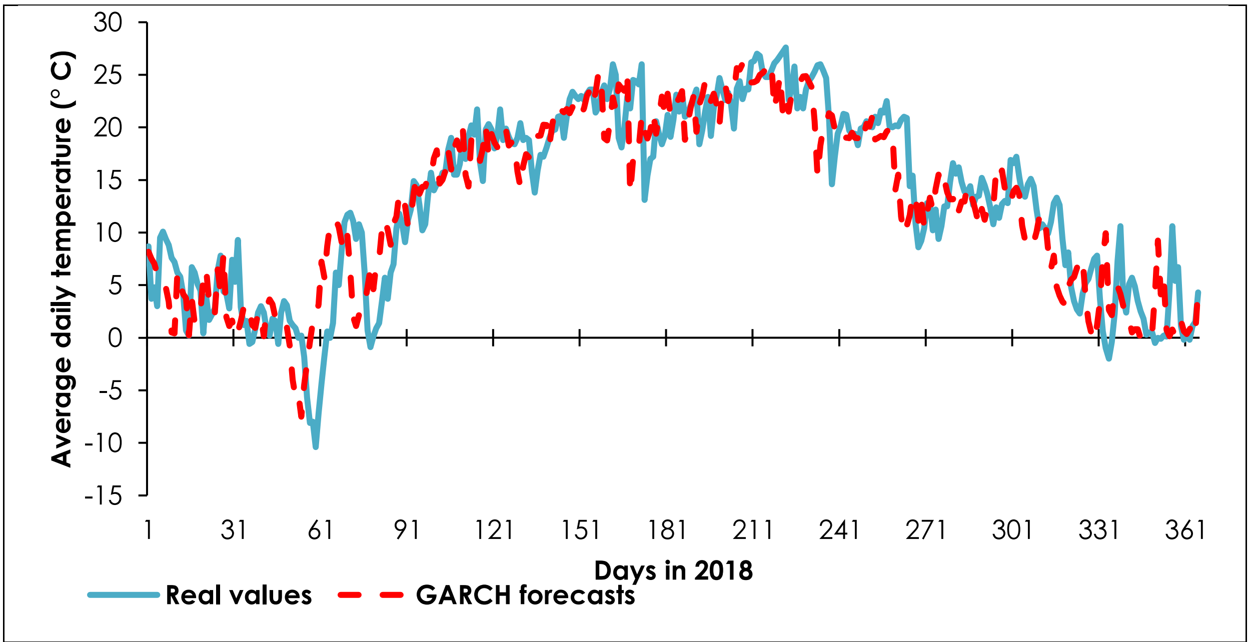

Figure 3 Comparison of actual values of average daily temperature for 2018 and GARCH forecast

Source: authors.

Despite seemingly accurate forecast presented on the graph, GARCH model must be equally examined via RMSE and MAPE indicators. RMSE for the GARCH model is 3.75, which means that the average deviation of the forecasted from the actual values is 3.75 °C. MAPE indicator for the GARCH model is 140.66%.

Possible further accuracy of GARCH model could be achieved by calculating regression coefficients and daily forecasts with help of supercomputers, same as in European Center for Medium Term Weather Forecasts (ECMWF) (Samso, 2018) and some other state meteorological offices. However, ECMWF supercomputers can forecast temperature with extremely high accuracy based on real-time observations and many calculations, making them the most successful platform for forecasts in the world so far. Such models, however, can hardly be used in scientific research because of their expensiveness and infrastructural requirements.

Forecasting results comparison

With values of RMSE and MAPE known for both models, all data for selecting the more accurate prognostic model is available. Table 3 gives an overview of the indicators for the observed models.

Table 3 Mean absolute percentage error (MAPE) and root mean square error (RMSE) of prognostic models

Indicator	Ornstein-Uhlenbeck process	GARCH model
RMSE	4.53 °C	3.75 °C
MAPE	144.55 %	140.66 %

Source: authors.

Bearing in mind that smaller value of the indicator presents a more accurate model, GARCH model is more successful in estimating the daily average air temperature than the O-U process with an average deviation of forecasted from the actual values of 3.75 °C versus 4.53 °C for the O-U process. On the other hand, MAPE indicator shows that the GARCH model is again slightly more successful than the O-U process, with a relative deviation of the forecasted from the actual values of 140.66% versus 144.55% for the O-U process. These indicator values confirm the hypothesis set before the calculation process, GARCH model can capture upward and downward trends of the temperature with regression equation much better than the fitting sine curve in O-U process.

Conclusion

Weather derivatives are financial instruments whose value is derived from the value of other assets, which in this case is the daily average air temperature. Meteorological phenomena such as rain, wind or fog can be also covered by this instrument which was created at the end of the 20th century to secure various industries such as energy sector or tourism whose business depends on the movement of air temperature. Businesses can bet on warm or cold weather using CDD and HDD indices that compare the average daily air temperature to a reference value of 18 °C and make payouts based on differences.

Weather derivatives pricing depends on their potential payments or how much can holders profit from them. Average air temperature is a specific "asset" since there is no other comparable asset to help with valuation. Therefore, valuing a weather derivative depends solely on a good forecasting model that models air temperature in the future. Stochasticity of temperature presents great obstacle in forecasting process, but some basic regularities in temperature movement exist: Firstly, autoregression speaks about the correlation of temperature to some periods close to the examined one; Secondly, sinusoidal movement shows a seasonal shift from the long-term average in summer and winter but also a return to the average in spring and autumn and finally yearly temperatures rise slightly as the years go by.

Models used in 2018 daily air temperature forecasting based on the 2000-2017 base period were O-U process and GARCH model. Daily forecasts were compared to the actual values of 2018 to determine the more precise model. GARCH model turned out to be a more accurate model for temperature forecasting, because the prognostic values, on average, deviate from the actual by 3.75 °C, while for the O-U process this indicator stands at 4.53 °C. GARCH model proved to be more successful according to the MAPE indicator as well, with relative deviation of forecasted to actual values of 140.66% compared to 144.55% in the O-U process. Weather derivative provider should, if choosing between these two models, opt for the GARCH model, which guarantees him a more accurate temperature forecast and, consequently, more stable earnings and cost planning when composing this financial instrument. GARCH model forecasts follow the regression line which proves to be more adequate and precise way of forecasting stochastic phenomenon of temperature as compared to average values and general trends in O-U process. Using two days temperature as independent variables $\boldsymbol{x_1}$ and $\boldsymbol{x_2}$, $\boldsymbol{t_{n+1}}$ day forecast is much more precise than using only one or none independent variables. Besides using short time span before the forecast period (that is, real daily temperature values), it is necessary to collect enough data for a base period to allow for the forecasting model to reveal potential patterns in historic data and replicate them on forecast data. Recent trends with supercomputers and real-time observations go along with these conclusions, as the most precise models from European Centre for Medium-Range Weather Forecasts (ECMWF) dispose with large amount of historical data as well as fast calculating programs which develop an improved forecast on incoming hourly data.

Nevertheless, as both observed indicators for the models show similar values and their difference is not large, the O-U process also deserves attention as a relevant model in air temperature forecasting. Limitations of this analysis such as the scope of database of 18 years, one temperature measuring spot (Maksimir), comparison of forecast values with MAPE and RMSE indicators and omission of detailed stochastic modelling in this work leave room for further research and possible different conclusions about the success of the O-U process and the GARCH model in forecasting air temperature using different methodologies or data. Although limitations of both models set them secondary to modern methods of forecasting using Artificial Intelligence, they still prove that standard mathematical procedures and statistical methods do provide solid forecasts of stochastic variables when chosen properly and executed with right assumptions.

References

1. Alaton, P., Djehiche, B., Stillberger, D. (2002). On Modelling and Pricing of Weather Derivatives. *Applied mathematical finance journal*, Vol. 9, No. 1, pp- 1-20.
2. Alexandridis, A. K., Zapranis, A. D. (2007). *Wavelet Neural Networks for Weather Derivatives Pricing*. 6th Hellenic Finance and Accounting Association Conference, Patra.
3. Bahovec, V., Erjavec, N. (2009). *Uvod u ekonometrijsku analizu*. Element, Zagreb.
4. Baković, T., Lazibat, T., Štulec, I. (2011). Specifičnosti trgovanja vremenskim derivativema ovisno o organiziranosti terminskog tržišta. *Zbornik Ekonomskog fakulteta u Zagrebu*, Vol. 9, No. 1, pp. 1-16.
5. Benth, F. E., Benth, J. Š. (2007). The volatility of temperature and pricing of weather derivatives. *Journal of Quantitative Finance*, Vol. 7, No. 5, pp. 553-561.

6. Berliner, L. M. (2001). Monte Carlo Based Ensemble Forecasting. *Statistics in Computing*, Vol. 11, No. 3, pp. 269-275.
7. Buizza, R., Taylor J. W. (2004). A Comparison of Temperature Density Forecast from GARCH and Atmospheric Models. *Journal of forecasting*, Vol. 23, No. 5, pp. 337-355.
8. Campbell, S. D., Diebold F. X. (2005). Weather Forecasting for Weather Derivatives. *Journal of American Statistical Association*, March 2005, pp. 1-11.
9. Cao, M., Li, A. (2003). *Weather derivatives: New Class of Financial Instruments*. Rotman School of Management.
10. Chai, T, Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)?. *Geoscientific Model Development*, Vol. 7, No. 1, pp. 1247-1250.
11. Considine, J. (2000). Introduction to Weather Derivatives. *Agroinsurance*, pp. 1-10.
12. Dall'Amico, M., Hornsteiner, M. (2006). A Simple Method for Estimating Daily and Monthly Mean Temperatures from Daily Minima and Maxima. *International Journal of Climatology*, Vol. 26, No. 1, pp. 1930-1936.
13. Frances, H. P., Dijk, D. (1996). Forecasting Stock Market Volatility Using (Non-Linear) Garch Model. *Journal of Forecasting*, Vol. 15, No. 1, pp. 229-235.
14. Gilks, W. R., Richardson, S., Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London.
15. Gillespie, D. T. (1990). Exact Numerical Simulation of the Ornstein-Uhlenbeck Process and its Integral. *Physical review*, Vol. 4, No. 2, pp. 1-8.
16. Hanke, J. E., Witchern, D. (2014). *Business Forecasting*. Pearson Education Limited, Essex.
17. Kölbl, F. (2016). *Aggregation of AR (2) process*. Diplomarbeit. Technische Universität Graz.
18. Lazibat, T., Županić, I. (2010). Vremenske izvedenice kao instrument upravljanja vremenskim rizikom. *Poslovna izvrsnost*, Vol. 4, No. 2, pp. 93-105.
19. Lewis, C.D. (1982). *Industrial and business forecasting methods*. Butterworths, London.
20. Manikandan, S. (2011). Measures of central tendency: Median and mode. *Journal of Pharmacology and Pharmacotherapeutics*, Vol. 2, No. 3, pp. 214-215.
21. Ornstein, L. S., Uhlenbeck, G. E. (1930). On the Theory of the Brownian Motion. *Physical review*, Vol. 36, No. 1, pp. 1-19.
22. Orsag, S. (2011). *Vrijednosni papiri: Investicije i instrumenti financiranja*. Revicon, Sarajevo.
23. Samso, J. (2018). *New weather supercomputer to be installed in Bologna*. Available at: https://www.euronews.com/2018/11/14/new-weather-supercomputer-to-be-installed-in-bologna [13 August 2019].
24. Schiller, F., Seidler, G., Wimmer, M. (2012). Temperature Models for Pricing Weather Derivatives. *Quantitative Finance Journal*, Vol. 12, No. 3, pp. 489-500.
25. Till, H. (2015). Why Haven't Weather Derivatives been more Successful as Futures Contracts? A Case Study. *Journal of Governance and Regulation*, Vol. 4, No. 4, pp. 367-371.
26. Vidić, S. (2019). *Zagreb Maksimir Temperature 2000-2018*. E-mail.

About the authors

Berislav Žmuk graduated with the major in accounting, obtained a post-graduate degree in the field of statistical methods for economic analysis and forecasting, and gained his PhD degree in business economics at the Faculty of Economics and Business, University of Zagreb. Currently he is an assistant professor at the Department of Statistics, Faculty of Economics and Business, University of Zagreb where he teaches the following courses: Statistics, Business Statistics and Business Forecasting. In 2013, he successfully completed Sampling Program for Survey Statisticians (SPSS) at the Survey Research Center (SRC) within the Institute for Social Research (ISR), University of Michigan in Ann Arbor, Michigan, USA. In 2015 and 2018, he completed several survey methodology courses (Introduction to Web Surveys, Introduction to Questionnaire Design, Mixed-Mode and Mixed Device Surveys, Applied Multiple Imputation, Web Survey Design, Sampling, Weighting, and Estimation) at Gesis, Leibniz Institute for Social Research in Cologne, Germany. In 2016, he also completed two survey methodology courses

(Introduction to Survey Data Analysis, Advanced Survey Data Analysis and Survey Experiments) at the Essex Summer School in Social Science Data Analysis in Colchester, United Kingdom. His main research fields include applications of statistics in business and economy, survey methodology and statistical quality control. The author can be contacted at *bzmuk@efzg.hr*.

Matej Kovač was born in Zagreb where he completed his high school education in the Fifth Gymnasium. In 2014 he enrolled in the Faculty of Economics, University of Zagreb and completed undergraduate studies in 2018. The same year he enrolled at the Master Study Programme "Analysis and Business Planning", which he finished in 2019 with highest honour "*Summa cum Laude*" and received Dean's award for an outstanding master's thesis. He spent six months in Ljubljana at FELU - Faculty of Economics and Business as a part of the Erasmus exchange programme, where he expanded his knowledge in the field of business economics. He is fluent in English and German, and is an experienced user of Microsoft Office programs and tools. During his college education he developed a preference for financial analysis and statistics, which he further refined by selecting the appropriate graduate programme and assisting the members of the Department of Statistics as a student assistant. From May 2019 he has worked in Addiko Bank. The author can be reached at *kovacmatej95@gmail.com*.