Stochastic Process Workshop

Juan Sebastián Cárdenas-Rodríguez
David Plazas Escudero
Mathematical Engineering, Universidad EAFIT

May 11, 2020

The code for this workshop can be found in source code.

Prediction 1

Question 1. Suppose a variable can be modelled by a homogeneous linear differential equation with parameters $\Theta = \{\mu, \sigma\}$. Simulate a trajectory for a total of $N = 365$ observations (days) and save the simulated trajectory (Series 1).

Question 2. Construct three trajectories out the sample modifying the parameter μ for a total of $N = 365$. The trajectories must consider three scenarios optimistic, pessimistic and constant. Each trajectory must have as initial point the last point of Series 1.

Question 3. Simulate multiple trajectories for Series 1 out of the historic information considering the three possible scenarios. Each trajectory will help to build a prediction band for each scenario. The bands can be constructed as combination of descriptive statistics obtained in the simulation of each trajectory longitudinally. Compare each confidence bands with data out the sample (scenarios). Determine the percentage of effectiveness of the bands. Conclude.

Question 4. Make a sensitivity analysis of the simulated scenarios and their forecast respect to σ. Conclude.

The homogeneous linear stochastic differential equation (SDE) is given by

$$dX_t = \mu X_t dt + \sigma X_t dW_t$$

with $\{W_t\}_{t\geq 0}$ a Weiner process, $\mu \in \mathbb{R}$ and $\sigma \in \mathbb{R}_+$.

The SDE was simulated using the Euler-Maruyama Method [6], with the parameters $\mu = 0.003$, $\sigma = 0.03$, $\Delta t = 1$, $t_f = 365$, $x_0 = 1$. The result of the simulation can be found in Figure 1. This simulation will be known as Series 1.

For the scenario simulation, we modified the parameter μ, in the optimistic case, $\mu = 0.007$ and, in the pessimistic case, $\mu = -0.001$. The scenarios where simulated with the same Δt, σ and $t_0 = 365$, $t_f = 730$, $x_0 = X_{t_0}$. The result of the simulation of these scenarios can be found in 2.

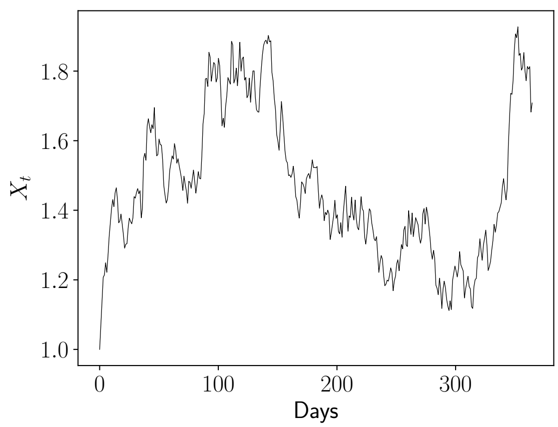


Figure 1: Series 1 simulation.

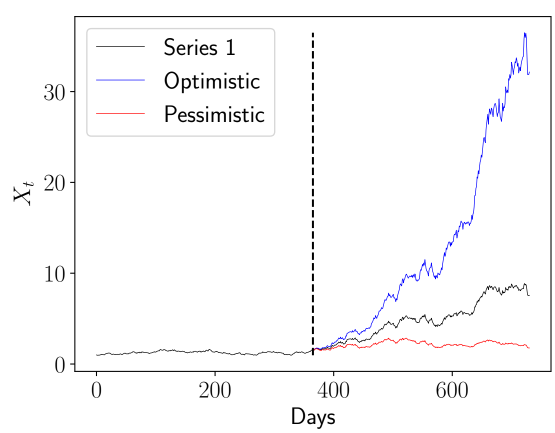
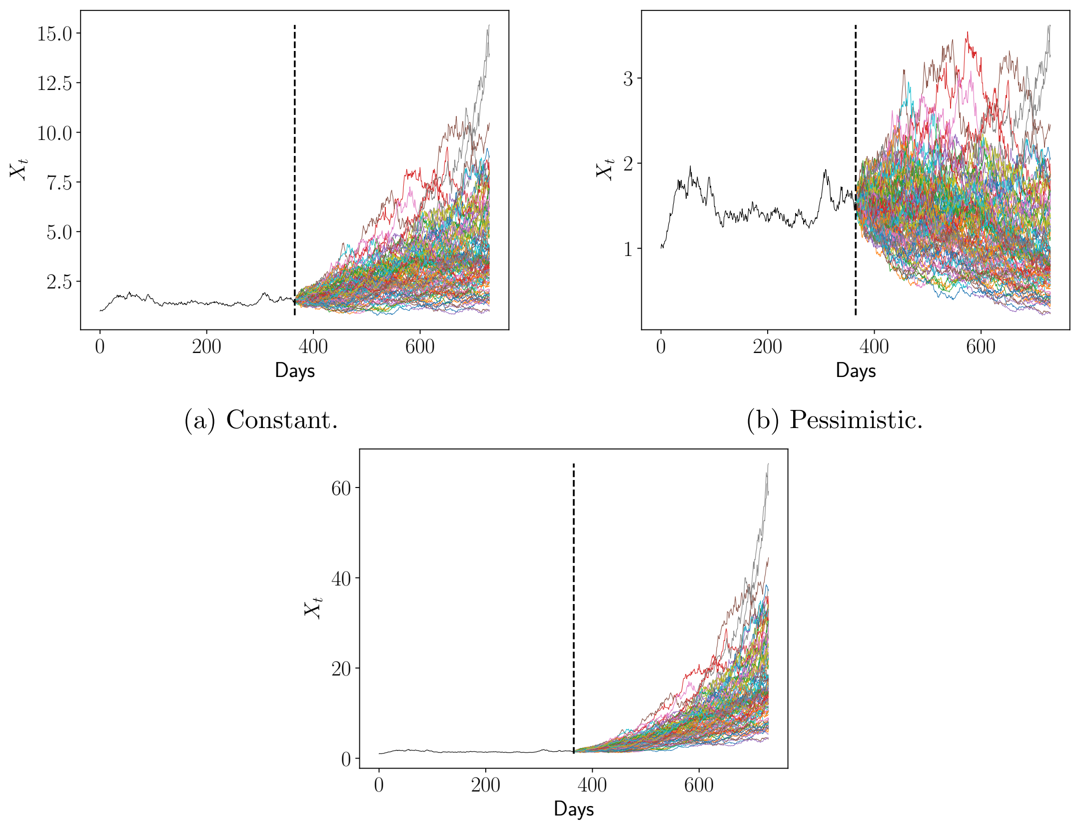


Figure 2: Initial simulation with scenarios trajectories.



(a) Constant. (b) Pessimistic. (c) Optimistic.

Figure 3: Prediction trajectories for each scenario.

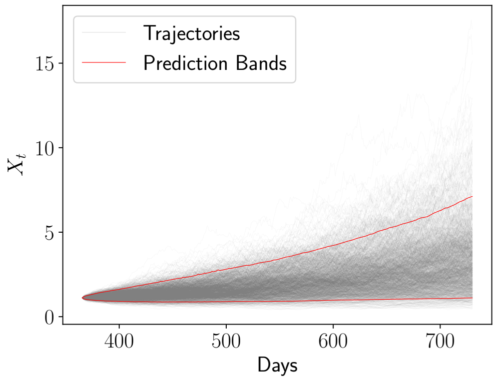
Using the defined parameters, multiple trajectories for each scenario were simulated. An example for 100 trajectories can be found in Figure 3. These figures show Series 1 along with the trajectories of the respective scenario.

The procedure to obtain the prediction bands with a $1-\alpha$ confidence is:

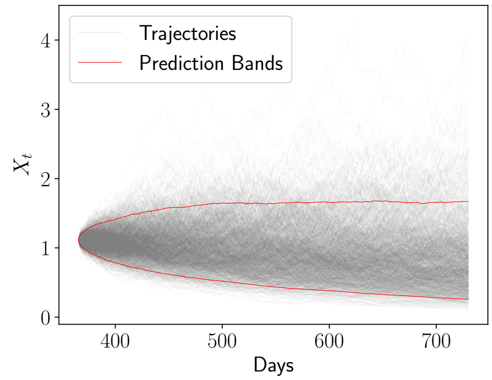
1. Simulate n different trajectories with Euler-Maruyama, using n different Weiner Processes.
2. For each $t \in (t_0, t_f]$ we fit a distribution $\hat{F}_t(\cdot)$ using the `scipy.stats` Python package.
3. For each $\hat{F}_t(\cdot)$ a goodness of fit test is done (Kolmogorov-Smirnov, Anderson-Darling, ...).
4. For each $\hat{F}_t(\cdot)$ calculate the $\alpha/2$ and $1-\alpha/2$ quantiles. These yield a confidence interval at time t.

Therefore, it is important to find the distribution for this process. It is well known that the SDE is a Geometric Brownian Motion and thus, it follows a lognormal distribution. Hence, the parameters are fitted but the goodness of fit test is not performed.

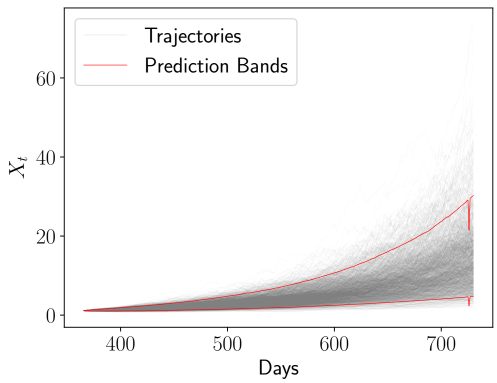
The prediction bands $\{B_t\}_{t\geq 0} = (L_t, U_t)$ for each scenario are presented in Figure 4, with $\alpha = 0.1$ and 1000 trajectories. It is important to notice that Series 1 is no longer shown, only the prediction



(a) Constant.



(b) Pessimistic.



(c) Optimistic.

Figure 4: Prediction bands for each scenario.

with the trajectories of each scenario.

It is clearly seen that the prediction bands enclose the majority of trajectories.

The percentage of effectiveness was calculated by checking how many points longitudinally are inside the bands for each scenario. The histogram of the effectiveness are shown Figure 5.

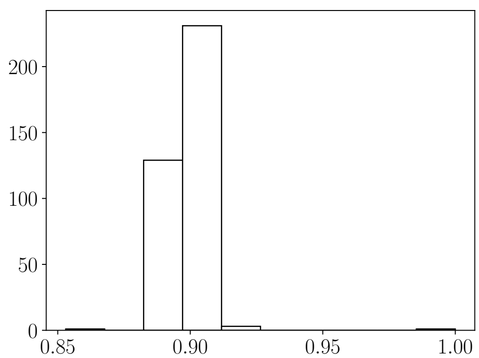
In conclusion, an algorithm to calculate a prediction band was developed and successfully implemented. Furthermore, the prediction band satisfies that approximately $1-\alpha$ percentage of the trajectories are inside the bands.

A sensitivity analysis was done for the parameter σ. The analysis changed the respective parameter in the following manner:

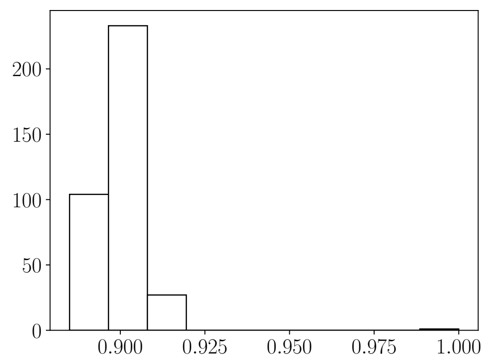
$$\sigma^* = (1+p)\sigma, \quad p \in [-0.5, 0.5]$$

For each value of σ^*, on each scenario, the prediction band was calculated and the last value of the band was saved. In Figure 6, these values for each σ^* are presented.

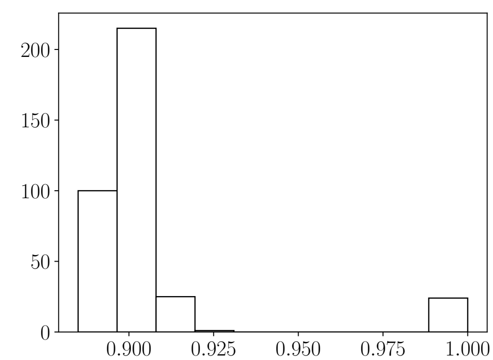
In conclusion, a sensitivity analysis on the σ parameter was successfully realized. As the parameter appears in the equation related to diffusion, it represents the variation of the process. Hence, it can be seen that the bands tend to get wider i.e. with more variance.



(a) Constant.

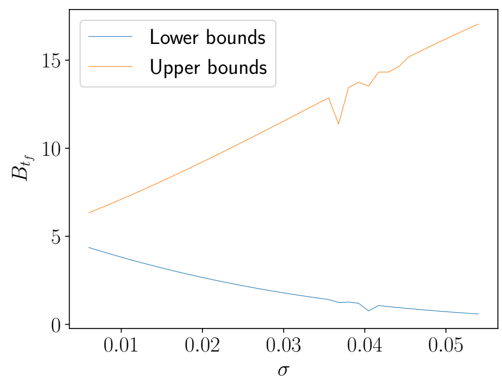


(b) Pessimistic.

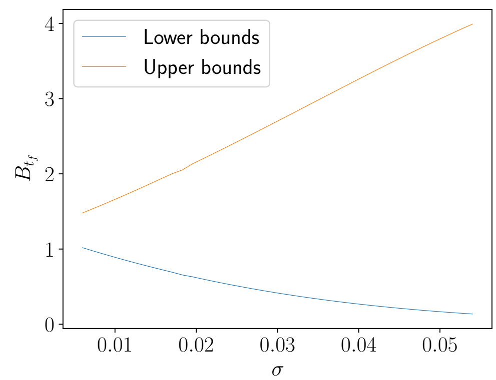


(c) Optimistic.

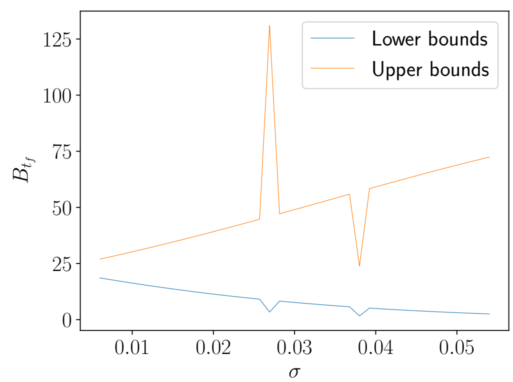
Figure 5: Effectiveness of bands for each scenario.



(a) Constant.



(b) Pessimistic.



(c) Optimistic.

Figure 6: Sensitivity on σ for each scenario.

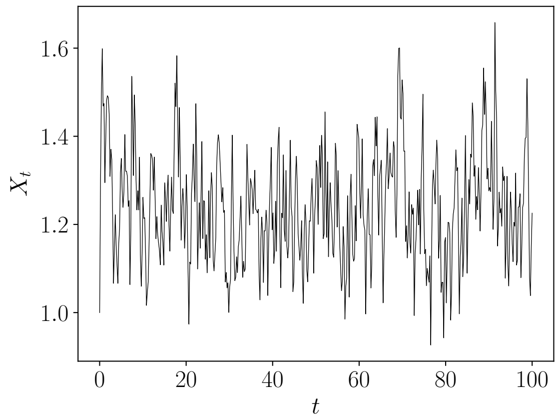


Figure 7: Simulation of mean reversion process.

Prediction 2

Question 5. Suppose that a variable can be modeled by a Ornstein-Uhlenbeck mean reversion process with parameters $\Theta = \{\mu, \sigma, \alpha, \gamma\}$. Simulate a trajectory with $N = 500$ and save this trajectory (Series 2).

Question 6. Analyze the statistical properties of Series 2. Justify the applicability of this equation in different knowledge areas.

Question 7. Based on the methodology from [7], estimate the parameters $\{\alpha, \mu, \sigma\}$ and make an efficient forecast for Series 2. Establish a confidence level for the forecast and conclude.

Question 8. Make a sensitivity analysis and its respective forecast for Series 2 for parameters $\{\alpha, \mu, \sigma\}$ and conclude.

The mean reversion processes of one factor with constant parameters can be written as [7]:

$$dX_t = \alpha(\mu - X_t)dt + \sigma X_t^\gamma dW_t$$

which is also known as the Chan–Karolyi–Longstaff–Sanders process [2], with $t_0 < t < t_f$ and $\gamma, \sigma, \alpha, \mu \in \mathbb{R}_+$.

The simulation was done using the following parameters: $\alpha = 2$, $\mu = 1.25$, $\sigma = 0.4$, $\gamma = 0.5$, $\Delta t = 0.2$, $t_f = 100$, $n = 1000$. The simulation can be found in Figure 7.

The Hurst Exponent of Series 2 was calculated to check its mean reversion property. A Hurst Exponent is mean reverting if the value H is less tan 0.5. The obtained value for this process is $H = 0.093$.

A variance ratio test was also performed. The variance ratio test checks if a process is a random walk. Hence, a rejection of the null hypothesis indicates evidence that the process is mean reverting. The obtained result for this test using different lags is presented in Table 1.

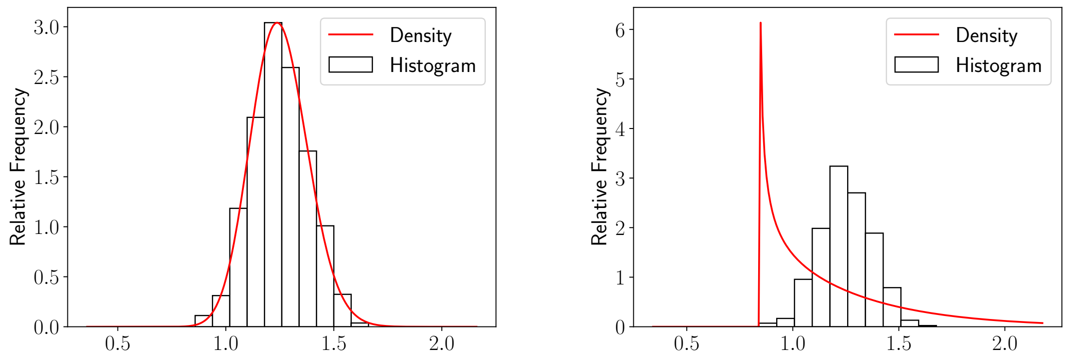


Figure 8: Best and worst non-central χ^2 fitting.

lag	p
2	0.039
4	0
8	0
16	0

Table 1

In this specific case where $\gamma = 0.5$, the process is known as Cox-Ingersoll-Ross (CIR) model [4]. It is well known that this process its imaginary party is 0 if $2\alpha\mu \geq \sigma^2$, i.e.

$$\mathrm{P}\left(\text{there is at least one value of } t > 0 \text{ for which } X_t = 0\right) = 0$$

This is the simplest model that allows for positive interest rates. Furthermore, this model is useful to simulate a derivative and future option [3].

A property of the CIR model is that the distribution for this process, longitudinally, is a non-central χ^2 distribution [5]. Although the process should follow this distribution, in the experiments realized only the 44% of points in time fitted the distribution. In Figure 8 the best and worst fitting can be seen.

This may occur since α is the mean reversion rate, hence for the parameters chosen and for the selected time frame the CIR process behaves as white noise. Hence, by fitting a normal distribution the 99.8% of points in time were fitted. In Figure 9 the best and worst fitting can be seen.

Furthermore, a normal distribution fitting was made transversally. The 99.7% of trajectories were fitted. In Figure 10 the best and worst fitting can be seen.

Let $M = \left\lfloor \frac{T}{\Delta t} \right\rfloor$. Following the ideas from [7], it is first calculated the following terms:

$$A = \sum_{i=1}^{M} \frac{X_i X_{i-1}}{X_{i-1}^{2\gamma}}, \quad B = \sum_{i=1}^{M} \frac{X_{i-1}}{X_{i-1}^{2\gamma}}, \quad C = \sum_{i=1}^{M} \frac{X_i}{X_{i-1}^{2\gamma}}$$

$$D = \sum_{i=1}^{M} \frac{1}{X_{i-1}^{|\gamma}}, \qquad E = \sum_{i=1}^{M} \left(\frac{X_{i-1}}{X_{i-1}^{\gamma}}\right)^2$$

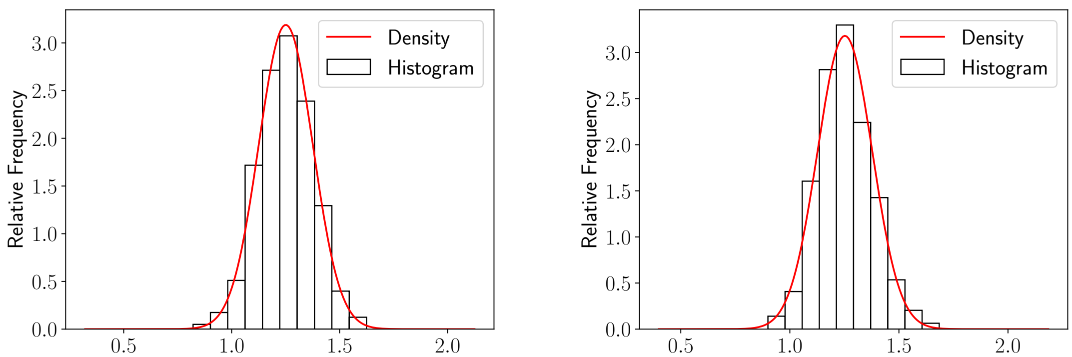


Figure 9: Best and worst normal fitting.

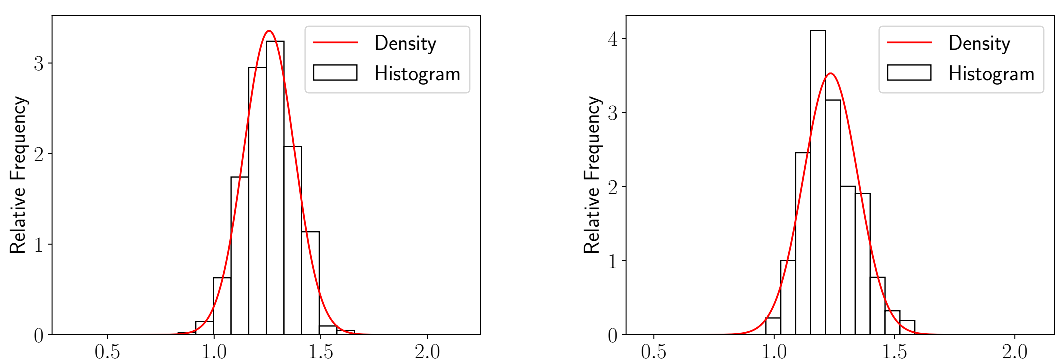


Figure 10: Best and worst normal transversal fitting.

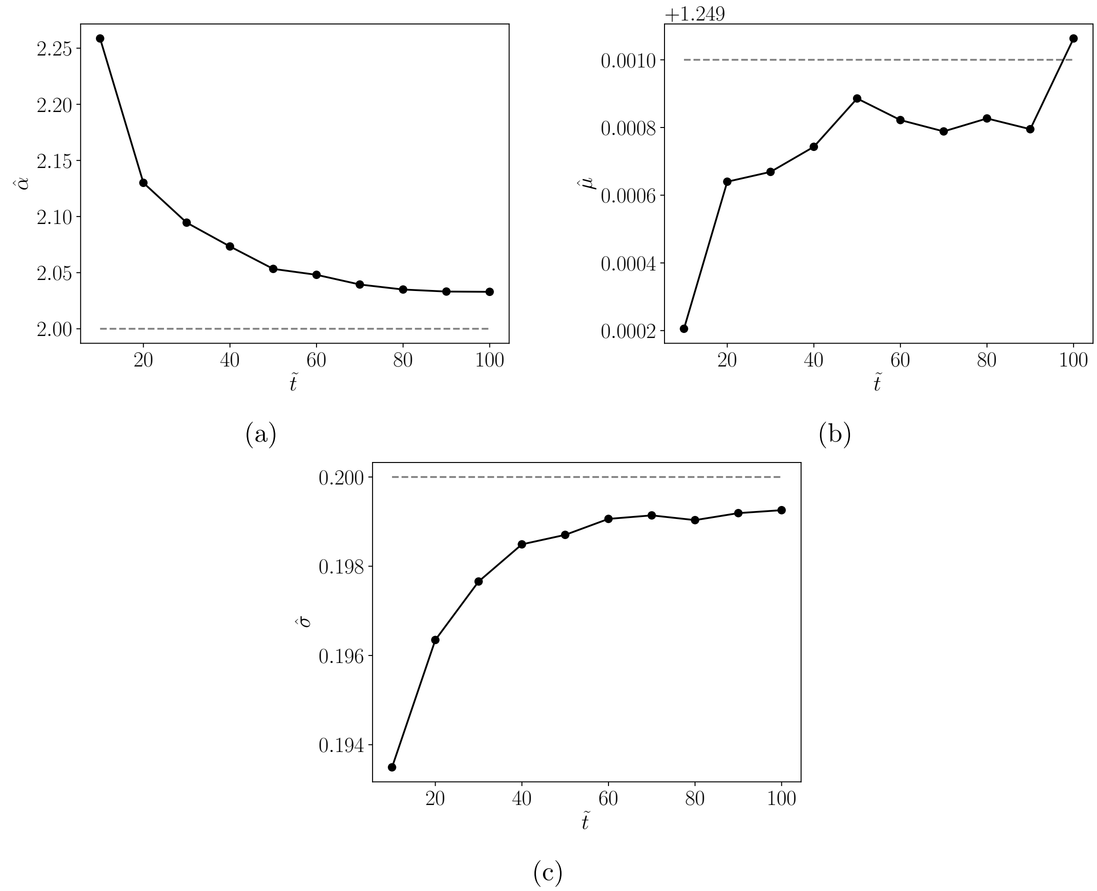


Figure 11: Convergence of parameter estimation.

Then, the estimation is done by:

$$\hat{\alpha} = \frac{ED - B^2 - AD + BC}{(ED - B^2)\Delta t}, \quad \hat{\mu} = \frac{A - E(1 - \hat{\alpha}\Delta t)}{\hat{\alpha} B \Delta t}$$
$$\hat{\sigma} = \sqrt{\frac{1}{M\Delta t}\sum_{i=1}^{M}\left(\frac{X_i - X_{i-1} - \hat{\alpha}(\hat{\mu} - X_{i-1})\Delta t}{X_{i-1}^{\gamma}}\right)^2}$$

To see the behavior of this estimation, a parameter estimation was done for each trajectory and the number of observation was modified. The parameter estimation was done in different time frames of the form $[0, \tilde{t}]$. In Figure 11 the average of the estimated parameters for each trajectory is seen.

Following the same procedure as the one in the previous section, prediction bands where constructed using the fitted distributed longitudinally. In Figure 12 the predictions bands are displayed.

Finally, a sensitivity analysis was made for parameters α, μ, σ using the same methodology already presented in prediction 1. In Figure 13, the sensitivity analysis for each parameter is seen.

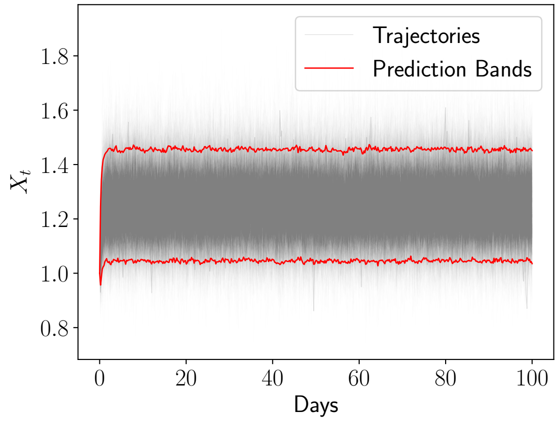
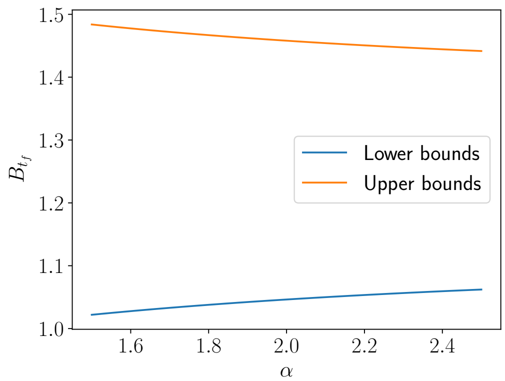
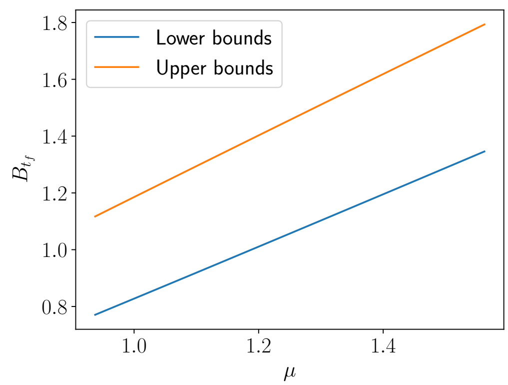


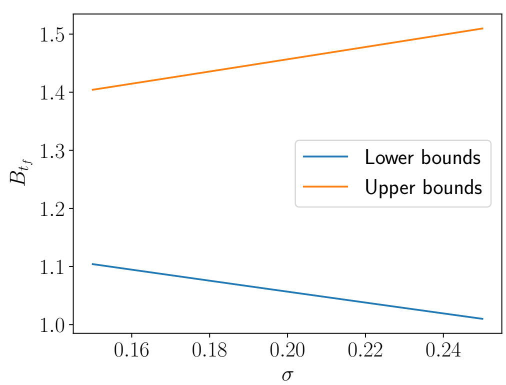
Figure 12: Prediction bands for CIR model.



(a)



(b)



(c)

Figure 13: Sensitivity analysis for parameters.

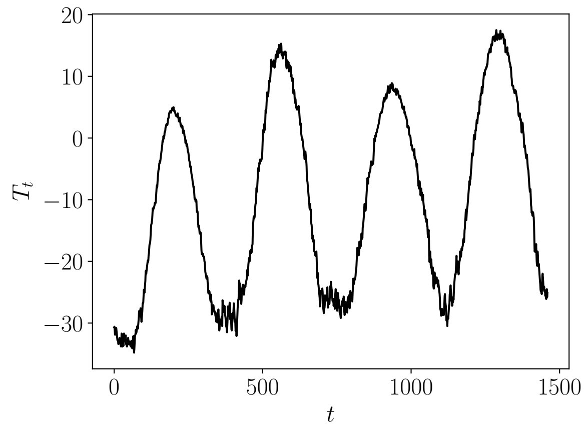


Figure 14: Daily temperatures.

Prediction 3

Question 9. Download the data of the daily temperatures of a region of the northern hemisphere during at least two years. In such series a trigonometric periodic functional trend must be visible.

Question 10. Analyze the statistical properties of the data and make a brief description of the region being analyzed.

Question 11. Following the methodology from [1] estimate the parameters for a mean reversion process.

Question 12. Make an efficient forecast for a time equal to the estimation period and conclude.

The data considered is given by Henry Laniado, for another subject. This data considered the average daily temperature in Canada for the last 35 years. It was only extracted the 4 more recent years. The data can be seen in Figure 14.

In first place, the same methodology to test mean reversion used in last section was applied with this series. The results are presented in Table 2.

lag	p
2	0
4	0
8	0
16	0

Table 2: Result of Variance Ratio Test.

It is clear that the ratio variance test shows that the process is mean reverting.

Following the ideas from [1], it was desired to see the behavior of the differences of adjacent daily temperatures which can be interpreted as the Driving Noise Process. In Figure 15 it can be found the time series for the difference.

The time series appears to be white noise. Hence, a normal distribution was fitted to the time series

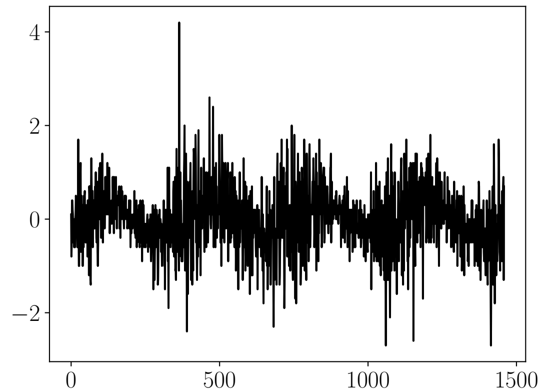


Figure 15: Time series for the Driving Noise Process.

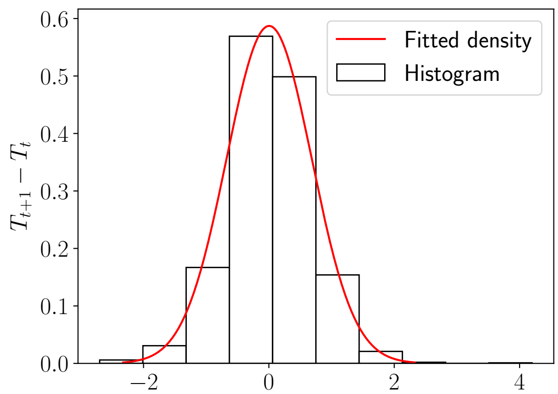


Figure 16: Fitted distrbution and histogram for Driving Noise Process.

to verify this. In Figure 16 it can be seen the fitted distribution and the histogram of the time series.

Based on [1], the daily temperatures trend was fitted to a process of the form:

$$Y_t = a_1 + a_2 t + a_3 \sin(\omega t) + a_4 \cos(\omega t)$$

The parameters where found using the least-squared optimization procedure. It was obtained:

$$\vec{a} = (-16.68, 9.01 \times 10^{-3}, 7.20, -18.87)$$

In Figure 17 the adjusted trend and time series can be seen.

The time series is desired to be fitted to a Ornstein-Uhlenbeck Process, given by:

$$dT_t = \alpha(T_t^m - T_t)dt + \sigma_t dW_t$$

with T_t^m being the adjusted trend. The procedure to estimate the parameters is:

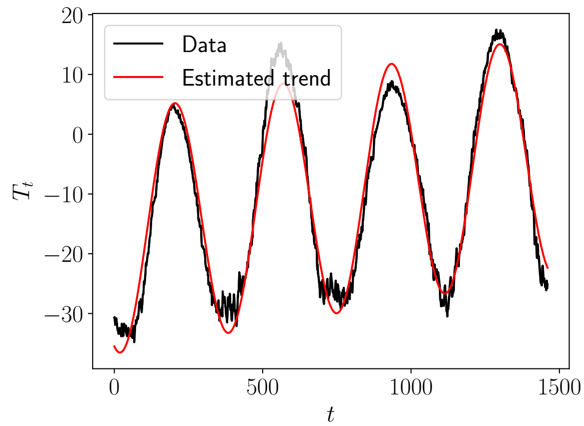


Figure 17: Trend and daily temperatures.

1. Let N_μ be the days in a month μ and T_j be the daily temperatures at day j. A initial estimation for σ in each month is calculated by:

$$\hat{\sigma}_\mu^2 = \frac{1}{N_\mu} \sum_{j=0}^{N_\mu - 1} (T_{j+1} - T_j)^2$$

2. Then, an estimation for $\hat{\alpha}$ is done by the following equation:

$$\hat{\alpha} = \frac{\sum_{i=1}^{n} Y_{i-1}(T_i - T_i^m)}{\sum_{i=1}^{n} Y_{i-1}(T_{i-1} - T_{i-1}^m)}, \quad \text{where } Y_i = \frac{T_i^m - T_i}{\sigma_i^2}$$

In Figure 18 the mean of 1000 simulations of the mean reverting process and the daily temperatures can be seen.

Using the process with the estimated parameters, it is possible to make a forecast of another 4 years. Furthermore, as the information of that 4 years is known a comparison can be made. In Figure 19 the prediction done by the SDE and the real data is shown.

It is important to notice that the estimated process is a very close approximation to the real data. In the second year forward the approximation differs from the real data. This occurs because the data has an unusual behavior, therefore if the real data continued its tendency the approximation would've been much better.

It is seen that the data does not follow the linear tendency when the predictions differs. In this manner, the authors suggest adding a term to the tendency equation, for example adding a higher degree polynomial.

In Figure 20 the prediction bands for the SDE and the process with 1000 trajectories is presented.

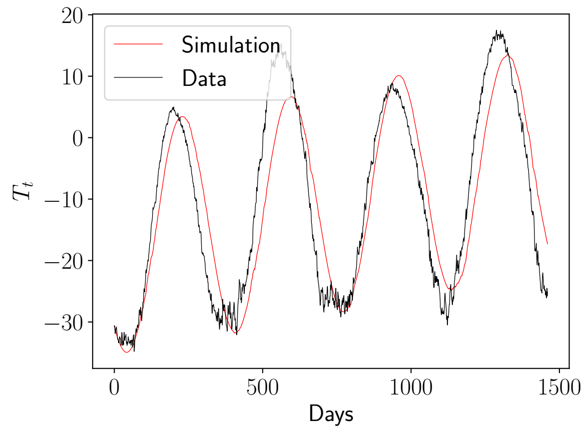


Figure 18: Simulation and temperatures.

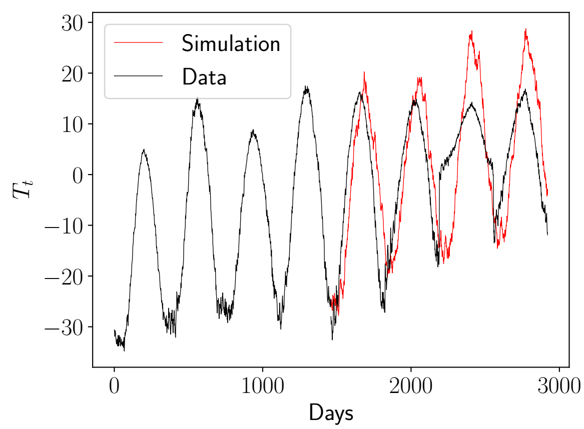


Figure 19: Mean reverting process and real data.

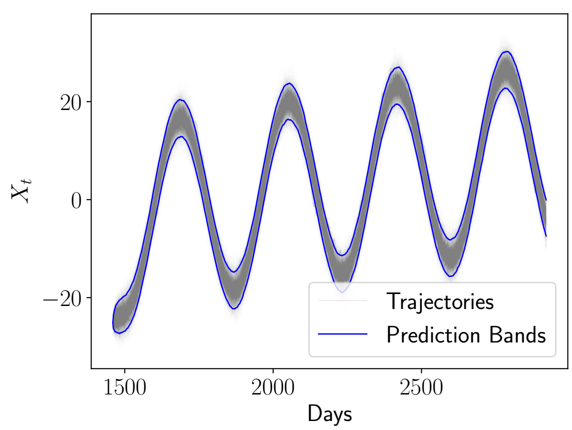


Figure 20: Prediction bands for temperatures prediction.

References

[1] Peter Alaton, Boualem Djehiche, and David Stillberger. "On Modelling and Pricing Weather Derivatives". In: *Applied mathematical finance* 9.1 (2002), pp. 1–20.

[2] Kalok C Chan et al. "An Empirical Comparison of Alternative Models of the Short-Term Interest Rate". In: *The journal of finance* 47.3 (1992), pp. 1209–1227.

[3] "Chapter 31: Cox-Ingersoll-Ross Model". In: (2017). URL: http://janroman.dhis.org/finance/IR/CIR%5C%20advanced.pdf.

[4] John C. Cox, Jonathan E. Ingersoll, and Stephen A. Ross. "A Theory of the Term Structure of Interest Rates". In: *Econometrica* 53.2 (1985), pp. 385–407. ISSN: 00129682, 14680262. URL: http://www.jstor.org/stable/1911242.

[5] Sigurd Dyrting. "Evaluating the Noncentral Chi-Square Distribution for the Cox-Ingersoll-Ross Process". In: *Computational economics* 24.1 (2004), pp. 35–50.

[6] Desmond J Higham. "An Algorithmic Introduction to Numerical Simulation of Stochastic Differential Equations". In: *SIAM review* 43.3 (2001), pp. 525–546.

[7] Freddy H Marín Sánchez and J Sebastian Palacio. "Gaussian Estimation of One-factor Mean Reversion Processes". In: *Journal of Probability and Statistics* 2013 (2013).