Stochastic Processes Workshop

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

May 12, 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 [**higham2001**] in `Python`, with parameters $\mu = 0.003$, $\sigma = 0.03$, $\Delta t = 1$, $t_f = 365$, $x_0 = 1$. It is important to mention that the implemented algorithm for Euler-Maruyama runs in different CPU cores simultaneously for speed execution and, additionally, the creation of Weiner processes is run in parallel as well. This fact implies that, regardless of a fixed seed for random number generation, the algorithm will return different trajectories on each run, since each core is accessing the random number generator simultaneously and not sequentially; therefore, the results here presented are those of a single run.

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.

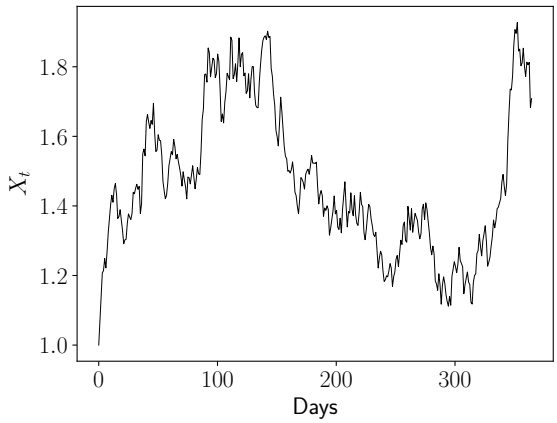


Figure 1: Series 1 simulation.

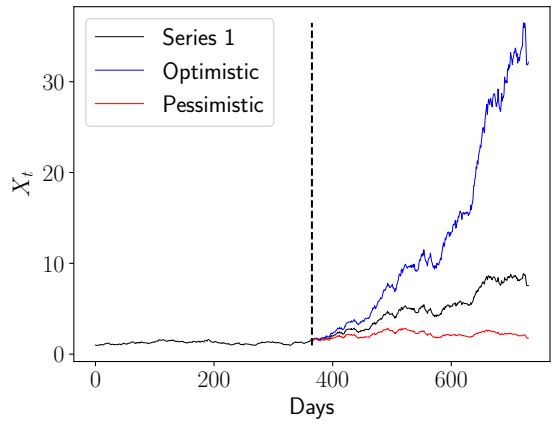
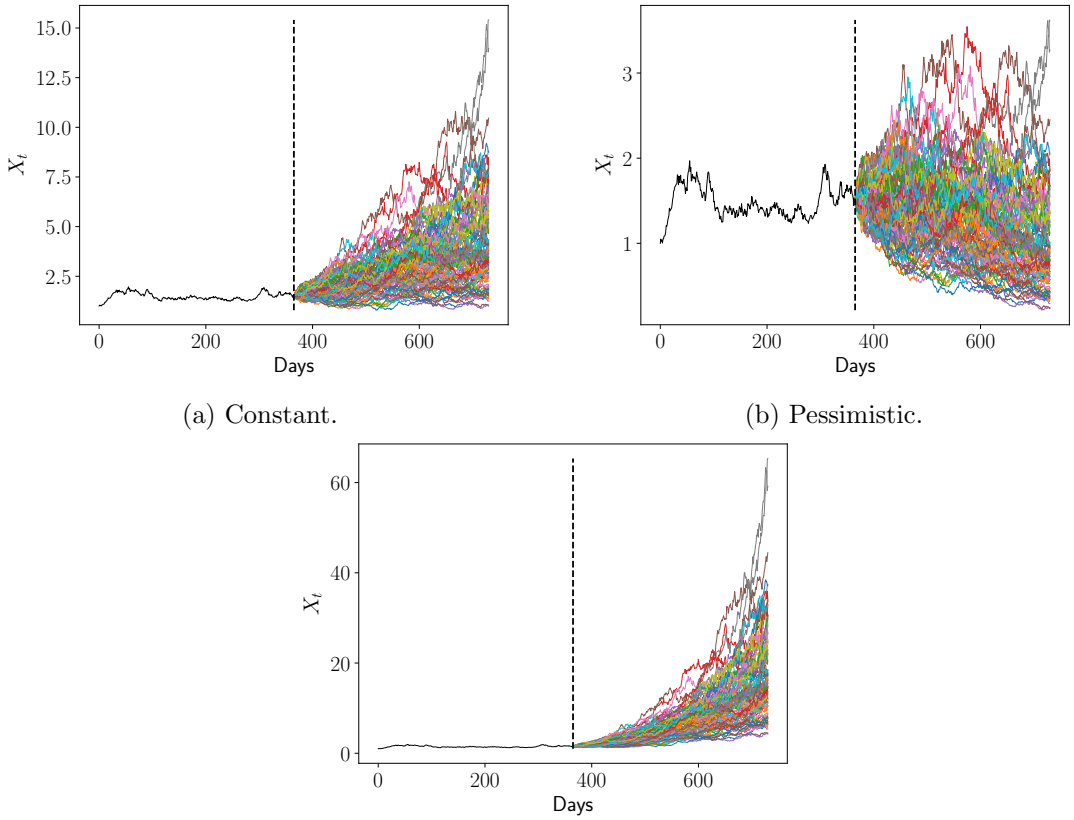


Figure 2: Initial simulation with scenarios trajectories.

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, ...).



(a) Constant.

(b) Pessimistic.

(c) Optimistic.

Figure 3: Prediction trajectories for each scenario.

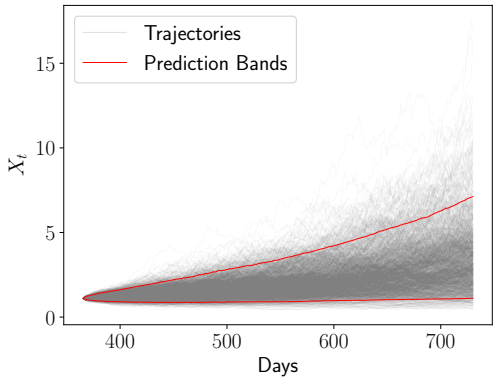
4. For each $\hat{F}_t(\cdot)$ calculate the $\alpha/2$ and $1-\alpha/2$ quantiles. These yield a confidence interval at time t.

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.

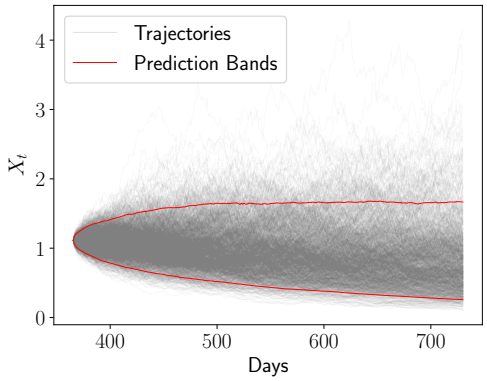
The prediction bands $\{B_t\}_{t\geq 0} = (L_t, U_t)$ for each scenario are presented in Figure 4, with $\alpha = 0.1$ and 1000 trajectories. It is important to notice that Series 1 is no longer shown, only the prediction with the trajectories of each scenario.

It is clearly seen that the prediction bands enclose the majority of trajectories; furthermore, it can be observed that these bands are not completely smooth, for example in Figure 4 (c) the bands have a spike near the end of the simulation time. These spikes are due to the presence of atypical points in that particular time; it is clear that these annomalies would be corrected when the number of trajectories considered is increased. 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.

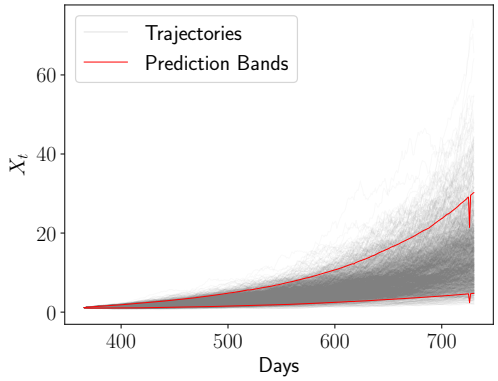
In conclusion, an algorithm to calculate a prediction band was developed and successfully imple-



(a) Constant.



(b) Pessimistic.



(c) Optimistic.

Figure 4: Prediction bands for each scenario.

mented. Furthermore, this band satisfies that approximately $1-\alpha$ percentage of the trajectories are inside the bands, according to the results obtained in 5 where the mode of enclosement is around $1-\alpha$.

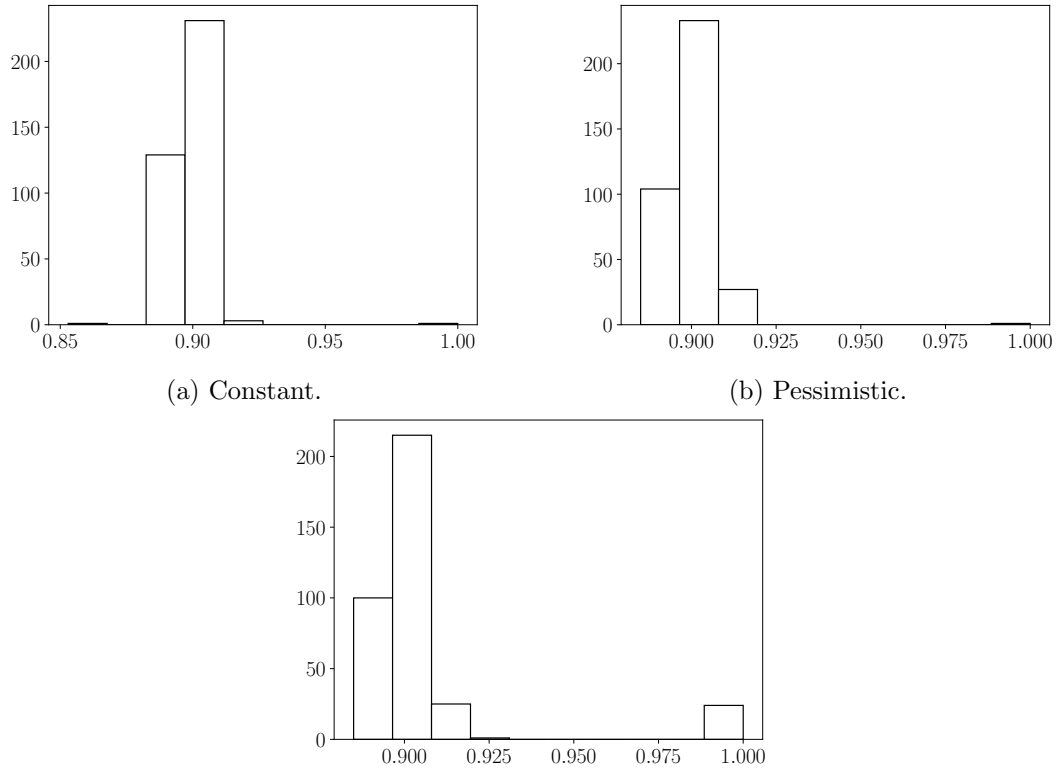
A sensitivity analysis was done for the parameter σ. The analysis changed the respective parameter in the following manner:

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

For each value of σ^*, on each scenario, the prediction band was calculated and the last value of the band was saved in order to summarize the behavior of the bands when the parameter changes. In Figure 6, these values for each σ^* are presented.

Note that the curves for the sensitivity analysis have some annomalies, this can be explained for the same reason why the prediction bands have spikes; the data has unusual behavior and, again, this would be corrected if the number of trajectories simulated is increased. Yet, the trend of these curves can still be appreciated and it is clear that the bands get wider as σ increases, due to the direct relation between the volatility and the variance of the process.

In conclusion, a sensitivity analysis on the σ parameter was successfully constructed and it showed the expected result: as the parameter appears in the part related to the diffusion, augmenting it



(a) Constant.

(b) Pessimistic.

(c) Optimistic.

Figure 5: Effectiveness of bands for each scenario.

implies wider bands.

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).

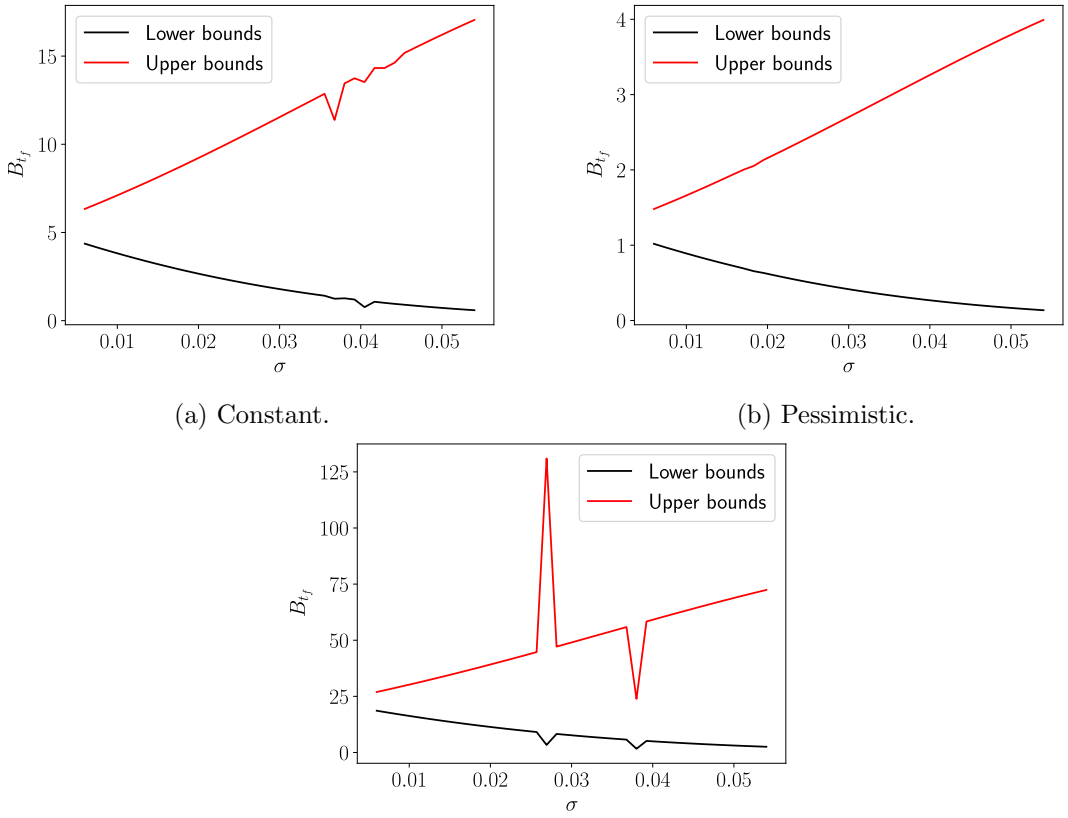
Question 6. Analyze the statistical properties of Series 2. Justify the applicability of this equation in different knowledge areas.

Question 7. Based on the methodology from [**marin2013**], 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 [**marin2013**]:

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



(a) Constant.

(b) Pessimistic.

(c) Optimistic.

Figure 6: Sensitivity on σ for each scenario.

which is also known as the Chan–Karolyi–Longstaff–Sanders process [**chan1992**], with $t_0 < t < t_f$ and $\gamma, \sigma, \alpha, \mu \in \mathbb{R}_+$.

The simulation was done using the already described Euler-Maruyama scheme with 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.

As for the statistical properties of Series 2, the Hurst Exponent of Series 2 was calculated to check its mean reversion property, following the ideas (and code) from [**quanstart**]. A time series is mean reverting if the value of the Hurst Exponent H is less than 0.5. the obtained value for this process is $H = 0.093$.

A variance ratio test was also performed, this is a well know test for random walks [**lo1989size**] and it has been widely used to test for mean reversion (see e.g. [**lo1988stock**, **risager1998random**, **lam2006new**]). The variance ratio test checks if a process is a random walk, using the quotient of a k-period return and the return for 1 period [**charles2009variance**]. Hence, a rejection of the null hypothesis gives evidence that the process is mean reverting. The obtained result for this test using different lags is presented in Table 1.

In this specific case where $\gamma = 0.5$, the process is known as Cox-Ingersoll-Ross (CIR) model [**cox1985**]. An important property of the CIR model is that it is, almost surely, strictly posi-

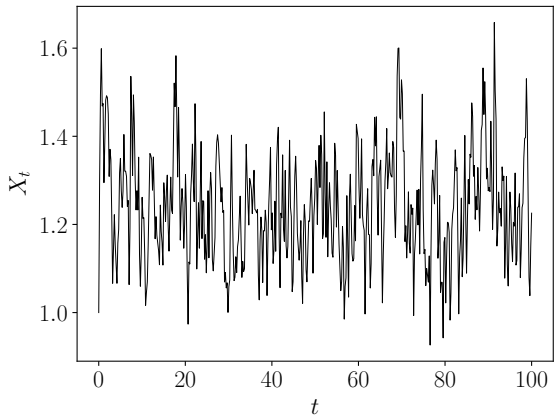


Figure 7: Simulation of mean reversion process.

lag	p
2	0.039
4	0
8	0
16	0

Table 1

tive if $2\alpha\mu \geq \sigma^2$ [**cox1985**, **unknown2017**], i.e.

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

For further extension of the CIR model see e.g. [**overbeck1997estimation**, **mishra2010study**, **li2015asymptotic**, **medvedev2019cox**] This is the simplest model that allows for positive interest rates. Furthermore, this model is useful to simulate a derivative and future option [**unknown2017**].

Another property of the CIR model is that the distribution of this process, longitudinally, is a non-central χ^2 distribution [**dyrting2004**]. Although the process should follow this distribution, in the experiments executed, only the 44% of points in time fitted this distribution according to a simple Kolmogorov-Smirnov test. In Figure 8 the best and worst (in a p-value sense) are presented.

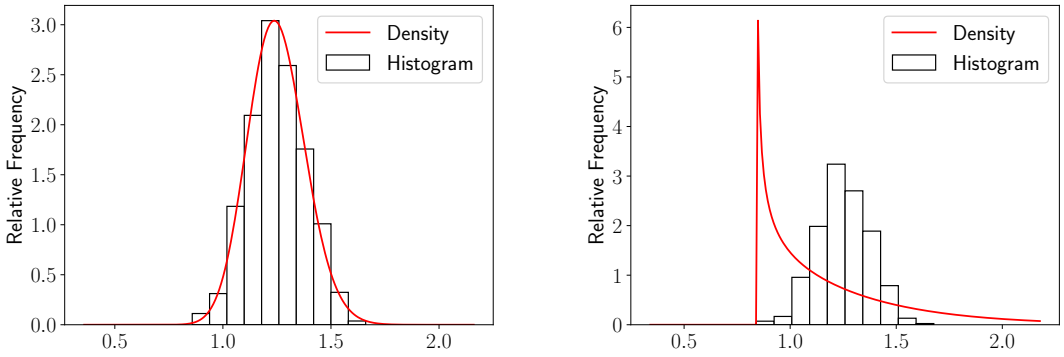


Figure 8: Best and worst non-central χ^2 fitting.

This may occur since α, that is the mean reversion rate, is relatively high, hence for the parameters chosen and for the selected time frame the CIR process behaves similar to a 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.

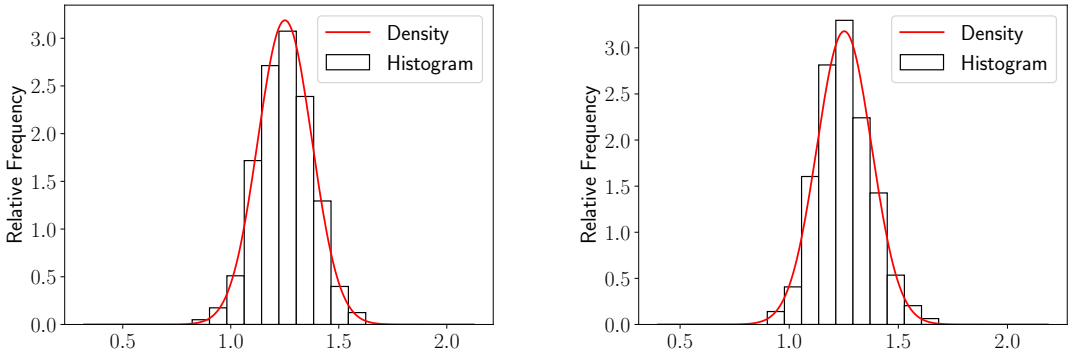


Figure 9: Best and worst normal fitting.

Furthermore, a normal distribution fitting was made transversally, which is a string evidence of a mean-reversion-like process. The 99.7% of trajectories were fitted. In Figure 10 the best and worst fitting can be seen.

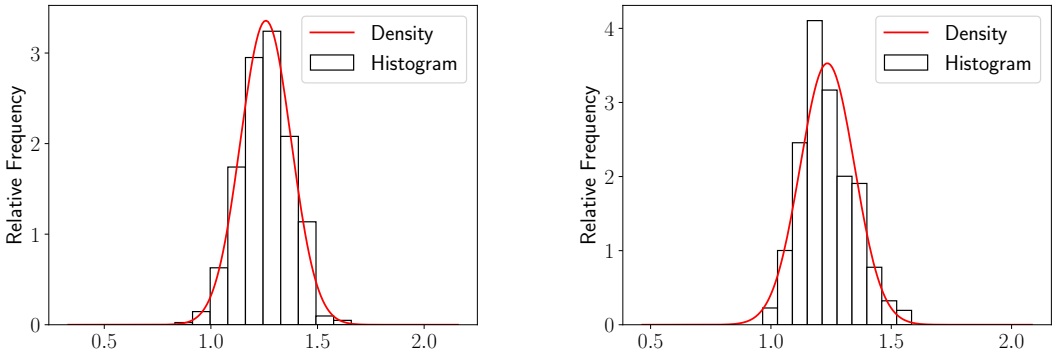


Figure 10: Best and worst normal transversal fitting.

The parameter estimation is done based on [**marin2013**]. Let $M = \lfloor \frac{T}{\Delta t} \rfloor$. The following terms are first calculated from a sample time series:

$$A = \sum_{i=1}^{M} \frac{X_i X_{i-1}}{X_{i-1}^{2\gamma}}, \quad B = \sum_{i=1}^{M} \frac{X_{i-1}}{X_{i-1}^{2\gamma}}, \quad C = \sum_{i=1}^{M} \frac{X_i}{X_{i-1}^{2\gamma}}$$
$$D = \sum_{i=1}^{M} \frac{1}{X_{i-1}^{2\gamma}}, \qquad E = \sum_{i=1}^{M} \left(\frac{X_{i-1}}{X_{i-1}^{\gamma}} \right)^2$$

Then, the estimation is given by:

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

In order to see the behavior of this estimation, it was done for each trajectory and the number of observations 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.

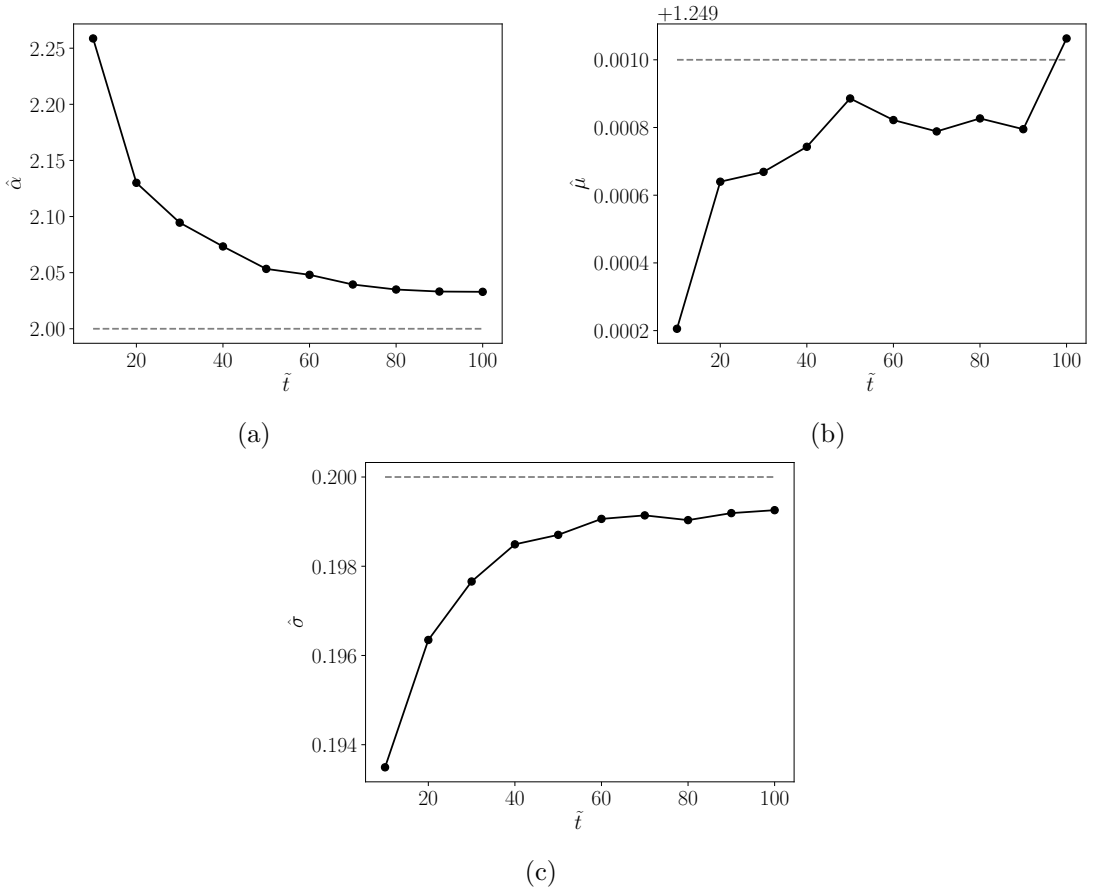


Figure 11: Convergence of parameter estimation.

It can be clearly observed that the parameters tend to converge on the true values as $\tilde{t}$ increases, and consequently, the sample size increases. Following the same procedure as the one in the previous section, prediction bands were constructed using the fitted distributions longitudinally. In Figure 12 the predictions bands are displayed.

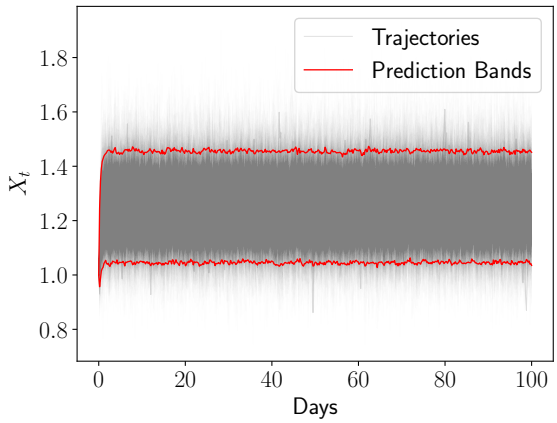
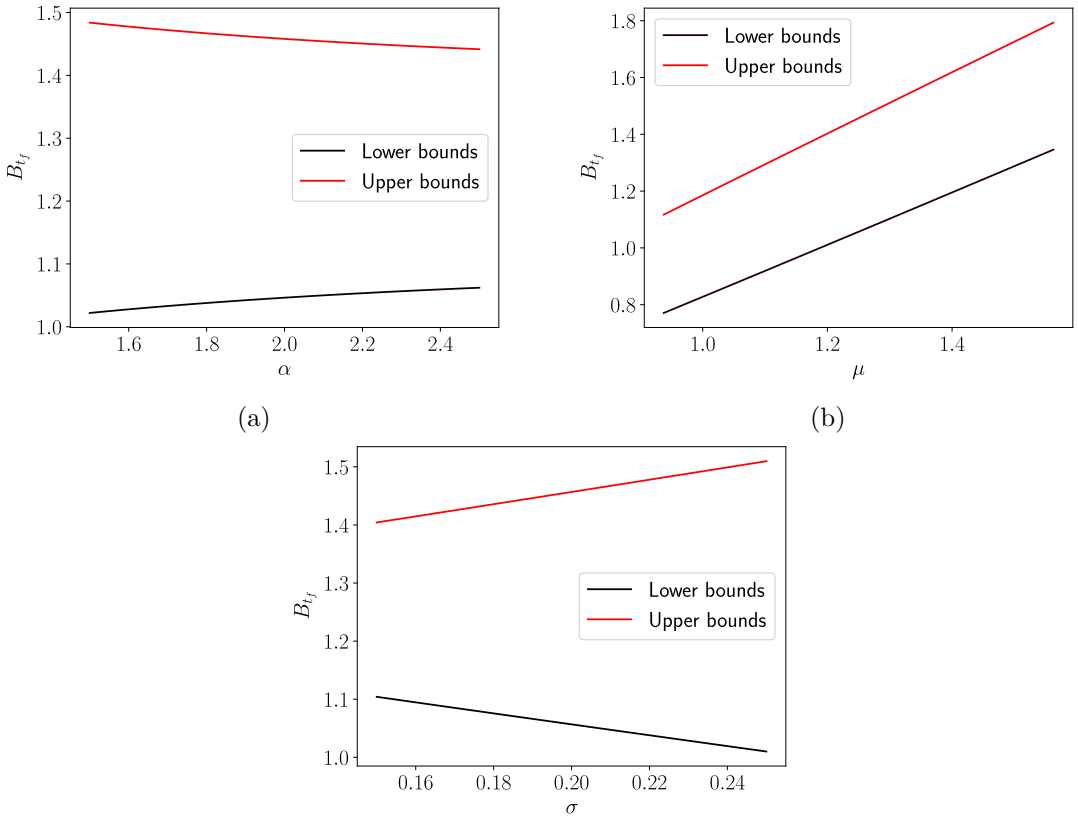


Figure 12: Prediction bands for CIR model.

Finally, a sensitivity analysis was made for parameters α, μ, σ using the same methodology already presented in prediction 1, but using $p \in [-0.25, 0.25]$. In Figure 13, the sensitivity analysis for each parameter is presented.



(a) (b)

(c)

Figure 13: Sensitivity analysis for parameters.

From Figure 13 (a), it can be concluded that the prediction bands tend to get thiner as the parameter α increases; recall that α represents the mean reversion rate, therefore, if it is increased, the process will "recover" faster implying that the trajectories will accumulate closer to the mean and, finally, the bands capture this behavior. Additionally, it is clear from Figure 13 (b) that μ is closely related to the mean that the process reverts to, since an almost linear relationship between μ and the prediction bands was obtained; the higher the μ, the higher the bands, with constant amplitude. Finally, Figure 13 (c) shows that the bands get wider as σ increases, and this can be justified with a similar argument than the one in the sensitivity analysis of prediction 1: as σ appears in the diffusion component, it has a proportional relation with the variance of the process.

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 [**alaton2002**] 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 considers the average daily temperature in Canada for the last 35 years. The four-most early years were extracted. The data can be seen in Figure 14.

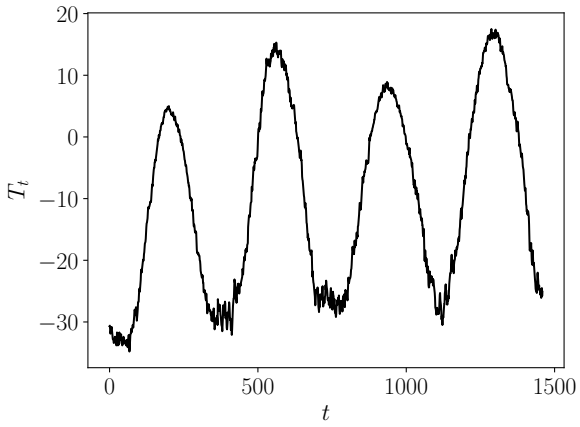


Figure 14: Daily temperatures.

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	$p - value$
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, as the null hypothesis of random-walk is rejected.

Following the ideas from [**alaton2002**], 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.

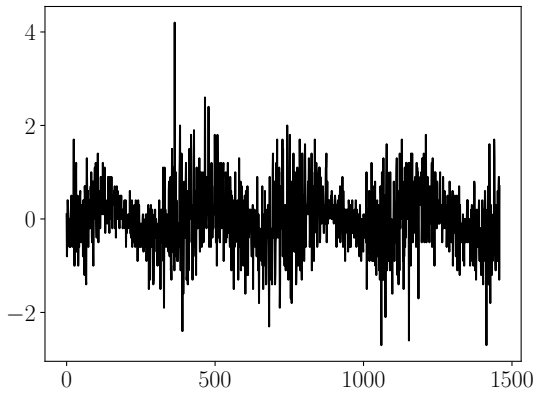


Figure 15: Time series for the Driving Noise Process.

The time series appears to be white noise with a small tendency. Hence, a normal distribution was fitted to the time series to verify this. In Figure 16 it can be seen the fitted distribution and the histogram of the time series.

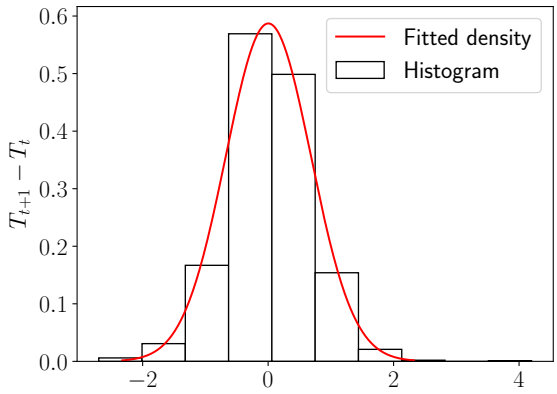


Figure 16: Fitted distrbution and histogram for Driving Noise Process.

Based on [**alaton2002**], the daily temperatures trend was fitted to a process of the form:

$$T_t^m = a_1 + a_2 t + a_3 \sin(\omega t) + a_4 \cos(\omega t)$$

The parameters were found using a least-squares optimization procedure using `SciPy`. 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.

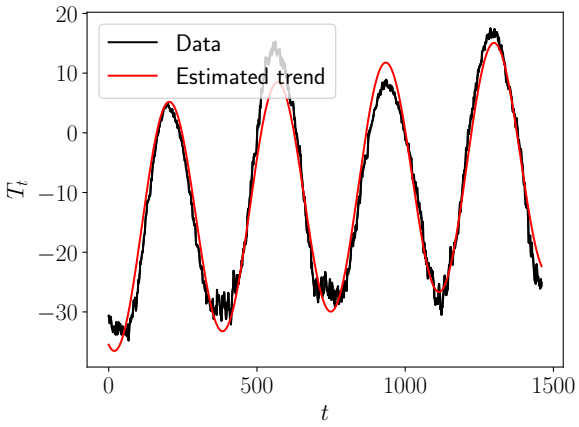


Figure 17: Trend and daily temperatures.

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:

1. Let N_μ be the days in a month μ and T_j be the temperatures in 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.

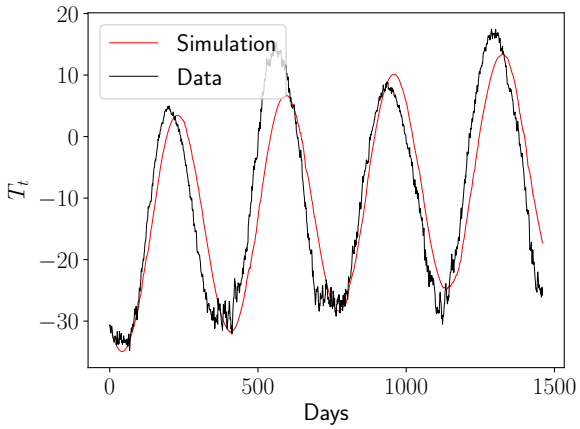


Figure 18: Simulation and temperatures.

Using the process with the estimated parameters, it is possible to make a forecast of another four years. Furthermore, a comparison can be made taking the real data of those corresponding four years and the forecast. In Figure 19 the prediction done by the SDE and the real data are shown.

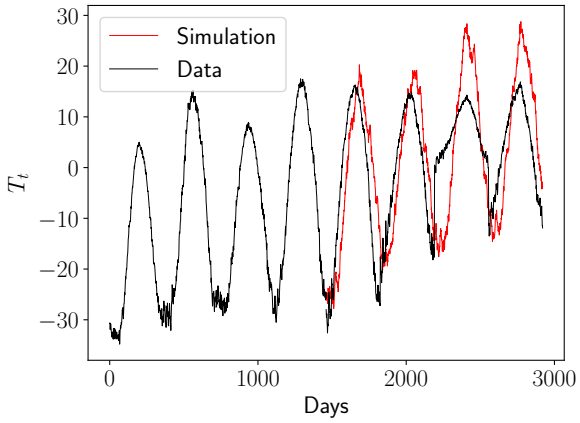


Figure 19: Mean reverting process and real data.

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 have 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.

Finally, the respective prediction bands can be constructed, using the fitted distribution of the

driving noise. Once more, 1000 trajectories were simulated and the obtained bands (using the same procedure already described) are presented in Figure 20.

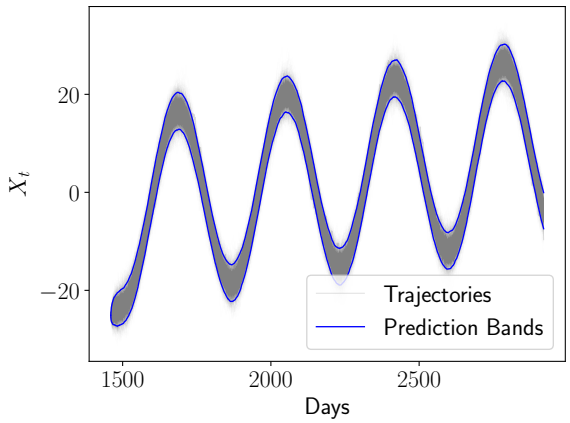


Figure 20: Prediction bands for temperatures prediction.