POLITECNICO MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE E DELL'INFORMAZIONE

Rating momentum in credit migrations

Project 5A

Financial Engineering 2022-23

Authors: **Algeri Matteo**

Cojutti Francesco

Fouad Davide

Abstract

This study focuses on estimating rating transition probabilities within Markov and non-Markov frameworks using real-world data from the Moody's Analytics Default & Recovery Database. In particular, the aim is to capture and assess the phenomenon of 'downward momentum', i.e the increased probability of a company experiencing additional rating downgrades following a previous downgrade, indicating a potential deterioration in its credit position.

Keywords: Rating momentum; Hawkes Process; Continuous Time Markov Chain.

Contents

Introduction

Credit risk modeling has gained significant attention since the 2008 financial crisis, as banks recognize the importance of effectively managing credit risk to maintain risk-adjusted returns, ensure regulatory compliance, and safeguard long-term success. From this perspective and according to Bladt-Sorensen (2009) [3], we study the impact of rating momentum in estimating rating transition probabilities employing two continuous-time models. In order to conduct this study, we take advantage of a very large portion of the Moody's Analytics Default & Recovery Database (DRD), so we handle real data and consequently deduct real-world transition probabilities.
After a brief introduction of the dataset's characteristics and the dataset cleaning techniques, we present and employ the Markovian model, a continuous time Markov chain. In that section, we report both the calibration of the model, i.e. how to obtain the transition intensities matrix, with its estimation, and the confidence intervals of these parameters, which are obtained via a parametric bootstrap following a specific simulation algorithm.
Furthermore, attempting to assess a non-Markovian effect while preserving some Markovian structure, we rely on point processes, in particular on a self-exciting process, the Multidimensional Hawkes process, which leads to capturing the rating momentum by adding only two new parameters. Also in this case we calibrate the model's parameters and evaluate their confidence regions.
Finally, in order to further investigate the rating momentum's implications we repeat the analysis for different scenarios, varying the starting rating of the momentum effect and assuming the default as a pure absorbing state.

1 | Dataset

The dataset employed is the Default & Recovery Database (DRD) detained by Moody's Investors System (MIS), which is a continuous record of ratings and defaults of 20701 companies from all over the world.
Each observation, i.e. the period of permanence in a specific rating, involves the following information:

- numerical id of the corporate
- initial date of the observation period
- final date of the observation period
- change in rating from the corporate previous observation
- rating during the observation period

The rating system used by MIS contemplates a range of 21 ratings and two more states: Withdrawal Rating (WR) and Default (D). With WR we refer to when MIS no longer rates an entity, debt or financial obligation, debt issuance program, preferred share, or other financial instrument for which it is previously assigned a rating, whereas D represents mostly the failure to make scheduled principal or interest payments.
However, the dataset presents some criticalities that do not allow the study to be developed directly, therefore we create a Matlab function `cleanDataset.m` where we apply the dataset cleaning guidelines in Baviera (2023) [2].
First of all, we take only observations completely contained between 1st January 1983 and 31st January 2020. The latter is also the date with which we replace all the `NaN`s present in the dataset. To increase reproducibility, we first simplify the rating system by reducing the number of ratings to 7, as reported in Table (1.1). Then, we further aggregate the ratings into a new rating system by grouping notches together based on predefined intervals and we merge consecutive observations that have the same new rating.
Finally, we can focus on the main issue, i.e. dealing with WR and D. The latter is considered in their soft version, i.e. at each D the company ceases to exist and a new one is created, which has the start rating of the next observation and continues its existence until

a new default happens or the final date of our analysis is reached. Then, before applying the conditions listed in Baviera [2], we delete all D's and WR's as first observations and also WR's as last observations of an issuer.
We complete the cleaning of the dataset by drawing a portion of 47403 observations relative to 23522 corporations and 2760 D.
As we will see in Section 4, we conducted the rating momentum analysis where D is considered in its strong version, i.e. a pure state of absorption from which neither new companies are born nor subsequent observations are considered. Applying this one modification, we obtain a dataset consisting of 44806 observations of 21915 firms, of which 2358 are D.

Old rating	**Old numerical rating**	**New numerical rating**
Aaa	1	1
Aa $(1,2,3)$	$2,3,4$	2
A $(1,2,3)$	$5,6,7$	3
Baa $(1,2,3)$	$8,9,10$	4
Ba $(1,2,3)$	$11,12,13$	5
B $(1,2,3)$	$14,15,16$	6
Caa $(1,2,3)$, Ca, C	$17,18,19,20,21$	7
D	23	8
WR	22	Deleted

Table 1.1: New credit rating system

2 | Continuous Time Markov Model

A continuous time Markov chain is a continuous stochastic process in which, for each state, the process changes state according to an exponential random variable and then move to a different state as specified by the probabilities of the transition matrix. We consider the homogeneous case, as in Lando and Skødeberg (2002) [6], Bladt and Sørensen (2009) [3] in order to estimate transition intensities.
We consider an 8-state Markov chain, where the states are the ratings defined as specified in the previous chapter. We collect transition intensities in a 7x8 q matrix, whose ij' element is the migration's intensity from rating i to rating j.
Exploiting the definition of a Continuous Time Markov chain, i.e.

$$\mathbb{P}\left(X_{t_n} = i_{t_n} \mid X_{t_{n-1}} = i_{t_{n-1}}, \ldots, X_{t_0} = i_{t_0}\right) = \mathbb{P}\left(X_{t_n} = i_{t_n} \mid X_{t_{n-1}} = i_{t_{n-1}}\right)$$

We can model the theoretical concept that a credit migration is independent of previous history since the credit rating itself does contain all the credit quality information for a given company.
Since transition probabilities are continuous we can give a definition for q

$$q_{ij} := \lim_{t \to 0^+} \frac{p_{ij}(0,t)}{t}, \qquad i \neq j$$

$$q_{ij} := \lim_{t \to 0^+} \frac{1 - p_{ii}(0,t)}{t}, \qquad i = j$$

where all q_{ij} are finite, and $q_{ii} = 0$.
Finally the intensity for the corporate $c \in \mathcal{C}$ is

$$\lambda_{ij}^c(t) := q_{ij}\mathbb{I}_{X_t^c = i}, \tag{2.1}$$

meaning that if the corporate c has a rating i at time t the intensity is non zero.
Carrying on the work of Bladt and Sørensen (2009) [3], we assume that each corporate in

the set $\mathcal{C}$ is modeled by an independent Markov process and we follow a notation similar to theirs.
The log-likelihood is

$$logL(p|t_1..., t_N) = -\sum_{i=1}^{L}\sum_{j\neq i} q_{ij}\sum_{c\in\mathcal{C}} R_i^c + \sum_{i=1}^{L}\sum_{j\neq i}\sum_{c\in\mathcal{C}}\sum_{n=1}^{N_{ij}^c} log(\lambda^c_{k_n=(ij)}(t_n)),$$

where

$$R_i^c := \int_0^T \mathbb{I}_{X_t^c=i}dt$$

is the total time spent in state i before time T by the process X_t^c, and $N_{ij}(T)$ is the number of transitions from state i to state j in the time interval $[0, T]$ of the process X_t^c.
Since the companies are independent from each other we can drop the dependence on c

$$N_{ij} := \sum_{c\in\mathcal{C}} N_{ij}^c, \qquad R_i := \sum_{c\in\mathcal{C}} R_i^c.$$

The log-likelihood then becomes

$$logL(p|t_1..., t_N) = \sum_{i=1}^{L}\sum_{j\neq i}\left[-q_{ij}R_i + N_{ij}log(q_{ij})\right]. \tag{2.2}$$

It is straightforward to conclude that the maximum likelihood estimator of the q matrix based on a continuous record of the credit rating histories is given (for all $i \neq j$) by

$$\hat{q}_{ij} = \frac{N_{ij}}{R_i}. \tag{2.3}$$

2.1. Calibration

In order to calibrate the CTM we implement the function `calibrationCTM.m`, which computes the matrix N_{ij}, the vector R_i and gives as output the q matrix.

$$\hat{q} = \begin{bmatrix} - & 7.82 & 0.45 & 0.08 & - & - & - & - \\ 1.07 & - & 9.22 & 0.21 & 0.02 & 0.02 & - & 0.00 \\ 0.07 & 2.74 & - & 6.12 & 0.23 & 0.05 & 0.01 & 0.02 \\ 0.02 & 0.13 & 4.15 & - & 4.93 & 0.41 & 0.06 & 0.05 \\ 0.00 & 0.05 & 0.41 & 6.60 & - & 10.12 & 0.54 & 0.30 \\ 0.00 & 0.04 & 0.14 & 0.43 & 5.66 & - & 10.79 & 1.54 \\ - & 0.01 & 0.04 & 0.12 & 0.33 & 7.30 & - & 11.83 \end{bmatrix}$$

2.2. Simulation

We aim to simulate a synthetic dataset that mimics the characteristics of the original one following the CTM model.
In order to simplify future dataset's manipulations we code the `findCorporates.m` function, thanks to which we save the first and last indices for each corporate $c \in \mathcal{C}$.
`simulationCTM.m` simulates the migration histories of each synthetic corporate within the same observation window as the corresponding real dataset.
To simulate the corporate migrations, according to Baviera 4.1 [2], we need to determine two essential elements at each migration time step: the interarrival time S_n, which represents the period that a corporate spends in the current rating, and the new rating j to which it migrates at the end of that period.
To efficiently generate the simulated migrations, our function uses only two uniform random variable extractions for each rating migration, one for the interarrival time S_n and the other one for the new rating j to be assigned.
In order to find the new rating we extract a uniform r.v in the interval $[0, \sum_j \lambda_{ij}]$ and then verify in which segment of the cumulative intensities vector it falls.
Using this method we have that the new rating j is:

$$\begin{cases} j = \inf \left\{ j \in I \text{ s.t.} \quad x \geq \sum_{m=1}^{j} \lambda_{im} \right\} \\ x \sim U\left(\left[0, \sum_j \lambda_{ij} \right] \right) \end{cases} \tag{2.4}$$

Then we model the interarrival time S_n with an exponential distribution. Indeed, we have that the exponential r.v describes an event that occurs continuously at a constant average rate $\sum_j \lambda_{ij}$ and, since the Markov property holds,

$$\mathbb{P}(X > t + s \mid X > t) = \mathbb{P}(X > s) \quad \forall s, t \geq 0.$$

Then the interarrivaltime is

$$\begin{cases} S_n = \frac{-log(1-u)}{\sum_j \lambda_{ij}} \\ u \sim U\left([0,1]\right) \end{cases} . \tag{2.5}$$

This approach results in a fast simulation algorithm, enabling us to simulate approximately 184,000 years of observations in just a few seconds.

Let us emphasize that the migration time interval S_n and the new rating j are assumed to be independent in this model.

2.3. Confidence Interval

In order to find a confidence interval for the $\hat{q}$ matrix, we calibrate the model over 1000 simulations and, since every q_{ij} has its own distribution, which is unknown, we take the α-quantiles with $\alpha = 0.025$, 0.975 in order to obtain an interval for q_{ij} with a level of confidence $\gamma = 95\%$.
The results are the following:

$$\begin{bmatrix} - & [7.26, 8.41] & [0.31, 0.60] & [0.03, 0.14] & - & - & - & - \\ [0.94, 1.19] & - & [8.86, 9.59] & [0.17, 0.27] & [0.00, 0.04] & [0.01, 0.04] & - & [0, 0.01] \\ [0.05, 0.09] & [2.61, 2.89] & - & [5.92, 6.31] & [0.19, 0.26] & [0.03, 0.07] & [0.00, 0.02] & [0.01, 0.03] \\ [0.01, 0.03] & [0.10, 0.16] & [3.99, 4.32] & - & [4.76, 5.10] & [0.36, 0.47] & [0.04, 0.08] & [0.03, 0.07] \\ - & [0.03, 0.08] & [0.35, 0.48] & [6.36, 6.88] & - & [9.82, 10.45] & [0.46, 0.61] & [0.24, 0.35] \\ - & [0.02, 0.06] & [0.11, 0.17] & [0.37, 0.50] & [5.46, 5.89] & - & [10.50, 11.08] & [1.43, 1.66] \\ - & [0, 0.02] & [0.02, 0.07] & [0.08, 0.15] & [0.27, 0.39] & [7.02, 7.59] & - & [11.48, 12.21] \end{bmatrix}$$

We can notice that the values of $\hat{q}$ calibrated on the original dataset in section 2.1 fall into the confidence region, which confirms the goodness of both our simulation and calibration methods.

3 | Multidimensional Hawkes Process

Point processes are generalizations of Markov processes, and they are dependent on their own history. Thus they are a natural choice for our model. The most common example of a self-exciting process is the Hawkes process, which is characterized by stochastic intensity. This process is a counting process with exponentially decaying intensity

$$\lambda_{ij}(t) = q_{ij}(t)\mathbb{I}_{X_t=i} + \sum_{t_n<t} \varphi^{ij,c}(t-tn)\mathbb{I}_{X_t=i} \tag{3.1}$$

where φ is the exponential kernel function

$$\varphi(t) := e^{-\frac{t}{\tau}}.$$

The couple ij denotes the starting and the arrival rating, respectively, and $c \in \mathcal{C}$ indicates the corporate for which there is the jump. Since $q_{ij}(t)$ does not depend on time, we can drop its dependence in the notation, leaving it just to q_{ij}.
Immediately after a jump the intensity has a spike, then reverts back to a baseline over time. Assuming again, like in the CTM model, that rating migrations are independent across different corporates, our new set of parameters is $p = (\alpha, \tau, q)$.
α and τ are the two parameters chosen to yield a better fit to the observed data, and they represent, respectively, the measure of increased intensities of downgrades and the time period of increased intensities of downgrades when a cascade of sequential downgrades occurs.
The most common technique for parametric estimation of Hawkes processes is the Maximum Likelihood Estimator.
For a non-homogeneous, multi-dimensional Poisson process the log-likelihood is (see Bacry

et al. 2015[1])

$$logL(p|t_1, ..., t_N) = -\sum_{i=1}^{L}\sum_{j\neq i}\int_0^T \lambda_{ij}(t)dt + \sum_{n=1}^{N} log(\lambda_{k_n}(t_n)) \tag{3.2}$$

where, following a notation similar to Baviera (2023) [2], L is the number of ratings without the default and the the pair $k_n = (ij)$ is the mark associated with the event in t_n.

For a single corporate c, the first term in (3.2) becomes

$$-\sum_{i=1}^{L}\sum_{j\neq i}\int_0^T \lambda_{ij}(t)dt = -\sum_{i=1}^{L}\sum_{j\neq i} q_{ij}R_i^c - \alpha\sum_{i>3}^{L}\sum_{j>i} q_{ij}I_i^c(\tau)$$

where

$$I_i^c(\tau) := \tau \sum_{k<i}\sum_{n\in\mathcal{D}_{ki}^c}\sum_{l\in\mathcal{D}^c(t_n)} (e^{-\frac{t_n-t_l}{\tau}} - e^{-\frac{t_{n+1}\wedge T-t_l}{\tau}}) \tag{3.3}$$

where $k_n = (ki)$, i.e. k_n is a downgrade for the corporate c to rating i, and 3 is the starting rating given for the momentum effect.

As far as the second term in (3.2) is concerned we have

$$\sum_{n=1}^{N^c} log(\lambda_{k_n}^c(t_n)) = \sum_{i=i}^{L}\sum_{j\neq i} N_{ij}^c log(q_{ij}) + \sum_{i>3}^{L}\sum_{j>i}\sum_{n\in\mathcal{D}_{ij}^c} log[1+\alpha J_{ij}^c(t_n;\tau)]$$

where

$$J_{ij}^c(t_n;\tau) := \sum_{l\in\mathcal{D}^c(t_{n-1})} e^{-\frac{t_n-t_l}{\tau}} \tag{3.4}$$

where $k_n = (ij)$, i.e. k_n is a downgrade for the corporate c to rating j from rating i.

Exploiting again the independence between the corporates $c \in \mathcal{C}$, we can obtain the total log-likelihood for the issuers summing the log-likelihood of each corporate

$$\begin{gathered} logL(p|t_1, ..., t_N) = -\sum_{i=1}^{L}\sum_{j\neq i} q_{ij}R_i - \alpha\sum_{i>3}^{L}\sum_{j>i} q_{ij}I_i(\tau) + \\ \sum_{i=1}^{L}\sum_{j\neq i} N_{ij}log(q_{ij}) + \sum_{i>3}^{L}\sum_{j>i}\sum_{c\in\mathcal{C}}\sum_{n\in\mathcal{D}_{ij}^c} log[1+\alpha J_{ij}^c(t_n;\tau)] \end{gathered} \tag{3.5}$$

where R_i and N_{ij} have been already defined in the previous sections, and

$$I_i(\tau) := \sum_{c \in \mathcal{C}} I_i^c(\tau).$$

Now we can obtain the MLE estimators $\hat{q}_{ij}$.
If the transition ij does not denote a downgrade of interest, then

$$\hat{q}_{ij} = \frac{N_{ij}}{R_i},$$

otherwise

$$\hat{q}_{ij} = \frac{N_{ij}}{R_i + \alpha I_i}.$$

Finally we can rewrite the log-likelihood expression as

$$\begin{aligned} logL(\alpha, \tau | t_1, ..., t_N) = -N + \sum_{i=1}^{L} \sum_{j \neq i} N_{ij} log(N_{ij}) - \sum_{(i,j) \notin \mathcal{D}} N_{ij} log(R_i) + \\ \sum_{(i,j) \in \mathcal{D}} \left[-N_{ij} log[R_i + \alpha I_i(\tau)] + \sum_{c \in \mathcal{C}} \sum_{n \in \mathcal{D}_{ij}^c} log[1 + \alpha J_{ij}^c(t_n; \tau)] \right]. \end{aligned} \tag{3.6}$$

3.1. Calibration

In order to calibrate the Multidimensional Hawkes Model we thus have to maximize the log-likelihood in (3.6). First of all we plot the log-likelihood in a neighborhood of the initial guesses for the parameters $(\alpha_0, \tau_0) = (2, 0.5)$.

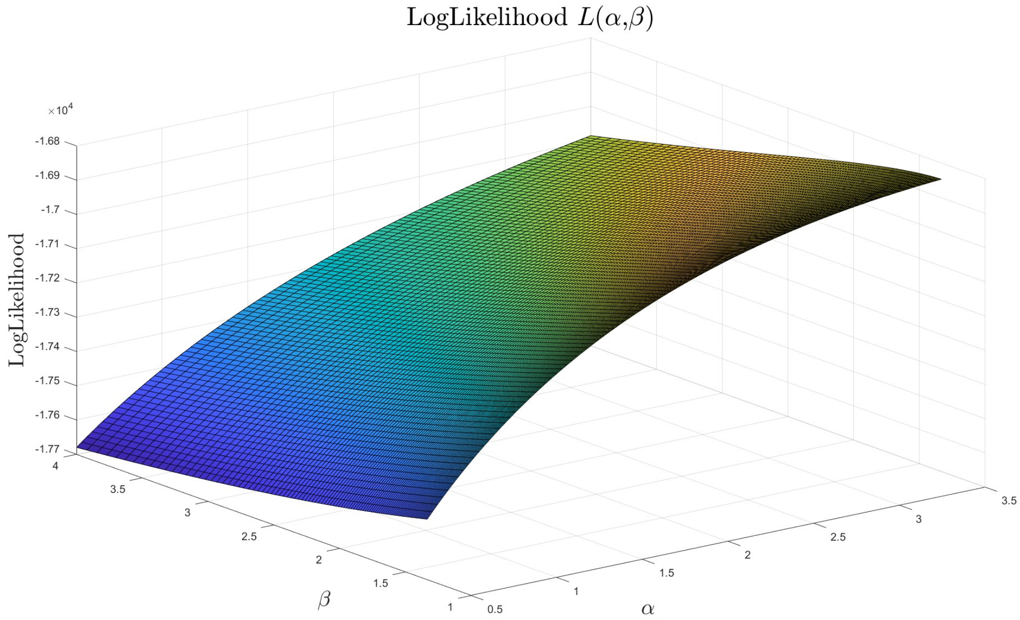

Figure 3.1: log-likelihood in a neighborhood of (α,τ)=(2,0.5)

In Figure (3.1) we observe that the surface is regular along all directions, but it seems to have a plateau as α increases, near the maximum. Moreover it looks like the log-likelihood could have the maximum for an α greater than the ones plotted above.

We implement the Matlab function `calibrationHawkes` to find the optimal set of parameters $p = (\alpha, \tau, q)$.

The base concept behind this calibration is quite simple: since q is a function of the other two parameters, we just need to find numerically the (α, τ) which maximizes the log-likelihood in (3.6). However, since the dataset is quite big, this simple operation can lead to a long computing time. So we need to find a way to speed up the calibration trying to compute all the downgrade cascades just once at the beginning of the algorithm, and then just maximize a function handle which contains only trivial operations between the previously computed values.

Our idea is to compute the Δt's at the exponent both in $I_i^c(\tau)$ (3.3) and in $J_{ij}^c(t_n; \tau)$ (3.4) and save them in matrices, respectively named `Itimes` and `Jtimes` in the code.

Let us briefly describe them:

- `Itimes` is a 5-dimensional matrix that stores the arrays Δt_1 and Δt_2, which we define as follows:

$$\Delta t_1 := t_n - t_l,$$

$$\Delta t_2 := t_{n+1} \wedge T - t_l.$$

  Along the first dimension of the matrix we save the downgrade cascade's Δt_1 for a given corporate $c \in \mathcal{C}$, and arrival rating i. We add a second dimension, so that we can add other downgrade cascades, if any, for both the same corporate c and arrival rating i.
  Now we append a second layer (third dimension), built in the same way of the one just described, but for the Δt_2's storage.
  Finally we add the fourth dimension to the matrix: for every corporate $c \in \mathcal{C}$, we have the 3-dimensional structure explained above.
  Lastly we introduce the dependence on the arrival rating i via the fifth dimension.

- `Jtimes` is a 5-dimensional matrix that stores the arrays Δt which we define as follows:

$$\Delta t := t_n - t_l.$$

  Similarly to the previous case the first and the second matrix' dimensions store the sets of downgrade cascade's Δt for a fixed rating couple ij and corporate c.
  The third dimension introduces the dependence on the arrival rating j.
  Finally, the forth and fifth dimensions give us respectively the dependence on the corporate $c \in \mathcal{C}$ and on the starting rating i.

Having these matrices filled up, we just have to plug the values in Equation (3.6) and maximize numerically. Thanks to this trick, one single calibration takes about a tenth of the total time elapsed in case we would have re-calculated all the sets of consecutive downgrades directly with the dependence on (α, τ). In order to improve even further the speed of the code we "slice" the matrices in order to reduce the number of elements.
Indeed both `Itimes` and `Jtimes` are initialized as full of `NaN`s. The idea is to loop on the corporates' dimension and remove the layers without any numerical value. In this way we can delete all the corporates that do not have any downgrade momentum effect, saving us a lot of useless computations.

We take as initial guesses for the calibration α_0=2 and τ_0=0.5 and we obtain the following result:

$$\hat{\alpha} = 4.4890$$

$$\hat{\tau} = 0.6493$$

$$\hat{q} = \begin{bmatrix} - & 7.82 & 0.45 & 0.08 & - & - & - & - \\ 1.07 & - & 9.22 & 0.21 & 0.02 & 0.02 & - & 0.00 \\ 0.07 & 2.74 & - & 6.12 & 0.23 & 0.05 & 0.01 & 0.02 \\ 0.02 & 0.13 & 4.15 & - & 4.93 & 0.41 & 0.06 & 0.05 \\ 0.00 & 0.05 & 0.41 & 6.60 & - & 8.45 & 0.45 & 0.25 \\ 0.00 & 0.04 & 0.14 & 0.43 & 5.66 & - & 8.95 & 1.28 \\ - & 0.01 & 0.04 & 0.12 & 0.33 & 7.30 & - & 8.38 \end{bmatrix}$$

where all the values in the matrix are expressed as a percentage.

The calibrated value $\hat{\alpha}$ is, as expected from the observations in Figure (3.1), outside of the plot. So we plot the log-likelihood again, centering the axis on the new parameters.

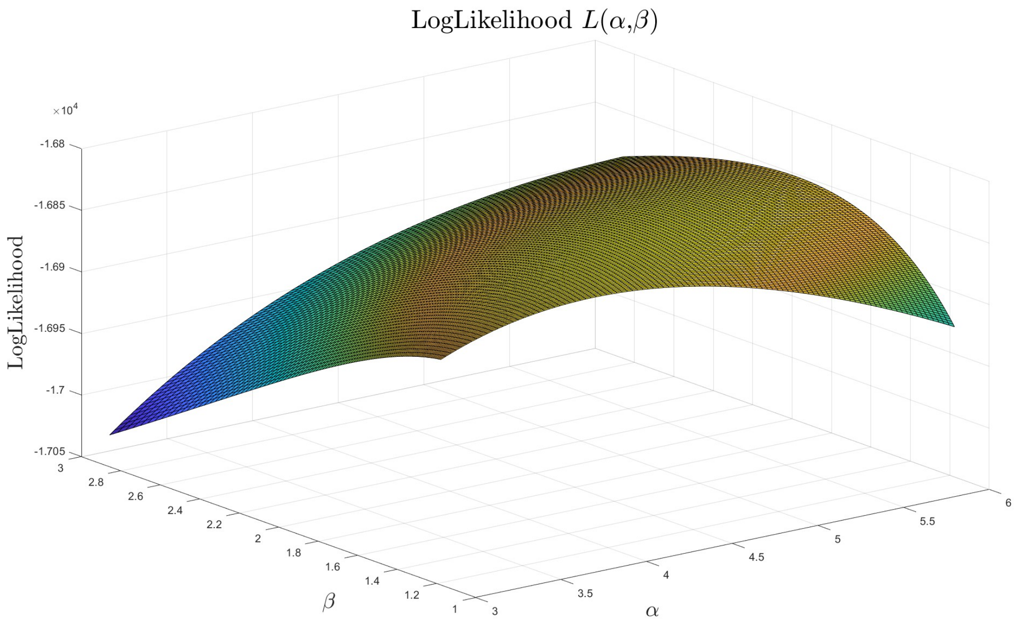

Figure 3.2: log-likelihood in a neighborhood of (α,τ)=(4.5,0.65)

Finally, in Figure (3.2) we can clearly see that the log-likelihood smoothly approaches a maximum in the plotted region, at the top of the plateau. Furthermore, from the plot we can deduce that the calibration should be stable.

3.2. Simulation

As in the CTM case, we aim to simulate a dataset according to the multidimensional Hawkes model using its calibrated parameters.
Unlike the simulation algorithm for the CTM process, it is crucial to take into account the dependence between the inter-arrival times and the intensities of the counting process.
In particular, for each corporate, we proceed in the following way:

- in case there is no previous rating or the current migration is not a downgrade of interest (i.e. the migration ij is s.t j $\leq$ minrating) we simulate the next rating grade using equations (2.4) and (2.5) from the CTM algorithm;
- in case the current state of the corporate is defaulted we extend the current rating period until T = 31th January 2020;
- In case the current rating i is reached through a donwngrade and $i \geq 3$ we modify the current intensity process in order to take into account the rating momentum phenomenon.
  In order to account for a bigger probability of a downgrade when the last migration is a downgrade, we have:

$$\lambda_{ij}^c(t) := \begin{cases} q_{ij}\mathbb{I}_{X_t^c=i} + \alpha \sum_{l\in\mathcal{D}^c(t^-)} \exp\left(-\frac{t-t_l}{\tau}\right) & (ij) \in \mathcal{D} \\ q_{ij}\mathbb{I}_{X_t^c=i} & \text{otherwise} \end{cases} \tag{3.7}$$

  where $\mathcal{D}$ is the set of downgrades ij of interest (i.e. $i < j$ and $i > 3$)) and $\mathcal{D}^c(t^-)$ is the set of consecutive downgrades of interest for the corporate c before time t.
  The main problem of implementing this in the simulation algorithm is related to the fact that the interarrival time and the new rating depend on each other so unlike the CTM simulation we cannot model the new rating and the interarrival time samplings through independent r.v.'s extractions. We manage to simultaneously model the two quantities of interest implementing the algorithm described by Ozaka (1979) [7].
  Let $\{t_n\}_{n=1}^{i}$ be the sequence of downgrades dates of interest for that corporate until time t_i, namely the cascade of downgrades that starts with a rating greater than minimum rating and ends at the current rating i.
  If we consider the definition of a non homogeneous Poisson process:

$$\mathbb{P}(\text{rating migration}(ij)) = e^{-\int_{t_i}^{x} \lambda_{ij}(t)dt}$$

and substitute the definition (3.7) we have that the new rating migration occurs at time x satisfies the following:

$$\begin{cases} \log(u) + \int_{t_i}^{x} \left(\sum_j q_{ij} + \sum_{j>i} q_{ij}\alpha \sum_{n=1}^{i} \varphi\left(t - t_n\right) \right) dt = 0 \\ u \sim U\left([0,1]\right) \end{cases}$$

where the second term in the integral represents the downgrade momentum. In the CTM the kernel function $\varphi(t)$ was equal to zero, while in the Hawkes model we have that the exponential kernel function $\varphi(t) := e^{-\frac{t}{\tau}}$.

If we suppose that no jumps occur in the time interval $[t_i, x]$ we exchange the integral with the series and develop the above equation in this way:

$$\begin{aligned}
&\int_{t_i}^{x} \left(\sum_j q_{ij} + \sum_{j>i} q_{ij}\alpha \sum_{n=1}^{i} \varphi\left(t - t_n\right) \right) dt = \\
&= \sum_j q_{ij}\left(x - t_i\right) + \alpha \sum_{j>i} q_{ij} \sum_{n=1}^{i} \int_{t_i}^{x} e^{-\frac{t-t_n}{\tau}} dt = \\
&= \sum_j q_{ij}\left(x - t_i\right) + \alpha \sum_{t>i} q_{ij}\left(-\tau\right) \left\{ \sum_{n=1}^{i} e^{-\frac{t_i - t_n}{\tau}} - \sum_{n=1}^{i} e^{-\frac{x-t_n}{\tau}} \right\} = \\
&= \sum_j q_{ij}\left(x - t_i\right) + a \sum_{j>i} q_{ij}\tau\left(-\psi\left(t_i\right) + \psi(x)\right) = -log(u)
\end{aligned}$$

where $\psi(x) := \sum_{n=1}^{i} e^{-\frac{x-t_n}{\tau}}$.

Using this method, we both model in the correct way the decaying intensity property of the Hawkes processes and we ensure that the interarrival time depends on the intensity not only through its current rating state but also through its previous rating history.

3.3. Confidence Region

We want to assess the behavior of the model calibrated in Section 3.1 in a simulated environment to check its stability. We run a series of $M = 1000$ simulations.
To draw a confidence region for α and τ, it is crucial to check their normality.

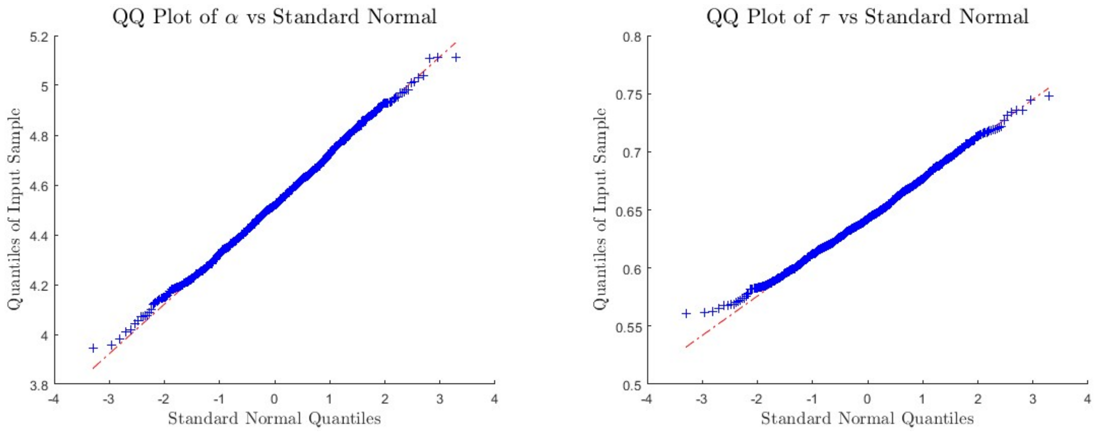

Figure 3.3: Normality check

Parameter	Result	p-value
α	0	0.9888
τ	0	0.4506

Table 3.1: One-sample Kolmogorov-Smirnov test

So we have that:

$$\frac{\alpha - \mathbb{E}[\alpha]}{\sqrt{\mathrm{Var}(\alpha)}} \sim \mathcal{N}(0,1) \tag{3.8}$$

In our case mean and variance of the distribution of α are not known so, if we estimate them through the unbiased sample mean and variance we obtain that:

$$\frac{\alpha - \bar{\alpha}_n}{\sqrt{\frac{s_n^2}{n}}} \sim t_{n-1} \tag{3.9}$$

where s_n^2 is the sample unbiased variance estimator.
The same hold for τ.
Thus, we can infer separately from the two distributions using the method described in [5].

To estimate a rectangular Bonferroni confidence region at the $\gamma = 95\% = 1 - \alpha$ level, the two individual confidence intervals have to be computed using the t-student distribution. As we are working within a bivariate distribution, it is necessary to apply the Bonferroni correction, which involves estimating the two single intervals at a 1-$\alpha/2$ confidence level.

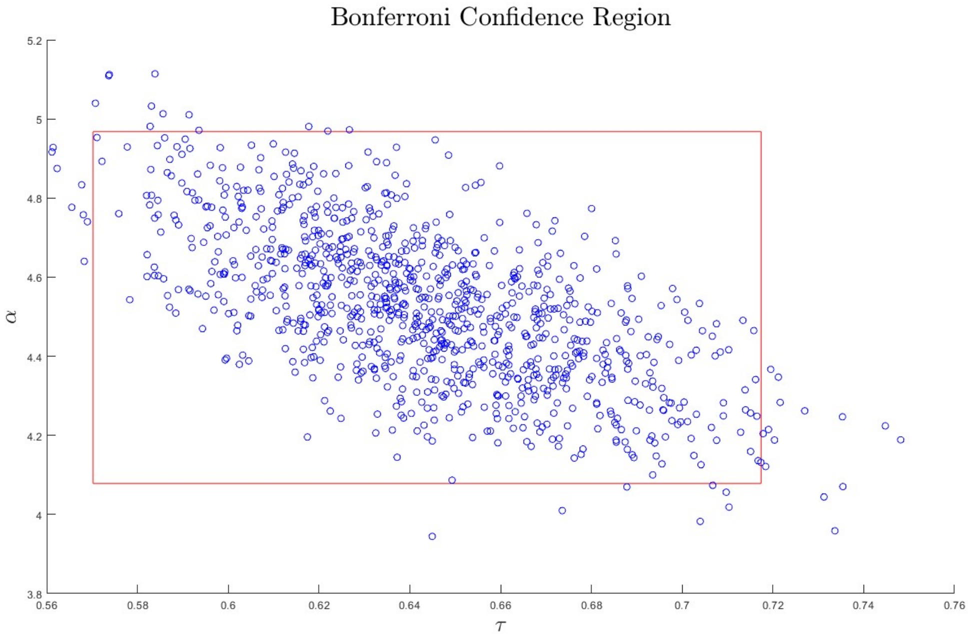

Figure 3.4: Bonferroni Confidence Region

We immediately notice a strong relative correlation between the two parameters. This suggest us the possibility of building an ellipsoidal confidence region to take into account this relation.

So we verify that the joint distribution $\mathbf{X} = \begin{bmatrix} \alpha \\ \tau \end{bmatrix}$ is normally distributed thanks to Henze-Zirkler's Multivariate Normality Test. Therefore if we recall that

$$X_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1) \Rightarrow \sum_{i=1}^{n} X_i^2 \sim \chi^2(n)$$

we can extend the theoretical result (3.9) using the Hotelling's T-Squared distribution, which is the multivariate analogous of the t-student distribution and obtain the following statistics:

$$T^2 = n(\mathbf{X} - \bar{\mathbf{X}})^{\mathbf{T}} V^{-1} (\mathbf{X} - \bar{\mathbf{X}}) \sim \chi^2(2)$$

where $\bar{\mathbf{X}} = \begin{bmatrix} \mathbb{E}[\alpha] \\ \mathbb{E}[\tau] \end{bmatrix}$ and V^{-1} is the covariance matrix.

It is worth mentioning that the approximation $T^2 \sim \chi^2(2)$ holds only for n $\to \infty$.
Finally, in order to plot the ellipsoidal confidence region for the bivariate distribution we get the spectral decomposition of the covariance matrix: the eigenvectors represent the directions along which the data have the highest variance. Thus they define the axes of the ellipsoidal shape of the confidence region. The eigenvalues correspond to the variances explained by each eigenvector direction. The larger the eigenvalue, the more spread the data have along the corresponding eigenvector.

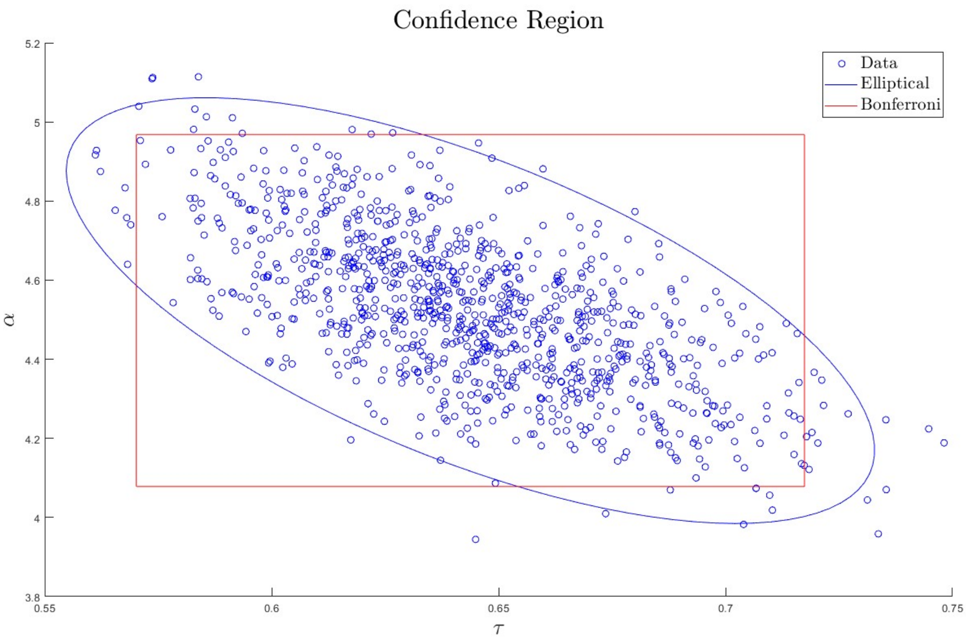

Figure 3.5: Confidence region (τ, α)

The ellipsoidal confidence region seems to be more reliable since it catches the correlation in the point cluod, while the Bonferroni region gives us a confidence interval for the parameters.

The confidence interval for q_{ij} calibrated with Hawkes model with level of significance $\alpha = 0.95$ is the following:

$$\begin{bmatrix}
- & [7.29, 8.41] & [0.32, 0.60] & [0.03, 0.15] & - & - & - & - \\
[0.95, 1.19] & - & [8.87, 9.54] & [0.17, 0.27] & [0.00, 0.03] & [0.01, 0.04] & - & [0, 0.01] \\
[0.05, 0.09] & [2.61, 2.88] & - & [5.92, 6.33] & [0.19, 0.27] & [0.03, 0.07] & [0.00, 0.02] & [0.01, 0.03] \\
[0.01, 0.03] & [0.10, 0.16] & [3.99, 4.32] & - & [4.75, 5.10] & [0.36, 0.47] & [0.04, 0.08] & [0.03, 0.07] \\
- & [0.03, 0.08] & [0.35, 0.48] & [6.35, 6.86] & - & [8.19, 8.74] & [0.36, 0.52] & [0.19, 0.30] \\
- & [0.02, 0.06] & [0.11, 0.17] & [0.38, 0.49] & [5.46, 5.89] & - & [8.68, 9.20] & [1.19, 1.37] \\
- & [0, 0.02] & [0.02, 0.07] & [0.08, 0.15] & [0.27, 0.39] & [7.04, 7.59] & - & [8.11, 8.69]
\end{bmatrix}$$

4 | Robustness check

In this section we want to test the robustness of the models, studying their performance when dealing with different scenarios.

4.1. Default as a pure absorbing state

Let us assume now that the default of a corporate in the original dataset is a pure absorbing state, i.e. we truncate each corporate's dataset at default time (see Section 1 for dataset cleaning details).

The q matrix for the CTM model is

$$\hat{q}_{CTM} = \begin{bmatrix} - & 7.82 & 0.45 & 0.08 & - & - & - & - \\ 1.07 & - & 9.23 & 0.22 & 0.02 & 0.02 & - & 0.00 \\ 0.07 & 2.74 & - & 6.12 & 0.23 & 0.05 & 0.01 & 0.02 \\ 0.02 & 0.13 & 4.14 & - & 4.91 & 0.41 & 0.06 & 0.05 \\ 0.00 & 0.05 & 0.41 & 6.57 & - & 10.13 & 0.53 & 0.29 \\ 0.00 & 0.04 & 0.14 & 0.45 & 5.59 & - & 10.77 & 1.50 \\ - & 0.01 & 0.05 & 0.09 & 0.31 & 7.11 & - & 11.30 \end{bmatrix}.$$

We can immediately notice that the q matrix is very close to the one calibrated in Section 2.1.
We carry on the analysis then, by calibrating the Multidimensional Hawkes model, as described in Section 3.1.
But first, let us look at the log-likelihood in a neighborhood of the previously calibrated parameters.

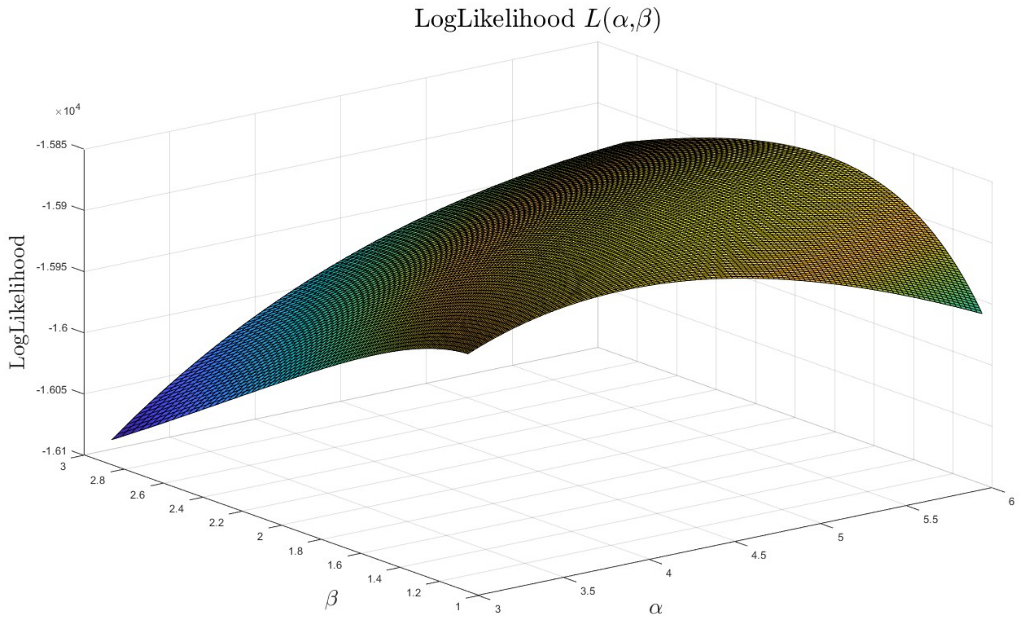

Figure 4.1: log-likelihood in a neighborhood of (α,τ)=(4.5,0.65)

This new function shows a very similar behaviour to the one plotted in Figure (3.2), letting us finally proceed to the model calibration, obtaining

$$\hat{\alpha} = 4.7252$$

$$\hat{\tau} = 0.6658$$

$$\hat{q}_{Hwk} = \begin{bmatrix} - & 7.82 & 0.45 & 0.08 & - & - & - & - \\ 1.07 & - & 9.23 & 0.22 & 0.02 & 0.02 & - & 0.00 \\ 0.07 & 2.74 & - & 6.12 & 0.23 & 0.05 & 0.01 & 0.02 \\ 0.02 & 0.13 & 4.14 & - & 4.91 & 0.41 & 0.06 & 0.05 \\ 0.00 & 0.05 & 0.41 & 6.57 & - & 8.34 & 0.43 & 0.24 \\ 0.00 & 0.04 & 0.14 & 0.45 & 5.59 & - & 8.77 & 1.22 \\ - & 0.01 & 0.05 & 0.09 & 0.31 & 7.11 & - & 7.61 \end{bmatrix}.$$

Also the Hawkes' parameters are quite close to the initial setting ones, even if both α and τ are a bit higher.

We finally verify α and τ normality using again the One-sample Kolmogorov-Smirnov test:

Parameter	Result	p-value
α	0	0.8464
τ	0	0.9364

Table 4.1: One-sample Kolmogorov-Smirnov test

Given the p-values above we cannot reject the null hypothesis, thus the two distributions are normal, as we can see in the following QQplots.

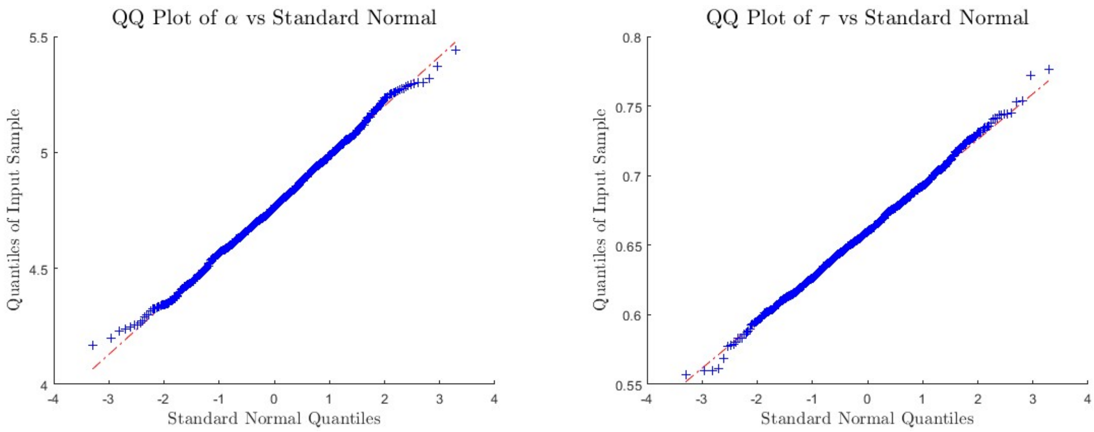

Figure 4.2: Normality check

so we can plot the confidence region for the parameters

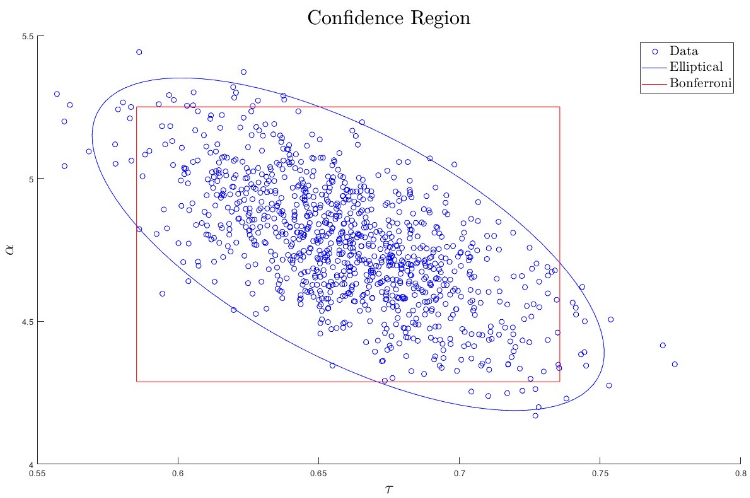

Figure 4.3: Confidence region (τ, α)

We notice that the p-value of the parameter α becomes smaller than the base case scenario

while the one of τ significantly increase. This results in a better in that the Bonferroni confidence region, which strongly relied on the normality assumption, could be more accurate.
Below are reported the confidence intervals for $\hat{q}$ obtained with the two models, and we can state that the calibrated q matrices are inside the confidence intervals bounds.

$$\hat{q}_{\text{CTM}} =$$

$$\begin{bmatrix} - & [7.27, 8.42] & [0.32, 0.59] & [0.03, 0.14] & - & - & - & - \\ [0.95, 1.19] & - & [8.88, 9.56] & [0.16, 0.27] & [0.00, 0.04] & [0.01, 0.04] & - & [0, 0.01] \\ [0.05, 0.09] & [2.61, 2.88] & - & [5.93, 6.31] & [0.19, 0.27] & [0.03, 0.07] & [0.00, 0.03] & [0.01, 0.03] \\ [0.01, 0.03] & [0.10, 0.16] & [3.97, 4.30] & - & [4.75, 5.10] & [0.36, 0.46] & [0.04, 0.09] & [0.03, 0.07] \\ [0, 0.01] & [0.03, 0.08] & [0.34, 0.47] & [6.31, 6.86] & - & [9.80, 10.48] & [0.46, 0.60] & [0.23, 0.35] \\ [0, 0.01] & [0.02, 0.06] & [0.11, 0.18] & [0.39, 0.51] & [5.39, 5.81] & - & [10.44, 11.07] & [1.39, 1.62] \\ - & [0, 0.01] & [0.02, 0.08] & [0.06, 0.13] & [0.25, 0.37] & [6.81, 7.43] & - & [10.93, 11.68] \end{bmatrix}$$

$$\hat{q}_{\text{Hwk}} =$$

$$\begin{bmatrix} - & [7.26, 8.42] & [0.32, 0.60] & [0.02, 0.14] & - & - & - & - \\ [0.96, 1.19] & - & [8.91, 9.60] & [0.16, 0.27] & [0.00, 0.03] & [0.01, 0.04] & - & [0, 0.01] \\ [0.05, 0.10] & [2.60, 2.87] & - & [5.92, 6.32] & [0.19, 0.26] & [0.03, 0.07] & [0.00, 0.02] & [0.01, 0.03] \\ [0.01, 0.03] & [0.10, 0.16] & [3.97, 4.30] & - & [4.74, 5.09] & [0.36, 0.46] & [0.05, 0.08] & [0.03, 0.07] \\ - & [0.03, 0.08] & [0.34, 0.47] & [6.32, 6.85] & - & [8.03, 8.60] & [0.37, 0.50] & [0.20, 0.29] \\ - & [0.02, 0.06] & [0.11, 0.18] & [0.39, 0.51] & [5.36, 5.80] & - & [8.52, 9.04] & [1.13, 1.32] \\ - & [0, 0.01] & [0.03, 0.08] & [0.06, 0.12] & [0.25, 0.37] & [6.81, 7.42] & - & [7.29, 7.91] \end{bmatrix}$$

4.2. Rating momentum starting from second rating

We want to repeat the analysis in a different scenario, assuming now that the rating downgrade momentum effect starts from rating equal to 2.
The dataset considered is the same as the base scenario, thus the results of the CTM model do not differ.
Once again, we first inspect the log-likelihood surface, observing that the behavior is still the same even in this scenario.

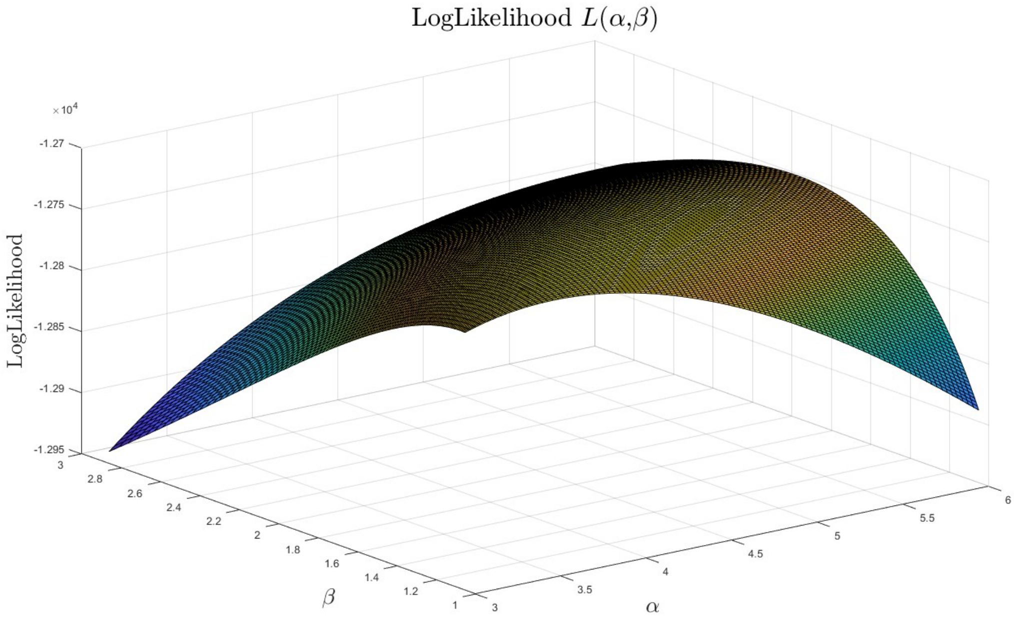

Figure 4.4: log-likelihood in a neighborhood of (α,τ)=(4.5,0.65)

Below we report the results of the Hawkes calibration:

$$\hat{\alpha} = 4.1254$$

$$\hat{\tau} = 0.7072$$

$$q_{\hat{Hwk}} = \begin{bmatrix} - & 7.82 & 0.45 & 0.07 & - & - & - & - \\ 1.06 & - & 9.21 & 0.21 & 0.01 & 0.02 & - & 0.00 \\ 0.07 & 2.74 & - & 6.11 & 0.22 & 0.05 & 0.00 & 0.01 \\ 0.01 & 0.13 & 4.14 & - & 4.26 & 0.35 & 0.05 & 0.04 \\ 0.00 & 0.05 & 0.41 & 6.59 & - & 8.29 & 0.44 & 0.24 \\ 0.00 & 0.04 & 0.13 & 0.43 & 5.65 & - & 8.92 & 1.27 \\ - & 0.01 & 0.04 & 0.11 & 0.32 & 7.29 & - & 8.40 \end{bmatrix}.$$

When comparing $\hat{q}$ with the base scenario, it is noticeable that the 'downgrade momentum' not only caused a fall in the intensities of low ratings (i.e., 6, 7), but also, as expected, of medium ratings. The inclusion of highly ranked companies has a significant impact on the evaluation of the cascade's momentum, which is reflected in a fall in $\hat{\alpha}$ and an increase in $\hat{\tau}$.

The confidence interval on the $\hat{q}$ is

$$\begin{bmatrix} - & [7.16, 8.56] & [0.29, 0.62] & [0.02, 0.16] & - & - & - & - \\ [0.93, 1.20] & - & [8.83, 9.63] & [0.15, 0.28] & [0.00, 0.04] & [0.00, 0.05] & - & [0.00, 0.01] \\ [0.05, 0.10] & [2.59, 2.90] & - & [5.88, 6.35] & [0.18, 0.28] & [0.03, 0.07] & [0.00, 0.02] & [0.01, 0.03] \\ [0.01, 0.03] & [0.10, 0.17] & [3.97, 4.33] & - & [4.08, 4.46] & [0.30, 0.41] & [0.04, 0.07] & [0.03, 0.07] \\ [0.00, 0.02] & [0.03, 0.08] & [0.34, 0.49] & [6.27, 6.93] & - & [7.99, 8.63] & [0.37, 0.52] & [0.19, 0.30] \\ [0.00, 0.01] & [0.02, 0.06] & [0.10, 0.18] & [0.36, 0.50] & [5.43, 5.93] & - & [8.61, 9.26] & [1.16, 1.39] \\ - & [0, 0.03] & [0.02, 0.07] & [0.07, 0.16] & [0.25, 0.41] & [6.93, 7.65] & - & [8.08, 8.76] \end{bmatrix}.$$

The $\hat{q}$ matrix is very similar to the one obtained using three instead of two as starting rating for the momentum cascades.

Finally we come to the normality check

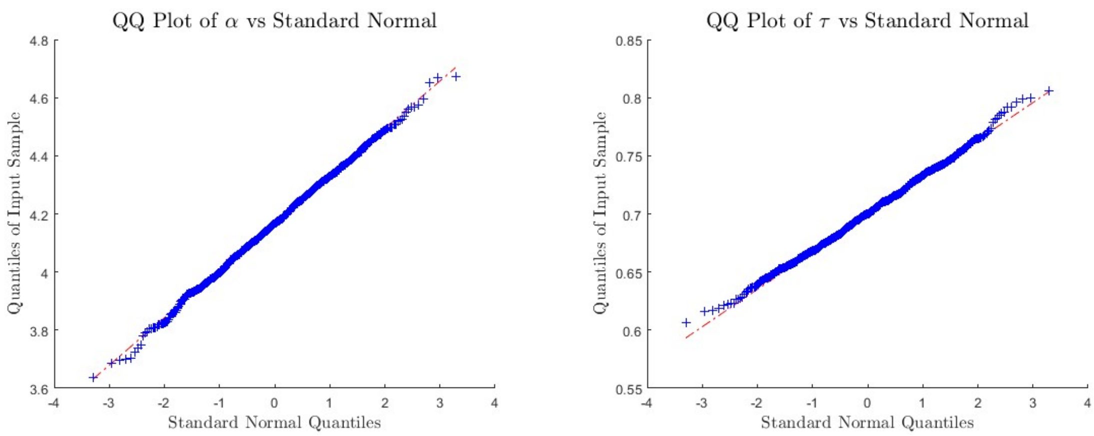

Figure 4.5: Normality check

Parameter	Result	p-value
α	0	0.9670
τ	0	0.7105

Table 4.2: One-sample Kolmogorov-Smirnov test

In conclusion we plot the confidence region for $\hat{\alpha}$ and $\hat{\tau}$

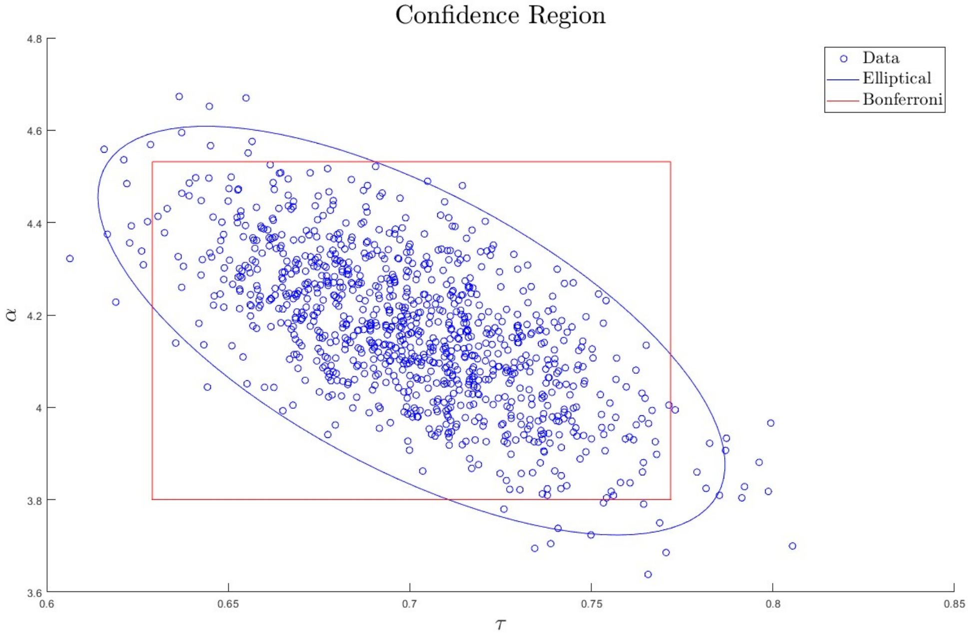

Figure 4.6: Confidence region (τ, α)

5 | Conclusions

In this research we made a comparison between the CTM and the Hawkes models' rating transition probabilities. First of all, it should be pointed out that the probability of migration in the two models is the same for the ratings that cannot be excited, i.e. the ratings lower than the minimum rating used to conduct the analysis. The Hawkes Model has larger probabilities of migration in the lower ratings (e.g. 6,7) since a company might be in either the excited or non-excited version of the state. The momentum effect becomes clear when the probabilities for both excited and non-excited states are added together, which results in greater probabilities in the Hawkes model. Similar findings are seen for the default category; in fact, the extended model shows slightly larger numbers because the momentum effect makes it more probable that there will be subsequent downgrades.

The Hawkes model seems to better take account of the rating migrations, reflecting the empirical non-Markovian behaviour of real-word data.
Both the models turned out to be consistent across different robustness checks, since the models' parameters do not show significant differences.

Bibliography

[1] E. Bacry, I. Mastromatteo, and J. Muzy. Hawkes processes in finance. *Market Microstructure and Liquidity*, 01, 02 2015. doi: 10.1142/S2382626615500057.

[2] R. Baviera. A note on rating momentum in credit migrations.

[3] M. Bladt and M. Sørensen. Efficient estimation of transition rates between credit ratings from observations at discrete time points. *Quantitative Finance*, 9:147–160, 03 2009.

[4] S. Fiumi. Modeling rating momentum in credit migrations. Master's thesis, Politecnico di Milano, 10 2022. URL `http://hdl.handle.net/10589/192121`.

[5] D. Johnson, Richard; Wichern. *Applied Multivariate Statistical Analysis*. Pearson, 2015.

[6] D. Lando and T. Skødeberg. Analyzing rating transitions and rating drift with continuous observations. *Journal of Banking Finance*, 26:423–444, 03 2002.

[7] T. Ozaki. Maximum likelihood estimation of hawkes' self-exciting point processes. *Ann Inst Stat Math 31*, page 145–155, 1979.