Hi Project team,

**1. Time series data**

For the purpose of the application to EM algorithm/MCMC, here are two potential options of data that I found.

**A) S&P Global annual report**

S&P Global previously published their default/rating studies annually, however, as discussed in a prior meeting, S&P seems to start restricting access to the reports for registered members only (whilst out subscription). I had stored the reports to my local for the last 5 years, and each report seems to have rating transition tables only for the last one year (e.g. table 20 for the latest report). I expect this to be able to applied for the purpose.

Please find an attachment, that includes the reports. I actually could find some older reports from web, but those are stored outside of S&P server, so I hold including them off.

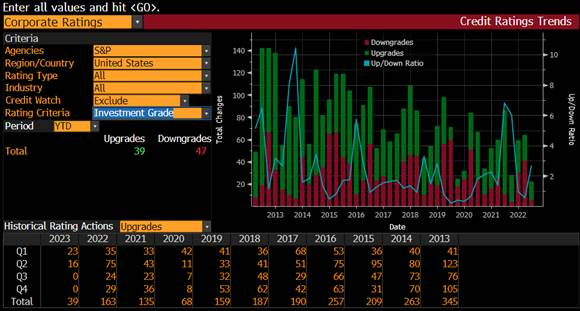
**B) Less granular data from BBG**

Bloomberg ticker, <RATT> provides historical data of

* the numbers that corporate ratings are upgraded/downgraded
* by rating agencies (e.g. S&P, Moody’s, Fitch, etc.)
* by rating criteria (e.g. IG or HY)

Here is an example of IG regarding S&P. The data is less granular; start/end ratings are not available, but I still expect

I assume you have some BBG access through the program, but let me know if you do not have BBG access and would like to take a look at the data.



**2. Comment of the jupyter notebook**

Haven’t had taken a deeper look at the accuracy for now, so will discuss in the meeting.

However, for EM algorithm, I initially saw a fluctuation of the error. That was because your initial guess of generator was set with larger numbers.

When it scaled by the dimension of the generator (e.g. 18), then convergence look much better.

Here is the convergence starting from the scaled generator, where y-axis is log scaled. The error is log-linearly reduced to 1.0e-16 at 1k iterations.

Another option is using another estimate (e.g. DA, WA, etc.) as an initial guess, which worked and of course a bit faster.

A picture containing line, plot, diagram, screenshot

Description automatically generated

Looking forward to seeing you on Thursday!

Best,

Tak

**--------------------------------------------------**

**Takuya Otani**

Vice President, Counterparty Portfolio Management Desk

Fixed Income Division

Mizuho Securities USA LLC

Mizuho Americas

1271 Avenue of the Americas, New York, NY 10020

T: +1 646-949-9817

[takuya.otani@mizuhogroup.com](mailto:takuya.otani@mizuhogroup.com)

[mizuhoamericas.com](https://urldefense.proofpoint.com/v2/url?u=http-3A__www.mizuhoamericas.com_&d=DwMFAg&c=-ib6oBKivNBwAXVVyaiAu6NBDGPkr1Jtp3dZ-MgcLP8&r=wC5PMKOH1dlcncSAQDzDs-sCkVK_ukFOEBTKkf1LsUM&m=VDM_W47i935Q7iJMk_R_HzsHrfeOwBeqKEYY5xEPXaA&s=-UyvweCD8J5-gXmF1aS7QjJcaR9L4ka1ANw3aw1Bq78&e=) | [Twitter](https://urldefense.proofpoint.com/v2/url?u=https-3A__twitter.com_mizuhoamericas&d=DwMFAg&c=-ib6oBKivNBwAXVVyaiAu6NBDGPkr1Jtp3dZ-MgcLP8&r=wC5PMKOH1dlcncSAQDzDs-sCkVK_ukFOEBTKkf1LsUM&m=VDM_W47i935Q7iJMk_R_HzsHrfeOwBeqKEYY5xEPXaA&s=LGaTw7zGiTOscwwd5FCqryOyu90q0SNiN0z6lEOtC60&e=) | [LinkedIn](https://urldefense.proofpoint.com/v2/url?u=https-3A__www.linkedin.com_company_mizuho-2Damericas&d=DwMFAg&c=-ib6oBKivNBwAXVVyaiAu6NBDGPkr1Jtp3dZ-MgcLP8&r=wC5PMKOH1dlcncSAQDzDs-sCkVK_ukFOEBTKkf1LsUM&m=VDM_W47i935Q7iJMk_R_HzsHrfeOwBeqKEYY5xEPXaA&s=pJelpy03_LIHVIkT-TD_-J5r04rR9lts2v0mkMtBPkU&e=) | [YouTube](https://urldefense.proofpoint.com/v2/url?u=https-3A__www.youtube.com_c_MizuhoAmericas&d=DwMFAg&c=-ib6oBKivNBwAXVVyaiAu6NBDGPkr1Jtp3dZ-MgcLP8&r=wC5PMKOH1dlcncSAQDzDs-sCkVK_ukFOEBTKkf1LsUM&m=VDM_W47i935Q7iJMk_R_HzsHrfeOwBeqKEYY5xEPXaA&s=eEUAhAfsCvYu1T8AKWoCUxEkkK3-176wKY6EXXz-tVE&e=)

This message is the property of Mizuho Americas and has been classified as Confidential

Dear Tak,

Thank you for the meeting and many helpful suggestions!

Attached is the MCMC update to date without the convergence check finalized.

We hope to get back to you with some more exciting discoveries with annual reports and Bloomberg <RATT> (if available).

(We will check the availability of the Bloomberg <RATT> function at our program’s terminals and get back to you!)

Thank you again and hope you have a wonderful remainder of the week.

Sincerely,

UChicago FINM Project Lab (Team Mizuho)

Hi Project Team,

Thank you for providing the updates. Here are some comments from my side.

Roughly speaking, the MCMC in the jupyter notebook includes a lot of things to be corrected, and still should have critical error whilst I am still on the way of investigation.

It is nicer if all of the team can work on reviewing the algorithm of MCMC and implementation in python.

**0. Next meeting**

Would you fill your availability in the following Doodle over the weekend?

[https://doodle.com/meeting/participate/id/aMJzVNmb](https://urldefense.com/v3/__https:/doodle.com/meeting/participate/id/aMJzVNmb__;!!AJslKrWaDbga!MMY46seSl2PdfT28buIyE5hGVf4f6tJugB0DdNWiEjYnoGDKVpsGoIM7QnRFCWyWDJBtZFR8ZsuXl82YSkVsQWDp0tdYRA$)

**1. MCMC: points to be corrected**

**1) Gamma parameters**

* Initial gamma parameters were fit to RTM rather its generator.
* Initial gamma parameters were fit to the data including diagonal (i, i) components, which should be excluded.
* For alpha estimation, alphas were originally set at qij/beta, but it may make more sense to set at **q\_ij\*beta\_i**, as mean of the gamma distribution is alpha/beta.

Also note I saw some corrected alphas negative, so technically floored at zero for the restriction for the gamma distribution (this must make the result of MCMC in the current code unreasonable, so open to do the better way to deal with the problem).

**2) MCMC algorithm**

* My understanding is that initial alpha/beta should keep being used for gamma distribution added by N\_tilde and R\_tilde.

However, the implemented algorithm overwrote them as alpha += N\_tilde, beta += R\_tilde.

I think we should keep using alpha\_0 and beta\_0, adjusted by new N\_tilde and R\_tilde at each step.

* For exterior approach, the implemented code always overwrote Q by gamma distribution with the given Q, so output series of Q was static.

**3) RTM chain**

* Why do you apply K-1 (excluding default) for transition simulation?

This makes mismatch between the diminution of arrange array and the dimension of Boolean array, then always caused exception.

* Through the jump simulation,  when a trial fails to make transition, it was recognized to change to the highest rating, which was the driver you saw huge probability at the left column in the original code.

A picture containing text, screenshot, font, line

Description automatically generated

**2. MCMC: further points I need to review**

Here is the quote from the Inamura paper p.19 for the note for rejection sampling, but I did not fully follow here.

I reviewed the referred paper of Bladt & Sorensen 2025, but I could not find description about the detail of this rejection sampling.

A picture containing text, screenshot, font, number

Description automatically generated

* Sk is said to be drawn by fk, but qk and deltat is given, so fk is a scalar rather distribution function.

In your code Sk is generated over exponential distribution, where I am still suspicious if this is what the paper intended.

* + Also, if this would be the right way, I would check if the numpy exponential distribution is implemented correctly.

If he scale parameter is defined as beta as follows, we need to input scale=1/qk, but the results look off.

A picture containing text, screenshot, font, document

Description automatically generated

* The step needs to continue “until the process reaches an observed rating grade” that I do not fully follow and I think not implemented in the code.
* In the implemented code, we use deltat only when judging Sk, which does not make sense to me. Sk is an estimate of holding time until transition, so if deltat gets granular the Sk should be smaller as well isn’t it?

**3. MCMC: convergence**

MCMC process is expected to include so-called ‘Burn-in period’, where stats have not converged yet.

For reference of convergence, here is an example stats of sum of absolute changes of Q  by step(to see convergence I used **DIRTY** code rather what I made corrections above).

In the figure below, total absolute value gets almost static after 8k iterations for example, so it makes sense to take average AFTER 8k of burn-in period for the final estimate.

To see some kind of stability of convergence, it would make more sense to evaluate stability in statistical ways such as **autocorrelation**.

A picture containing plot, line, screenshot

Description automatically generated

Best,

Tak

**--------------------------------------------------**

**Takuya Otani**

Vice President, Counterparty Portfolio Management Desk

Fixed Income Division

Mizuho Securities USA LLC

Mizuho Americas

1271 Avenue of the Americas, New York, NY 10020

T: +1 646-949-9817

[takuya.otani@mizuhogroup.com](mailto:takuya.otani@mizuhogroup.com)

[mizuhoamericas.com](https://urldefense.proofpoint.com/v2/url?u=http-3A__www.mizuhoamericas.com_&d=DwMFAg&c=-ib6oBKivNBwAXVVyaiAu6NBDGPkr1Jtp3dZ-MgcLP8&r=wC5PMKOH1dlcncSAQDzDs-sCkVK_ukFOEBTKkf1LsUM&m=VDM_W47i935Q7iJMk_R_HzsHrfeOwBeqKEYY5xEPXaA&s=-UyvweCD8J5-gXmF1aS7QjJcaR9L4ka1ANw3aw1Bq78&e=) | [Twitter](https://urldefense.proofpoint.com/v2/url?u=https-3A__twitter.com_mizuhoamericas&d=DwMFAg&c=-ib6oBKivNBwAXVVyaiAu6NBDGPkr1Jtp3dZ-MgcLP8&r=wC5PMKOH1dlcncSAQDzDs-sCkVK_ukFOEBTKkf1LsUM&m=VDM_W47i935Q7iJMk_R_HzsHrfeOwBeqKEYY5xEPXaA&s=LGaTw7zGiTOscwwd5FCqryOyu90q0SNiN0z6lEOtC60&e=) | [LinkedIn](https://urldefense.proofpoint.com/v2/url?u=https-3A__www.linkedin.com_company_mizuho-2Damericas&d=DwMFAg&c=-ib6oBKivNBwAXVVyaiAu6NBDGPkr1Jtp3dZ-MgcLP8&r=wC5PMKOH1dlcncSAQDzDs-sCkVK_ukFOEBTKkf1LsUM&m=VDM_W47i935Q7iJMk_R_HzsHrfeOwBeqKEYY5xEPXaA&s=pJelpy03_LIHVIkT-TD_-J5r04rR9lts2v0mkMtBPkU&e=) | [YouTube](https://urldefense.proofpoint.com/v2/url?u=https-3A__www.youtube.com_c_MizuhoAmericas&d=DwMFAg&c=-ib6oBKivNBwAXVVyaiAu6NBDGPkr1Jtp3dZ-MgcLP8&r=wC5PMKOH1dlcncSAQDzDs-sCkVK_ukFOEBTKkf1LsUM&m=VDM_W47i935Q7iJMk_R_HzsHrfeOwBeqKEYY5xEPXaA&s=eEUAhAfsCvYu1T8AKWoCUxEkkK3-176wKY6EXXz-tVE&e=)