In NumerCon 2022, BENCHMARK MODELS was announced as a new way that Numerai could help participants to facilitate their research process.

I was watching this announcement with other Japanese participants and we were all very excited by this idea that Numerai puts together all the available ideas and code and stuff in one place and even backfills TC for those models.

[

image

1100×756 59.2 KB

](https://forum.numer.ai/uploads/default/original/2X/a/aa72dceadbc4346f4ad9bb07a5fe0348e7d5fcb6.jpeg)

I believe that this BENCHMARK MODELS is something that we are all looking forward to. However, I wonder if this new feature comes relatively soon or not. I can imagine that making this happen would take serious amount of time—not just because Numerai data team should have a lot of things to do already but making this new feature available itself would not be easy.

To make this amazing idea happen sooner, how about incentivizing the community to contribute to it?

Basically what I think of is:

- Numerai first makes a github repository to collect code for modeling in a specified format (e.g., python environment, library requirement etc)

- Numerai then makes a list of benchmark models to be implemented in the issue

- Participants can contribute by working on the TODO list and get rewarded by NMR

- Participants who are willing to put their models to be part of the benchmark models can also simply format their code and make a pull request (if accepted, rewarded)

In this way, Numerai does not have to implement everything on their own and can focus on calculating backtest scores (e.g., TC) and deploying them for the live era. I hope then the release of the BENCHMARK MODELS would be earlier.

I have a study group with other motivated Japanese Numerai participants (almost weekly), and there we make kind of our own benchmark models to facilitate our research process. I guess others do something similar too: making some benchmark/baseline models and submit without stakes for AB testing.

So if incentivized correctly, participants would be happy to contribute as they wouldn’t have to waste their model slots just for AB testing.

That’s essentially what we discussed in our study group. What do you think?