sneaky:

1. for every live era choose the model with the highest score and its predictions are now labels of the era

What is your plan for calculating a score for the live eras without targets? Do you mean instead "choose the model corresponding to the nearest era cluster for the given live era, predict with it, and consider those predictions as the new targets for that live era"?

If I'm right about that being your plan, let's assume your models get single-digit correlation percentages like most seem to do around here. Then you're creating targets for the live era that are honestly not very accurate. If you train models on those eras I wouldn't be very confident about any results you see.

The Numerai folks did say they were going to release target info for live eras in March. Maybe the best move would be to wait and train against the true targets when they appear? That's my plan.