

Some comments have been made regarding the possible causes of the models' poor performance in the last year. The theory that the change in the payment system using TC has promoted stacking to models that have deviated from CORR as the target in favor of diversity has been suggested as one of them.

I would like to point out another possible cause, which is a significant change in market timing, among other things causing an increase in interest rates. Here, I would like to reflect on the features we use in the models. Since all features are normalized for each era, it is impossible for a model to know if a company is, for example, expensive in terms of P/E.

Its value would be, for example, 4 (max value) in a feature that captures this data. But if three years later, the P/E of the top 20% most expensive companies is tripled, the value of the feature will still be 4 for them. Intra-era normalization is good for many things, but it specifically hinders the ability to see the market scenario.

I am not saying that a new feature not normalized by era would solve this, as with only about 2000 eras, any model could easily overfit certain periods of such a feature. The problem is complex, but I believe it is crucial for models to generalize across any market timing.

How do you think Numerai should do feature engineering so some features can give the market context to the models? Introducing not normalized by era features?