

V5 "Atlas" Data is here. Don't worry, nothing is changing right now. You cannot submit on v5 yet. Live data will be released in September.

This release expands and improves the universe we use to craft our dataset, thus evolving our features and targets to be more predictive. Because features and targets significantly changed, you will need to retrain your models as any model trained on v4.x data will soon be obsolete. The new data offers more diversification in rows in each era, higher correlation, and lower variance. Our research shows that most models will simply have higher CORR when retrained with this data.

For example, when we compare the performance of our [hello\\_numerai](#) tutorial notebook running on v4.3 vs v5 we can see a drastic improvement in performance with no change to the underlying model:

CORR

v4.3

v5.0

mean

0.0245

0.0293

std

0.021

0.021

sharpe

1.1379

1.335

max drawdown

0.0637

0.0329

Furthermore, since our hedge fund is now trading a different universe than v4, we must have predictions on all of these new stocks which aren't in the old data version. We had hoped that old models would be able to predict the new universe just as well, but instead our internal research found that there is a steep decline in performance when v43 models attempt to predict on v5.

It's clear that the new universe forces a breaking change, but we are aiming to give everyone plenty of time to re-train their models on the new dataset. For now, we have only released the v5 training and validation datasets so you can start exploring. Here is what is currently available in the API:

"v5.0/features.json", "v5.0/train.parquet", "v5.0/train\_benchmark\_models.parquet", "v5.0/validation.parquet",  
"v5.0/validation\_benchmark\_models.parquet", "v5.0/validation\_example\_preds.csv",  
"v5.0/validation\_example\_preds.parquet"

Here is the current roadmap for this data release:

July 19

- Support v5 data in diagnostics

September 13

- Release v5 live data:

"v5.0/live.parquet" "v5.0/live\_benchmark\_models.parquet" "v5.0/live\_example\_preds.csv"  
"v5.0/live\_example\_preds.parquet"

- Support v5 data in Model Uploads
- Start accepting v5 submissions, but v5 submissions will not be scored

September 17

- Update example scripts and tutorial notebooks
- Start submitting benchmark models to website

September 27

- Change all scores and payouts to v5
- Stop supporting v4 data
- Stop accepting v4 submissions

## FAQ

What's different about v5?

The universe (the list of stocks we are willing to trade) has changed. This means features and targets have also changed due to the nature of ranked residual returns. Because of this, we have changed the feature names, but the features are still in the same order, so you should be able to map the features from v4 to v5 easily. In our research, we have found that models trained on v5 data are significantly better at predicting residual returns than v4 models.

Can I still train and submit on v4 data?

Yes. All v4 data (v4 through v4.3) is still supported for the next 2 months. You can train and predict on all v4.x until September 27. After this date, all v4.x data will no longer be updated and you will no longer be able to use it for live prediction.

What will happen to my models?

Nothing for now. No breaking changes will happen until September 27. On September 27, if your model is still submitting on v4, it will fail to submit.

Where is the live data?

It's coming in September. Live data is not publicly available because you will not be able to submit predictions on v5 until September.

What is v5? Can I use v4 and v5 together?

V5 changes the universe of stocks used to craft the dataset, thus changing IDs, features, and targets. You will not be able to use v4 and v5 in tandem (hence the major version number change).

Where are all of the targets?

Upon release, the datasets only have 4 targets: cyrus\_20, cyrus\_60, teager\_20, and teager\_60. After we some more research and development, we will include the following targets in the v5 datasets:

- Ralph
- Victor
- Tyler
- Waldo
- Alpha
- Bravo
- Charlie
- Delta
- Echo
- Jeremy (instead of Jerome)
- Teager
- Cyrus
- Caroline

- Sam
- Xerxes