
# The Winton Stock Market Challenge

Join a multi-disciplinary team of research scientists

$50,000 · 1,300 teams · 2 years ago

**Tsakalis Kostas**
15th place

# Solution Sharing

posted in The Winton Stock Market Challenge 2 years ago

12

My model which had relatively consistent performance through leaderboard, CV and public is simple hierarchical bayesian linear regression with student-t noise using only the day returns. Essentially I only predicted Ret +1 and Ret+2; the minute returns are set to zero.

The Stan model is this:

```
 stan_model = """
 data {
 int<lower=0> N;
 vector[N] t; // Ret Plus One / Ret Plus two
 vector[N] x_1; //Ret Minus Two
 vector[N] x_2; //Ret Minus + First 120 mins
 real df; // Degrees of freedom (2.6)
}
 parameters {
 real w_1_0;
 real w_1_1;
 real w_1_2;
 real<lower=0> alpha;
 real<lower=0> beta;
 }
 transformed parameters {
 real<lower=0> sigma_w;
 real<lower=0> sigma_t;
 sigma_w <- sqrt(alpha);
 sigma_t <- sqrt(beta);
 }
 model {
 w_1_0 ~ normal(0,sigma_w);
 w_1_1 ~ normal(0,sigma_w);
 w_1_2 ~ normal(0,sigma_w);
 alpha ~ inv_gamma(1E-2, 1E-4);
 beta ~ inv_gamma(.3, .0001);
 t ~ student_t(df, w_1_0+w_1_1*x_1+w_1_1*x_2, sigma_t);
 } """
```

Options

Comments (46)

Sort by Hotness

Please sign in to leave a comment.

Μαριος Μιχαηλι... • (49th in this Competition) • 2 years ago • Options

9

> ***Blacksou wrote***
>
> Blacksou, could you please clarify this thought? Why MAE resulted in many competitors scored above zero, and why MSE wouldn't?

My 2 cent.

ALWAYS find a model that optimizes the metric you are being tested on.

(Unless it is not possible to find one) , It can make all the difference.

A linear regression that optimizes for RMSE (the classic linear regression implementation - all features, no exclusions) scored worse (higher) than the zero benchmark for me.

A Linear regression that optimizes for MAE (with exact same parameters ) scored conveniently better than the zero benchmark ( e.gt my current place 49th)

Μαριος Μιχαηλι... • (49th in this Competition) • 2 years ago • Options

9

I used SGD Linear regression that optimizes for MAE. I used all features and predicted all columns.

**Humberto Bran...** • (2nd in this Competition) • 2 years ago • Options

3

Hi chenhan zhang,

I hope you are fine.

It was not a data leak. It is important to say. It was about an important and public information in the dataset that we could use it without to know about returns. People from Winton explained a little about the main feature. It was the Feature 7. The mistake was in the data split and they changed the dataset for this reason. After this, everything was ok.

In my opinion, Feature 7 represented a mapping between days and sectors. For example, 03/04/2012 x Gas and Oil -> 678965 F7 value. It is only an example. It does not represent a right value.

I found it using my excel and a person correlation between stocks (using intraday series). With this information, you can get many points if you understand about stock exchanges.

Many people knew that Feature 7 was important. However, they do not know why and how to use it. I saw some people in the forum recommending to drop F7. It was a big mistake.

Other people created wrong groups to validate their models. You cannot use a "future" information to validate your training. And F7 could help you to spilt data in the right way. For this reason, people were talking about very good CV results but bad results in the public LB. Maybe the majority saw it in the home tests. I saw it in the first days. And I realized that something was wrong and I started a process to find the "secret".

There were many ways to work with this information. I tried to assume high risks to get high returns. The best that I find was change few points in D+1. The equation to evaluate quality of solution was not good, in my opinion, and I explored it as well. I did thousands of tests to explore this kind of "mistake".

A wrong equation + a good information in the dataset (F7) made me to change only 74 points.

It was the "secret". Many people said bullshits about my strategy in the forum. Anyway, they have failed. I only studied a lot the dataset and the function.

There were many ways to work with F7 information. I only chose one of them. My strategy was to predict volatility, controlling risks. It was not about returns, that is impossible, in my opinion.

Best regards...

**Mike Kim** • (25th in this Competition) • 2 years ago • Options

6

I just bagged XGBoost for PlusOne and PlusTwo (and a few others). I used 0.1*xgboostpreds + 0.9*colmedian. I used raw and diff features. I selected features via KS test between train column and test column. I joined late (2 weeks ago) and have no idea what the leak was. So I decided not to risk any crazy train+test merged approaches.

xgb2f39bagPart.R (1.95 KB)

Willie Liao • (52nd in this Competition) • 2 years ago • Options

3

One of my submissions is a complicated ensemble of xgboost and regularized regression. It worked great in cv but failed on both the public and private LBs.

The other submission is just a "weighted median" of the d+1 and d+2 returns. It's a completely useless model in real life but good enough for #52 on LB.

```
library(data.table)
ds <- fread('train.csv', select=c('Ret_PlusOne', 'Ret_PlusTwo', 'Weight_Daily'))
sub <- fread('sample.csv')
sub[, Predicted:=0.0]
sub[seq(61, .N, 62), Predicted:=ds[,
median(Ret_PlusOne*Weight_Daily)/median(Weight_Daily)]]
sub[seq(62, .N, 62), Predicted:=ds[,
median(Ret_PlusTwo*Weight_Daily)/median(Weight_Daily)]]
write.csv(sub, 'med_day.csv', row.names=F, quote=F)
```

chenhan zhang • (63rd in this Competition) • 2 years ago • Options

1

wow, really really thank you for your answers. best regards

> ***Humberto Brandã wrote***
>
> Hi chenhan zhang,
>
> I hope you are fine.
>
> It was not a data leak. It is important to say. It was about an important and public information in the dataset that we could use it without to know about returns. People from Winton explained a little about the main feature. It was the Feature 7. The mistake was in the data split and they changed the dataset for this reason. After this, everything was ok.
>
> In my opinion, Feature 7 represented a mapping between days and sectors. For example, 03/04/2012 x Gas and Oil -> 678965 F7 value. It is only an example. It does not represent a right value.
>
> I found it using my excel and a person correlation between stocks (using intraday series). With this information, you can get many points if you understand about stock exchanges.
>
> Many people knew that Feature 7 was important. However, they do not know why and how to use it. I saw some people in the forum recommending to drop F7. It was a big mistake.
>
> Other people created wrong groups to validate their models. You cannot use a "future" information to validate your training. And F7 could help you to spilt data in the right way. For this reason, people were talking about very good CV results but bad results in the public LB. Maybe the majority saw it in the home tests. I saw it in the first days. And I realized that something was wrong and I started a process to find the "secret".
>
> There were many ways to work with this information. I tried to assume high risks to get high returns. The best that I find was change few points in D+1. The equation to evaluate quality of solution was not good, in my opinion, and I explored it as well. I did thousands of tests to explore this kind of "mistake".
>
> A wrong equation + a good information in the dataset (F7) made me to change only 74 points.
>
> It was the "secret". Many people said bullshits about my strategy in the forum. Anyway, they have failed. I only studied a lot the dataset and the function.
>
> There were many ways to work with F7 information. I only chose one of them. My strategy was to predict volatility, controlling risks. It was not about returns, that is impossible, in my opinion.
>
> Best regards...

**Alexey Golyshev** • (295th in this Competition) • 2 years ago • Options

3

feature_7 - stock ID, we can group the same stocks by ID. We have 824 stocks in the train and 2592 stocks in the test.

feature_5 - month (Jan. - Oct.)

I have trading experience. It's very hard to predict the next 1 minute candle without an additional information: volume, price levels. Example picture in attachment (60 min. chart, 5 min., 1 min.): the next 1 min. candle (Ret_121) will be red, on the real chart the last green candle can have a big volume and high upper shadow. But in the competition's data we have no information about max and min return, only return at the end of a minute.

Congrats to the winners. And thank Winton, in any case it was very interesting competition.

109.jpg (46.18 KB)

**chenhan zhang** • (63rd in this Competition) • 2 years ago • Options

1

Can you share us the process that you found the data leakage at the beginning of the competition? I am always want to know how to do that. thank you.

> ***Humberto Brandã wrote***
>
> I only changed 74 points in the D+1... All the rest I sent as 0.0...
>
> 2nd place in this competition...
>
> I would like to know if Bill performed a similar strategy...

**pmarelas** • (690th in this Competition) • 2 years ago • Options

4

> ***Alexey Golyshev wrote***
>
> feature_7 - stock ID, we can group the same stocks by ID. We have 824 stocks in the train and 2592 stocks in the test.
>
> feature_5 - month (Jan. - Oct.)

I found a pattern between Minus+Plus+Feature_5 but I didn't understand why until now. Thanks

download (1).png (101.45 KB)

Humberto Bran... • (2nd in this Competition) • 2 years ago • Options

1

I only changed 74 points in the D+1... All the rest I sent as 0.0...

2nd place in this competition...

I would like to know if Bill performed a similar strategy...

Tamas • (339th in this Competition) • 2 years ago • Options

1

The reason why this makes big difference in this challenge is that daily stock returns are positively skewed since positive information spreading on the market fast while companies tend to hold back bad news. This is well documented phenomena in the literature thus predicting negative median was a safe bet. If you predict the median of the first two days of the new data (-0.0005) for Day1 and Day2 you end up in the top 30.

Blacksou • (23rd in this Competition) • 2 years ago • Options

1

> *rakhlin wrote*
>
> > *Blacksou wrote*
> >
> > In this competition it was important to notice that we had to minimize MAE not MSE. This is why most traditional methods or "black box" failed. I believe this is the main explanation why so many persons scored more than the zero benchmark.
>
> Blacksou, could you please clarify this thought? Why MAE resulted in many competitors scored above zero, and why MSE wouldn't?

Most regression tools use MSE as a default error metric because it makes all computations easier and usually guarantees a unique solution. The issue is that it gives too much weight to outliers. For most problems the MSE solution can be a good proxy for the MAE solution but here that was not the case as the data was mainly noise.

One quote that I like is that you don't get paid in squared dollars so why should you use MSE?

all_random • (13th in this Competition) • 2 years ago • Options

1

You could take the example of mean and median. With MSE, the mean provides the best estimation of the target values. On the other hand, the median comes in the case of MAE. And the twos do not always have the same value.

As you could see in the PB, there are a lot of submits using the median for DAY_PLUS_ONE, and DAY_PLUS_TWO, which are better than the 0-submit.

**rakhlin** • 2 years ago • Options

1

> ***Blacksou wrote***
>
> In this competition it was important to notice that we had to minimize MAE not MSE. This is why most traditional methods or "black box" failed. I believe this is the main explanation why so many persons scored more than the zero benchmark.

Blacksou, could you please clarify this thought? Why MAE resulted in many competitors scored above zero, and why MSE wouldn't?

**Marc** • (637th in this Competition) • 2 years ago • Options

1

here's what I tried - I think that I really messed up the model fitting :D but anyway - it was quite some fun and I enjoyed learning something new - my thanks to the organizers.

winton.html (816.8 KB)

**Blacksou** • (23rd in this Competition) • 2 years ago • Options

1

In this competition it was important to notice that we had to minimize MAE not MSE. This is why most traditional methods or "black box" failed. I believe this is the main explanation why so many persons scored more than the zero benchmark.

My one and only submission was extremely simple with only 8 parameters and had very stable results both on the train and test datasets.

**alexeymosc** • (612th in this Competition) • 2 years ago • Options

1

My approach that yielded the result slightly below the zero benchmark in both public and private consisted of making feature selection suing a rather complex mutual information-based fitness function, and using discretized selected variables to build a set of rather simple rules involving a median measure as the expectation of the predicted value.

The reason I pursued this way - instead of trying to teach a complex nonlinear machine - is because in practice I trade currencies and learning simple rules for making trade decisions is valuable when one writes a code of a trading robot instead of referring to a 'black-box' external trained model. This gives one ability to understand the market behavior too. However at this competition the challenge appeared to be quite hard for me, although I get solidly better-than-zero results on single instruments of FOREX.

**vwood** • (9th in this Competition) • 2 years ago • Options

1

I figured that Feature 7 was the time period the returns were from. Clearly each value of Feature 7 was correlated, and no values were shared between train and test (as you would expect if they were different times).

But the give-away was the effect if you did CV such that no values of Feature 7 were shared between folds. This killed most increases. Now it could still be the case that Feature 7 was something else, but I figured it was that and started to disregard most local CV results. I figured at least the public LB came from the same distributions and could be trusted.

**innovaitor** • (207th in this Competition) • 2 years ago • Options

2

I wrote up my contribution and thought process, using 5fold CV gbm to model out of sample days (with fully reproducible R code attached).

http://intelligenttradingtech.blogspot.com/

If I am in any violation of rules, please let me know asap. thanks.

**cydonia** • (29th in this Competition) • 2 years ago • Options

2

My approaches are simple.

I did predict just one target variable, Ret_PlusOne, leaving all of Ret_121 - Ret_180 and Ret_PlusTwo zero.

Based on my guess that positions in the Private leaderboard would shake up and down, I tried to make the complexity of models as low as possible, hoping to keep my position stable between Leaderboards.

So my final solution is a blend of two models:

- ensembles of 30 predictions of ExtraTreesRegressor(scikit-learn) using all features(I did not try XGBoost)
- c(constant) * "sum of Ret_2 - Ret_120"

An approach that did not work:

Weight_Intraday, Weight_Daily are correlated with Feature_13 and Feature_20. So I made new daily and intraday weights by taking medians of the Weight variables grouped by the two Features. Then I built recurrent neural network models using the new weights * Ret_x features. These models beated zero benchmark on Public LB a little bit, but were worse than the simple linear model above. So I discarded them.

Hope this helps.

yilisg • (18th in this Competition) • 2 years ago • Options

2

After trying many models from linear regression to neural net, I gave up trying to predict intraday returns. In my case, nothing beats a zero benchmark for the intraday columns (not even median intraday returns), despite having an extremely good local CV of the MAE. Everything falls apart for me on the public LB (and presumably the private LB too).

I've also tried a few other ideas such as momentum trading (it's Winton after all, but the data seems to exhibit a weak mean-reversion tendency), volatility clustering and correlation binning -- again, some worked great locally but was way off in the public LB (they *might* work in the private LB). Again, simple noise / outliers removal technique helped in my local CV but wasn't in my final solution because it degrades my public LB score... (At one point I was totally confused if I should trust my CV or the LB).

All the above was done with just the price/return data and I experimented separately with the features using xgboost and nnet -- long story short, they worked *great* with Feature 7 included (even for intraday returns) but not having much gains with Feature 7 removed. From the charts attached, it was clear that Feature 7 is very different in the training and test set (contrast the pictures with another random feature) -- I wonder what Feature 7 could be...

Therefore, my final submission ignores all features and uses only the T-2, T-1 and known T intra-day returns. I computed a weighted average of these known returns (tried both scaling down known T since it is only half a day, and scaling up T when using an exponentially-weighted idea by giving more weights to most recent observations), binned them systematically with bins 2..30, and found that somewhere between 7-9 bins worked best. My final submission is a simple average of the median predictions of bin 7, bin 8 and bin 9 for Ret+1 and Ret+2.

Maybe the trick to this competition is knowing what you cannot predict and stay simple. To be honest, I was a little annoyed by the discrepancies in the features between the training and test data, and I hope this was a honest mistake (and not by design to trick the contestants).

Winton_Fet7_1.jpg (245.73 KB)
Winton_Fet7_2.jpg (49.57 KB)
Winton_Fet17_1.jpg (153.93 KB)
Winton_Fet17_2.jpg (42.52 KB)
Winton-Feature-Importance.JPG (62.01 KB)

**Mike Kim** • (25th in this Competition) • 2 years ago • Options

0

> *M* ***wrote***
>
> > ***Blacksou wrote***
> >
> > Blacksou, could you please clarify this thought? Why MAE resulted in many competitors scored above zero, and why MSE wouldn't?
>
> My 2 cent.
>
> ALWAYS find a model that optimizes the metric you are being tested on.
>
> (Unless it is not possible to find one) , It can make all the difference.
>
> A linear regression that optimizes for RMSE (the classic linear regression implementation - all features, no exclusions) scored worse (higher) than the zero benchmark for me.
>
> A Linear regression that optimizes for MAE (with exact same parameters ) scored conveniently better than the zero benchmark ( e.gt my current place 49th)

I actually somewhat disagree with this. First, I did (and usually always do) find a model that minimizes the target in question (wMAE). In fact I found a model that minimizes weighted MAE. I found two models. One being R's gbm with Laplace objective with weights specified by the training data. I also tried nlm / minimization by R with weights given by the training data. Neither worked particularly well. A single value for each target via nlm will get you around top 60-70 depending on your optimization tolerances and starting position. I also tried various robust models such as R's rlm, etc. but all of these methods were sub optimal.

My opinion based upon my placements (and reading forums posts) is that you should find whatever works. This means trying a variety of objectives and transformations. It means you don't always have to use wMAE, or RMSE, or what have you. Either through human error, or other factors (overfitting) direct minimization of the objective (which you can do with XG given you have both the gradient and hessian) is not always optimal. It won't get you the best score. I can't tell you exactly why, but I can say there are many cases where it happens (e.g. see Yr's Crowdflower (I wrote code based upon his to directly max Kappa via XG), Winton me, and many others where ensembles of different objectives > single). Again this isn't a single observation, but many. As a result I admit there are many confounding factors here which make reality != theory.

There are so many theoretical things I learned in the class room many years ago. They mostly don't hold up in Kaggle. Perhaps the influences of how to attack problems sort of hold up, but mostly no. You, alone , must figure out what works. Often it doesn't match up exactly with what some professor told you in some class room long ago.

Note if I'm incorrect and you can show me the data with regards to R's nlm, gbm (Laplace with weights), rlm, etc. Please let me know and show me the code and output. It is possible to make a mistake(s). Thanks.

tmp.R (644 B)

**Nerotulip** • (752nd in this Competition) • 2 years ago • Options

0

Most likely, feature 7 was the date: it was a categorical variable, had different values between train and test set, and returns for a given value of feature 7 were highly correlated.

If you did a standard random split of the train set, you were giving your model information about the future, so of course you got real good CV scores, but they did not hold on the LB. You had to pick train and validation sets with different values for feature 7.

**javifalces** • (779th in this Competition) • 2 years ago • Options

0

I developed a too much complicated model that didnt beat the zero benchmark....

Just to explain ,i combine a lot of arquitectures(MultilayerNN , convolutional NN, bayesian with a Randomforest feature selection,and sVr) and i applied this to each of the columns intraDay and Daily... i choose the final testing arquitecture based on results on the testing set(subsampling training data). I have done this spllitting all this mess by a feature selection(using only the integers features(0-10)), So in my set i have around 10*62=620 arquitectures.

I dont know if my idea is very complicated, or very simple , it was my first challenge and i couldnt beat zero benchmark in any of the submissions Any help will be grateful! because i would like to use it in real trading in the future , or something similar

**Mendrika Rama...** • (3rd in this Competition) • 2 years ago • Options

-2

**lux** • (332nd in this Competition) • 2 years ago • Options

0

I only predict 10% of plusOne data, and I checked my solution on real market . I shouldn't remove high risk points :(

**stanley** • (770th in this Competition) • 2 years ago • Options

0

I had trained models with unique Feature_7s from the folds and got good results in CV but still couldn't generalize to the test data.

**btmd** • (184th in this Competition) • 2 years ago • Options

0

Since I wanted to leave the puddle of oneliners, I used the power of zero prediction as an ensemble together with the l1 regression.

**Tamas** • (339th in this Competition) • 2 years ago • Options

0

My svm, linear and quantile regressions on last day return and realized minute volatility explained around 0.18% of Day1, 0.04% of Day2 and 0.0002% of the minute return variations on the public set. The fit was about 30-50% better on the local set. The minute and Day2 prediction turned out to be robust on the private set while the Day1 bounce back and risk-return relationship has changed quite a bit degrading my score behind some naive benchmarks. What I regret that I haven't created my own surely independent validation set from recent stock prices but trusted the persistence of these relationships. I know the rules prohibit the use of external data but we have learned that having fun has priority over following the rules.

Congrats to the winners and thanks to the hosts for this exciting challenge

**ignl** • (93rd in this Competition) • 2 years ago • Options

0

I tried various models like others here just last day I realized that probably arima & co models could be used to predict minute data however didn't have time to try it (and I don't have experience with them)...

**m** • (638th in this Competition) • 2 years ago • Options

0

I thought SVM was the way to go, however my parameters obviously were wrong.

I did also a FFT analyze, which gave really different results, but classifying them was much too complex. Had no chance to finish it.

Multiple regression seemed to work on train data for some parts, but failed on test data.

tried out indicators MACD, Momentum, CCI and others (also combined and using overbought/oversold signals) but that result was even more worse than using simply the trend channel slope from 2nd hour for prediction of RetPlusOne.

**Piotr Golach** • (741st in this Competition) • 2 years ago • Options

0

> ***maralski wrote***
>
> > ***Alexey Golyshev wrote***
> >
> > feature_7 - stock ID, we can group the same stocks by ID. We have 824 stocks in the train and 2592 stocks in the test.
> >
> > feature_5 - month (Jan. - Oct.)
>
> I found a pattern between Minus+Plus+Feature_5 but I didn't understand why until now. Thanks

Hi, how have you generated this kind of plot, its looks really interesting ? Please share your code of give name of function used to generate it.

> ***xvnv wrote***
>
> > ***maralski wrote***
> >
> > > ***Alexey Golyshev wrote***
> > >
> > > feature_7 - stock ID, we can group the same stocks by ID. We have 824 stocks in the train and 2592 stocks in the test.
> > >
> > > feature_5 - month (Jan. - Oct.)
> >
> > I found a pattern between Minus+Plus+Feature_5 but I didn't understand why until now. Thanks
>
> Hi, how have you generated this kind of plot, its looks really interesting ? Please share your code of give name of function used to generate it.

Use seaborn package.

sns.lmplot(data=train_df, x='Ret_MinusOne', y='Ret_PlusTwo', hue='Feature_5', palette=sns.cubehelix_palette(10, start=2, rot=0, dark=0, light=.95, reverse=True), x_jitter=.15, size=10, hue_order=range(1,11,1))

Christopher Hef... • (350th in this Competition) • 2 years ago • Options

0

alexeymosc • (612th in this Competition) • 2 years ago • Options

0

> ***rakhlin wrote***
>
> Blacksou, could you please clarify this thought? Why MAE resulted in many competitors scored above zero, and why MSE wouldn't?

- I would like to know Blacksow's opinion on the role of error metric.

**Marc** • (637th in this Competition) • 2 years ago • Options

0

@ innovaitor thanks for the post :)

> I believe two big factors were differences in train and test datasets, as well as a much larger hold out dataset that was different enough that the true results were much worse than 5 fold CV would predict.

I have the feeling that the differences between the train and the test data were causing the standard CV approach to fail. After removing the outliers with awkward correlation patterns between the past return (see attachment in my previous post) and splitting the data based on some of the features, I could continuously increase the performance on the train data. Even leaving more than 30% of the train data out (random sampling), the performance based on the CV was - depending on the method - really good (far better than zeroes). So - not sure, but I guess that it is either the way one splits the data for CV which did not work in this case, or a subtle difference between the train and the test data which I missed :)

and then the metrics - thanks for the hints

**Marc** • (637th in this Competition) • 2 years ago • Options

0

hm - tempted to try it - looks like one can still submit sth - but should also get some work done ^^

**Mendrika Rama...** • (3rd in this Competition) • 2 years ago • Options

0

> *Nerotulip wrote*
>
> Most likely, feature 7 was the date: it was a categorical variable, had different values between train and test set, and returns for a given value of feature 7 were highly correlated.
>
> If you did a standard random split of the train set, you were giving your model information about the future, so of course you got real good CV scores, but they did not hold on the LB. You had to pick train and validation sets with different values for feature 7.

I believed so as well. I believe the train/test set was created using two-stage random sampling without replacement based on feature_7, but chronological order was not taken into account. If that's the case look-ahead bias is built into this competition's evaluation method. Since train set contains information from the future, and test set contains information from the past. My conclusion, local CV was useless in this competition.

**alexeymosc** • (612th in this Competition) • 2 years ago • Options

0

> ***Mendrika Ramarlina wrote***
>
> If that's the case look-ahead bias is built into this competition's evaluation method. Since train set contains information from the future, and test set contains information from the past. My conclusion, local CV was useless in this competition.

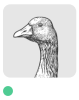
In a real-life automated trading application a researcher is always forced to deal with the "look-ahead bias". One has to construct a model on some past train data, and evaluate it on another block of past data which does not overlap with the train data because when real money will be put into trading, the model will deal with the block of data (the future) that does not overlap with the past.

For people who have never tried stocks this sort of bias seems painful, but in the stock domain it is natural and unavoidable.

**Aindriú** • (241st in this Competition) • 2 years ago • Options

0

I will briefly share what I did. First, I wondered why we are working on returns (which look like random noise) and not on prices as the latter have some sort of predictable structured trends going up or down. Some papers in the literature used models to predict prices. So in a big chunk of my work (before data renovation), I converted returns to prices, assuming that what we have is (Min2-Min1)/Min1 (now I think what we have is log(M2/M1) ) and setting the initial value to, say, 100, I managed to "recover" price trends which demeaned were well modelled by ARIMA/GARCH models (according to the training error). In fact it worked better than zero returns (which in prices would mean the fixed price for all 60 minutes). At least for a few first minutes. Then I clustered the initial space based on features, and search for a good fit for each cluster minute data. Using the model parameters for each cluster I used the testing minute data as initialisation for the model to predict future values. Then, convert prices back to returns with the original formula. Did not work on leaderboard AT ALL. Then played with model residuals, trying to incorporate those to correct predictions, especially later predictions. NOPE. Well, everything done on prices did not work. Then I read a few papers from Humberto and figured out that high frequency trading is a completely different beast :). Second and higher order stats (volatility) are of importance here. Later better than never. With the new data, I gave up on minutes and did common silly things on Day Returns (weighted stats, clusters). Yep, based on my private score there are a few subms that would have put me in the top 50-60, their public score was not good enough to be selected by me :). Interesting experience thanks to Winton and Kaggle.



**Camzzz** • 2 years ago • Options

0

Hi,

I didn't get time to actually submit any models or finish my work, so bear that in mind when reading - my thoughts are purely hypothesis! Thought it would share some of the things I was planning to do / comments I had. The use cases I've suggested are simple to help illustrate the points, obviously I suspect something clever would be needed to get a top score!

First point is regarding the minute by minute data. For any given minute do we really think that returns

early in the day will help us predict returns in the afternoon? That seems pretty unlikely to me so focus on using the recent few minutes, and possibly a few summary statistics from the morning. Does it make sense that the previous two returns would not be relevant but, the two before that would be relevant? Probably no so restrict yourself to trying to use all of the previous N returns in your model.

This leaves us with a problem, how can we predict the last return in the day, since we dont know the returns before that? My thought was to try and predict the return for the whole afternoon, using the aggregated return for the morning (i.e. combine the minute by minute returns into one) and the previous days returns. We are now in a situation where we can predict a return using the previous returns of comparable horizons, which makes more sense to me. You can then 'spread out' this return evenly over each minute. You could extend this by gradually increasing the window you are combining the returns over. I.e use this method to predict the windows [0,1], [0,2], [0,4] ... [0, 60]. You might use your predictions for the first half of the afternoon (a), and the whole afternoon (b) in two different ways. Either you predict (a) and (b) for the two segments or you predict (a) and (b - a) i.e. guessing what the residual return will be in the second half of the period.

Another thought I had is regarding the nature of the rows. Each on is a different stock or time horizon, and could have reasonably different behavior. Some may be more volatile that other stocks for example. I think it would be worth trying to normalise the stocks by dividing through by the standard deviation. This should make each row a bit more comparable, and now your model can more easily judge how to react to each type of return. Here's a simple example where it might help. Suppose large (2 s.d.) returns of a stock normally mean revert quickly. If you look at all stocks at once you are more likely to find the most volatile stocks, if you normalize first you can easily search for the peak returns, and compare to which ever feature or return to look for a pattern in behavior. You could still include the standard deviation in your model however, as that can be important information (e.g. larger vol stocks might be more likely to have longer trends, or something like that)

In general I'd advise a hypothesis driven approach. I think some fancy machine learning at this sort of data is going to be pretty problematic, I would say its pretty key to understand why your methods aren't working and try to improve them. With a complex model that is made much harder.

People who are asking about whether these results are useful or not, think about it this way. This is a toy problem to allow it to fit into a competition which can be attempted in a reasonable amount of time. We aren't building trading algorithms we are merely doing time series analysis in a slightly unusual format. We dont have nearly the full data available to Winton. The best results would not be profitable if converted to an algorithm, but combining the sort of methods/data used here with many other features and ideas, could make something profitable. Winton are going to be interested in the types of skills people are showing.

Im also seeing high frequency trading mentioned a bit. Be careful if you're looking this up for ideas, high frequency is typically much shorter horizons and depends on much more granular data. Their prediction is also very tied to execution, not the case here as we're purely going for prediction scores. You're probably better off looking up general time series predictions / stats as well as a bit of financial time series context if you're looking for inspiration.

qamly • 2 years ago • Options

0

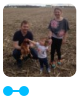
Interesting competition. For my side I didn't sumit new solution since the new datas. But for my previous solution I use stratified cross validation using feature_7 and I build my own features. This approach worked both on training and and test set to improve the 0 bench.

**bingdu** • 2 years ago • Options

0

My, strategy was: 1. see if features are correlated with price movements (they are not) 2. do not use features at all 3. multiple regression (x3, missing values: average interpolation)

I managed to get 1818 the best.



**rerun** • (325th in this Competition) • 2 years ago • Options

0

Cool to see that a few top finishers ignored the intra day data. I spent most of my time looking at intra day series :|

In short, I wanted to see how partial least squares could predict intra day and +1, +2 day returns using only continuous features. I found that I could summarize about 8-9% of the variation in y. My local CV results were better than leaderboard, but this could be explained by the lack of independence between samples (day correlation mentioned by organizers). I also used alternating least squares PCA to fill missing returns. After reading about the median prediction, I scaled my predictions to be near the median. A multiple of 10^-5 for intra day and 10^-4 for +1 +2 day returns was used ( I think ).

-Brian

**vwood** • (9th in this Competition) • 2 years ago • Options

0

I didn't, I just predicted the mean for all of them. Those predictions ended up pretty small.

**Glimmung** • (364th in this Competition) • 2 years ago • Options

0

I used "ret = c + Sum(retI x (aI + bI x feature))" and minimized the MSE. Then put an adjustable "Clamp" on it and adjusted the aI and bI to minimize the MAE. I did this for first minute return and got 364th spot. Then I "cleaned up" my code and lost all. The feature was the one that returned the best MSE.