What machine learning models do we want to consider?

The models they implement

| features to focus on in our extension

Representation: linear model
Objective: Loss: squared loss, regulizer: none
evaluation metric: (adjusted) students t statistic, adjusted R^2
model selection: none

Model 1: $r^e_{t \to t+k} = \beta_0 + \beta_1 X_t(a) + \varepsilon_{t \to t+k}$

with $X_t(a)$: VRP^U, VRP^D, VRP, skewness (one at a time)

Model 2: $r_{t \to t+k} = \beta_0 + \beta_1 VRP^U_t(a) + \beta_2 VRP^D_t(a) + \varepsilon_{t \to t+k}$

(both VRP together)

Model 3: repeat 1 & 2 but instead of taking VRP use only variance

a) with alpha-implied method (Q)
b) with realized method (P)

(h.. aggregation horizon, k.. prediction horizon)

aim: compare upside and downside effects but also contrast contribution of P and Q

result: downside variance has a significant positive effect on equity premium, risk-neutral measure driven the result

robustness tests

- include more variables that are known to predict equity returns
- split sample period according to crisis

Possible criticism and extensions of their models

1. They use an in-sample model and evaluation metric

- estimate an out-of-sample model and use some kind of prediction accuracy evaluation metric instead of R^2, e.g. use RMSE (Root mean squared error)

$$RMSE = \frac{1}{T}\sum_{t=1}^{T}\sqrt{(y_{t+1} - \hat{y}_{t+1|t})}$$

- use model selection, e.g. k-fold cross validation to obtain parameter estimates

2. They don't use regularizes that penalize higher estimates

- apply ridge regression or lasso

Note: In our models I think both would work. Lasso has the advantage of preferring sparsity but in model 1-3 we don't have many parameters anyway. I think lasso would be interesting in the robustness together with the other equity parameters

→ I think both points are not so hard to implement and make sense, just have to deal with time series issue (see below)

Considering well-known ML techniques

1. Neural Networks

⊕ learn what features to use and to which fraction to fit them simultaneously

⊖ Not sure if it serves our purpose. I would interpret the aim of the study more as "what information content can we get from volatility measures", rather than "what is the best model to predict equity returns"

⊖ might give problems with time series → there are some ANNs are "made for" time series analysis e.g. LSTM network

→ would not recommend ANNs

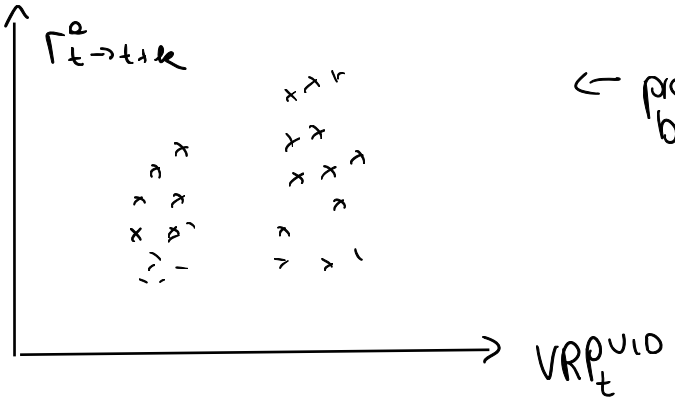
2. Turn towards unsupervised learning: Clustering

⊕ take VRP and do not label it as "upside" and "downside" but use our ML model to determine what cluster fits best when looking at time equity returns

⊖ challenge: we can not simply plot our VRP against equity returns because the clustering happens "before"

What we could do

a) "Double check" whether the VRP^{VIX} clustering is the one our model would find



← probably it will not look like this but just as a sketch

b) look at basic relationship of returns and RV to determine if sign of returns determines the cluster

c) ...

Models: k-means (search for spherical clusters)
Kernel-k-means (adjustment that allows for non-spherical patterns)

→ unsure, might be interesting but could not figure out best way to implement it

Challenges with time series data

Autocorrelation (non-stationarity, not i.i.d)

- accuracy metrics can give a good fit when our model for $t+1$ simply predicts the value from t
- occurs with accuracy measures such as R^2
- possible improvements: time-difference data to see if the model can predict the change in variable
- possible checks to do on time series
  - check autocorrelation
  - stationarity tests
  - see if a model that predicts t for $t+1$ does best ("persistence model")

→ these are problems with ts data and tbh also the authors do not seem to address them too much in their regression models.

For the ridge/lasso we could maybe address these issues with the above mentioned points (check stationarity, differencing, carefully choose evaluation metric)

For the clustering I don't think we assume i.i.d. data (I will double check!)

Questions

1. Do you agree with using ridge/lasso and maybe clustering?
2. Do you think the challenges (and "solution") related to time series make sense?
3. What "variables" (x, y axis) would you propose for clustering (if we do it)