Discussion

Model diagnosis

Having seen the model performance in Table~\ref{tab:res}, it is natural to ask why certain models perform better than others. While logistic regression models are expected to perform less well than more complex models, it is worthwhile to explore why the RF-RW model and the RNN-BiLSTM-focal model do not perform as well as the RF-CV model.

After further investigation of the model performance on the training data, I find that there is an overfitting problem for the RF-RW model. The model has a 99.5\% balanced accuracy rate on the training data but only 75\% on the testing data. The model performs worse on the testing data than on the training data. On the contrary, the RF-CV model does not see this problem.

Why would the RF-RW overfit but not the RF-CV? It might be that the re-weighting method makes the model over-trained for the bubble cases in the training data. The re-weighting method replicates the bubble cases, which could cause the model to fit too much on peculiar bubble patterns in the training data. The patterns are not universal for all bubble cases, and they are likely not presented in the testing data. This is perhaps why the RF-RW model performs worse on the testing data. On the other hand, the RF-CV alters the decision threshold. This method makes the model focuses on the bubble cases generally but not restricting to cases seen in the training data. Without the over-training on the peculiar bubble patterns in the training data, the RF-CV model avoid the overfitting problem.

For the RNN-BiLSTM-focal model, the prediction is no better than a naïve predictor. I have tried various techniques including adjusting hypermeters, e.g., number of neurons, changing layers, e.g., using LSTM rather than BiLSTM, and changing the loss function, e.g., changing the hypermeters of the focal function. None of these techniques works. This is quite a disappointment given the relevant research that endorses the model.

I investigate the probability output of the model, and its distribution is shown in Figure~\ref{fig:rnn}. It shows that the probability output of the model does not vary at all. This suggests that the model barely trains. I think it is likely because the model is too complex for our data size. The training data is insufficient to train this model.

It is worth noting that the above model diagnoses for the RF-RW and the RNN-BiLSTM-focal are largely speculative. Due to the complexities of the models, there does not seem to be an easy way to check my explanations.

Robustness analysis

As discussed in the data section, the quantitative definitions of market crashes and bubbles are arbitrary. I alter the “1\% quantile” and “six months” part of my definition to check the performance of the best performing model (RF-CV). The results are shown in the Table~\ref{tab:rob}.

We see the model performance is robust.

Limitations

It should be noted that a balanced accuracy rate should not be confused with a precision rate. A balance accuracy rate means that the category-averaged percentage of cases being correctly identified. On the other hand, a precision rate means among all cases that are predicted as “bubble”, the percentage of cases that are actual bubbles. The