The Covered Interest Parity condition ("CIP") should dictate prices on the trillion-dollar foreign exchange market. Yet the CIP fails. When CIP fails, cross-currency basis ("basis") appears. As illustrated below, a non-zero basis is an arbitrage opportunity, as all quantities in the basis are known today and a positive payoff can be guaranteed without any cash outflow.

Our project seeks to predict the bases in the Post-Financial Crisis period of the Australian dollar ("AUD") and of the Japanese yen ("JPY"), both relative to the US dollar ("USD").

Employing Regression Trees, Random Forests, and three versions of Regularized Regressions, we predict the bases with reasonable success: our best test MSEs of ~70 for AUD and ~200-270 for JPY compare favorably to the standard deviation of the Post-Crisis bases, which are 11.2 and 23.8 for AUD and JPY, respectively.

We collect 32 data series for each of the three relevant currencies: AUD, JPY, and USD. Our data capture activities in the financial markets, conditions of the economy, the state of international trade, and the stance of economic policies. With the exception of the Economic Policy Uncertainty ("EPU") Index, which is compiled by three leading economists and available at www.policyuncertainty.com, all data are obtained from Bloomberg:

For the AUD basis, the regression tree and the regularized regressions perform similarly. Comparing the MSEs from the two specifications of regularized regressions (linear vs. polynomial feature space), we note that the prediction error in the AUD basis is likely caused by a bias problem, as the MSE decreases with the inclusion of the higher-dimensional features in both the training and the test set. In predicting the JPY basis, we find indications of a variance problem, i.e. overfitting, since the polynomial improves the training error but increases the test error.

Overall, the random forest algorithm delivers the best performance, with the lowest train and test MSE. Note that in the second contiguous test period (2017), no algorithm seems to fit very well. One potential reason is that the cross-currency bases were on an elevated level with low variance, a state that has not been observed in the main data.

We also explore the set of variables with the most predictive power. The figure below shows how often the trees in the random forest (as applied to the Complete AUD sample) split on each of the considered features within the first seven splits. The horizontal line indicates how often a feature would appear if no feature had predictive power.

Given the observed bias issue with the AUD data, we would collect more economic features and use higher order polynomial features in the regularized linear regressions in the future.

To improve the performance on the JPY data, we will expand the set of algorithms employed. Specifically, we will apply the boosting technique, and we will consider training a neural network.

Finally, we want to extend the analysis to a larger set of currency bases.

References

Variable Importance AUD Random Forest

In the chart below, we exemplarily show a fitted tree for the JPY basis. Interestingly, the JPY basis is predicted by mostly US features, in contrast with the prediction of the AUD basis, which relies on both Australian and US features. This raises the question whether we have included the appropriate set of features for JPY basis.

Future Work
Given the observed bias issue with the AUD data, we would collect more economic features and use higher order polynomial features in the regularized linear regressions in the future.

To improve the performance on the JPY data, we will expand the set of algorithms employed. Specifically, we will apply the boosting technique, and we will consider training a neural network.

Finally, we want to extend the analysis to a larger set of currency bases.

Variable Importance JPY Random Forest