Untitled

Karel Räpp, Henry Enno Turu

29 11 2020

1 Methodology

1.1 Linear regressions

The first part of our methodology consists of replicating the methodology conducted by Cremers et al. (2017). In other words, we test whether the treasury implied volatility can be used to predict the future macroeconomic real activity.

To quantify the predictability of macroeconomic real activity using the treasury implied volatility, we run an ordinary least squares (OLS) regression. In the first regression, we take the 5-year YIV and use it to predict the logarithmic values of year-on-year growth rate of the real GDP. In the equation, H is equal to the periods predicted - e.g. if H=4, it means that we are taking the rolling overlapping average of GDP growth over the 4 quarters.

$$\sum_{j=1}^{j=H} log(1 + GDP_{i,t+j})/H = \alpha_H + \beta_H \sigma_{IV,t}^{INT} + \varepsilon_{t+H} \tag{1}$$

In addition, by using an autoregressive model, an additional regression with the lagged values of macroeconomic variables is run to see whether the predictive ability will improve. The number of lagged variables included is decided by the Akaike information criterion (AIC).

We then proceed to include different control variables to the regression model such as term spreads, credit spreads, stock returns, stock market implied volatility.

$$\sum_{j=1}^{j=H} log(1 + GDP_{i,t+j})/H = \alpha_H + \beta_H \sigma_{IV,t}^{INT} + Controls + \varepsilon_{t+H} \tag{2}$$

Next, to determine the direction of the causality, we run a Vector Autoregressive model (VAR) Granger Causality test on every variable to make sure that e.g. YIV indeed granger-causes movements in GDP.

1.2 Adjusting for different economic cycles

In the second part of our research we build upon the research conducted by Cremers et. al. We test whether the model holds when accounting also for the possible business cycle performance asymmetries as suggested by Siliverstovs & Wochner (2020). We perform separate regressions on the subsamples i.e. forecasting performance during expansionary and recessionary periods. More specifically, we compare the root-mean-square-forecasting-error (RMSFE) of the predictive model during recession and expansion with full sample forecast to find out whether full sample forecast's relative RMSFE forecasts are robust during the expansionary and recessionary subsamples (in other words, we check whether a distributional shift occurs during the recessionary and/or expansionary periods).

1.3 Macroeconomic Random Forest (MRF)

As Machine Learning methods successfully tackle nonlinearities that cannot be accounted for in simple linear models, the forecasting gains tend to be the highest during times of high economic uncertainty (Coulombe et al., 2020). As our research includes many different financial crises, we expect to find forecasting gains by utilizing Random Forest whose benefits were elaborated in review of literature.

However, in our research, we use an algorithm developed by Coulombe (2020) named Macroeconomic Random Forest. His proposed model further develops off-the-shelf random forest with regards to two areas:

- **Statistical efficiency** - due to the nature of random forest approximation, achieving smooth linear relationships takes many splits which in case of a smaller time series can result in the loss of degrees of freedom which in turn increases the error terms. MRF, however, includes a linear component which can capture these less complex relationships with less splits, saving the degrees of freedom for more complicated sequences.

- **Interpretation** - random forest model itself is often considered as a black box model in terms of interpretability, and external interpreter algorithms are used to understand it. However, MRF tackles this problem through adding the linear part - after splitting the initial parameter into many pieces of the whole forest, it allows to directly interpret the coefficients of the underlying surrogate models[1].

The general model (see Equation (3)) proposed by Coulombe (2020). The main difference between MRF and RF is that the latter is a restricted model ,where Xt=1.

$$y_t = X_t \beta_t + \varepsilon_t \tag{3}$$

$$\beta_t = F(S_t) \tag{4}$$

The author has provided us with an R-package[2], which dramatically simplifies our research process and makes it more feasible for a Bachelor level thesis.

Coulombe, P. G. (2020). *The Macroeconomy as a Random Forest* [PhD thesis]. https://doi.org/10.2139/ssrn.3633110

Cremers, M., Fleckenstein, M., & Gandhi, P. (2017). Treasury Yield Implied Volatility and Real Activity. *Journal of Financial Economics (JFE), Forthcoming*. https://doi.org/10.2139/ssrn.3006473

Siliverstovs, B., & Wochner, D. (2020). State-Dependent Evaluation of Predictive Ability †. *Journal of Forecasting*. https://doi.org/10.1002/for.2715

[1] The full derivation & optimization can be found in Coulombe (2020)

[2] Macroeconomic Random Forest R package can be downloaded form https://philippegouletcoulombe.com/code