'Bond Risk Premiums with Machine Learning', Bianchi et al.(2020, RFS)

Younghae Kim

Studies in Financial Machine Learning

June 7, 2021

Replicate Bianchi, Buchner, Tamoni(2020, RFS)

- Data : Yield (Liu, Wu(2020)), Macro variables (McCracken, Ng(2015))
- Sample-period : 1971.08 - 2018.12
- Using expanding windows for estimation : re-estimate parameters every month (with training : validation = 85% : 15% split)
- $R^2_{OOS} = 1 - \frac{\sum_{t=1}^{T-1}(xr_{t+1}^{(n)} - \hat{xr}_{t+1}^{(n)}(M_s))^2}{\sum_{t=1}^{T-1}(xr_{t+1}^{(n)} - \bar{xr}_{t+1}^{(n)})^2}$ (Campbell and Tompson(2007))
- To test R^2_{OOS} is greater than zero, implement MSPE-adjusted Clark and West(2007) statistic

Replicate Bianchi, Buchner, Tamoni(2020, RFS)

- To test whether there exists macro-economic risk that is not spanned by yield curve information, implement Diebold and Mariano(1995) (with same models but different x-variables)
- To test whether non-linearity (GBRT, RF, NN) methods improve predictive accuracy of forecast, implement Diebold and Mariano(1995) (with same x-variables but different models)

Results from forward rate information

Table 1
Forecasting annual holding period returns with forward rates

	R^2_{oos}						P-value					
	$xr^{(2)}_{t+1}$	$xr^{(3)}_{t+1}$	$xr^{(4)}_{t+1}$	$xr^{(5)}_{t+1}$	$xr^{(7)}_{t+1}$	$xr^{(10)}_{t+1}$	$xr^{(2)}_{t+1}$	$xr^{(3)}_{t+1}$	$xr^{(4)}_{t+1}$	$xr^{(5)}_{t+1}$	$xr^{(7)}_{t+1}$	$xr^{(10)}_{t+1}$
1. PCA & PLS												
PCA (10 components)	-58.1%	-48.7%	-42.5%	-38.9%	-33.7%	-26.9%						
PCA (5 components)	-59.1%	-50.8%	-45.6%	-40.5%	-38.3%	-31.4%						
PCA (3 components)	-26.0%	-24.1%	-22.1%	-19.0%	-18.5%	-16.3%						
PLS (5 components)	-58.8%	-48.7%	-43.1%	-39.2%	-32.4%	-25.0%						
PLS (3 components)	-59.3%	-51.0%	-45.7%	-41.2%	-36.8%	-29.3%						
2. Penalized linear regression												
Lasso	-11.3%	-16.8%	-17.9%	-17.1%	-15.7%	-14.7%						
Ridge	-32.1%	-30.2%	-27.0%	-23.7%	-20.9%	-15.4%						
Elastic-net	-15.0%	-15.6%	-13.4%	-13.5%	-15.1%	-11.5%						
3. Regression trees & Neural networks												
Gradient boosted tree	-7.0%	-0.9%	23.6%	24.6%	13.9%	18.7%			0.000(***)	0.000(***)	0.000(***)	0.000(***)
Random Forest	-1.5%	2.3%	0.59%	1.2%	0.8%	1.4%		0.013(**)	0.016(**)	0.015(**)	0.014(**)	0.010(**)
NN – 2 hidden layer (5, 5)	-4.0%	-24.1%	-23.5%	-12.8%	-12.4%	-13.8%						

R^2_{oos} : Campbell and Tompson (2007)
P-value : Clark and West (2007), null hypothesis $R^2_{oos} \le 0$,

Results from forward rate & macro information

Table 2
Forecasting annual holding period returns with forward rates & macroeconomic variables

	R^2_{oos}						P-value					
	$xr^{(2)}_{t+1}$	$xr^{(3)}_{t+1}$	$xr^{(4)}_{t+1}$	$xr^{(5)}_{t+1}$	$xr^{(7)}_{t+1}$	$xr^{(10)}_{t+1}$	$xr^{(2)}_{t+1}$	$xr^{(3)}_{t+1}$	$xr^{(4)}_{t+1}$	$xr^{(5)}_{t+1}$	$xr^{(7)}_{t+1}$	$xr^{(10)}_{t+1}$
1. PCA & PLS												
PCA (8 components)	-5.2%	-9.3%	-14.6%	-17.3%	-19.1%	-23.4%						
PLS (8 components)	15.8%	8.3%	11.1%	11.0%	12.1%	16.9%		0.000(***)	0.000(***)	0.000(***)	0.000(***)	0.000(***)
2. Penalized linear regression												
Lasso (using fwd rate directly)	8.9%	10.7%	3.6%	1.6%	6.8%	5.3%	0.003(***)	0.003(***)	0.006(***)	0.001(***)	0.014(**)	0.014(**)
Ridge (using fwd rate directly)	51.4%	48.6%	46.5%	40.5%	37.4%	37.8%	0.000(***)	0.000(***)	0.000(***)	0.000(***)	0.000(***)	0.000(***)
Elastic-net (using fwd rate directly)	23.6%	25.6%	21.5%	26.7%	29.8%	31.1%	0.000(***)	0.000(***)	0.000(***)	0.000(***)	0.000(***)	0.000(***)
3. Regression trees & Neural networks												
Gradient boosted tree	77.5%	69.5%	69.7%	67.9%	73.0%	64.4%	0.000(***)	0.000(***)	0.000(***)	0.000(***)	0.000(***)	0.000(***)
Random Forest	70.0%	65.4%	62.6%	62.0%	58.8%	56.0%	0.000(***)	0.000(***)	0.000(***)	0.000(***)	0.000(***)	0.000(***)
NN – 2 hidden layer (32, 16), fwd direct	-46.3%	-20.0%	-10.0%	0.2%	-8.6%	-9.4%				0.071(*)		
NN – 2 hidden layer (32, 16), fwd 1 hidden layer (3)	-0.0%	-0.0%	-0.0%	-0.0%	-0.0%	-0.0%						

R^2_{oos} : Campbell and Tompson (2007)
P-value : Clark and West (2007), null hypothesis $R^2_{oos} \leq 0$,

DM test (same model, different input)

Table 3
Pairwise tests of predictive accuracy (fwd rate only with fwd rate + macroeconomic variables)

	PCA (8 components)	PLS (8 components)	Lasso	Ridge	Elastic-net	Gradient boosted tree	Random Forest	NN – 2 hidden layer (32, 16), fwd direct	NN – 2 hidden layer (32, 16), fwd 1 hidden layer (3)
PCA (3 components)	0.525								
PLS (3 components)		0.049 (**)							
Lasso			0.0424 (**)						
Ridge				0.001 (***)					
Elastic-net					0.004 (***)				
Gradient boosted tree						0.000 (***)			
Random Forest							0.000 (***)		
NN – 2 hidden layer (5, 5)								0.12	0.84

Same Machine learning model, tests of predictive accuracy Forward rate only model with Forward & macro model
P-value : Diebold and Mariano(1995) extended by Harvey et al.(1997), (Null hypothesis : the pairs have the same performance)

- Except PCA and NN models, We can reject null hypothesis under 5% significance level
- That is, there exists macro-economic risk that is not spanned by yield curve information

DM test (same input, different model)

Table 4
Pairwise tests of predictive accuracy (Forecast with fwd rates)

	PCA (3 components)	PLS (3 components)	Lasso	Ridge	Elastic-net	Gradient boosted tree	Random Forest	NN – 2 hidden layer (5, 5)
PCA (3 components)		0.910	0.436	0.577	0.369	0.013 (**)	0.114	0.426
PLS (3 components)			0.150	0.066 (*)	0.113	0.002 (***)	0.024 (**)	0.187
Lasso				0.648	0.064 (*)	0.000 (***)	0.054 (*)	0.450
Ridge					0.259	0.001 (***)	0.033 (**)	0.368
Elastic-net						0.001 (***)	0.093 (*)	0.552
Gradient boosted tree							0.993	0.995
Random Forest								0.940

P-value : Diebold and Mariano(1995) extended by Harvey et al.(1997), (Null hypothesis : the pairs have the same performance)

- GBRT, Random forest models have more prediction power than linear models

DM test (same input, different model)

Table 5
Pairwise tests of predictive accuracy (Forecast with fwd rates + macroeconomic variables)

	PCA (8 components)	PLS (8 components)	Lasso	Ridge	Elastic-net	Gradient boosted tree	Random Forest	NN – 2 hidden layer (32, 16), fwd direct	NN – 2 hidden layer (32, 16), fwd 1 hidden l(3)
PCA (8 components)		0.043(**)	0.045(**)	0.000(***)	0.000(***)	0.000(***)	0.000(***)	0.266	0.845
PLS (8 components)			0.648	0.032(**)	0.218	0.007(***)	0.021(**)	0.826	0.845
Lasso				0.007(***)	0.001(***)	0.000(***)	0.000(***)	0.945	0.845
Ridge					0.844	0.004(***)	0.030(**)	0.999	0.845
Elastic-net						0.000(***)	0.006(***)	0.999	0.845
Gradient boosted tree							0.989	0.999	0.845
Random Forest								0.999	0.845
NN – 2 hidden layer (32, 16), fwd direct									0.845

P-value : Diebold and Mariano(1995) extended by Harvey et al.(1997), (Null hypothesis : the pairs have the same performance)

- GBRT, Random forest models have more prediction power than linear models (So usefulness of non-linearity)
- Penalized regression and PL (using y information when extracting state variables) have more prediction power than PCA