Introduction to Forecasting Models

Exercise 2: Prediction with Regression

Adam Cihlář[1]

14. February 2022

Abstract: The goal of this exercise is to estimate models for predicting the time series of BAA10Y (Moody's Seasoned Baa Corporate Bond Yield Relative to Yield on 10-Year Treasury Constant Maturity). We employ standard linear regression models and consider three different specifications. Their performance is compared based on properties of their prediction errors and prediction metrics. Model with least explanatory variables has the worst performance, while the two other models perform very similarly.

Keywords: Linear Regression, Time Series, Financial Market Prediction

Introduction

Purpose of this paper is to construct three various regression models to predict BAA10Y (Moody's Seasoned Baa Corporate Bond Yield Relative to Yield on 10-Year Treasury Constant Maturity). Firstly, we briefly look at all the available data - the dependent variable BAA10Y and the explanatory variables VIX (Volatility Index) and T10Y3M (10-Year Treasury Constant Maturity Minus 3-Month Treasury Constant Maturity). Secondly, we split the data to save a portion for the out-of-sample predictions and evaluation of the models. Thirdly, we estimate the models and compare them based on how they can fit the data, based on properties of their prediction errors and of course based on their ability to predict future values - forecast metrics. The whole analysis is implemented in statistical software R.

[1]Masaryk University, Faculty of Economics and Administration, Field: Mathematical and Statistical Methods in Economics, 468087@mail.muni.cz

1 Data

Our dataset consists of three variables BAA10Y (Moody's Seasoned Baa Corporate Bond Yield Relative to Yield on 10-Year Treasury Constant Maturity), VIX (Volatility Index) and T10Y3M (10-Year Treasury Constant Maturity Minus 3-Month Treasury Constant Maturity). The available time period is from 1. 1. 2018 to 31. 12. 2021.

None of the variables seems to contain a trend component, thus we will work in the models with these original time series. For the estimate data from 1. 1. 2018 to 31. 12. 2020 will be used and the year 2021 will be used for out-of-sample predictions and evaluation of the models.

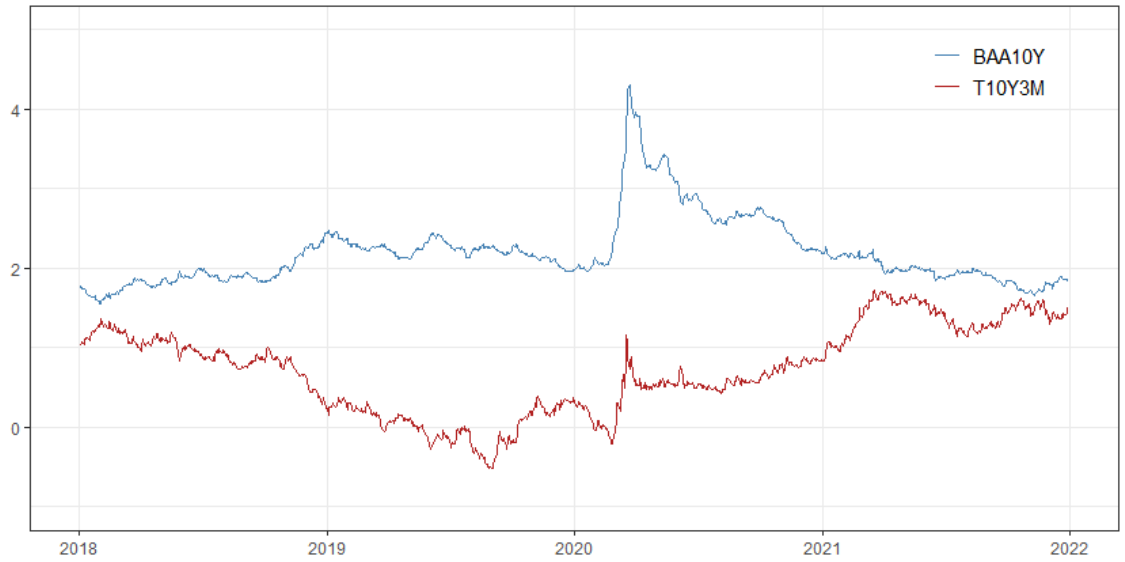

Figure 1: BAA10Y and T10Y3M

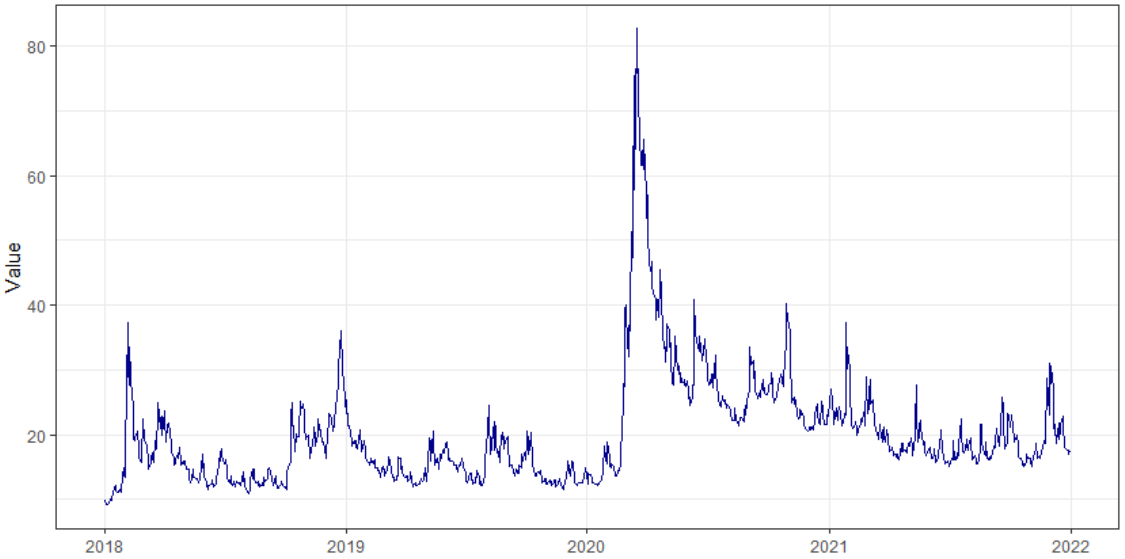

Figure 2: VIX

2 Models and estimates

We define and estimate following three models to predict the BAA10Y time series:

$$BAA10Y_t = \beta_0 + \beta_1 T10Y3M_{t-1} + \beta_2 VIX_{t-1} + \varepsilon_t \tag{2.1}$$

$$BAA10Y_t = \beta_0 + \beta_1 VIX_{t-1} + \varepsilon_t \tag{2.2}$$

$$BAA10Y_t = \beta_0 + \beta_1 T10Y3M_{t-1} + \beta_2 VIX_{t-1} + \beta_3 T10Y3M_{t-2} + \beta_4 VIX_{t-2} + \varepsilon_t \tag{2.3}$$

	Dependent variable:		
	BAA10Y		
	(1)	(2)	(3)
$T10Y3M_{t-1}$	-0.329^{***}		-0.647^{***}
	(0.019)		(0.172)
VIX_{t-1}	0.039^{***}	0.038^{***}	0.007^{**}
	(0.001)	(0.001)	(0.003)
$T10Y3M_{t-2}$			0.311^{*}
			(0.172)
VIX_{t-2}			0.033^{***}
			(0.003)
Constant	1.642^{***}	1.507^{***}	1.623^{***}
	(0.021)	(0.022)	(0.019)
Observations	749	749	748
R^2	0.759	0.666	0.789
Adjusted R^2	0.759	0.666	0.788
Residual Std. Error	0.229 (df = 746)	0.270 (df = 747)	0.215 (df = 743)
F Statistic	$1{,}177^{***}$ (df = 2; 746)	$1{,}489^{***}$ (df = 1; 747)	693^{***} (df = 4; 743)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 1: Estimated models

F Statistics show that all models have statistically significant ability to explain some part of variance of the dependent variable $BAA10Y_t$, moreover every explanatory variable is significant at least on 10 % level.

As model 1 and 3 are extensions of model 2, we, of course, get higher R^2 for the more complex models. More interestingly, we get also higher Adjusted R^2 when we include more variables.

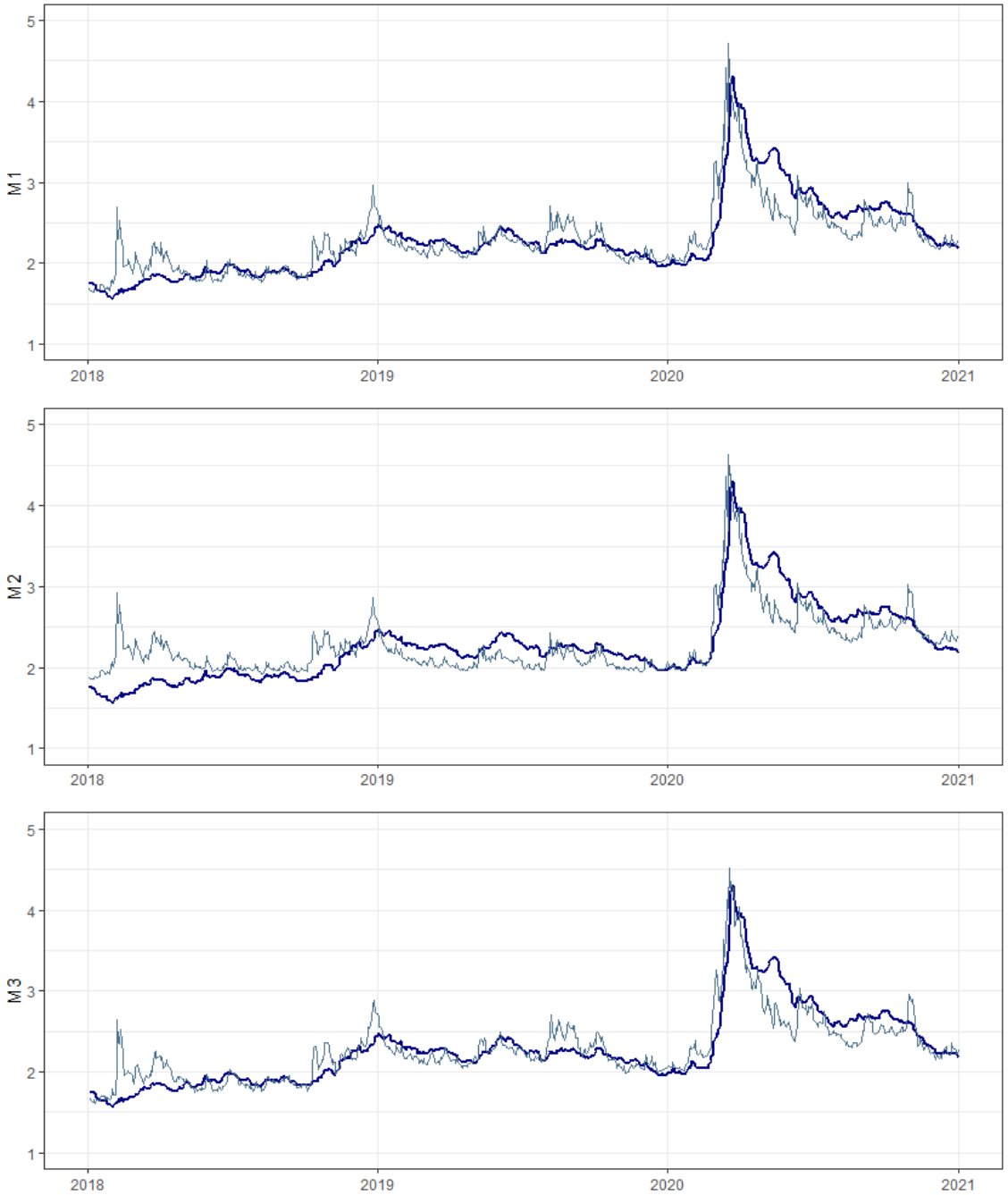

Figure 3: Fitted and true (bold) values

3 Predictions

In this section we will predict the values of BAA10Y for year 2021 using the three estimated models.

Firstly, let's have a look on various prediction metrics and plot the predicted and true values. Clearly, the first and third model outperform the most simple second model with only one independent variable (VIX_{t-1}). To evaluate difference between predictions of model M1 and M3 we would need to employ a statistical test but that is not in scope of this exercise.

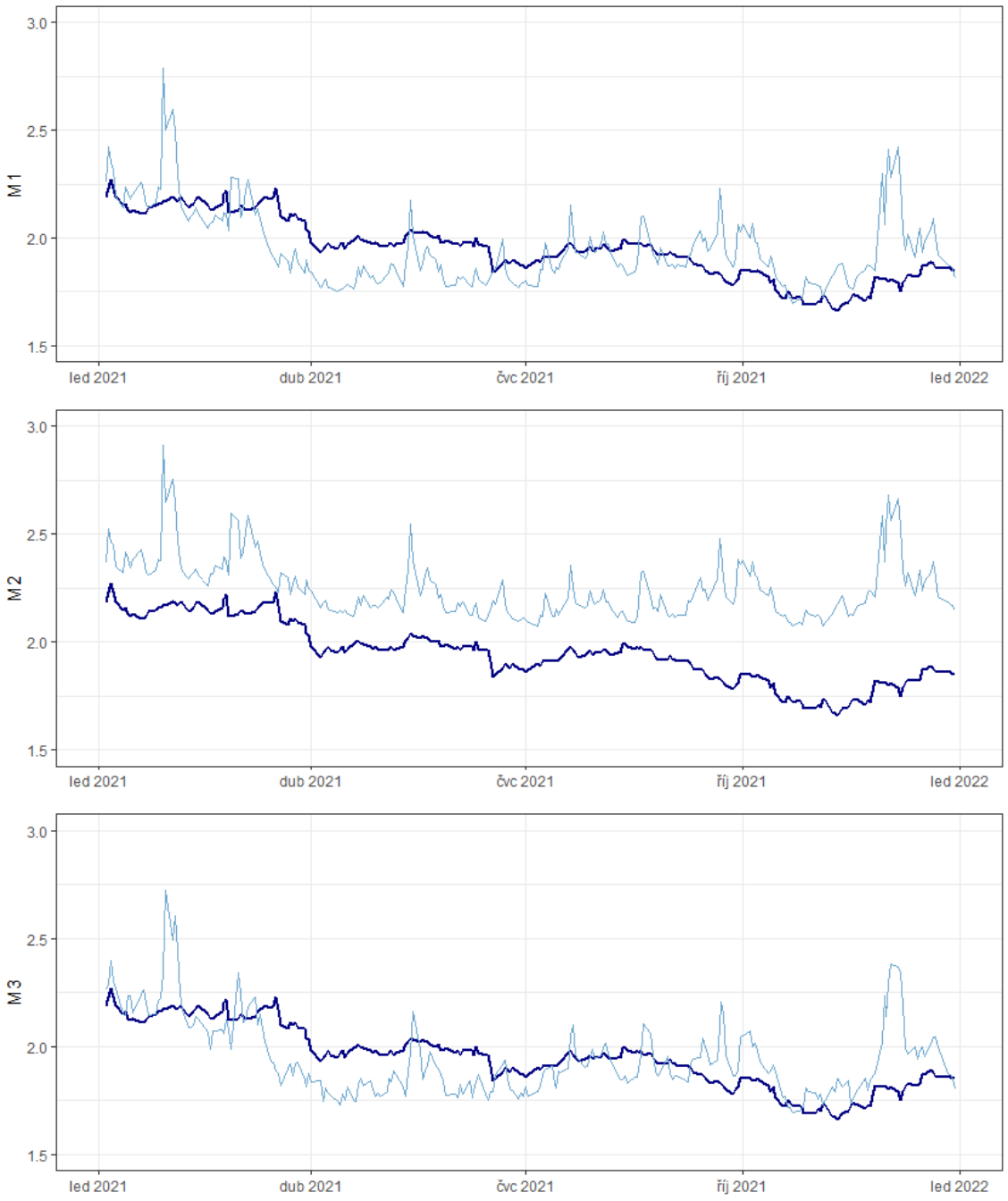

Figure 4: Predictions

	M1	M2	M3
MAE	0.127	0.316	0.125
RMSE	0.168	0.353	0.164
MAPE	6.6%	16.7%	6.5%

Table 2: Prediction Metrics

Secondly, we will inspect properties of prediction errors of all models. None of the prediction errors is normally distributed, which we can observe also from the prediction errors histograms (Figure 5).

	M1	M2	M3
X-squared	72.704	87.476	63.539
p-value	< 0.0001	< 0.0001	< 0.0001

Table 3: Jarque-Bera test of normality of prediction errors

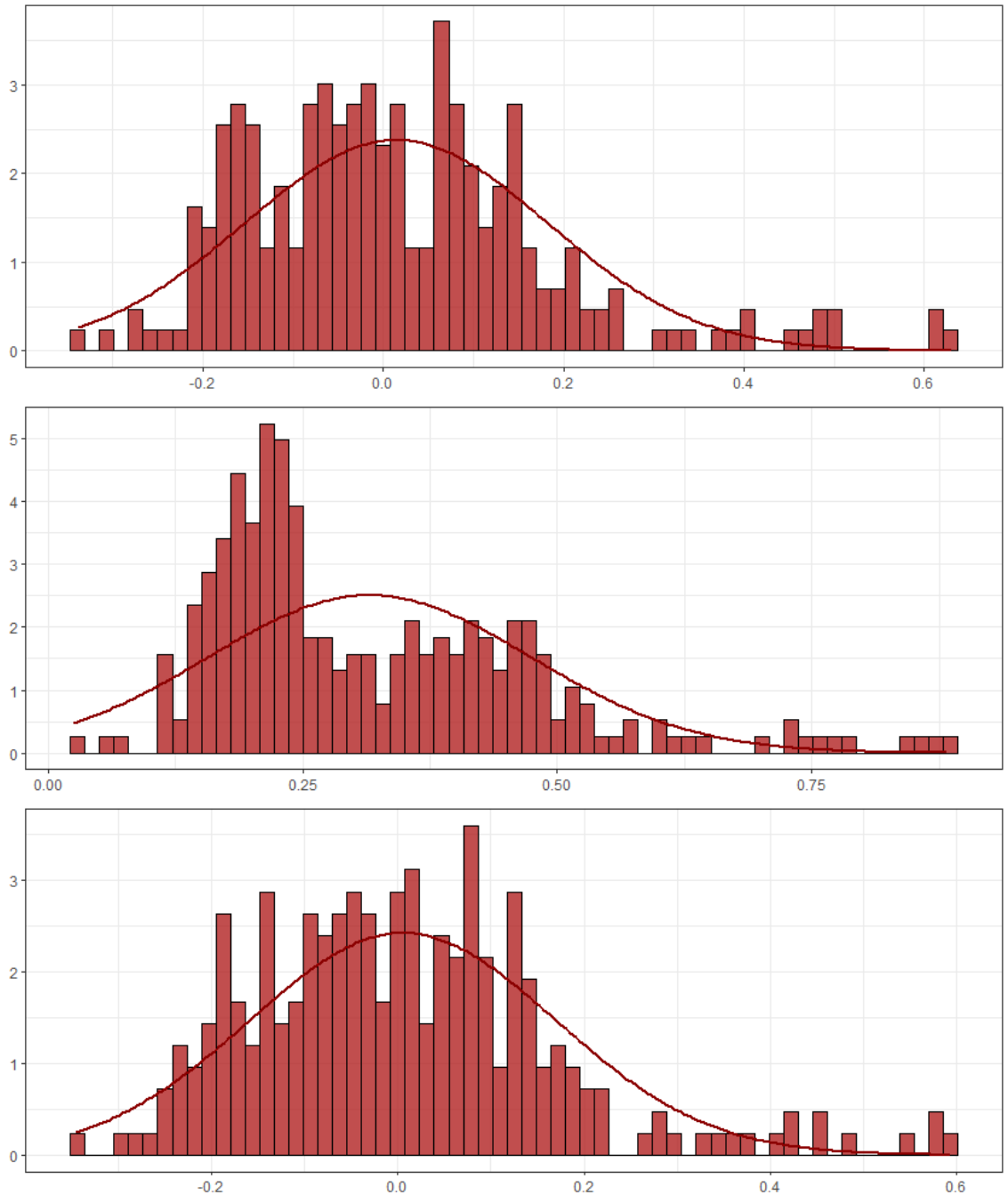

Figure 5: Prediction errors distributions

4 Specification tests

Ultimately, we will have a closer look on the prediction errors and test if the specification of the model is proper or if we missed something in our models.

The following table presents regressions of prediction errors on selected transformations of original explanatory variables. If we can explain the error term with them, we did not specify our model correctly.

	Dependent variable:		
	$M1_{error}$	$M2_{error}$	$M3_{error}$
$(\text{T10Y3M}_{t-1})^2$	-0.377^{*}		1.109
	(0.201)		(0.907)
$(\text{VIX}_{t-1})^2$	0.010	0.023^{***}	-0.032
	(0.013)	(0.002)	(0.021)
$(\text{T10Y3M}_{t-2})^2$			1.467
			(0.918)
$(\text{VIX}_{t-2})^2$			-0.008
			(0.024)
$\text{T10Y3M}_{t-1} * VIX_{t-1}$	0.013		-0.180^{***}
	(0.009)		(0.051)
$\text{T10Y3M}_{t-1} * T10Y3M_{t-2}$			-0.983^{***}
			(0.193)
$\text{T10Y3M}_{t-1} * VIX_{t-2}$			0.172^{***}
			(0.063)
$\text{VIX}_{t-1} * T10Y3M_{t-2}$			0.182^{***}
			(0.051)
$\text{T10Y3M}_{t-2} * VIX_{t-2}$			-0.175^{***}
			(0.060)
$\text{VIX}_{t-1} * VIX_{t-2}$			0.002^{***}
			(0.0004)
Constant	-0.019	-0.147^{***}	-1.484^{**}
	(0.283)	(0.043)	(0.697)
Observations	267	267	267
R^2	0.456	0.309	0.570
Adjusted R^2	0.450	0.307	0.553
Residual Std. Error	0.125 (df = 263)	0.132 (df = 265)	0.110 (df = 256)
F Statistic	73.4^{***} (df = 3; 263)	118.5^{***} (df = 1; 265)	33.9^{***} (df = 10; 256)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 4: Specification test - interactions

All of the models can still explain some part of the error term, so none of the models is perfect and we would need to respecify them to reach better results.

Second specification test tests whether there is a serial correlation in the prediction errors. Again, we reject the null hypothesis of no serial correlation for all models as the F statistics are very high.

	Dependent variable:		
	$M1_{error}$	$M2_{error}$	$M3_{error}$
$error_{t-1}$	0.759^{***}	0.766^{***}	0.897^{***}
	(0.062)	(0.062)	(0.062)
$error_{t-2}$	0.151^{*}	0.124	0.042
	(0.077)	(0.077)	(0.083)
$error_{t-3}$	-0.011	0.023	-0.042
	(0.063)	(0.063)	(0.063)
Constant	0.003	0.030^{***}	0.002
	(0.005)	(0.012)	(0.005)
Observations	264	264	264
R^2	0.754	0.760	0.793
Adjusted R^2	0.751	0.758	0.790
Residual Std. Error (df = 260)	0.084	0.079	0.076
F Statistic (df = 3; 260)	265.828^{***}	274.954^{***}	331.441^{***}
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 5: Specification test - prediction errors autocorrelation

From the prediction metrics and properties of the prediction errors we would prefer using model 1 because it performs almost the same as model 3 but is simpler. However, none of the models is satisfies the requirements for prediction errors and every model would benefit from further improvement.

Conclusion

In this exercise we constructed three various regression models to predict BAA10Y (Moody's Seasoned Baa Corporate Bond Yield Relative to Yield on 10-Year Treasury Constant Maturity).

Firstly, we visualized all the available data - the dependent variable BAA10Y and the explanatory variables VIX (Volatility Index) and T10Y3M (10-Year Treasury Constant Maturity Minus 3-Month Treasury Constant Maturity). Secondly, we split the data to save a portion for the out-of-sample predictions and evaluation of the models. Thirdly, we estimated the models and compared them based on how they can fit the data, based on properties of their prediction errors and of course based on their ability to predict future values - forecast metrics.

We decided to prefer the first model over the others but all of them would require further improvement as they all fail the specification tests.