

# 1 Appendix:

## 1.1 Grid search

### 1.1.1 XGBOOST

	Max Depth	Learning Rate	Reg Lambda	Reg Alpha	Min split Loss	Base Score
XGBOOST	2.000E+00	5.000E-03	1.000E+00	1.000E+00	0.000E+00	3.000E-01
	3.000E+00	1.000E-02	5.000E+00	2.000E+00	1.000E+00	4.000E-01
	4.000E+00		1.000E+01	1.000E+01	3.000E+00	5.000E-01
	5.000E+00		2.000E+01	2.000E+01	4.000E+00	6.000E-01
	6.000E+00		1.000E+02	1.000E+02	5.000E+00	7.000E-01
	7.000E+00				6.000E+00	1.000E+01

- Max Depth designates the depth of the trees.
- Learning rate is shrinkage intensity of the features weights during the boosting process.
- Reg Lambda is a regularization term on the l2 norm.
- Reg Alpha is a regularization term on the l1 norm.
- Min Split Loss is used to create additional partitions of the data at the end node of a leaf.
- Base score is the starting point of the algorithm.

More information can be found on this website:

<https://xgboost.readthedocs.io/en/latest/parameter.html>

### 1.1.2 Neural Network

	Architecture	Kernel	Alpha	Learning Rate
Neural Network	(32,50,32)	"Logistic"	1.000E-02	1.000E-03
	(100,100)	"Relu"	1.000E-01	5.000E-03
	(50,50)	"Tanh"	1.000E+00	1.000E-02
	(50,100,50)		5.000E+00	
			1.000E+01	
			2.000E+01	
		1.000E+02		

- Architecture defines the number of layers and the neurons per layer. For instance, (32,50,32) means that there are three layers composed of 32, 50 and 32 neurons.
- Kernel is the activation function for the hidden layers that introduce non-linearity in the algorithm.
- Alpha denotes the strength of the l2 regularization
- Learning Rate is used to update the weights of the neurons

More information can be found on this website: [https://scikit-learn.org/stable/modules/generated/sklearn.neural\\_network.MLPRegressor.html](https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPRegressor.html)

### 1.1.3 Support Vector Regression

	C	Kernel	Tol	Epsilon	Coef0	Degree
Support Vector Machine	1.000E-05	"Polynomial"	1.000E-04	1.000E-03	-2.000E+00	1.00
	1.000E-03	"Radial basis"	1.000E-03	1.000E-02	-1.500E+00	2.00
	1.000E-02	"Sigmoid"	1.000E-02	1.000E-01	-1.200E+00	3.00
	1.000E-01		1.000E-01	1.000E+00	-1.000E+00	4.00
	1.000E+00		1.000E+00	1.100E+00	-3.000E-01	5.00
	1.000E+01			1.200E+00	-2.000E-01	6.00
	1.500E+01			1.300E+00	-1.000E-01	7.00
	1.000E+02			1.500E+00	0.000E+00	8.00
				1.600E+00	5.000E-01	
				1.700E+00	1.000E+00	
				2.000E+00		

- C is a regularization term that adds a l2 penalty. The regularisation is inversely proportional to the value of C.
- Kernel is the type of functions used to define the kernel of the algorithm.
- Tol is the tolerance acceptance in the algorithm. It is a stopping criterion.
- Epsilon is the tube tolerance in which the errors are not penalized.
- Coef0 is a parameter associated with the kernel functions "Polynomial" and "Sigmoid".
- Degree is a parameter associated with the kernel function "Polynomial".

More information can be found on this website: <https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVR.html>

## 1.2 Weekly log returns

### 1.2.1 Moments of the distribution, Jarque-Bera and Augmented Dickey-Fuller tests

	Mean	Variance	Skewness	Kurtosis	JB		ADF	
					Test value	P-value	Test value	P-value
Pre-MSR	7.10E-03	4.93E-03	-0.16	1.14	45.17	1.56E-10	-5.56	1.56E-06
Post-MSR	6.07E-03	4.32E-03	-0.94	5.26	1414.52	6.92E-308	-7.54	3.38E-11

### 1.2.2 Performances of the samples with number of PCs and variance explained

#### 1.2.2.1 Full sample

	Full	RMSE	MSE	MAPE	MAE	R2	# PCs	Variance explained
XGBOOST	with	4.68E-02	2.19E-03	2.20E+02	3.46E-02	0.59	13	0.886
	without	4.63E-02	2.14E-03	2.36E+02	3.45E-02	0.60	7	0.875
NN	with	4.00E-02	1.60E-03	1.07E+03	3.02E-02	0.70	13	0.886
	without	4.00E-02	1.60E-03	2.86E+03	2.97E-02	0.70	7	0.875
SVR	with	4.75E-02	2.26E-03	1.68E+03	3.64E-02	0.57	13	0.886
	without	3.86E-02	1.49E-03	3.70E+03	2.84E-02	0.72	7	0.875
Naive Strategy	N/A	4.86E-02	2.36E-03	1.34E+03	3.51E-02	0.56	N/A	N/A

### 1.2.2.2 Pre-MSR

Pre-MSR		RMSE	MSE	MAPE	MAE	R2	# PCs	Variance explained
XGBOOST	with	5.50E-02	3.03E-03	1.12E+02	3.89E-02	0.49	10	0.856
	without	5.18E-02	2.69E-03	1.11E+02	3.69E-02	0.55	7	0.906
NN	with	7.68E-02	5.90E-03	1.21E+02	5.66E-02	0.01	10	0.856
	without	7.68E-02	5.90E-03	1.21E+02	5.66E-02	0.01	10	0.856
SVR	with	7.23E-02	5.23E-03	1.13E+02	5.30E-02	0.13	10	0.856
	without	4.98E-02	2.48E-03	1.03E+02	3.54E-02	0.59	7	0.906
Naive Strategy	N/A	5.10E-02	2.60E-03	1.40E+02	3.60E-02	0.57	N/A	N/A

### 1.2.2.3 Post-MSR

Post-MSR		RMSE	MSE	MAPE	MAE	R2	# PCs	Variance explained
XGBOOST	with	4.04E-02	1.63E-03	1.00E+03	3.29E-02	0.59	11	0.861
	without	4.01E-02	1.61E-03	1.58E+03	3.23E-02	0.60	7	0.895
NN	with	4.30E-02	1.85E-03	4.63E+03	3.43E-02	0.54	11	0.861
	without	4.89E-02	2.39E-03	1.48E+03	3.87E-02	0.41	7	0.895
SVR	with	5.91E-02	3.49E-03	8.57E+02	4.51E-02	0.13	11	0.861
	without	5.71E-02	3.26E-03	2.31E+03	4.39E-02	0.19	7	0.895
Naive Strategy	N/A	3.88E-02	1.50E-03	1.36E+03	3.12E-02	0.63	N/A	N/A

## 1.3 Monthly log return

### 1.3.1 Moments of the distributions, Jarque-Bera and Augmented Dickey-Fuller tests

	Mean	Variance	Skewness	Kurtosis	JB		ADF	
					Test value	P-value	Test value	P-value
Full sample	2.69E-02	1.63E-02	-0.58	1.09	195.26	3.98E-43	-6.96	9.46E-10
Pre-MSR	2.97E-02	1.98E-02	-0.50	0.42	36.81	1.02E-08	-3.80	2.92E-03
Post-MSR	2.49E-02	1.40E-02	-0.69	1.72	216.80	8.37E-48	-4.67	9.40E-05

### 1.3.2 Performances of the samples with number of PCs and variance explained

#### 1.3.2.1 Full sample

Full		RMSE	MSE	MAPE	MAE	R2	# PCs	Variance explained
XGBOOST	with	5.57E-02	3.10E-03	9.72E+01	4.06E-02	0.82	10	0.894
	without	4.93E-02	2.43E-03	9.56E+01	3.70E-02	0.86	4	0.892
NN	with	8.17E-02	6.67E-03	1.30E+02	6.31E-02	0.61	10	0.894
	without	6.33E-02	4.01E-03	1.27E+02	5.05E-02	0.76	4	0.892
SVR	with	1.19E-01	1.41E-02	1.89E+02	9.12E-02	0.17	10	0.894
	without	5.12E-02	2.62E-03	9.62E+01	3.87E-02	0.85	4	0.892
Naive Strategy	N/A	4.83E-02	2.34E-03	1.04E+02	3.62E-02	0.86	N/A	N/A

#### 1.3.2.2 Pre-MSR sample

Pre-MSR		RMSE	MSE	MAPE	MAE	R2	# PCs	Variance explained
XGBOOST	with	5.22E-02	2.72E-03	8.22E+01	3.81E-02	0.82	9	0.890
	without	5.35E-02	2.86E-03	8.12E+01	3.90E-02	0.81	5	0.933
NN	with	1.18E-01	1.39E-02	1.57E+02	9.36E-02	0.09	9	0.890
	without	6.08E-02	3.70E-03	8.25E+01	4.51E-02	0.76	5	0.933
SVR	with	1.02E-01	1.05E-02	1.26E+02	8.28E-02	0.31	9	0.890
	without	5.49E-02	3.02E-03	8.62E+01	4.05E-02	0.80	5	0.933
Naive Strategy	N/A	4.86E-02	2.36E-03	8.24E+01	3.52E-02	0.85	N/A	N/A

### 1.3.2.3 Post-MSR sample

Post-MSR		RMSE	MSE	MAPE	MAE	R2	# PCs	Variance explained
XGBOOST	with	4.11E-02	1.69E-03	8.22E+01	3.24E-02	0.89	9	0.876
	without	7.99E-02	6.38E-03	1.06E+02	5.82E-02	0.57	5	0.918
NN	with	1.04E-01	1.08E-02	1.02E+02	8.30E-02	0.27	9	0.876
	without	5.77E-02	3.33E-03	8.86E+01	4.30E-02	0.78	5	0.918
SVR	with	5.33E-02	2.85E-03	1.00E+02	4.06E-02	0.81	9	0.876
	without	1.00E-01	1.00E-02	1.22E+02	7.28E-02	0.33	5	0.918
Naive Strategy	N/A	3.96E-02	1.57E-03	8.87E+01	3.16E-02	0.89	N/A	N/A

## 1.4 95-99% variance explained

### 1.4.1 Weekly log returns

#### 1.4.1.1 XGBOOST with 95% of variance explained

		RMSE	MSE	MAPE	MAE	R2	# PCs	Variance explained
Full	with	4.95E-02	2.45E-03	7.71E+02	3.44E-02	0.54	22	0.950
	without	4.63E-02	2.14E-03	2.36E+02	3.45E-02	0.60	12	0.952
Pre-MSR	with	5.69E-02	3.24E-03	1.12E+02	4.02E-02	0.46	20	0.953
	without	5.18E-02	2.69E-03	1.11E+02	3.69E-02	0.55	11	0.951
Post-MSR	with	4.05E-02	1.64E-03	1.90E+03	3.22E-02	0.59	22	0.951
	without	4.18E-02	1.75E-03	1.37E+03	3.31E-02	0.56	12	0.952

#### 1.4.1.2 XGBOOST with 99% of variance explained

		RMSE	MSE	MAPE	MAE	R2	# PCs	Variance explained
Full	with	4.95E-02	2.45E-03	7.71E+02	3.44E-02	0.54	40	0.991
	without	4.50E-02	2.02E-03	6.75E+02	3.30E-02	0.62	23	0.991
Pre-MSR	with	5.69E-02	3.24E-03	1.12E+02	4.02E-02	0.46	41	0.990
	without	5.18E-02	2.69E-03	1.11E+02	3.69E-02	0.55	26	0.991
Post-MSR	with	4.06E-02	1.65E-03	2.58E+03	3.23E-02	0.59	42	0.990
	without	3.95E-02	1.56E-03	1.92E+03	3.18E-02	0.61	26	0.991

### 1.4.2 Monthly log returns

#### 1.4.2.1 XGBOOST with 95% of variance explained

		RMSE	MSE	MAPE	MAE	R2	# PCs	Variance explained
Full	with	4.99E-02	2.49E-03	9.53E+01	3.73E-02	0.85	17	0.950
	without	4.93E-02	2.43E-03	9.56E+01	3.70E-02	0.86	8	0.959
Pre-MSR	with	6.28E-02	3.94E-03	7.48E+01	4.43E-02	0.74	15	0.952
	without	4.97E-02	2.47E-03	8.10E+01	3.67E-02	0.84	7	0.956
Post-MSR	with	6.82E-02	4.66E-03	8.95E+01	5.00E-02	0.69	18	0.952
	without	7.99E-02	6.38E-03	1.06E+02	5.82E-02	0.57	8	0.953

#### 1.4.2.1.1 XGBOOST with 99% of variance explained

		RMSE	MSE	MAPE	MAE	R2	# PCs	Variance explained
Full	with	4.99E-02	2.49E-03	9.53E+01	3.73E-02	0.85	36	0.990
	without	4.93E-02	2.43E-03	9.56E+01	3.70E-02	0.86	22	0.990
Pre-MSR	with	5.07E-02	2.57E-03	8.06E+01	3.75E-02	0.83	37	0.990
	without	4.97E-02	2.47E-03	8.10E+01	3.67E-02	0.84	24	0.991
Post-MSR	with	7.05E-02	4.97E-03	9.91E+01	5.28E-02	0.67	40	0.990
	without	4.09E-02	1.67E-03	8.24E+01	3.23E-02	0.89	26	0.991

#### 1.4.2.1.2 Two-sided Diebold-Mariano test

		95% variance explained		99% variance explained	
		Diebold-Mariano		Diebold-Mariano	
		Test value	P-value	Test value	P-value
Full	with	3.41	0.00	3.41	0.00
	without	0.00	1.00	1.40	0.16
Pre-MSR	with	-2.76	0.01	1.61	0.11
	without	2.05	0.04	2.05	0.04
Post-MSR	with	-6.24	0.00	-6.84	0.00
	without	0.00	1.00	6.93	0.00