Investing through Economic Cycles with Ensemble Machine Learning Algorithms

Thomas Raffinot

Silex Investment Partners

Big Data in Finance Conference

Turning points detection in real time: Ensemble ML algorithms

- In theory, investment strategies based on growth cycle turning points outperform not only passive buy-and-hold benchmarks, but also business cycles' strategies

Turning points detection in real time: Ensemble ML algorithms

- In theory, investment strategies based on growth cycle turning points outperform not only passive buy-and-hold benchmarks, but also business cycles' strategies
- Nowcasting growth cycle turning points in real time in the euro area and in the United States to time markets

Turning points detection in real time: Ensemble ML algorithms

- In theory, investment strategies based on growth cycle turning points outperform not only passive buy-and-hold benchmarks, but also business cycles' strategies
- Nowcasting growth cycle turning points in real time in the euro area and in the United States to time markets
- Non parametric model to avoid local maxima in the likelihood

Turning points detection in real time: Ensemble ML algorithms

- In theory, investment strategies based on growth cycle turning points outperform not only passive buy-and-hold benchmarks, but also business cycles' strategies
- Nowcasting growth cycle turning points in real time in the euro area and in the United States to time markets
- Non parametric model to avoid local maxima in the likelihood
- Ensemble machine learning algorithms:
  - Random forest (Breiman (2001))
  - Boosting (Schapire (1990))

Ensemble Machine Learning Algorithms

- Machine learning adapts statistical methods to get better results in an environment with much more data and processing power

Ensemble Machine Learning Algorithms

- Machine learning adapts statistical methods to get better results in an environment with much more data and processing power
- Ensemble algorithms: making decisions based on the input of multiple people or experts

Ensemble Machine Learning Algorithms

- Machine learning adapts statistical methods to get better results in an environment with much more data and processing power
- Ensemble algorithms: making decisions based on the input of multiple people or experts
- Entertain a large number of predictors and perform estimation and variable selection simultaneously

Ensemble Machine Learning Algorithms

- Machine learning adapts statistical methods to get better results in an environment with much more data and processing power
- Ensemble algorithms: making decisions based on the input of multiple people or experts
- Entertain a large number of predictors and perform estimation and variable selection simultaneously
- Random forest (Breiman (2001)): simple averaging of models

Ensemble Machine Learning Algorithms

- Machine learning adapts statistical methods to get better results in an environment with much more data and processing power
- Ensemble algorithms: making decisions based on the input of multiple people or experts
- Entertain a large number of predictors and perform estimation and variable selection simultaneously
- Random forest (Breiman (2001)): simple averaging of models
- Boosting (Schapire (1990)): iterative process where the errors are kept being modelled

Random forest

- Each decision tree is built from a bootstrapped sample of the full dataset and then, at each node, only a random sample of the available variables is used
- Algorithm:
  I Given that a training set consists of N observations and M features, choose a number $m \leq M$ of features to randomly select for each tree and a number K that represents the number of trees to grow.
  II Take a bootstrap sample Z of the N observations. So about two third of the cases are chosen. Then select randomly m features.
  III Grow a CART using the bootstrap sample Z and the m randomly selected features.
  IV Repeat the steps 2 and 3, K times.
  V Output the ensemble of trees T_1^K
  VI For regression, to make a prediction at a new point x:

$$\hat{y}_{RF}(x) = \frac{1}{K} \sum_{i=1}^{K} T_i(x)$$

The gradient descent view of boosting (Friedman (2001))

- The task is to estimate the function $\hat{f}(\mathbf{x})$, that minimizes the expectation of some loss function, $\Psi(y, f)$, i.e.,

$$\hat{f}(\mathbf{x}) = \arg\min_{f(\mathbf{x})} \mathbf{E}(\Psi(y, f(\mathbf{x}))$$

- One has to provide the choices of functional parameters $\Psi(y, f)$ and the weak learner $h(\mathbf{x}, \theta)$
- The function estimate $\hat{f}(\mathbf{x})$ is parameterized in the additive functional form:

$$\hat{f}(\mathbf{x}) = \sum_{m=1}^{M_{stop}} \beta_m h(\mathbf{x}, \theta_m)$$

- The original function optimization problem has thus been changed to a parameter optimization problem
- The size of the ensemble is determined by M, which is determined by cross-validation

Boosting: loss-functions

- The most frequently used loss-functions for classification are the following:
  - y typically takes on binary values $y \in 0, 1$. To simplify the notation, let us assume the transformed labels $\bar{y} = 2y - 1$ making $\bar{y} \in -1, 1$
  - Adaboost loss function: $\Psi(y, f(\mathbf{x})) = \exp(-\bar{y} f(\mathbf{x}))$
  - Binomial loss function: $\Psi(y, f(\mathbf{x})) = -\log(1 + \exp(-2\bar{y} f(\mathbf{x})))$
- The most frequently used loss-functions for regression are the following:
  - Squared error loss: $\Psi(y, f(\mathbf{x})) = (y - f(\mathbf{x}))^2$
  - Absolute loss: $\Psi(y, f(\mathbf{x})) = |y - f(\mathbf{x}|$

GBM algorithm with shrinkage

Step 1 Initialize $\hat{f}_0(\mathbf{x}) = \arg\min_\rho \sum_{i=1}^N \Psi(y_i, \rho)$, $m = 0$.

Step 2 $m = m + 1$

Step 3 Compute the negative gradient

$$z_i = -\frac{\partial}{\partial f(\mathbf{x}_i)} \Psi(y_i, f(\mathbf{x}_i)) \Bigg|_{f(\mathbf{x}_i) = \hat{f}_{m-1}(\mathbf{x}_i)}, i = 1, \ldots, n$$

Step 4 Fit the base-learner function, $h(\mathbf{x}, \theta)$ to be the most correlated with the gradient vector.

$$\theta_m = \arg\min_{\beta, \theta} \sum_{i=1}^n z_i - \beta h(\mathbf{x}_i, \theta_m)$$

Step 5 Find the best gradient descent step-size ρ_m

$$\rho_m = \arg\min_\rho \sum_{i=1}^N \Psi(y_i, \hat{f}(\mathbf{x}_i)_{m-1} + \rho h(\mathbf{x}, \theta_m))$$

Step 6 Update the estimate of $f_m(\mathbf{x})$ as

$$\hat{f}_m(\mathbf{x}) \leftarrow \hat{f}(\mathbf{x})_{m-1} + \lambda \rho_m h(\mathbf{x}, \theta_m))$$

Step 7 Iterate 2-6 until $m = M_{stop}$.

Variables: almost non-revised series

- Financial series:
  Government bonds, Yield curves, investment-grade and high-yield corporate spreads, stock markets (Large caps, large caps sectors, small caps, mid caps, the growth and value version of those indexes), Assets volatility, VIX index and the VSTOXX index, commodities (crude oil, natural gas, gold, silver and CRB index),...
- Economic surveys:
  European Commission, the Institute for Supply Management, the Conference Board and the National Association of Home Builders (NAHB)
- Real economic data:
  Initial claims
- Different lags of differentiation were considered: 1 to 18 months
- More than 1000 variables

Different models

- Boosting:
  - Combination of a binomial loss function with decision trees ("*BTB*") as in Ng (2014)
  - Combination of a squared error loss function with P-splines ("*SPB*") as in Berge (2015) or Taieb et al. (2015)
- Random forest *RF*
- Competitive models:
  - *Acc* classifies all data as "acceleration"
  - *Slow* classifies all data as "slowdown"
  - *Random* randomly assigns classes based on the proportions found in the training data
  - *Prob* refers to the probit model based on the term spread
  - *MS* refers to the Markov-switching dynamic factor model
  - *EN* refers to the elastic-net logistic model

Real time issues

- To implement the ensemble algorithms, a classification of economic regimes is needed
- Applied to the context of nowcasting, it can be summarized as follows:

$$R_t = \begin{cases} 1, & \text{if in acceleration} \\ 0, & \text{otherwise} \end{cases}$$

- A recursive estimation is computed:
  The ensemble algorithms are trained each month on a sample that extends from the beginning of the sample through month $T-12$, over which the turning point chronology is assumed known
- The estimation windows is thus expanding as data accumulates, over the period from January 2002 to December 2013

Data snooping

- Data snooping occurs when a given set of data is used more than once for purposes of inference or model selection. It leads to the possibility that any successful results may be spurious because they could be due to chance (White (2000))
- Model Confidence Set (Hansen et al. (2011)): Model selection algorithm, which filters a set of models from a given entirety of models. The MCS aims at finding the best model and all models which are indistinguishable from the best

Classical criteria

- The Brier's Quadratic Probability Score (QPS):

$$QPS = \frac{1}{F}\sum_{t=1}^{F}(\hat{y}_t - y_t)^2$$

- The Area Under the ROC curve (AUROC), defined by:

$$\text{AUROC} = \int_0^1 \text{ROC}(\alpha) d\alpha$$

where the Receiver Operating Characteristics (ROC) curve describes all possible combinations of true positive ($T_p(c)$) and false positive rates ($F_p(c)$) that arise as one varies the threshold c used to make binomial forecasts from a real-valued classifier. As c is varied from 0 to 1, the ROC curve is traced out in ($T_p(c)$, $F_p(c)$) space that describes the classification ability of the model.

Investment strategies

- Disconnection between econometric predictability and actual profitability (Cenesizoglu and Timmermann (2012))
- Very basic investment strategies:
  - Equity portfolio: if acceleration: 120% of his wealth is invested on the asset and 20% of cash is borrowed, otherwise 80% of his wealth is invested on the asset and 20% is kept in cash
  - Asset allocation; if acceleration: 80% of the portfolio is allocated to equities and 20% to bonds, otherwise 40% of the portfolio is allocated to equities and 60% to bonds

Classical evaluation criteria in the United States, January 2002 to December 2013

	QPS	AUROC
SPB	0.13	
RF	0.07**	0.94
BTB	0.05**	0.94
Prob	0.22	
MS	0.21	
EN	0.18	
Acc	0.21	
Slow	0.79	
Random	0.25	

Note: ** indicates the model is in the set of best models $\widehat{M}^*_{75\%}$.

Turning point signals of the reference cycle in the United States

	SPB	*RF*	*BTB*
Trough: February 2003	0	-1	-2
Peak: October 2007	1	-2	-1
Trough: September 2009	1	2	3
Peak: June 2011	-	3	2
Trough: December 2011		1	1

Note: Value shown is the model-implied peak/trough calculated using a 0.5 threshold. The minus sign refers to the lead in which the models anticipate the turning point dates. "-" indicates that the model did not generate any signal. *SPB* refers to a boosting model based on squared error loss with P-splines, *RF* refers to a random forest model, *BTB* refers to a boosting model based on binomial loss function with decision trees.

United States: 120/80 equity strategy, January 2002 to December 2013

	Average returns	Volatily	SR	MDD
SPB	0.110	0.149	0.74^{**}	-0.43
RF	0.107	0.147	0.72	-0.43
BTB	0.109	0.146	0.75^{**}	-0.44
Prob	0.094	0.173	0.54	-0.57
MS	0.101	0.171	0.59	-0.56
EN	0103	0.161	0.64	-0.51
Acc	0.099	0.177	0.56	-0.58
Slow	0.066	0.118	0.56	-0.43
Random	0.092	0.155	0.59	-0.51
Benchmark	0.083	0.147	0.56	-0.51

Note: ** indicates the model is in the set of best models $\widehat{M}^*_{75\%}$.

United States: dynamic asset allocation, January 2002 to December 2013

	Average returns	Volatily	SR	MDD
SPB	0.091	0.090	1.0**	-0.18
RF	0.088	0.088	0.98	-0.18
BTB	0.091	0.087	1.0**	-0.20
Prob	0.074	0.113	0.66	-0.39
MS	0.075	0.101	0.74	-0.28
EN	077	0.098	0.79	-0.25
Acc	0.075	0.116	0.65	-0.42
Slow	0.060	0.058	1	-0.18
Random	0.076	0.095	0.79	-0.30
Benchmark	0.068	0.085	0.79	-0.31

Note: ** indicates the model is in the set of best models $\widehat{M}^*_{75\%}$.

Classical evaluation criteria in the euro area, January 2002 to December 2013

	QPS	**AUROC**
SPB	0.12**	0.90
RF	0.11**	0.91
BTB	0.12**	0.90
Prob	0.25	
MS	0.20	
EN	0.15	
Acc	0.45	
Slow	0.54	
Random	0.48	

Note: ** indicates the model is in the set of best models $\widehat{M}^{*}_{75\%}$.

Turning point signals of the reference cycle in the euro area

	SPB	*RF*	*BTB*
Trough: September 2003	1	1	0
Peak: May 2004	11	9	10
Trough: May 2005	4	3	4
Peak: October 2007	-1	1	-2
Trough: August 2009	1	3	2
Peak: June 2011	-1	-2	-2
Trough: March 2013	2	2	3

Note: Value shown is the model-implied peak/trough calculated using a 0.5 threshold. The minus sign refers to the lead in which the models anticipate the turning point dates. "-" indicates that the model did not generate any signal. *SPB* refers to a boosting model based on squared error loss with P-splines, *RF* refers to a random forest model, *BTB* refers to a boosting model based on binomial loss function with decision trees.

Euro area: 120/80 equity strategy, January 2002 to December 2013

	Average returns	**Volatily**	**SR**	**MDD**
SPB	0.085	0.161	0.53**	-0.46
RF	0.083	0.160	0.52**	-0.46
BTB	0.079	0.158	0.50	-0.46
Prob	0.075	0.182	0.41	-0.48
MS	0.076	0.178	0.43	-0.47
EN	078	0.169	0.46	-0.47
Acc	0.077	0.207	0.37	-0.61
Slow	0.051	0.138	0.37	-0.43
Random	0.076	0.182	0.42	-0.53
Benchmark	0.064	0.173	0.37	-0.54

Note: ** indicates the model is in the set of best models $\widehat{M}^*_{75\%}$.

Euro area: dynamic asset allocation, January 2002 to December 2013

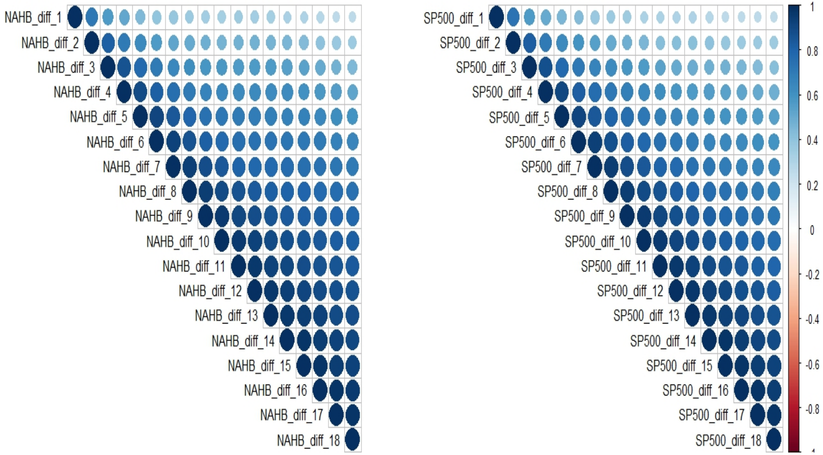
	Average returns	**Volatily**	**SR**	**MDD**
SPB	0.081	0.094	0.86^{**}	-0.21
RF	0.080	0.093	0.86^{**}	-0.22
BTB	0.075	0.091	0.83	-0.22
Prob	0.064	0.114	0.56	-0.25
MS	0.069	0.105	0.66	-0.24
EN	071	0.098	0.72	-0.23
Acc	0.060	0.137	0.44	-0.44
Slow	0.052	0.070	0.75	-0.21
Random	0.064	0.115	0.55	-0.32
Benchmark	0.06	0.100	0.55	-0.34

Note: ** indicates the model is in the set of best models $\widehat{M}^*_{75\%}$.

Conclusion

- Timing the market based on the indicators is possible in real time
- Ensemble machine learning algorithms are effective
- Depending on the data and the objective, random forest sometimes performs better than boosting, sometimes not
- Further work:
  - Economic turning points forecasting (business cycles ?)
  - New features (google trends, news-based sentiment values,...)
  - Deep learning

Appendix: Correlations between lagged variables



References I

Berge, T. (2015). Predicting Recessions with Leading Indicators: Model Averaging and Selection over the Business Cycle. *Journal of Forecasting*, 34(6):455–471.

Breiman, L. (2001). Random forests. *Machine Learning*, 45:5–32.

Cenesizoglu, T. and Timmermann, A. (2012). Do return prediction models add economic value? *Journal of Banking and Finance*, 36(11):2974 – 2987. International Corporate Finance Governance Conference.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29:1189–1232.

Hansen, P., Lunde, A., and Nason, J. (2011). The model confidence set. *Econometrica*, 79(2):453–497.

Ng, S. (2014). Viewpoint: Boosting recessions. *Canadian Journal of Economics*, 47(1):1–34.

References II

Schapire, R. E. (1990). The strength of weak learnability. In *Machine Learning*, pages 197–227.

Taieb, S. B., Huser, R., Hyndman, R. J., and Genton, M. G. (2015). Probabilistic time series forecasting with boosted additive models: an application to smart meter data. Technical report.

White, H. (2000). A Reality Check for Data Snooping. *Econometrica*, 68(5):1097–1126.