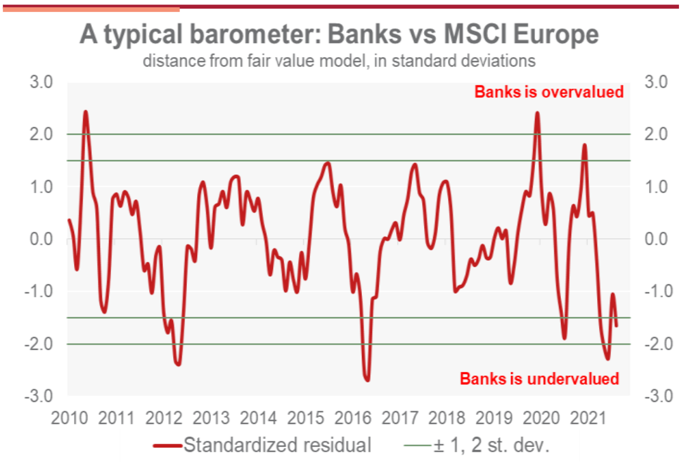
Core Matters

A Machine Learning approach to equity quantitative models

Mattia Mammarella, Michele Morganti, Vladimir Oleinikov, Federica Tartara

14 January 2022

Our Core Matters series provides thematic research on macro, investment and insurance topics



Source: Datastream, GIAM calculations

- This paper presents a real-life application of Machine Learning (ML) through Genetic Algorithm (GA): a step forward from traditional econometric modelling.
- Efficient search-based analytics, implemented with the use of GA, results in a considerable reduction of time needed to develop models, and an increased predictive ability. This is done by maximising results from the out-of-sample (OOS) back testing, which is the theoretical profit derived from investment positions based on the signals coming from the models.
- Prior to 2020, we had developed "barometer models" for 16 European equity sectors and 12 styles, using a classical regression-based approach, to support equity investment decisions. By construction, the models are set up to exploit large deviations from the fair value in trading over/under-valued sectors against the market (MSCI Europe), thus justifying the description given to them – barometer models.
- The performance is measured by the profits under a simple stylized trading rule: Sell (buy) the over-valued (under-valued) sectors. In the traditional approach, finding a well-specified model has proved to be rather time consuming. Furthermore, there is also no guarantee that a *good* fair value model would bring a satisfactory (not to speak of the best) investment performance. On the other hand, GA are used for solving optimisation problems in machine learning, helping to find the best possible model in terms of out-of-sample results. In our initial application of the ML procedure, we obtained models for 31 European equity sectors, 12 European equity styles and 12 equity emerging markets.
- Our investment decisions are not solely based on relative over/under-valuation coming from our models; we also use additional complementary quantitative tools as well qualitative analysis.
- Next steps: increase the number of asset classes to cover with the described approach.

CONTENT

1. Introduction

In this paper we describe a new more efficient selection approach applied to quantitative equity models. We draw a comparison to a “traditional fundamental approach”, which is based on the econometric methodology of general-to-specific modelling. The latter is intrinsically time consuming, and, by construction, does not necessarily produce the best results in out-of-sample (OOS) back testing. In this regard, the process of evaluation and model selection using machine learning (ML) approach proves to be superior. Through automation, it handles data more efficiently, significantly cutting the workload and development time (from weeks to 1-2 days) while simultaneously maximising the results of the models in an out-of-sample back test. In addition to the new model selection approach, we highlight qualitative analysis used to complement ML results and thus, to finalise portfolio choices.

2. Traditional approach to Fair value barometer models

Prior to 2020, we developed “barometer models” for 16 European equity sectors and 12 for styles, using a classical regression-based approach, to support equity strategic investment decisions. These are traditional fair value econometric models that link the equity sector/style relative performance versus the broader EU index (MSCI Europe) to a set of relevant macroeconomic and financial variables. Our dataset covers more than 200 explanatory variables, which include, among others, activity indicators (orders, industrial production, both actual and surveys, inflation, GDP, labour market measures), government bond yields (mainly German and US), corporate spreads, forex, commodities, volatility and equity-related financial variables (earnings’ estimates, etc.).

The equity regression-based models

The models consider relative returns. Acceptable specifications are either of the kind

$$log(sector/market) = \alpha + \sum_k \gamma_k x_k + u$$

or of the (equivalent, transformed) kind

$$log(sector) = \alpha + \beta \cdot log(market) + \sum_k \gamma_k x_k + u$$

Specifications explicitly modelling a sector but having the market as an independent variable are therefore theoretically acceptable as they can always be transformed into a relative variable. However, in our regressions, the dependent variable is of a relative kind.

The basic idea behind our approach is that the regression’s fitted value represents the fair value to which the actual value must revert, “sooner or later” [1]. As a result, (large) deviations from the fair value can be exploited in trading over/under-valued sectors against other sectors or the market, thus justifying the description given to the models – barometer models.

In statistical terms, a model possesses a “mean-reversion” property, if there is a **cointegrating relationship** among the dependent and independent variables**,** i.e. the model’s residuals are stationary. Stationarity means that the statistical properties of a time series do not change over time. On the other hand, if a time series has a unit root, it shows a systematic pattern that is unpredictable. A stationarity test is typically performed using an Augmented Dickey-Fuller, or ADF test: rejection of the null hypothesis assuming the presence of a unit root is a necessary requirement of the validity of barometer models. In other words, a model with unit root residuals would not represent any kind of “long-run equilibrium” and will not guarantee any kind of reversion towards the fair value.

2.1. Formalised rule

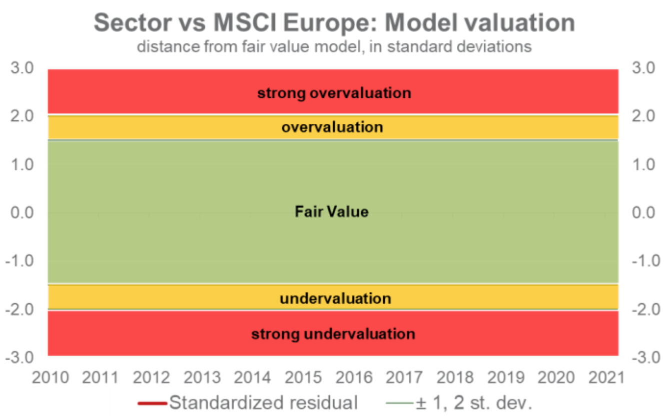
The behaviour of a rule-based trader can be mimicked as follows. Estimate a fair value model satisfying cointegration conditions. Define a “large deviation” from the fair value as a residual beyond a given boundary level $\bar{u}$. Then, if a residual $\bar{u}t$ crosses the upper (lower) boundary - which is what we call

[1] measures of the average mean-reversion speed, better interpreted with the concept of half-life, can be computed

an over- (under-) valuation signal - a trigger is pulled, and the relevant sector is sold (bought) in *t* and kept for a predetermined time of *s* periods.

Model residuals give the signal for our strategy

In practice, the residuals of the regression (i.e. the deviation of the relative index from its fair value) are used as a measure of over/under-valuation (OV/UV). To ensure comparability across models we use standardized residuals, which are plotted in the “barometer” chart.



Source: Datastream, GIAM calculations

Residual values are considered to be large when they are above 2 standard deviations or below -2 standard deviations. If its standardized residuals lie between +1.5 and -1.5 standard deviations, a sector/style is considered **fairly valued**. If the barometer shows values above +2 (below -2) standard deviations the sector/style is judged to be **strongly overvalued (undervalued).** To make sure we do not miss signals, we put on our **watch list** sectors and styles with standardized residuals approaching extreme values (i.e. above +1.5 standard deviation or below -1.5 standard deviation[2]).

In our monetisation strategy we assume that the sale of the sector/style is invested in (its purchase is financed by selling) the market. The return produced by the trade can be measured by the average relative return on the sector between *t* and *t+s*, where *s* is the time for which we keep the position open (*s* is 3, 6 and 12 months in our models).

We also consider a modification, implemented here as well, which consists in waiting for the signal to persist for *p* periods before taking action. In this case, if there happens an UV (or OV) signal in *t*, the sector will be “bought” (sold) in *t+p,* only if all residuals between *t* and *t+p* also produce an UV signal (else no action is taken), and held until *t+p+s*.

So, the strategy will be initiated if there is a signal, no action (and therefore lead to zero relative returns) **over all other periods**

3. Ex-post assessments of performance

To measure the performance of a model-based trading rule, we should apply the above stylized investment rule and calculate the average of returns from each investment decision. This can be done very simply, checking the residuals of the fair value model at *t* and “investing” accordingly. If there are multiple consecutive signals, the trade would be triggered only at the first period. We got a completely out-of-sample procedure of back testing to obtain more reliable results. One way of replicating the state of information of an investor at time *t* is to estimate the model recursively; at each period we would make investment decisions based on data **which would be known to somebody having to make the choice in *t*.** For every *t*, the returns are based on strategies based on **projecting out of** the current training sample.

4. The ML / OOS-based model selection

The traditional approach described above runs into **two issues**: first, it is **time consuming**. For any sector/style, an econometric model plus back test procedure must be estimated, satisfying the statistical and economic rules that make it valuable. Second, there is **not any guarantee that a good fair value model brings a satisfactory (a fortiori optimal) performance**. By construction, the fair value models are chosen according to their economic consistency and statistical significance, which does not mean that their back test will automatically bring optimal results.

Traditional approach is time consuming and does not give any guarantee of a satisfactory performance.

It can then make sense to base a model selection procedure on the specifications that **do best at OOS.**

In practice, after having spelled out the OOS criteria for evaluating model performance ex-post, we endeavour in proactively selecting the most effective model specification using the OOS statistic (such as R^2 or other indicators of fit). Again**, time becomes crucial.**

[2] which are nevertheless proven to signal trades with positive returns

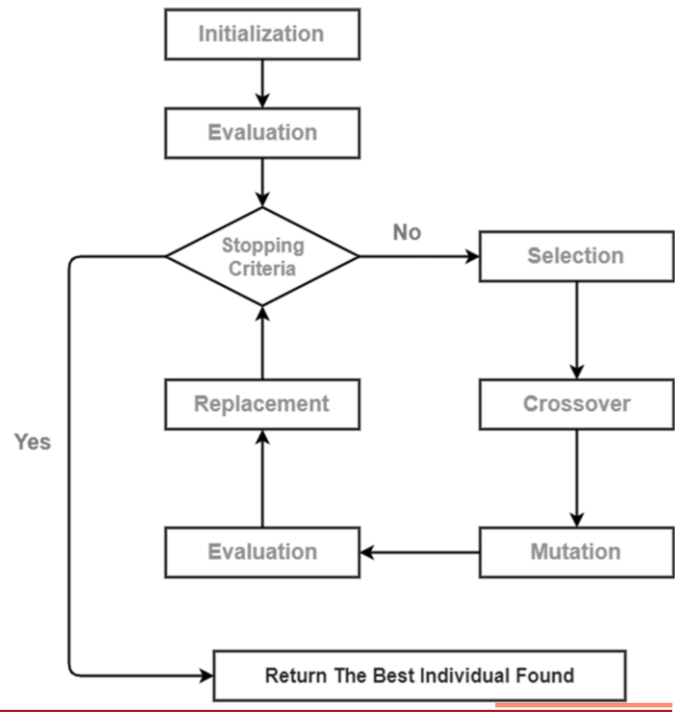
5. Brief Genetic algorithm (GA) description

In the last 15 years in economic and financial literature there has been growing attention to the problem of managing financial datasets with thousands of potentially relevant time series. Computationally straightforward solutions have been proposed. One of them is to apply a Machine Learning (ML) framework to tackle the problem.

ML is a branch of artificial intelligence that use statistics to find patterns in huge amounts of data. **A genetic algorithm (henceforth GA) is a search-based algorithm used for solving optimisation problems in Machine Learning**. It is a stochastic procedure[3] that uses the biological paradigm of evolution to solve optimisation problems and is usually applied in complex systems with great dimensions and several constraints.

In general, **the algorithm operates on a set of randomly generated potential solutions, called chromosomes, applying the concept of survival of the fittest to produce incrementally improved approximations of the best solution via cycles of differential replication, recombination, and mutation.**

Generally, a Genetic Algorithm works as follow. **1.** An **initial sample** of potential solutions is generated and evaluated based on fitness function. A **stopping criterion** is then chosen to define when the cycle will stop: according to literature, 200-300 repetitions are sufficient[4]. **2.** In the **selection phase** couples of parents are created from the initial population. Couples are randomly selected; from each couple the best individuals are chosen; then the best individuals are combined in couples of parents. **3.** In the **crossover phase**, couples of parents generate a new individual. **4.** In the **mutation phase**, the initial parents are modified randomly. The totality of the population is evaluated according to the fitness function selected and ranked. **The best models become the new initial population, and the cycle restarts.**



At the end of the cycle, when the stopping criteria is met, the models are ranked one last time to return the best individual.

6. GA out-of-sample maximisation

To identify, in a reasonable computational time, models that are suitable to forecast over/under-valued sectors, as well as guarantee the statistical soundness, we implemented a Genetic Algorithm technique. **The procedure finds the best possible model in terms of out-of-sample results that satisfies a series of constraints.**

In particular, **the criteria for the decision rule are**:

- Simultaneously maximize the mean return for the UNDER-valued signals at 3, 6 and 12-months
- Simultaneously minimize the mean return for the OVER-valued signals at 3, 6 and 12-months

These objectives are maximized under **the following constraints**:

- All regressors in the model must be significant (β with a low p-value)
- Residuals must be stationary (ADF test with a low p-value)
- Suitable goodness-of-fit (adjusted $R^2 \geq$ given threshold)

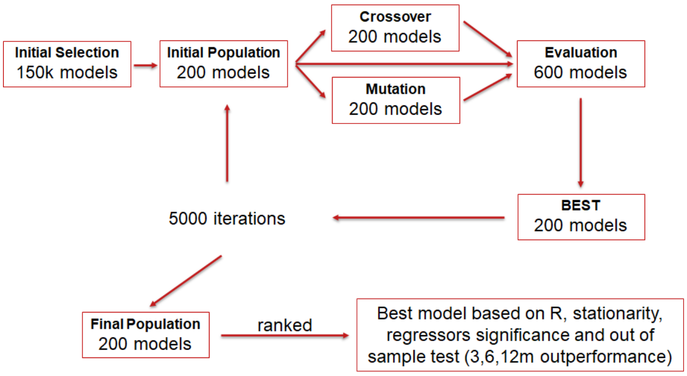
[3] Predicting EU Energy Industry Excess Returns on EU Market Index via a Constrained Genetic Algorithm, Massimiliano Kaucic, Comput Econ (2009).

[4] but for our purpose we decided to set this number up to 4,000-5,000 repetitions

We define the fitness function by combining the **six out-of-sample performance measures** (3M, 6M, 12M applied to over/under-valuation signals) through TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution), assigning equal weight to each goal (i.e. the 3M performance has the same weight as of the 12M performance in the out-of-sample testing).

We set the objectives and the constraints.

The GA procedure comes back with the best (maximum OOS gain) model



TOPSIS Technique

The basic principle of TOPSIS (technique for order preference by similarity to an ideal solution) is that the chosen alternative should have the shortest distance from the ideal solution and the farthest distance from the negative-ideal solution. It consists of the following steps:

1. Calculate the normalised decision matrix
2. Calculate the weighted normalised decision matrix
3. Determine the ideal and negative-ideal solution
4. Calculate the separation measures, using the n-dimensional Euclidean distance.
5. Calculate the relative closeness to the ideal solution.
6. Rank the preference order.

Our GA procedure

The algorithm needs an initial population of at least one individual that satisfies all the constraints. We first generate a large number (150,000) of K-sized candidate models according to the following strategies:

- 50k random sampling
- 50k sampling based on the k-means clustering for the regressors (variables are clustered into 8 main groups)
- 50k sampling based on the correlation between each regressor and the dependent variable

Next, stepwise regression (for each model we test every sub-model until all the variables included are significant) and feasibility tests (for each model we test for the fulfilment of the constraints set) are used to identify the initial population of 200 models.

If the initial population is too large (>200), we select the subset of the best individuals by TOPSIS. If the initial population is too small (<200), we randomly duplicate the selected individuals.

The procedure (STEPWISE-GA-TOPSIS) runs. In the Selection phase from the initial population, we create couples of parents. We randomly select 400 couples; from each couple we select the best individual (400 individuals); we then combine the 400 individuals in 200 couples of parents. In the Crossover phase from the 200 couples of parents selected, we create 200 new individuals. As a result, we have 200 new children. In the Mutation phase from the initial 200 individuals, we generate 200 mutated individuals. We mutate the initial population as it represents the best individuals from all the previous iteration, so we look for small local modification that improve the already evaluated models. As a result, we get 200 new mutated individuals. Then we evaluate all the individuals we created (200 initial + 200 from crossover + 200 from mutation = 600). Lastly, we replace the initial 200 models with the best 200 models from the evaluation (from 600 to 200).

After G generations (4,000-5,000 iterations), the algorithm has identified the best model - within the last 200 models resulted from the iterations - that satisfies the in-sample prescription and has the best out-of-sample performance. In the end one model is chosen among over 1 mln scrutinised.

7. Further procedure checks: results validation and stability tests

We take additional steps to control for the goodness of models which were found by the above procedure. To this end, an assessment is performed ex-post, using the brute-force model search (see box).

Brute-force model search

Consider a set of N potential explanatory variables, and set the number of regressors we want to include in the model at K. A brute-force search can be conducted over the space of **all possible** models, which are the number of combinations choosing subset K from N. All possible specifications are then, ranked by OOS.

The search problem scales linearly with the number of models to be estimated (the number of available combinations), which is a function of N and K; and with the number of recursions for each model, which is a function of T.

Once a parsimonious algorithm has selected the optimal models with *K* regressors out of a set of *N*, the brute-force search procedure can be employed as a validation device as follows. Take a subset *M* (20 in our case) of feasible dimension from the original *N*, including all "optimal" regressors; run the brute-force search; the latter should return the very same optimal choice of regressors found by GA procedure. This was indeed the case.

In addition to this, we compared the out-of-sample results of the models found by the described procedure with the "old" classical approach ones. All of them show higher levels of performance. Consider the example of the Energy sector: the classical model had an **out-of-sample** average gain of +3.7%, while the new model provided on average +21.7%.

We also performed statistical stability tests on out-of-sample performance, to validate the use of the final models, out of the training sample in particular. According to the Chow's forecasting ability test, over a six-month forecasting horizon, only 6 models out of 43 fail the test at the 5% confidence level; another 5 fail at the 10% confidence level.

8. Actual results of the model

In our initial application of the ML procedure, we obtained a total of 43 models: **31 European equity sectors** (all M1 and M2 sectors) and **12 European equity styles**.

Moreover, we applied the same methodology for the equity emerging markets, and in the process created the barometer of the **EM equity markets** (relative to the MSCI EM) **for 12 indices**: China (MSCI & A-Share), Taiwan, Korea, India, Brazil, South Africa, Russia, Thailand, Mexico, Poland and Czech Republic.

31 European equity sectors, 12 European equity styles, 12 EM equity market

Index	ML Approach Current ST. DEV.	 Over/Undervalued OV/UV
MSCI Europe	-	-
Banks	-1.47	N
Insurance	-1.48	N
Div. Financials	3.12	Strong OV
Capital Goods	0.32	N
Transportation	1.67	OV
Comm. & Prof. Services	2.81	Strong OV
Pharma	1.07	N
HC Equip. & Services	-1.17	N
Energy	1.42	N
Telecom. Services	-4.36	Strong UV
Media & Entertainment	-4.07	Strong UV
Utilities	-0.07	N
Materials	-2.15	Strong UV
Food & Staples Retailing	1.74	OV
Food, Beverages & Tobacco	-1.06	N
Household & Pers. Products	-2.66	Strong UV
Cons. Durables and Apparel	-2.12	Strong UV
Cons. Services	-4.19	Strong UV
Retailing	-3.16	Strong UV
Automobiles & Components	-1.87	UV
Software	0.44	N
Tech. Hardware & Equipment	2.38	Strong OV
Semiconductors	3.52	Strong OV
Real Estate	-1.63	UV
Cyclicals	-1.53	UV
Defensive	0.94	N
Value	-0.56	N
Growth	3.49	Strong OV
Quality	0.60	N
Momentum	-0.56	N
Large Cap	-1.23	N
Large Cap Value	-1.63	UV
Small Cap	-2.02	Strong UV
Small Cap Growth	1.24	N
Low Leverage	0.87	N
Minimum Volatility	0.45	N

UV = Undervaluation OV = Overvaluation

We checked for the fair value change in the recent past (1-3-6 months) as a confirmatory signal. If we see an over-valued sector with decreasing fair value, the signal will turn even stronger in the near future. If fair value is rising, then the

OV signal gets weaker. A similar reasoning could be done for under-valued sectors/styles.

At the present the fair value of the models changed as follows:

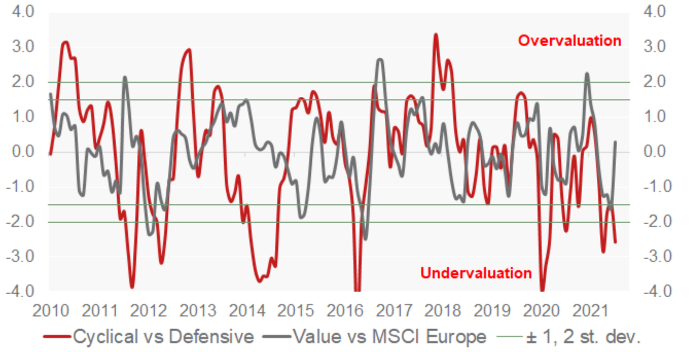
Index	Fair Value Change ML Approach		
	1M chg	3M chg	6M chg
MSCI Europe	-	-	-
Banks	2.7%	3.7%	14.9%
Insurance	1.4%	1.3%	2.2%
Div. Financials	-3.3%	-9.5%	-16.1%
Capital Goods	0.2%	-0.3%	-1.4%
Transportation	-0.4%	-5.2%	-3.3%
Comm. & Prof. Services	1.2%	-0.1%	-1.8%
Pharma	1.8%	2.2%	3.7%
HC Equip. & Services	1.5%	1.7%	3.0%
Energy	-1.1%	-0.2%	-1.3%
Telecom. Services	0.6%	5.5%	5.5%
Media & Entertainment	2.8%	11.3%	12.3%
Utilities	1.2%	1.1%	-0.1%
Materials	0.3%	-1.8%	-1.4%
Food & Staples Retailing	1.2%	1.2%	1.1%
Food, Beverages & Tobacco	1.4%	0.6%	0.5%
Household & Pers. Products	2.6%	8.5%	6.7%
Cons. Durables and Apparel	0.1%	4.9%	13.6%
Cons. Services	0.0%	2.3%	6.9%
Retailing	-0.6%	-6.8%	-7.2%
Automobiles & Components	-0.6%	0.3%	7.3%
Software	1.9%	2.3%	0.7%
Tech. Hardware & Equipment	0.2%	-5.0%	-12.6%
Semiconductors	-1.5%	-4.5%	-16.3%
Real Estate	0.2%	-5.0%	-12.6%
Cyclicals	-0.3%	-1.6%	3.5%
Defensive	-0.1%	-2.3%	-0.3%
Value	-0.8%	-1.2%	-2.5%
Growth	0.4%	-2.1%	-3.6%
Quality	-0.4%	3.1%	3.9%
Momentum	1.0%	2.6%	2.9%
Large Cap	0.0%	0.0%	0.0%
Large Cap Value	-0.4%	-1.6%	-2.3%
Small Cap	0.2%	-0.4%	-3.0%
Small Cap Growth	-0.4%	-0.8%	-7.9%
Low Leverage	-0.9%	-2.7%	-1.8%
Minimum Volatility	-1.0%	0.3%	3.8%

We also monitor **the trend of the standardised residuals to detect possible trend reversals**.

Furthermore, we combine the result of pairs of models to have **relative valuations of a sector vs another one**. Illustrative examples for some selected sector models (incl. cyclicals vs defensive and value vs growth) are presented below.

MSCI Europe Styles: Value and Cyclicals

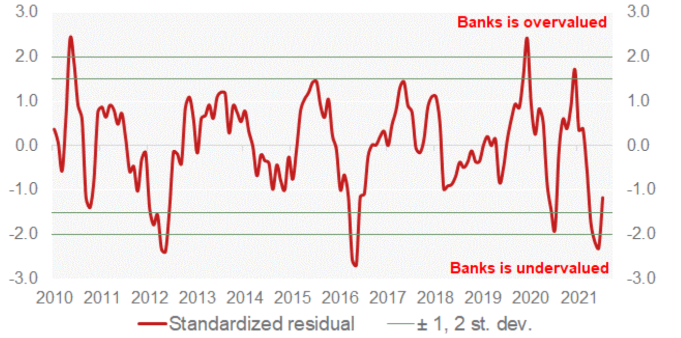
distance from Machine Learning fair value model, in SD



Source: Datastream, GIAM calculations

Banks vs MSCI Europe: Model valuation

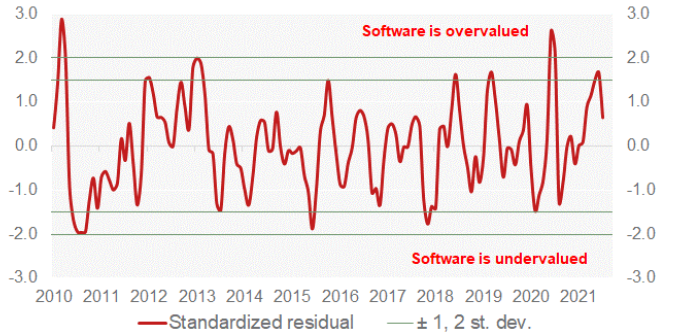
distance from fair value model, in standard deviations



Source: Datastream, GIAM calculations

Software vs MSCI Europe: Model valuation

distance from fair value model, in standard deviations



Source: Datastream, GIAM calculations

Again, **in itself an indication of relative over/under-valuation is not sufficient to make investment decisions**, as the models use historical and current inputs of the macroeconomical and financial variables, without making considerations of how they can develop going forward. Still, the results are useful as they narrow down the number of sectors/styles to be further analysed, using a more qualitative assessment and a bottom-up analysis.

We started this GA optimisation for EU equity sectors and styles; then we moved to EM equity markets. Next, we will use the procedure for other developed equity markets and for cross asset classes (such us equities versus bonds).

In the next chapter we propose a brief review of other tools and analysis we take into consideration to complement this approach.

9. Complementary Qualitative analysis

First, we want to know in which **phase of the cycle** we stand. For this purpose, we developed a tool based on NIPA profits on GDP (corporate margins). It is based on the trend in the pre-tax US NIPA profits/GDP, transformed into the distance from its 5Y rolling peak; this approach breaks the cycle into four phases: recovery, expansion, slowdown and recession. Knowing in which phase of the cycle we are helps us giving a tilt to different styles in terms of potential outperformance or under-performance versus the benchmark. Like Growth outperforming during slowdowns or Value during recoveries.

Second, we developed a proprietary **combination of different valuation methods**, to try to have a lower bias towards value (trap) than is usually the case when using valuation tool. Indeed, we take into good consideration also the **expected earnings growth** development, the specific sector risk (**beta**) and a standardization of relative market multiples versus the specific sector history. We use different **market multiples** compared to historical average: the Shiller PE or CAPE (cyclically adjusted PE, where earnings are adjusted for the cycle, using the last 10-year average instead of the current estimate), the total return based on the expected dividend yield plus earnings growth as well as the PE adjusted for growth, beta, cost of capital and ROE.

Finally, for every sector we consider the relative performance achieved in the last year (proxy for **positioning**) as well as the current trend in relative earnings and sales **revisions** versus the MSCI Europe.

We combine the qualitative measures to come out with a rank across markets or sectors. The outcome is used to assess our degree of confidence in markets/sectors to be put on OW (UW) based on GA models.

10. Conclusions

We have delivered a new tool to obtain signals on over/under-valuation of a market, a sector, a style, and whichever asset class we want to put on radar screen. The signals come from models that are now built with the goal to produce best OOS results. In the past, signals from traditional models of over/under-valuation came from the best econometric model, which was no guarantee of satisfactory OOS performance.

Through automation, data get handled more efficiently and the procedure cuts the workload and development time quite appreciably. This is where ML techniques come in, considerably reducing the time spent (from weeks to days) on building the quantitative tools an analyst needs to perform her job. However, they still require that a proper framework, inside which the process has to run, is set and that results are interpreted, calibrated and possibly complemented by qualitative analysis.

The method presented can be extended to more asset classes and may be improved in setting more stringent goals and constraints for the asset class under scrutiny.

IMPRINT

Issued by:	**Generali Insurance Asset Management S.p.A.** **Società di gestione del risparmio, Research Department**
Head of Research:	**Vincent Chaigneau**
Head of Macro & Market Research:	**Dr. Thomas Hempell, CFA**
Team:	**Elisabeth Assmuth |** Research Operations **Elisa Belgacem |** Senior Credit Strategist **Radomír Jáč |** GI CEE Chief Economist **Jakub Krátký |** GI CEE Financial Analyst **Michele Morganti |** Head of Insurance & AM Research, Senior Equity Strategist **Vladimir Oleinikov, CFA |** Senior Quantitative Analyst **Dr. Martin Pohl |** GI CEE Economist **Dr. Thorsten Runde |** Senior Quantitative Analyst **Dr. Christoph Siepmann |** Senior Economist **Dr. Florian Späte, CIIA |** Senior Bond Strategist **Guillaume Tresca |** Senior Emerging Market Strategist **Dr. Martin Wolburg, CIIA** | Senior Economist **Paolo Zanghieri, PhD |** Senior Economist
Head of Insurance and AM Research:	**Michele Morganti**
Team:	**Raffaella Bagata |** Research Operations **Alberto Cybo-Ottone, PhD |** Senior Economist **Mattia Mammarella |** Research Analyst **Roberto Menegato |** Senior Insurance Research Analyst **Antonio Salera, PhD** | Economist, Pension Expert **Federica Tartara, CFA |** Senior Economist
Head of Credit Research:	**Vivek Tawadey**