A Machine-Learning View of Quantitative Finance

Prof. Stéphan Clémençon - Institut Mines Telecom
LTCl UMR Telecom ParisTech & CNRS No. 5141
Agenda

• The Big Data Era - Challenges for Applied Maths
• Machine-Learning - State of the Art
• An Illustrative Example: the PRISMS Project (with BNP Exane)
Machine-Learning & The Era of Big Data

- Ubiquity of sensors

- Gargantuan data repositories
  - web, finance, genomics, marketing, (e-) commerce, industry, defense/security ...

- Data deluge everywhere!
  - high dimensional, massive, heterogenous, (semi-) structured, streams, on-line processing, ...

- Many traditional statistical methods are obsolete:
  - no scalability, batch, a lengthy preprocessing is often required, strong/rigid model assumptions are involved, ...

- Machine-Learning algorithms are efficient generic and automated approaches to process such data
Machine-Learning & The Era of Big Data

- Ubiquity of sensors

- Gargantuan data repositories
  - web, finance, genomics, marketing, (e-) commerce, industry, defense/security ...

- Data deluge everywhere!
  - high dimensional, massive, heterogenous, (semi-) structured, streams, on-line processing, ...

- Many traditional statistical methods are obsolete:
  - no scalability, batch, a lengthy preprocessing is often required, strong/rigid model assumptions are involved, ...

- Machine-Learning algorithms are efficient generic and automated approaches to process such data

This can be successfully applied to Quantitative Finance!
The Present Data Era
## Employment statistics (1)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Femmes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15-24 ans</td>
<td>45,5</td>
<td>45,7</td>
<td>45,2</td>
<td>43,9</td>
<td>44,2</td>
<td>42,9</td>
<td>42,1</td>
<td>41,87</td>
</tr>
<tr>
<td>25-49 ans</td>
<td>58,6</td>
<td>60,3</td>
<td>62,1</td>
<td>62,8</td>
<td>64,7</td>
<td>65,4</td>
<td>66,2</td>
<td>67,55</td>
</tr>
<tr>
<td>50 ans et plus</td>
<td>42,9</td>
<td>43,1</td>
<td>44,4</td>
<td>43,9</td>
<td>44,8</td>
<td>45,9</td>
<td>45,2</td>
<td>43,47</td>
</tr>
<tr>
<td>Ensemble</td>
<td>51,5</td>
<td>52,5</td>
<td>53,6</td>
<td>53,6</td>
<td>54,8</td>
<td>55,1</td>
<td>55,1</td>
<td>55,29</td>
</tr>
<tr>
<td>Hommes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15-24 ans</td>
<td>55,6</td>
<td>54,7</td>
<td>53,7</td>
<td>52,2</td>
<td>52,5</td>
<td>52,0</td>
<td>50,4</td>
<td>45,02</td>
</tr>
<tr>
<td>25-49 ans</td>
<td>97,0</td>
<td>97,1</td>
<td>96,9</td>
<td>96,9</td>
<td>96,9</td>
<td>97,1</td>
<td>96,9</td>
<td>96,75</td>
</tr>
<tr>
<td>50 ans et plus</td>
<td>79,5</td>
<td>78,8</td>
<td>79,5</td>
<td>78,8</td>
<td>79,4</td>
<td>78,3</td>
<td>75,4</td>
<td>71,65</td>
</tr>
<tr>
<td>Ensemble</td>
<td>82,5</td>
<td>82,2</td>
<td>82,1</td>
<td>81,6</td>
<td>81,8</td>
<td>81,5</td>
<td>80,4</td>
<td>78,14</td>
</tr>
</tbody>
</table>

http://www.insee.fr/
Employment statistics (2)

males vs. females

http://www.insee.fr/
### Financial Data (1)

<table>
<thead>
<tr>
<th>Dern. Cours:</th>
<th>13.820,19</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heure:</td>
<td>21 sept.</td>
</tr>
<tr>
<td>Variation:</td>
<td>↑53,49 (0,39%)</td>
</tr>
<tr>
<td>Clôture Préc:</td>
<td>13.766,70</td>
</tr>
<tr>
<td>Ouverture:</td>
<td>13.768,33</td>
</tr>
<tr>
<td>Var. sur 1 an:</td>
<td>11.926,80 - 14.121,00</td>
</tr>
<tr>
<td>Volume:</td>
<td>419.389.397</td>
</tr>
</tbody>
</table>

http://fr.finance.yahoo.com/
Financial Data (2)
Netscape Proxy format:

format=%Ses->client.ip% 146.127.62.22 %Req->vars.pauth-user% [%SYSDATE%] "%Req->reqpb.proxy-request% %Req->srvhdrs.clf-status% %Req->vars.p2c-cl% %Req->vars.remote-status% %Req->vars.r2p-cl% %Req->headers.content-length% %Req->vars.p2r-cl% %Req->vars.c2p-hl% %Req->vars.p2c-hl% %Req->vars.p2r-hl% %Req->vars.r2p-hl% %Req->vars.xfer-time% %Req->vars.actual-route% %Req->vars.cl-status% %Req->vars.svr-status% %Req->vars.cch-status%
146.127.123.16 146.127.62.22 - [10/Dec/1997:00:30:09 -0500] "GET http://www.nba.com/bulls/ HTTP/1.0" 200 8816 -- 321 164 359 164 1 SOCKS(146.127.11.3:1080) FIN FIN NON-CACHEABLE
146.127.253.84 146.127.62.22 - [10/Dec/1997:00:30:12 -0500] "GET http://www.pathfinder.com/NY1/images/steel.gif HTTP/1.0" 304 - 304 -- - 443 142 468 142 0 SOCKS(146.127.11.3:1080) FIN FIN UP-TO-DATE
Sequencing the human genome

A DNA sequence identified in the human liver - Database Ensembl
Barcoding of Life Data Systems

http://www.barcodinglife.org/
Statistical Datasets

- Arrays/matrices of numbers
- Strings of characters
- Graphs/Networks
- Functions/Curves/Time Series
Statistical learning - Historical milestones

- 1943: Artificial neuron model - McCullough, Pitts
- 1958: Perceptron algorithm - Rosenblatt
- 1971: Uniform laws of large numbers - Vapnik, Chervonenkis
- 1974, 1986: Backpropagation algorithm
- 1984: CART - Breiman, Friedman, Stone, Olshen
- 1984: Theory of the learnable - Valiant
- 1995: Statistical learning theory - Vapnik
Statistical learning - Main topics

- **Approach:** Statistical Learning Theory ($M$-estimation)
- **Problems:** Classification, Regression, Anomaly Detection, Ranking, Source Separation, ...
- **Frameworks:** Supervised, Unsupervised, Semi-supervised, Transductive, Sequential, Reinforcement, ...
- **Methods:** Decision Trees, Support Vector Machines, Boosting, Random Forests, Kernel Methods, Minimum Volume Sets, ...
- **Concepts:** Risk Minimization, Complexity Regularization, Sparsity
Related fields

- **Computer science**:
  - Information systems
  - Algorithmic analysis

- **Machine Learning**:
  - Efficient methods for high-dimensional data

- **Mathematics**:
  - Linear algebra
  - Stochastic modeling,
  - Probability /Statistics
  - Statistical learning theory
  - Optimization
  - Signal processing
  - Computational harmonic analysis
How to implement Machine-Learning solutions?

- Typology of learning problems: Task - Performance - Experiment
- No Free Lunch!
- Choice of adequate performance criteria
- Decision theory and Risk
- Control the complexity of models
- Validation of predictive models
- Role of resampling
- Monitoring predictive models
Ressources

- **Web:**
  - http://www.learningtheory.org/
  - http://www.kernel-machines.org/

- **Mathematical Programming**
  - C libraries: LIBSVM, ...
  - Python libraries: SciKitLearn, ...
  - MATLAB + SVM Toolbox: Spider, ...
  - The R Project for Statistical Computing
    http://cran.r-project.org/web/views/MachineLearning.html
  - WEKA
  - Orange
  - RapidMiner
Main features of machine-learning methods

- Prediction based on past data
- Nonparametric approach - high dimensional data
- Statistical performance - consistency, rates of convergence
- Complexity calibration and risk assessment
- Optimization methods with efficient implementation
- Access mode of the data: batch, online, one-pass, ...
Tasks Carried Out by Machine-Learning Methods
Unsupervised Learning: Indexing/Filtering Data

- Find **sparse representations** $\mathcal{P}(X)$ of massive data $X$

- **Fast search** among a dictionary $\mathcal{D}$ (splines, wavelets, time-frequency atoms, . . .)

- Choose a loss function, ex: $L_p$ distance

\[
\arg\min_{\mathcal{P} \in \mathcal{D}} \mathbb{E}[(\mathcal{P}(X) - X)^p]
\]

- Matching/projection pursuit, wavelet shrinkage, compressed sensing, ...

- **Ex:** fast segmentation of financial return series
  - Automatically identify (linear) trends and their durations
  - Describe residuals: variance, skewness, kurtosis, etc.
Indexing financial series by adaptive segmentation
Features produced by the unsupervised learning algorithm can be used

- to represent successive market regimes efficiently (clustering)
- to detect market anomalies/opportunities (MV-set estimation vs VaR analysis)
- to feed a supervised learning algorithm:
  - prediction of (trends of) returns
  - portfolio optimization
Clustering based on descriptive features
Describing market regimes
Temporal evolution of market regimes

![Graphs showing the temporal evolution of market regimes with different clusters and proportions over time.](image-url)
Anomaly/novelty detection - Minimum Volume Sets
Multivariate VaR analysis
Summarizing dependence across assets

Conditional dependence structure
- Bayesian networks
- Markov networks
Beyond description: predictive learning

Data → Preprocessing → Descriptors → Descriptive tool

Predictive tool → Learning algorithm → Prediction rule

Monitoring tool

Post processing → Indicators Alarms

Prediction + Confidence + Interpretation
Generic setup for supervised learning

- Random pair $= (X, Y) \sim P$ unknown
- $X =$ observation vector in $\mathcal{X}(\mathbb{R}^d)$
- $Y =$ label in $\mathcal{Y} \subset \mathbb{R}^d$
- Predictor: $g : \mathcal{X} \rightarrow \mathcal{Y}$ in a class $\mathcal{G}$
- Loss function: $\ell : \mathcal{Y} \times \mathcal{Y} \rightarrow \mathbb{R}^+$
- Risk functional (unknown!) $= \text{Generalization error}$

$$L(g) = \mathbb{E} (\ell(Y, g(X)))$$

to minimize over $g \in \mathcal{G}$. 
Empirical Risk Minimization (ERM)

- Data = $D_n = \{(X_1, Y_1), \ldots, (X_n, Y_n)\}$
- Learner: $f : \mathcal{X} \to \mathbb{R}$ in a class $\mathcal{F}$
- Practical loss function: $\tilde{\ell} : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}^+$
  - In general, convexified/penalized surrogate of $\ell$
- Empirical risk functional = Training error

\[
R_n(f) = \frac{1}{n} \sum_{i=1}^{n} \tilde{\ell}(Y_i, f(X_i))
\]

to minimize over $f \in \mathcal{F}$.

- Solution: $\hat{f}_n = \arg \min_{f \in \mathcal{F}} R_n(f)$
Issues - Overfitting
Issues

- Rates of convergence

\[ L(\hat{g}_n) - \min_{g \in \mathcal{G}} L(g) \leq C(n, \mathcal{G})? \]

- Empirical estimation of \( L(\hat{g}_n) \) and validation
  \( \rightarrow \) Cross-validation, leave-one-out, bootstrap estimates

- Interpretation of prediction rules

- Sparse representations

- Online vs. batch learning, parallelization

- Multitask prediction
Reinforcement Learning and Market Impact

Agent décisionnel

Etat

Action

Renforcement

Environnement

incertain
partiellement observable
complexe
P.R.I.S.M.S.*

Risk monitoring and investment recommendations

*Portfolio Risk Identification and Selection Methods based on Statistical learning
Contents

PRISMS in three questions

Process and functionalities
Risk monitoring
Investment recommendations
Appendix
What is it?

- PRISMS is the result of leading-edge research
- PRISMS provides investors with efficient decision-making tools in quantitative portfolio management: selection, prediction, monitoring, risk and optimisation
What does it do?

1 Risk monitoring

- Characterise market dynamics
  - Is there a market change happening?
  - How (a)typical is the current market configuration?

- Assess risks
  - What is the risk-aversion level?

- Identify market drivers
  - Should we first look at inflation, real rate, credit?
  - Is the market showing any arbitrage?

2 Investment recommendations

- Return buy / sell signals
  - On both relative and absolute basis
  - Do we expect a trend reversal?

- Compute confidence indicators for each prediction
  - How likely is a good prediction?

- Identify main drivers for the prediction
  - What drives foremost my prediction?
Why is it different?

- Efficient
  - Good hit rates (55-60%)
  - Emerging market factors captured efficiently

- Transparent
  - Not a black box
  - Automatic calibration
  - Can be used every day

- Innovative
  - Statistical learning
  - Proprietary results
  - Access to the best researchers in the field

- Objective
  - Data-driven
  - Extraction of relevant information
  - No “model” assumed

PRISMS
Contents

PRISMS in three questions

Process and functionalities

Risk monitoring

Investment recommendations

Appendix
Process

1. Market data
   - Several asset classes (FI, FX, commodities, etc.)
   - Relevant information is extracted (filtering)
   - Each predictor is characterised by its momentum and volatility

2. Learning
   - Learn from market data without assumptions
   - Analyse market configurations
   - Monitor large datasets

3. Risk monitoring
   - Characterize the current market configuration
   - Rank market drivers
   - Produce global risk indicators

4. Recommendations
   - Predict trends and reversals (absolute or relative)
   - Return confidence indicators for each signal
In this example:
- The whole market is characterized by only two variables: the Brent price and the Euro-dollar rate.
- The stock we want to predict is Total.

The following tree produces a “map” of the market:
- When the Dollar is up (>0.15%), Total is up with a frequency of 87%.
- When the Dollar is down, there are two cases:
  - Either the Brent is down (<0.45%) : then Total is down with a frequency of 82%.
  - Either the Brent is up (>0.45%) : then Total is up with a frequency of 86%.
Contents

PRISMS in three questions

Process and functionalities

Risk monitoring

Investment recommendations

Appendix
Market dynamics

Main indicators
Latest indicator values

Target asset
STOXX Europe 600 index

Predictability
Assesses at which extent current market configuration can be used to make predictions.
Range: 0 - 100%

Features
Model settings

Market confidence
Measures how (a)typical the current market configuration is.
Range: 0 - 100%

Momentum vs Volatility
Compares the ranks of market drivers when categorised as either volatility drivers or momentum drivers.
Range: (100%) - 100%

Key points
Main conclusions of the market dynamics analysis
Market dynamics
Confidence, predictability and market crashes

Market changes are characterised by divergences between confidence and predictability.

Confidence and predictability indicators

STOXX Europe 600 index

Exaggeration

Shock

New regimes

Sovereign debt crisis

Burst of the money bubble

Burst of the internet bubble

Asian crisis (Oct 1997)

Russian crisis (Aug 1998)

LTCM crash (Sep 1998)
Market dynamics
Momentum vs volatility balance

Momentum-volatility balance steadily switches to the volatility side in crisis periods.

Key events:
- Asian crisis (Oct 1997)
- Russian crisis (Aug 1998)
- LTCM crash (Sep 1998)
- Burst of the internet bubble
- Burst of the money bubble
- Sovereign debt crisis

Exane BNP Paribas | QUANTITATIVE RESEARCH TEAM
Market drivers

- **Driver category mapping**: Represents the average weights of driver classes setting apart momentum and volatility drivers.
- **Top ten market drivers**: Market drivers ranked by predictive power.

### Key points
- Main conclusions on market drivers:
  - The market drivers now reflect the complete interest rate structure for Southern European countries (real rates, inflation and risk premium).
  - Inflation momentum has lost 10% ranks, while credit and currency volatilities are gaining predictive power.
Market drivers
Breakeven inflation in EMU and volatility of the USD-EUR cross-rate

Breakeven inflation and STOXX Europe 600

Volatility of USD and STOXX Europe 600
Contents

PRISMS in three questions

Process and functionalities

Risk monitoring

Investment recommendations

Appendix
Predictions on the STOXX Europe 600 index
Historical hit ratio and market crashes

57% of overall performance on a 25-day horizon

Hit rate alternates between well performing periods (around 63%) and poorly performing periods (around 47%)

Increasing performance periods after shocks or bubbles show the ability to learn on any new market regime, regardless of its trend.
Predictions on the STOXX Europe 600 index

Historical P&L

- Prediction target
  - STOXX Europe 600 index
  - Up/down on a 25-day horizon

- Predictors
  - 64 assets (global indices, commodities, rates, etc.)
  - Trend and volatility for each
  - 128 variables in total

- Back test features
  - Data start in 1990 (STOXX Europe 600 inception)
  - Daily models tested on 25 subsequent days after model training
Sector recommendations

Performance since inception
This chart shows the actual track record of our method since the first publication.

Hit ratio vs confidence chart
The overall increasing relationship between prediction confidence and realised hit ratio shows confidence is a relevant indicator to select recommendations.

Recommendations on sectors
Recommendation table on sector indices relative to the STOXX 600 index, with the optimal horizon:
• Most confident outperforming sectors
• Sectors without any relevant recommendation
• Most confident underperforming sectors

Historical returns of recommendations
This chart gives gross recommendations’ returns implemented with simple positions, without further money management or strategy.
Stock recommendations

**Global performance**

- **Universe:** STOXX Europe 200 Large
- **Portfolio:** Long-short, leverage of 2, weekly rebalancing
- **Hedging:** sectors

<table>
<thead>
<tr>
<th>Performance</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hit rate</td>
<td>55.4%</td>
</tr>
<tr>
<td>Annual return</td>
<td>4.2%</td>
</tr>
<tr>
<td>Volatility</td>
<td>4.1%</td>
</tr>
<tr>
<td>IR</td>
<td>1.02</td>
</tr>
<tr>
<td>Average holding time</td>
<td>15</td>
</tr>
<tr>
<td>Avg # of stocks held</td>
<td>30</td>
</tr>
</tbody>
</table>

**Performance per year**

- **IR (rhs)**
- **Annual performance (lhs)**
- **Volatility (lhs)**

- **Return**
  - 3
  - 5
  - 7
  - 9
  - 11
  - 13
  - 15
  - 17
  - 19

- **Holding time (days)**
  - (2%)
  - (4%)
  - (6%)

- **Good trades**
- **Bad trades**
Contents

PRISMS in three questions
Process and functionalities
Risk monitoring
Investment recommendations

Appendix

Statistical learning at a glance
Inputs and outputs
Risk analysis and portfolio optimization
Statistical learning at a glance
What is statistical learning?

- Learn from data:
  - Analyse great amounts of data.
  - Infer dependence structures from observations.
  - Detect patterns in massive databases.
  - Focus on prediction accuracy and optimization of statistical performance.

- Types of tasks:
  - Supervised: considering known labels associated to data, one wants to learn the relationship between observations and labels.
  - Unsupervised: one wants to assign labels to data elements by clustering them into consistent groups.

- Examples:
  - Medical diagnosis from clinical data, object recognition in images, etc.
  - Speech separation, homogeneous regions clustering, social network peer groups, etc.
Problems can be formulated as learning tasks:

- Market prices representation as successive periods of homogeneous trend is a typical unsupervised problem.
  - Segmenting a single signal into regimes implies clustering dates into homogeneous groups.
  - Market regime identification assigns to each period a cluster label, the regime it belongs to.
  - Such indexing provides a market description.

- Predicted return is a challenging supervised problem.
  - One is interested in a few categories of asset return (up/down, possibly strong/weak trend).
  - Given an observation of the market, one wants to classify the future return into one of the categories.
  - Based on past observations of the dependence between the market and the future return, a classifier model can be learnt.

- Risk factor analysis is a matter of source separation.
  - Analyzing a portfolio behaviour relatively to the market involves breaking it down to predominant factors.
  - Those factors must be independent, as if the portfolio value originated from distinct signal sources.
  - Statistical learning provides robust techniques to perform such a task.
Statistical learning at a glance
Is it well-suited and beneficial?

- **Ability to analyse high dimensional data**
  - Market history is an overwhelmingly large database containing information.
  - Learning algorithms can handle problems with highly multivariate behaviours.
  - Underlying information is automatically analysed and used.

- **Early arbitrage opportunities**
  - We believe there are inefficiencies in the market and that mean reversion is not instantaneous.
  - A model that efficiently learnt from historical data can reveal these phenomena.
  - Such a tool is likely to detect arbitrage opportunities before other observers.

- **Rational information for decision making**
  - Data-driven algorithms bear the minimum hypotheses possible, making them fully objective.
  - Automatically learnt models are complementary to other models and heuristics.
  - Human decision making is enhanced.
Statistical learning at a glance
Methods used in our approach (1/3)

- **Signal indexing representation**
  - Minimization criterion: least square regression error
  - Dyadic Coifman-Wickerhauser (bottom-up)
    - Initial state: consider all dates separated
    - Merge consecutive pairs of dates in case the criterion decreases
    - Iterate until the criterion cannot be improved or the maximum merging degree is reached
  - Unsupervised CART (Classification and Regression Trees, top-down)
    - Initialisation: consider the whole history as one period
    - Find the best splitting date so as to minimise the criterion
    - Iterate until the criterion cannot be improved or the maximum splitting degree is reached

- **Market regime identification**
  - K-Means:
    - Minimise intra-cluster distances
    - Iteratively find cluster “centroids” and data labels

---

Exane BNP Paribas | QUANTITATIVE RESEARCH TEAM
Classification with Random Forests

- Base classifier: decision trees (CART)
  - Objective: find an optimal partition of the input space, defined by a tree.
  - Criterion to minimise: impurity, known as the Gini discriminant index, measuring how mixed a set is.

- Randomization and aggregation
  - In order to increase robustness, several subsets of data are sampled.
  - At each node, the best splitting variable is chosen among a random sub-sample of all the variables.
  - Build trees on the subsets and make them vote.

Extension to several trend levels

- The algorithm is intrinsically designed for any number of classes.
- Pre-processing is necessary to define levels such as strong/weak or downtrend/uptrend.
### Backtesting

- **The method is tested with past events**
  - On each backtest date, a model with the same parameters is built and tested on following days.
  - Performance is assessed by backtest date and number of consecutive test days.
- **Predominant factors are identified.**
  - Each model selects the important variables.
  - Variable ranks evolve in time following the economic environment.
- **Parameters are calibrated to maximise expected performance.**

### Monitoring

- **Predictions are always associated to confidence measures such as:**
  - Similarity to training data,
  - Local known performance in the neighbourhood of the new observation.
- **Those measures can provide alerts on the model’s reliability.**
Possible predictors:
- Economic variables: currencies (4), FI (25), commodities (5), indices (30).
- Stocks (index constituents).
- In the near future, financial variables such as P/E, dividend yields and EV/EBITDA.

Variables to predict:
- Stock prices (trends).
- Index, commodity, currency and FI prices.
- Potentially, volumes, volatilities, and correlations.

Volumetric:
- 650 market data series
- 84 segmentations
- 277 weekly calibrated models
• **Analysis**
  - Data indexing
  - Market regime mapping

• **Prediction**
  - Predicted directions with confidence indicators
  - Market scenarios
  - Assets and predictors ranking

• **Backtesting and monitoring**
  - Monitoring alerts
  - Strategy backtests

**Some outputs:**
- Ready-to-use Excel files: scoreboards, ranking lists
- Saved Matlab models: market regimes, predictive models, backtest analysis
Inputs and outputs
Task complexity

- **Input space dimension:**
  - Input variables: 128 dimensions
    - 64 predictors
    - 2 dimensions per predictor (last trend and volatility)
  - Usual training data: 250 dates
    - 2-year initial daily data
    - 1-year sliding window market representation

- **Usual computing time:**
  - Standard model learning (5 minutes*)
    - Indexing a usual set of predictors: 10 minutes* at the first run, 4 minutes* for subsequent use
    - Training models for 5 classes: 6 seconds* per model
    - Prediction: < 1 second per predicted value
  - Backtesting
    - Learning and testing on 3-year input data
    - 250 models trained and tested: 20 minutes*

*Hardware for time evaluation:
- Processor: Pentium E2180, dual-core, 2 GHz
- RAM: 2 GB
- Storage device: Hard Disk Drive
**Risk analysis and portfolio optimisation**

**Hedging**

- **PRISMS identifies and analyses systematic risk factors**
  - Identify robust market risk factors as shocks
  - Identify key risk factors in portfolio
  - Assess betas as well as systematic and residual risks

- **PRISMS also allows optimising the hedging**
  - Analyse the efficiency of a hedging strategy
  - Find best hedging directions
Why we don’t use the CAPM:
- Its ability to capture the market is questionable.
- Estimates as well as betas are unstable.
- The CAPM relies on a LT equilibrium that does not account for local variations.

We prefer the Independent Component Analysis (ICA):
- ICA detects market chocks and identifies the systematic factors as arbitrage portfolios.
- Systematic factors in ICA are much more robust than those in CAPM.
- ICA identifies not only the overall LT systematic factors but also the local systematic factors that vary around.
Risk analysis and portfolio optimisation

Process

1. Variable to explain
2. Explanatory variables
3. Systematic space

Systematic component of the variable to explain

Systematic components of the explanatory variables

Risk breakdown

Betas

Risk analysis

Process
Here, we analyse the efficiency of a simple hedging strategy for a long-only portfolio.

- Long leg: about 30 European stocks.
- Short leg: each long position is hedged using the corresponding index future.
- The portfolio has been initiated with a zero value.

The results show limited impact of the global hedging strategy.

- Even though the systematic risk has decreased from 67% to 36%, it remains high.

Residual risk analysis shows that the portfolio is not immune to the euro-dollar, for instance.

- Euro-dollar contributes by 29% to the risk of the hedged portfolio.
- Of the residual systematic risk, the proportion of risk attributed to the euro-dollar is 80%.
Hedging is deficient concerning the Health Care sector
Optimisation could be performed in two ways:

- Either hedging factors are fixed (as in the previous example), and the optimisation is an allocation problem:
  - At what price should we buy/sell index futures?
  - How do we rebalance the portfolio?
- Either hedging factors are not determined, and the optimisation consists in finding the best hedging assets:
  - Which assets should be in the hedging leg? And what should be the amounts?
  - Powerful algorithms such as genetic algorithms may be used.

Our methodology is based on powerful algorithms such as genetic algorithms.

Identification of the best hedging assets for a portfolio of stocks, using genetic algorithms

<table>
<thead>
<tr>
<th>Représentation du panier dans les indices</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interlocuteur</td>
</tr>
<tr>
<td>---------------</td>
</tr>
<tr>
<td>ABC</td>
</tr>
<tr>
<td>DEF</td>
</tr>
<tr>
<td>GHI</td>
</tr>
<tr>
<td>JKL</td>
</tr>
<tr>
<td>MNO</td>
</tr>
<tr>
<td>PQR</td>
</tr>
<tr>
<td>STU</td>
</tr>
<tr>
<td>VWX</td>
</tr>
<tr>
<td>YZT</td>
</tr>
<tr>
<td>ABC</td>
</tr>
<tr>
<td>DEF</td>
</tr>
<tr>
<td>GHI</td>
</tr>
<tr>
<td>JKL</td>
</tr>
<tr>
<td>MNO</td>
</tr>
<tr>
<td>PQR</td>
</tr>
<tr>
<td>STU</td>
</tr>
<tr>
<td>VWX</td>
</tr>
<tr>
<td>YZT</td>
</tr>
<tr>
<td>ABC</td>
</tr>
<tr>
<td>DEF</td>
</tr>
</tbody>
</table>


BNP PARIBAS has acquired an interest in VERNER INVESTISSEMENTS the parent company of EXANE. VERNER INVESTISSEMENTS is controlled by the management of EXANE. BNP PARIBAS’s voting rights as a shareholder of VERNER INVESTISSEMENTS will be limited to 40% of overall voting rights of VERNER INVESTISSEMENTS.

Exane BNP Paribas research is also available on the website (www.exanebnpparibas-equities.com) as well as on Bloomberg (EXEQ), First Call, Reuters and The Markets.

Important notice: Please refer to our complete compliance page available on www.exane.com/compliance that includes Exane’s policy for managing conflicts of interest, Rating definitions and complete disclosures per company.

Exane is independent of BNP Paribas (BNPP) and the agreement between the two companies is structured to guarantee the independence of Exane’s research, published under the brand name “Exane BNP Paribas”. Nevertheless, to respect a principle of transparency, we separately identify potential conflicts of interest with BNPP regarding the company(ies) covered by this research document.

This research is produced by EXANE SA and/or EXANE LTD ("EXANE") on behalf of themselves. EXANE SA is regulated by the "Autorité des Marchés Financiers" (AMF) and EXANE LTD is regulated by the "Financial Services Authority" (FSA). In accordance with the requirements of FSA COB 7.16.7R and associated guidelines "Exane’s policy for managing conflicts of interest in relation to investment research" is published on Exane’s web site (www.exane.com). Exane also follows the guidelines described in the code of conduct of the AFEI (Association Francaise des Entreprises d’Investissement) on “managing conflicts of interest in the field of investment research”. This code of conduct is available on Exane’s web site (www.exane.com).

This research is solely for the private information of the recipients. All information contained in this research report has been compiled from sources believed to be reliable. However, no representation or warranty, express or implied, is made with respect to the completeness or accuracy of its contents, and it is not to be relied upon as such. Opinions contained in this research report represent Exane's current opinions on the date of the report only. Exane is not soliciting an action based upon it, and under no circumstances is it to be used or considered as an offer to sell, or a solicitation of any offer to buy.

While Exane endeavours to update its research reports from time to time, there may be legal and/or other reasons why Exane cannot do so and, accordingly, Exane disclaims any obligation to do so.

This report is provided solely for the information of professional investors who are expected to make their own investment decisions without undue reliance on this report and Exane accepts no liability whatsoever for any direct or consequential loss arising from any use of this report or its contents.

ANALYST CERTIFICATION: all of the views expressed in the research report accurately reflect the research analyst’s personal views about any and all of the subject securities or issuers of this research report. No part of the research analyst’s compensation was, is, or will be, directly or indirectly, related to the specific recommendations or views expressed by the research analyst in this research report.

This report may not be reproduced, distributed or published by any recipient for any purpose. Any United States person wishing to obtain further information or to effect a transaction in any security discussed in this report should do so only through Exane Inc., which has distributed this report in the United States and, subject to the above, accepts responsibility for its contents.

BNP PARIBAS has acquired an interest in VERNER INVESTISSEMENTS the parent company of EXANE. VERNER INVESTISSEMENTS is controlled by the management of EXANE. BNP PARIBAS’s voting rights as a shareholder of VERNER INVESTISSEMENTS will be limited to 40% of overall voting rights of VERNER INVESTISSEMENTS.

Exane BNP Paribas research is also available on the website (www.exanebnpparibas-equities.com) as well as on Bloomberg (EXEQ), First Call, Reuters and The Markets.