



HAL
open science

Asset Allocation, Economic Cycles and Machine Learning

Thomas Raffinot

► **To cite this version:**

Thomas Raffinot. Asset Allocation, Economic Cycles and Machine Learning. Economics and Finance. Université Paris sciences et lettres, 2017. English. NNT : 2017PSLED067 . tel-01872176

HAL Id: tel-01872176

<https://theses.hal.science/tel-01872176>

Submitted on 11 Sep 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE DE DOCTORAT

de l'Université de recherche Paris Sciences et Lettres
PSL Research University

Préparée à l'Université Paris-Dauphine

Asset Allocation, Economic Cycles and Machine Learning

École Doctorale de Dauphine — ED 543

Spécialité **Sciences économiques**

**Soutenue le 28.09.2017
par RAFFINOT Thomas**

Dirigée par **Mme EPAULARD Anne**

COMPOSITION DU JURY :

Mme EPAULARD Anne
Professeure, Paris Dauphine
Directrice de thèse

M. HURLIN Christophe
Professeur, Université d'Orléans
Rapporteur

M. FERRARA Laurent
Professeur associé, Paris Nanterre
Rapporteur

Mme BRIERE Marie
Professeure associée, Paris Dauphine
Membre du jury

Mme MIGNON Valérie
Professeure, Paris Nanterre
Membre du jury

M. VILLENEUVE Bertrand
Professeur, Paris Dauphine
Président du jury

L'Université n'entend donner aucune approbation ni improbation aux opinions émises dans les thèses : ces opinions doivent être considérées comme propres à leurs auteurs.

Acknowledgments

I would first like to thank my thesis advisor Prof. Anne Epaulard. Her guidance helped me in all the time of research and writing of this thesis.

Besides my advisor, I would like to thank Prof. Marie Bessec and my thesis committee, Prof. Marie Brière, Prof. Christophe Hurlin, Prof. Laurent Ferrara, Prof Valérie Mignon and Prof. Bertrand Villeneuve for their insightful comments, encouragements and criticisms, witch incentivized me to widen my research from various perspectives.

I must express my very profound gratitude to my parents, brother and sister for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them.

Last but not least, I would like to dedicate this thesis to Raphaëlle and to my children Camille and Zélie.

Contents

Introduction	1
1 Time-varying risk premiums and economic cycles	9
1.1 Cyclical framework	11
1.1.1 Economic cycles definitions	11
1.1.2 Turning point chronology	13
1.2 Asset classes and economic cycles	17
1.2.1 The economic rationale	17
1.2.2 Historical facts	19
1.3 Dynamic investment strategies	21
1.3.1 Active portfolio management and economic cycles	21
1.3.2 Comparison criteria	24
1.3.3 Empirical results	27
1.4 Strategic asset allocation and economic cycles	31
1.4.1 Correlations and economic cycles	32
1.4.2 A truly diversified strategic asset allocation	35
Appendices	41
1.A The double Hodrick-Prescott filter	41
2 Can macroeconomists get rich nowcasting output gap turning points with a simple machine-learning algorithm?	43

2.1	Learning Vector Quantization	46
2.2	Empirical setup	49
2.2.1	Real time recursive estimation	49
2.2.2	Data set	50
2.2.3	Model evaluation	52
2.2.4	Model selection	57
2.2.5	Competitive models	58
2.3	Empirical results	59
2.3.1	United States	59
2.3.2	Euro area	63
Appendices		70
2.A	Explanatory variables	70
2.A.1	Economic Surveys	70
2.A.2	Financial series	73
3	Ensemble Machine Learning Algorithms	75
3.1	Ensemble Machine Learning Algorithms	78
3.1.1	Random forest	79
3.1.2	Boosting	82
3.2	Empirical setup	87
3.2.1	Turning point chronology in real time	87
3.2.2	Data set	89
3.2.3	Alternative classifiers	90
3.2.4	Model evaluation	92
3.3	Empirical results	97
3.3.1	United States	97
3.3.2	Euro area	100

4 Hierarchical Clustering based Asset Allocation	107
4.1 Risk Budgeting Approach	110
4.1.1 Notations and definitions	111
4.1.2 Risk budgeting portfolios	111
4.2 Hierarchical clustering and asset allocation	113
4.2.1 Notion of hierarchy	113
4.2.2 Hierarchical clustering	114
4.2.3 Asset allocation weights	116
4.3 Investment strategies comparison	118
4.3.1 Datasets	118
4.3.2 Comparison measures	123
4.4 Empirical results	127
4.4.1 S&P sectors	127
4.4.2 Multi-assets dataset	128
4.4.3 Individual stocks	129
4.5 Future research	130
Appendices	135
4.A Bond market	135
Conclusion	137

Introduction

Asset allocation is an investment strategy that attempts to balance risk versus reward by adjusting the percentage of each asset in an investment portfolio according to the investor's risk tolerance, goals and investment time frame. Essentially, asset allocation is not putting all of your eggs in one basket when it comes to investing. Having all investments in a single security or issuance can result in the entire portfolio being wiped out if the investment goes bad. Markowitz (1952) quantifies with mathematics the benefits of this concept, by developing the workhorse theory of mean-variance efficiency.

The seminal paper by Brinson et al. (1986) reports that asset allocation (measured as the average quarterly exposure to stocks, bonds, and cash) explained 93.6% of the variability of returns for the total portfolio holdings. Many citations of Brinson et al. (1986) falsely suggest that their analysis makes conclusions about return attribution. Yet, explaining 93.6% of the monthly variance in total returns is not the same thing as saying that the portfolio mix determines 93.6% of the returns.

Ibbotson and Kaplan (2000) recognize the omnipresence of misperception around Brinson et al. (1986) and set out to correct this in their paper. They confirm that asset allocation is the main factor to explain the variability of returns over time: market evolution of asset classes dictates 90% of the movement of your portfolio. Moreover, they answered two related questions: to what degree does asset allocation explain the variability of performance between funds and institutions, and; to what degree does asset allocation explain the level of long-term performance?

To determine how well asset allocation explained the dispersion in returns across funds,

the authors performed a cross sectional regression of returns from funds and institutions against respective policy benchmarks. They determined that 40% of the difference in returns across funds is explained by differences in asset allocation policy, with the balance determined by a combination of tactical shifts, sector bets, security selection, and fees.

Lastly, Ibbotson and Kaplan (2000) perform an attribution analysis to determine the percent of long-term performance explained by asset allocation. They calculated the long-term performance of each fund's policy portfolio and compared it against actual long-term fund returns. They state that, on average, asset allocation explained about 104% of long-term returns. It is surprising: how can asset allocation explain greater than 100% of total returns? as a matter of fact, the total return to portfolios were decomposed into the total return to the fund's policy portfolio using asset class benchmarks, plus the active return, minus trading frictions. So the results of this study demonstrate that, over the periods studied in the analyses, the average institutional investor lost 4% of total return to fees, ineffective active management or poor manager selection.

To sum up, those studies emphasize the importance of asset allocation for investors: diversification is the only free lunch in investing.

In practice, Ilmanen (2011) highlights that institutional practices have evolved from the traditional 60/40 equity/government bond split, *i.e.* 60% of the portfolio allocated to equities and 40% to fixed income securities (bonds), to globally diversified portfolios, often including emerging markets and alternative assets, which were seen as having almost no correlation with traditional stocks and bonds. Large institutional losses in 2008 raised significant questions about the best way to pursue asset allocation: Page and Taborsky (2011) showed that diversification does not accomplish its goals: diversification often disappears when most needed:

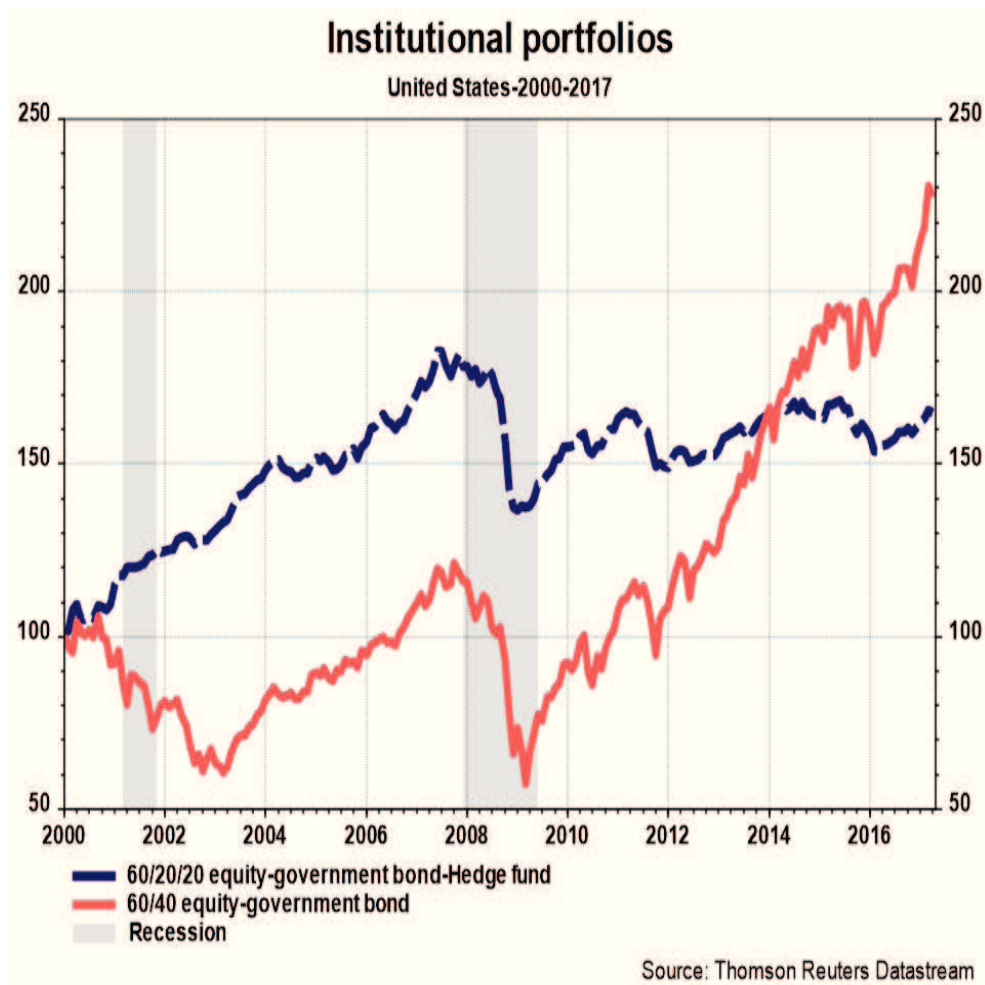


Figure 1: Traditional institutional portfolios (Author's calculations)

To alleviate the difficulties encountered in the context of the Great Recession, this thesis proposes certain topics for reflection and discussion on the measures to be taken to truly diversify portfolios. The first one is to deepen the knowledge of interaction dynamics between financial markets the macroeconomy (Chapters 1 to 3). The second one is to explore a new way of capital allocation (Chapter 4).

With a view to a better understanding of the complex relationship between financial markets and the macroeconomy, a well-worked theory of macro-based asset allocation is introduced in the Chapter 1. The objective is to illustrate that asset returns are not correlated with the business cycle but are primarily caused by the economic cycles. To demonstrate this claim, economic cycles are first rigorously defined, namely the classi-

cal business cycle and the growth cycle, which is better known as the output gap. The description of different economic phases is refined by jointly considering both economic cycles. The theoretical influence of economic cycles on time-varying risk premiums is then explained based on two key economic concepts: nominal GDP and adaptive expectations. It is exhibited over the period from January 2002 to December 2013: dynamic investment strategies based on economic cycles turning points emphasize the importance of economical cycles, especially the growth cycle, for euro and dollar-based investors. An empirical analysis in the United States over the period from January 2002 to December 2013 highlight that this economic cyclical framework can improve strategic asset allocation choices.

In theory, dynamic macro-based regime-switching asset allocations achieve thus superior risk adjusted returns. Yet, economic turning points detection in real time is a notorious difficult task. One stylised fact of economic cycles is the non-linearity: the relationship between variables is not simply static and stable, but instead is dynamic and fluctuating. For example, Phillips (1958) concludes the last sentence of his first paragraph with: "The relation between unemployment and the rate of change of wage rates is therefore likely to be highly non-linear".

Real-time regime classification and turning points detection require thus methods capable of taking into account the non-linearity of the cycles. In this respect, many parametric models have been proposed, especially Markov switching models (see Piger (2011) for a review) and probit models (see Liu and Moench (2016) for a review). Parametric models are effective if the true data generating process (DGP) linking the observed data to the economic regime is known. In practice, however, one might lack such strong prior knowledge.

Non-parametric methods, such as machine-learning algorithms¹, do not rely on a specification of the DGP. The machine-learning approach assumes that the DGP is complex

¹Machine learning generally refers to the development of methods that optimize their performance iteratively by "learning from the data". Machine learning is broadly understood as a group of methods that analyse data and make useful discoveries and inferences from the data.

and unknown and attempts to learn the response by observing inputs and responses and finding dominant patterns. This places the emphasis on a model's ability to predict well and focuses on what is being predicted and how prediction success should be measured². Machine learning is used in spam filters, ad placement, credit scoring, fraud detection, stock trading, drug design, and many other applications, but it is largely unknown in economics (with the exception of Giusto and Piger (2014), Ng (2014) and Berge (2015)).

The real-time ability of several machine learning algorithms (from very simple to quite complex) to nowcast economic turning points is gauged in Chapter 2 and Chapter 3. The aim is to quickly and accurately detect economic turning points in the United States and in the euro area.

In Chapter 3, probabilistic indicators are created from a simple and transparent supervised machine-learning algorithm known as Learning Vector Quantization (Kohonen (2001)). Those indicators are robust, interpretable and preserve economic consistency. In Chapter 3, a more complex approach is evaluated: ensemble machine learning algorithms, referred to as random forest (Breiman (2001)) and as boosting (Schapire (1990)), are applied. The two key features of those algorithms are their abilities to entertain a large number of predictors and to perform estimation and variable selection simultaneously.

In both chapters, profit maximization measures are computed in addition to more standard criteria to assess the value of the models to take into account the disconnection between econometric predictability and actual profitability (see, among others, Cenesizoglu and Timmermann (2012) or Brown (2008)).

Importantly in this Thesis, when comparing predictive accuracy and profit measures, the model confidence set procedure proposed by Hansen et al. (2011) is applied to avoid

²For prediction, the most common form of machine learning is supervised learning. Imagine that we want to build a system that can classify images as containing, say, a house, a car, a person or a pet. We first collect a large data set of images of houses, cars, people and pets, each labelled with its category. During training, the machine is shown an image and produces an output in the form of a vector of scores, one for each category. We want the desired category to have the highest score of all categories, but this is unlikely to happen before training. We compute an objective function that measures the error (or distance) between the output scores and the desired pattern of scores. The machine then modifies its internal adjustable parameters to reduce this error. These adjustable parameters, often called weights, are real numbers that can be seen as "knobs" that define the input-output function of the machine.

data snooping. Data snooping occurs when a given set of data is used more than once for purposes of inference or model selection and leads to the possibility that any results obtained in a statistical study may simply be due to chance rather than to any merit inherent in the method yielding the results (White (2000))³.

Both approaches are effective to detect economic turning points in real time over the period from January 2002 to December 2013. Strategies based on the turning points of the growth cycle induced by the models achieve thus excellent risk-adjusted returns in real time: timing the market is possible.

At last, modern and complex portfolio optimisation methods are optimal in-sample, but out-of-sample underperform alternative methods that are suboptimal in-sample. For instance, DeMiguel et al. (2009) demonstrate that the equal-weighted allocation, which gives the same importance to each assets, beats the entire set of commonly used portfolio optimization techniques.

López de Prado (2016a) points out that these methods lack the notion of hierarchy, thereby allowing weights to vary freely in unintended ways. Indeed, Nobel prize laureate [Herbert Simon](#) has demonstrated that complex systems can be arranged in a natural hierarchy, comprising nested sub-structures (Simon (1962)): *"the central theme that runs through my remarks is that complexity frequently takes the form of hierarchy, and that hierarchic systems have some common proper-ties that are independent of their specific content. Hierarchy, I shall argue, is one of the central structural schemes that the architect of complexity uses"*.

Building upon the fundamental notion of hierarchy, a hierarchical clustering based asset allocation method, which uses unsupervised machine learning techniques⁴, is in-

³Researchers conducting multiple tests on the same data tend to publish only those that pass a statistical significance test, hiding the rest. Because negative outcomes are not reported, readers are only exposed to a biased sample of outcomes. This problem, called "selection bias", is caused by multiple testing combined with partial reporting. It appears in many different forms: analysts who do not report the full extent of the experiments conducted ("file drawer effect"), journals that only publish "positive" outcomes ("publication bias"), managers who only publish the history of their (so far) profitable strategies ("self selection bias"), etc. What all these phenomena have in common is that critical information is hidden from the decision maker, with the effect of a much larger than anticipated Type I Error probability.

⁴Unsupervised learning is a type of machine learning algorithm used to draw inferences from datasets

troduced in Chapter 4. The out-of-sample performances of hierarchical clustering based portfolios and more traditional risk-based portfolios are evaluated across three disparate datasets, which differ in terms of number of assets and composition of the universe ("S&P sectors", multi-assets and individual stocks). The empirical results indicate that hierarchical clustering based portfolios are robust, truly diversified and achieve statistically better risk-adjusted performances than commonly used portfolio optimization techniques.

The rest of the thesis proceeds as follows. Chapter 1 provides a clear, precise and efficient framework for macro-based investment decisions. Chapter 2 underlines that a very simple machine-learning algorithm known as Learning Vector Quantization appears very competitive with commonly used alternatives. Chapter 3 points out the interest of ensemble machine learning algorithms, referred to as random forest and as boosting. Chapter 4 presents the hierarchical clustering based asset allocation method.

consisting of input data without labeled responses.

Chapter 1

Time-varying risk premiums and economic cycles

Abstract

Asset returns are not correlated with the business cycle but are primarily caused by the economic cycles. To validate this claim, economic cycles are first rigorously defined, namely the classical business cycle and the growth cycle, better known as the output gap. The description of different economic phases is refined by jointly considering both economic cycles. The theoretical influence of economic cycles on time-varying risk premiums is then explained based on two key economic concepts: nominal GDP and adaptive expectations. Simple dynamic investment strategies confirm the importance of economical cycles, especially the growth cycle, for euro and dollar-based investors. At last, this economic cyclical framework can improve strategic asset allocation choices.

Introduction

The willingness of investors to bear risk varies over time, larger in good times, and less in bad times, leading to time-varying risk premiums (Cochrane (2016)). Yet, there is still no consensus on the definition of good and bad time.

The most common approach is to consider the business cycle expansions and recessions (see, among others, Lustig and Verdelhan (2012)). Cooper and Priestley (2009) choose a slightly different way and point out the importance of the output gap for investment decisions. They note that the output gap is a classical business cycle variable that begins to fall before and throughout every recession. At last, some authors sometimes consider four distinct phases of the business cycle: "expansion", "peak", "recession" and "recovery" (see, for example, Ahn et al. (2016)). The business cycle is thus a fundamental yet ambiguous concept, since it can refer to conceptually different economic fluctuation.

To deepen the knowledge of interaction dynamics between financial markets the macroeconomy, this ambiguity needs above all to be removed. To this end, economic cycles are rigorously defined, namely the classical business cycle and the growth cycle, which seeks to represent the fluctuations around the trend. If we consider the trend growth rate as the potential growth rate, the growth cycle is better known as the output gap. The description of different economic phases is then refined by jointly considering both economic cycles. It improves the classical analysis of economic cycles by considering sometimes two distinct phases and sometimes four distinct phases.

To explain the theoretical the influence of economic cycles on the time-varying risk premiums, the key concept is nominal growth expectations, which is the same thing as saying that income expectations are crucial. Indeed, a fall in nominal GDP growth tends to lead to mass unemployment, lower profits and sharply higher debt defaults (Keynes (1936) and Sumner (2014)). Forward-looking investors adjust thus their portfolios according to their ever-changing current expectations of future events. The theory of adaptive expectations (Fisher (1911)) makes the link between the cyclical framework and nomi-

nal growth expectations. For example, when the real growth rate is above its potential, inflation pressures surge: the nominal growth rate of the economy increases. Adaptive expectations imply that nominal growth rate expectations should increase.

To gauge the potential value of this cyclical framework, dynamic investment strategies based on economic cycles turning points are created. Empirical results highlight the importance of economical cycles, especially the growth cycle, for euro and dollar-based investors. Indeed, strategies based on output gap turning points statistically outperform not only passive buy-and-hold benchmarks, but also business cycles' strategies, in the United States and in the euro area.

In other words, asset prices and returns are not correlated with the business cycle (Cochrane (2016)) but are primarily caused by the economic cycles. Assessing if the current growth rate of the economy is above or under the trend growth rate is thus the most crucial task for investors.

At last, the presence of regimes with different correlations and assets' characteristics can enhance strategic asset allocation, which is the most important determinant of long run investment success (Campbell and Viceira (2002)). Since correlations should theoretically vary during economic regimes, the main idea is to build a portfolio that would stay diversified when needed. Empirical results illustrate the influence of the correlation matrix on strategic asset allocation. In particular, investment-grade corporate bonds are not substitute to government bonds and risk-averse investors should select an asset allocation based on a correlation matrix whose elements are generated from bad times periods.

1.1 Cyclical framework

1.1.1 Economic cycles definitions

The classical business cycle definition is due to Burns and Mitchell (1946): "*Business cycles are a type of fluctuation found in the aggregate economic activity of nations that*

organize their work mainly in business enterprises: a cycle consists of expansions occurring at about the same time in many economic activities, followed by similarly general recessions, contractions, and revivals which merge into the expansion phase of the next cycle". The business cycle is meant to reproduce the cycle of the global level of activity of a country. The turning points of that cycle (named B for peaks and C for troughs) separate periods of recessions from periods of expansions.

Burns and Mitchell (1946) point out two main stylised facts of the economic cycle. The first is the co-movement among individual economic variables: most of macroeconomic time series evolve together along the cycle. The second is non-linearity: the effect of a shock depends on the rest of the economic environment. In other words, economic dynamics during economically stressful times are potentially different from normal times. For instance, small shock, such as a decrease in housing prices, can sometimes have large effects, such as recessions.

The growth cycle, introduced by Mintz (1974), seeks to represent the fluctuations of the GDP around its long-term trend. Mintz (1974) indicates that the rationale for investigating the growth cycle is that absolute prolonged declines in the level of economic activity tend to be rare events when the economy grows at a sustained and stable rate, so that in practice many economies do not very often exhibit recessions in classical terms. As a consequence, other approaches to produce information on economic fluctuations have to be proposed.

Growth cycle turning points (named A for peaks and D for troughs) have a clear meaning: peak A is reached when the growth rate decreases below the trend growth rate and the trough D is reached when the growth rate overpasses it again. Those downward and upward phases are respectively named slowdown and acceleration. A slowdown signals thus a prolonged period of subdued economic growth though not necessarily an absolute decline in economic activity. In other words, all recessions involve slowdowns, but not all slowdowns involve recessions.

If the long-term trend is considered as the estimated potential level¹, then the growth cycle equals the output gap. A turning point of the output gap occurs when the current growth rate of the activity is above or below the potential growth rate, thereby signalling increasing or decreasing inflation pressures.

The ABCD approach (Anas and Ferrara (2004)) refines the description of different economic phases by jointly considering the classical business cycle and the growth cycle. Let us suppose that the current growth rate of the activity is above the trend growth rate (acceleration phase). The downward movement will first materialize when the growth rate will decrease below the trend growth rate (point A). If the slowdown gains in intensity, the growth rate could become negative enough to provoke a recession (point B). Eventually, the economy should start to recover and exits from the recession (point C). As the recovery strengthens, the growth rate should overpass its trend (point D). However, a slowdown will not automatically translate into a recession: if the slowdown is not severe enough to become a recession, then point A will not be followed by point B, but by point D.

This framework improves thus the classical analysis of economic cycles by allowing sometimes two distinct phases, if the slowdown is not severe enough to become a recession, and sometimes four distinct phases, if the growth rate of the economy becomes negative enough to provoke a recession.

1.1.2 Turning point chronology

A cycle turning point chronology is required for empirical studies to create and validate real-time detection and forecasting methods. The turning point chronology is only suitable for ex post explanatory analyses and not for ex ante decision making.

In the United States, the NBER's Business Cycle Dating Bureau's Committee determines the peaks and troughs of the classical business cycle. In the euro area, the CEPR euro area Business Cycle Dating Committee establishes the chronology of recessions and

¹The potential output is the maximum amount of goods and services an economy can turn out at full capacity.

expansions. The European chronology is only available on a quarterly basis. To refine the chronology, the monthly GDP introduced by Raffinot (2007)² is exploited in this article. For instance, the monthly GDP allows to select the month with the lowest level within the quarter selected by the CEPR to be chosen as the through of the recession.

If dating the classical business cycle is not an easy task, then dating the growth cycle is quite challenging since the series must first be de-trended. Several growth cycle extraction methods have been proposed in statistical literature, ranging from filtering techniques to parametric modelling, mainly based on state-space models and Markov switching models (see Anas et al. (2008) for a review). As advocated by Nilsson and Gyomai (2011), a double Hodrick-Prescott filter (18 months-96 months) is used on the monthly GDP³ (see Appendix 1.A for more information on this filter).

The turning points of the growth cycle are then estimated by the non-parametric procedure introduced by Harding and Pagan (2002). The algorithm first identifies peaks as observations that are lower over a two-sided window of five months and troughs are points associated with observations in the five month window that are higher. The algorithm then applies censoring rules to narrow the turning points of the reference cycle: the duration of a cycle must be no less than 15 months, while the phase (peak to trough or trough to peak) must be no less than 5 months.

The complete chronology is contained in the table 1.1⁴. The turning point chronology highlights the persistence of the regimes and the non-linearity of the cycles: the sequence of up and down phases is recurrent but not periodic. Expansions are longer than recessions. Moreover, even if the American and euro area chronologies are linked, they are quite distinct. For example, there was no double dip in the United States following the Great Recession and there was no recession in the euro area following the dot-com bubble burst.

²A temporal disaggregation based on business surveys of the non revised values of gross domestic product GDP is used to develop a monthly indicator of GDP.

³In unreported results, others filters were tested, such as Christiano and Fitzgerald (2003) and Baxter and King (1999). The empirical results are qualitatively similar.

⁴The American chronology starts in 1985, since Stock and Watson (2003) demonstrate that approximately 40% of 168 macro variables have significant breaks in their conditional variance during 1983-1985.

Table 1.1: Turning point chronology

Euro area (Jan 1999-December 2013)		United States (Jan 1985-Dec 2013)	
Trough D	February 1999	Peak A	November 1985
Peak A	December 2000	Trough D	April 1987
Trough D	September 2003	Peak A	December 1989
Peak A	May 2004	Peak B	July 1990
Trough D	May 2005	Trough C	March 1991
Peak A	October 2007	Trough D	August 1991
Peak B	March 2008	Peak A	January 1993
Trough C	April 2009	Trough D	July 1993
Trough D	August 2009	Peak A	September 94
Peak A	June 2011	Trough D	March 1996
Peak B	August 2011	Peak A	June 2000
Trough C	November 2012	Peak B	March 2001
Trough D	March 2013	Trough C	November 2001
		Trough D	February 2003
		Peak A	October 2007
		Peak B	December 2007
		Trough C	June 2009
		Trough D	September 2009
		Peak A	June 2011
		Trough D	December 2011

Figures 1.1 and 1.2 illustrate the monthly GDP, its trend, the growth cycle and exhibit the turning points in the euro area between 2001 and 2009. The euro area experienced a slowdown without recession in 2000-2003. In 2003, the recovery started to materialise (Point D), but the activity fell apart in May 2004 (Point A). The slowdown was short-lived: starting in May 2005 (point D) a building boom got under way. The slowdown starting in October 2007 (Point A) translated into the Great Recession a few months later (Point B). In April 2009, the recession was over (Point C).

Figure 1.1: Euro area monthly GDP and its trend over 2001-2009

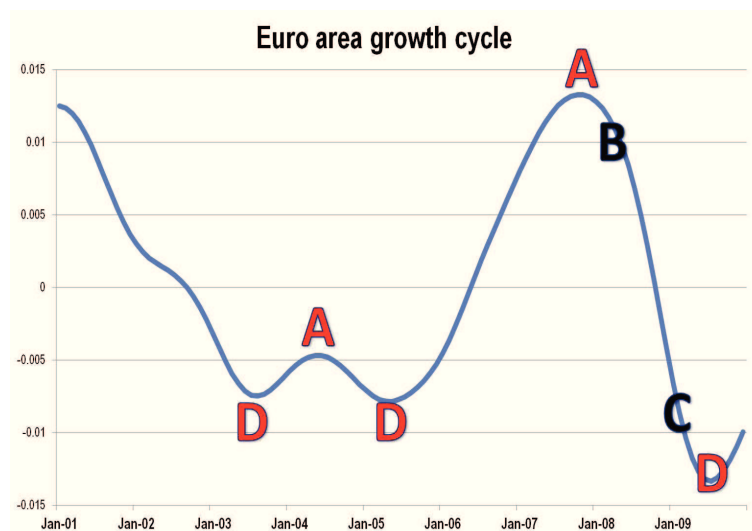
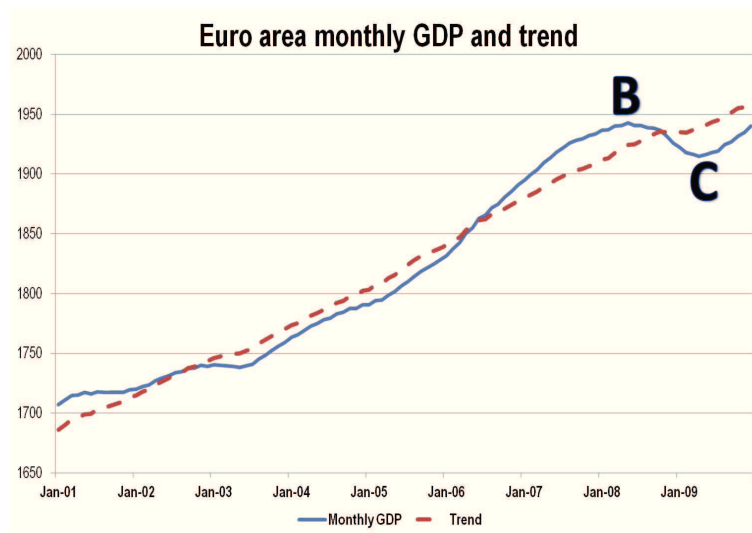


Figure 1.2: Euro area growth cycle over 2001-2009

1.2 Asset classes and economic cycles

1.2.1 The economic rationale

Key economic concepts

Two key basic economic concepts are needed to deepen the knowledge of interaction dynamics between financial markets the macroeconomy: nominal GDP and adaptive expectations.

The economy is simply the sum of all transactions - the exchange of money and credit for goods, services, and financial assets - among individuals, banks, and governments. The technical term for the value of everything a country produces is nominal Gross Domestic Product (nominal GDP). It is total output (real GDP) times the current prices paid. Since all income is derived from production (including the production of services), the gross domestic income of a country should exactly equal its gross domestic product. Indeed, the gross domestic income is the total income received by all sectors of an economy within a State.

When nominal GDP falls, there is no longer enough spending to sustain the same number of jobs unless wages fall. Because wages are slow to adjust, unemployment rises instead (see Keynes (1936) and Sumner (2014)). Moreover, since most debts are not indexed to inflation, nominal income is the best measure of a person's ability to repay their debts. When determining how much debt to take on, borrowers consider their ability to repay that debt. If income gains falls short of these expectations, interest and principal payments will be more burdensome than what was planned for. Problems of debt overhang become that much worse for the economy and debt defaults rise. To sum up, a fall in nominal GDP growth tends to lead to mass unemployment, lower profits and sharply higher debt defaults.

Adaptive expectation models are ways of predicting an agent's behaviour based on their past experiences and past expectations for that same event. They are introduced by

(Fisher (1911)) and most famously used by Friedman (1957). For example, if consumers begin to actually see prices rising, they will begin to form robust expectations of inflationary expectations. The same theory might claim that consumers will expect an economic recovery to begin only after ample evidence that the turning point has been passed. The expectations-augmented Phillips curve introduces adaptive expectations into the Phillips curve. This equation appears in many recent New Keynesian dynamic stochastic general equilibrium models (Roberts (1995)).

Interactions between economic cycles and asset classes

The existing link between asset classes and the economy is not straightforward, especially for equities.

The stock price or the value of a company is the sum of all future dividend payments discounted to its present value. Under a constant payout ratio, the dividend growth rate will equate to the growth rate in earnings. In other words, investors buy a stock for its future earnings potential. Since the sum of all money earned in an economy each year is the nominal GDP, equity prices should thus be linked to nominal growth expectations (not real growth expectations). When the real growth rate is above its potential, inflation pressures surge: the nominal growth rate of the economy increases. Adaptive expectations imply that nominal growth rate expectations should increase, which is the same thing as saying that income expectations should rise⁵. In theory, equities should thus perform well during acceleration phases and suffer during slowdowns. Since slowdowns signal a prolonged period of subdued economic growth though not necessarily an absolute decline in economic activity, equities performances can thus be negative when real growth rate are positive.

In theory, government bonds should perform well during slowdowns and recessions. Indeed, the expectations theory of the term structure holds that the long-term interest

⁵A link with the classical drivers of the equity risk premium can be done. For example, market participants' expectations about the future economic activity affect the determination of dividends and thus stock return premiums.

rate is a weighted average of present and expected future short-term interest rates plus a term premium (the latter captures the compensation that investors require for bearing interest rate risk). If the growth rate of the activity is lower than the potential growth rate, inflation pressures trend down and the central bank is more likely to cut rates. In consequence, investors should forecast a lower path of future short-term interest rates. Long-term rates should thus decrease (and bond prices increase).

Because a company's capacity to service its debt is uncertain, corporate bond should offer a higher expected return compared to sovereign bonds. Corporate bonds can thus be decomposed as the sum of a government bond plus a spread, which compensates for the expected losses due to default. The probability of default for an investment grade firm is quite small, but raises when corporate income expectations decline. Investment grade bonds should benefit from falling rates during slowdowns, but to a lesser extent than government bonds. During acceleration and expansion, investment grade bonds should take advantage of the declining spread and the coupon. High-yield bonds should behave more like equities, since defaults have always been commonplace.

For all assets, volatility is higher during bad times. One plausible explanation is that the uncertainty around the likely course of monetary policy rises during bad times, leading to an increase in the volatility of the safest and riskiest assets.

1.2.2 Historical facts

Table 1.2 details the four asset classes in consideration. The investment universe is as stripped-down and simple as possible without raising concerns that the key results will not carry over to more general and intricate asset classes or factors.

Table 1.2: **Asset classes**

Asset class	Index	
Equities	Euro area	Euro Stoxx 50 (total return)
	United States	S&P 500 (total return)
Government bonds	Euro area	IBOXX SOVEREIGNS EUROZONE ALL MATS (total return)
	United States	BOFA Merrill Lynch Treasury Master Index (total return)
Investment Grade bonds	Euro area	IBOXX Euro Corp All Mats (total return)
	United States	Barclays Capital U.S. Corporate Investment Grade Index (total return)
High-yield bonds	Euro area	IBOXX EUR High Yield Index (total return)
	United States	BOFA Merrill Lynch High Yield 100 Index (total return)

All series are provided by Datastream.

Table 1.3 confirms that asset classes perform differently during different stages of both economic cycles and no single asset class dominates under all economic conditions, in line with the theory.

The returns of the riskiest assets, equities and high yield bonds, are pro-cyclical and contrasted. For instance, in the United States, equities increase at a rate of 18.7% during expansions, whereas they fall by almost 17% during slowdowns. The amplitude of expected returns equals thus 35%.

The expected returns of government bonds and investment grade bonds are always positive, even if they are more attractive during bad times. For example, in the euro area, the performance of government bonds is four times higher during slowdowns (7.9%) than during expansions (1.9%).

Moreover, the presence of asymmetric volatility for risky assets is most apparent during bad times (slowdowns) and very bad times (recessions) when a large decline in stock prices or high yield prices is associated with a significant increase in market volatility. For example, in the euro area, the stock market volatility increases from 14.3% during accelerations to 22.9% during slowdowns.

In the euro area, the performance of high yield bonds is surprisingly the same during accelerations and slowdowns. This is not in line with the theory and with the results observed in the United States. The limited size of this market and the short sample period may partly explain this anomaly.

Importantly, the performance to volatility ratio, which compares the expected returns

of an investment to the amount of risk undertaken to capture these returns, is statistically different at a 95% confidence level in each macroeconomic environment (except for the American government bonds as regards the business cycle).

1.3 Dynamic investment strategies

1.3.1 Active portfolio management and economic cycles

To study time-varying risk premiums, the traditional approach is to test whether assets excess returns are predictable. The literature on return predictability is voluminous (see, for example, Rapach et al. (2013) for a review on stocks returns, Zhou and Zhu (2017) for government bonds returns and Lin et al. (2016) for corporate bonds returns). One drawback with this approach is the disconnection between econometric predictability and actual profitability (see, among others, Cenesizoglu and Timmermann (2012) or Brown (2008)). This paper only focusses on profitability⁶.

To examine the economic value of market forecasts, the current practice is to calculate realized utility gains for a mean-variance investor who optimally allocates across equities (or bonds or corporate bonds) and the risk-free asset on a real-time basis. Portfolio weights are usually constraint to lie between -50% and 150% each month.

Another approach has been preferred, mainly for two reasons. First, the mean-variance optimisation requires the investor to forecast the excess returns and the variance of asset returns. But, even small estimation errors can result in large deviations from optimal allocations in an optimizer's result (Michaud (1989)). This is why, academics and practitioners have developed risk-based portfolio optimization techniques (minimum variance, equal risk contribution, risk budgeting,...), which do not rely on return forecasts (Roncalli (2013)). Yet, even the variance is hard to estimate. A ten-year rolling window or five-year

⁶To test the predictability of the proposed economic framework, a regression between asset returns and the evolution of the output gap should be done. The evolution of the output gap exhibits if the current growth rate of the economy is above the trend or not. It differs from Cooper and Priestley (2009) because the sign and the magnitude of the output gap are not taken into account.

Table 1.3: Summary of returns and risk measures in each macroeconomic environment, 1999-2013

Euro area					
	Growth cycle		Business cycle		Full Period
	Slowdown	Acceleration	Recession	Expansion	
Equities					
Average returns	-8.9*	18.1	-14.8*	7.2	4
Volatility	22.9*	14.3	25.6*	18.1	19.5
Performance to volatility ratio	-0.4*	1.3	-0.6*	0.4	0.2
Government bonds					
Average returns	7.9*	1.9	10.8*	3.6	4.7
Volatility	4.1*	3.4	8.2*	3.4	3.8
Performance to volatility ratio	1.9*	0.6	1.3*	1.1	1.2
Investment Grade bonds					
Average returns	7*	2.3	4.4	4.7	4.7
Volatility	4.6*	2.9	6.4*	3.3	3.9
Performance to volatility ratio	1.5*	0.8	0.7*	1.4	1.2
High-yield bonds					
Average returns	9.9	9.8	-2.1*	13.3	9.9
Volatility	15*	5.3	19.8*	7.2	11.1
Performance to volatility ratio	0.7*	1.8	-0.1*	1.8	0.9
United States					
	Growth cycle		Business cycle		Full Period
	Slowdown	Acceleration	Recession	Expansion	
Equities					
Average returns	-16.9*	18.7	-16.1*	9.8	5.9
Volatility	20.4*	11.2	23.2*	13.6	15.6
Performance to volatility ratio	-0.8*	1.7	-0.7*	0.7	0.4
Government bonds					
Average returns	10.1*	2.6	6.8*	4.7	4.9
Volatility	5.5*	3.9	6.4*	4.2	4.6
Performance to volatility ratio	1.8*	0.7	1.1	1.1	1.1
Investment Grade bonds					
Average returns	8.1*	5.5	5.4*	6.4	6.3
Volatility	7.9*	4.6	10*	4.6	5.7
Performance to volatility ratio	1*	1.2	0.5*	1.4	1.1
High-yield bonds					
Average returns	0.0*	9.4	-6.7*	8.9	6.7
Volatility	13.8*	5.6	16.8*	6.7	8.9
Performance to volatility ratio	0.0*	1.7	-0.4*	1.3	0.8

Note: This table reports annualized average monthly returns, annualized standard deviation (volatility) and performance to volatility ratio of asset classes during different economic regimes over the period 1999-2013. The performance to volatility ratio compares the expected returns of an investment to the amount of risk undertaken to capture these returns, the higher the better. * signals that the null hypothesis of equal mean of the Wilcoxon rank sum is rejected at a 95% confidence level.

rolling window is often applied to estimate the variance. In a regime switching framework, another solution has to be found.

Second, the relative profitability of a dynamic strategy using econometrically superior forecasts shrinks significantly if the investor is highly risk averse and heavily constrained (Baltas and Karyampas (2016)). As a matter of fact, investors typically face a number of constraints, either by mandate or regulation, which put hard thresholds on minimum and maximum allocation across risky assets. For instance, by law in France, equity funds managers must have at least 60% of their portfolio invested in equities.

To address the same problem with another approach, simple hypothetical dynamic trading strategies are created⁷. Dynamic strategies should take advantage of positive economic regimes, as well as withstand adverse economic regimes and reduce potential drawdowns.

We consider an investor managing a portfolio consisting of an unique asset class (either stocks or government bonds) investing 100€ or 100\$ on January 1, 1999. Each month the investor decides upon the fraction of wealth to be invested based on the current state of the economy. If the asset class should perform well, then the investor can leverage his portfolio (120% of his wealth is invested on the asset and 20% of cash is borrowed), otherwise he only invests 80% of his wealth and 20% is kept in cash. The active strategies are then compared among them and with the buy-and-hold strategy - henceforth a passive strategy. These strategies are named 120/80 hereafter⁸.

To avoid look-ahead bias, the reallocation takes place at the beginning of the month following the turning point. As a matter of fact, an investor could not know at the beginning of any month whether a turning point would occur in that month.

In contrast to the standard approach, asset returns are not predicted but only condi-

⁷These strategies reflects the investment process in place in the asset management companies where I used to work.

⁸Since asset classes perform differently during different stages of the growth cycle, it might be reasonable to rebalance the portfolio (shifting allocation weights) based on the stage of the economic cycles. For brevity reasons, the results of such strategies are not presented. Yet if the 120/80 strategies work for both bonds and equities, it seems reasonable to conclude that dynamic asset allocation strategies should perform well.

tioned on the stage of the economic cycle.

The main weakness is that the strategies described so far are not implementable because investors do not know turning point dates in real time. Yet, recent academic studies such as Ng (2014) and Berge (2015) apply a machine learning algorithm referred to as boosting ((Schapire (1990))) to the problem of identifying business cycle turning points in the United States. Even if forecasting turning points remains challenging, they conclude that nowcasting turning points in real time is feasible.

1.3.2 Comparison criteria

Performance and risk measures

To compare the different strategies, four different performance and risk measures are computed:

- The annualized average returns (μ)
- The annualized standard deviation (Volatility)(σ)
- The certainty-equivalent return (CEQ) is the risk-free rate of return that the investor is willing to accept instead of undertaking the risky portfolio strategy.

$$CEQ = (\mu - r_f) - \frac{\gamma}{2}\sigma^2$$

where r_f is the risk-free rate⁹ and γ is the risk aversion. Results are reported for the case of $\gamma = 1; 3; 5$. More precisely, the *CEQ* captures the level of expected utility of a mean-variance investor, which is approximately equal to the certainty-equivalent return for an investor with quadratic utility (DeMiguel et al. (2009)).

- The Max drawdown (MDD) is an indicator of permanent loss of capital. It measures the largest single drop from peak to bottom in the value of a portfolio. In brief, the MDD offers investors a worst case scenario.

⁹A risk-free interest rate of zero is assumed when calculating the *CEQ*.

Pesaran and Timmermann (1994) and Han et al. (2013) demonstrate that the total cost of transactions appears to be low, less than 1% (around 50 basis points when trading in stocks while the cost for bonds is 10 basis points). To simplify, since economic turning points are rare, no transaction costs are considered.

Data snooping

To avoid data snooping, which occurs when a given set of data is used more than once for purposes of inference or model selection (White (2000)), the model confidence set (MCS) procedure proposed by Hansen et al. (2011) is computed.

The MCS procedure is a model selection algorithm, which filters a set of models from a given entirety of models. The resulting set contains the best models with a probability that is no less than $1 - \alpha$ with α being the size of the test.

An advantage of the test is that it not necessarily selects a single model, instead it acknowledges possible limitations in the data since the number of models in the set containing the best model will depend on how informative the data are.

More formally, define a set M_0 that contains the set of models under evaluation indexed by: $i = 0, \dots, m_0$. Let $d_{i,j,t}$ denote the loss differential between two models by

$$d_{i,j,t} = L_{i,t} - L_{j,t}, \forall i, j \in M_0$$

L is the loss calculated from some loss function for each evaluation point $t = 1, \dots, T$. The set of superior models is defined as:

$$M^* = \{i \in M_0 : E[d_{i,j,t}] \leq 0 \forall j \in M_0\}$$

The MCS uses a sequential testing procedure to determine M^* . The null hypothesis being tested is:

$$\begin{cases} H_{0,M} : E[d_{i,j,t}] = 0 \forall i, j \in M \text{ where } M \text{ is a subset of } M_0 \\ H_{A,M} : E[d_{i,j,t}] \neq 0 \text{ for some } i, j \in M \end{cases}$$

When the equivalence test rejects the null hypothesis, at least one model in the set M is considered inferior and the model that contributes the most to the rejection of the null is eliminated from the set M . This procedure is repeated until the null is accepted and the remaining models in M now equal $\widehat{M}_{1-\alpha}^*$.

According to Hansen et al. (2011), the following two statistics can be used for the sequential testing of the null hypothesis:

$$t_{i,j} = \frac{\bar{d}_{i,j}}{\sqrt{\widehat{var}(\bar{d}_{i,j})}} \text{ and } t_i = \frac{\bar{d}_i}{\sqrt{\widehat{var}(\bar{d}_i)}}$$

where m is the number of models in M , $\bar{d}_i = (m - 1)^{-1} \sum_{j \in M} \bar{d}_{i,j}$, is the simple loss of the i^{th} model relative to the averages losses across models in the set M , and $\bar{d}_{i,j} = (m)^{-1} \sum_{t=1}^m d_{i,j,t}$ measures the relative sample loss between the i^{th} and j^{th} models. Since the distribution of the test statistic depends on unknown parameters a bootstrap procedure is used to estimate the distribution.

In this paper, the MCS is applied with a profit maximization loss function (CEQ).

As regards investment strategies, the MCS aims at finding the best model and all models which are indistinguishable from the best, not those better than the benchmark. To determine if models are better than the benchmark, the stepwise test of multiple reality check by Romano and Wolf (2005) and the stepwise multiple superior predictive ability test by Hsu et al. (2013) should be considered. However, if the benchmark is not selected in the best models set, investors can conclude that their strategies "beat" the benchmark.

1.3.3 Empirical results

Tables 1.4 and 1.5 highlight that active investment strategies based on the growth cycle statistically outperform not only passive buy-and-hold benchmarks, but also business cycles' strategies, in the United States and in the euro area.

For both government bonds and equities, strategies based on growth cycle turning points are always the only constituent of the best models set $\widehat{M}_{75\%}^*$, whatever the degree of risk aversion.

In comparison with the benchmark, 120/80 equities strategies based on output gap turning points reduce the volatility and the losses in extreme negative events (MDD). For instance, the MDD decreases from 59.9% to 52.4% in the euro area and from 50.9% to 43% in the United States. To sum up, strategies based on the growth cycle improve the returns and reduce risk measures.

As regards governments bonds, 120/80 strategies based on the growth cycle improve returns while taking almost the same risk. In the euro area, the MDD declines from 5.7% to 5.2%, whereas the MDD slightly increases from 4.8% to 5% in the United States.

The business cycles' strategies produced mixed results. In particular, for bond investors, they do not add any value and the CEQ of the strategy is lower than the CEQ of the benchmark for all the degrees of risk aversion. As regards equities, avoiding the worst times is undoubtedly a good idea to get better returns, especially in the United States. The performance progresses from 4% to 5.4% in the euro area and from 5.9% to 8.3% in the United States. Yet risk measures increase. In particular,, the MDD surge from 59.9% to 65.4% in the euro area and the volatility rises from 15.5% to 16.8% in the United States.

In the end, expansions, which are composed of both acceleration and slowdown periods, can not be considered as good times, especially for the safest assets. The only suitable definition of good and bad times is thus acceleration and slowdown periods.

Table 1.4: **Equities: 120/80 investment strategies**

Euro area			
	Growth cycle 120/80	Business cycle 120/80	Full period
Average returns	6.8	5.4	4.0
Volatility	18.0	21.5	19.5
<i>CEQ</i> : $\gamma = 1$	0.052*	0.031	0.021
<i>CEQ</i> : $\gamma = 3$	0.019*	-0.015	-0.017
<i>CEQ</i> : $\gamma = 5$	-0.013*	-0.062	-0.055
MDD	-52.4	-65.4	-59.9
United States			
	Growth cycle 120/80	Business cycle 120/80	Full period
Average returns	9.7	8.3	5.9
Volatility	15.1	16.8	15.5
<i>CEQ</i> : $\gamma = 1$	0.086*	0.069	0.047
<i>CEQ</i> : $\gamma = 3$	0.063*	0.041	0.023
<i>CEQ</i> : $\gamma = 5$	0.040*	0.012	-0.001
MDD	-43.0	-51.0	-50.9

Note: This table reports the characteristics of active strategies based on the state of the economic cycle over the period from January 1999 to December 2013. A 120/80 equity strategy is computed. Returns are monthly and annualized. The volatility corresponds to the annualized standard deviation. The certainty-equivalent return (CEQ) is the risk-free rate of return that the investor is willing to accept instead of undertaking the risky portfolio strategy. The Max drawdown (MDD) measures the largest single drop from peak to bottom in the value of a portfolio. * indicates the model is in the set of best models $\widehat{M}_{75\%}^*$.

Table 1.5: **Government bonds: 120/80 investment strategies**

Euro area			
	Growth cycle 120/80	Business cycle 120/80	Full period
Average returns	5.2	4.4	4.7
Volatility	4.0	3.6	3.8
<i>CEQ</i> : $\gamma = 1$	0.051*	0.043	0.046
<i>CEQ</i> : $\gamma = 3$	0.050*	0.042	0.045
<i>CEQ</i> : $\gamma = 5$	0.048*	0.041	0.043
MDD	-5.2	-5.2	-5.7
United States			
	Growth cycle 120/80	Business cycle 120/80	Full period
Average returns	5.2	4.3	4.9
Volatility	4.8	4.4	4.6
<i>CEQ</i> : $\gamma = 1$	0.051*	0.042	0.048
<i>CEQ</i> : $\gamma = 3$	0.049*	0.040	0.046
<i>CEQ</i> : $\gamma = 5$	0.046*	0.038	0.044
MDD	-5.0	-5.4	-4.8

Note: This table reports the characteristics of active strategies based on the state of the economic cycle over the period from January 1999 to December 2013. A 120/80 bond strategy is computed. Returns are monthly and annualized. The volatility corresponds to the annualized standard deviation. The certainty-equivalent return (CEQ) is the risk-free rate of return that the investor is willing to accept instead of undertaking the risky portfolio strategy. The Max drawdown (MDD) measures the largest single drop from peak to bottom in the value of a portfolio. * indicates the model is in the set of best models $\widehat{M}_{75\%}^*$.

These results have also implications for the risk management and hedging. Especially, in the options market one can utilize the current state of the economy to hedge the portfolio against the possible price declines. For example, besides following one of the

the previous strategy, writing an out-of-the-money covered call or buy a put option when the stock market is expected to decrease (slowdown or recession) would limit the losses.

Timing of turning point detection

A cyclical framework with learning gives content to the idea of an economy moving gradually from one regime to another, particularly if the central bank as well as the public is assumed to be updating its beliefs (Bernanke (2007)).

It implies that the strategies described in the previous sections should not rely on an exact timing of turning points detection. To verify the validity of this claim, the 120/80 strategies are computed based on different timings of the turning points detection: up to three months in advance, right in time or up to three months late.

Tables 1.6 and 1.7 illustrate that timing is an important issue, but an exact timing is not needed.

Table 1.6: **Government bonds: 120/80 investment strategies and timing of growth cycle turning points detection**

Euro area							
	Right timing	One month ahead	Two months ahead	Three months ahead	One month late	Two months late	Three months late
Average returns	5.2	5.1	5.0	5.0	5.2	5.1	5.0
Volatility	4.0	4.0	4.0	4.0	4.1	4.1	4.1
$CEQ : \gamma = 1$	0.051*	0.050*	0.049	0.049	0.051*	0.050*	0.049
$CEQ : \gamma = 3$	0.050*	0.049*	0.048	0.048	0.049*	0.048	0.047
$CEQ : \gamma = 5$	0.048*	0.047*	0.046	0.046	0.048*	0.047*	0.046
MDD	-5.2	-5.2	-5.2	-5.2	-5.2	-5.7	-6.0
United States							
	Right timing	One month ahead	Two months ahead	Three months ahead	One month late	Two months late	Three months late
Average returns	5.2	5.3	5.1	5.1	5.1	5.0	4.9
Volatility	4.8	4.8	4.7	4.8	4.8	4.8	4.7
$CEQ : \gamma = 1$	0.051*	0.052*	0.050	0.050	0.050	0.049	0.048
$CEQ : \gamma = 3$	0.049*	0.050*	0.048	0.048	0.048	0.047	0.046
$CEQ : \gamma = 5$	0.046*	0.047*	0.045	0.045	0.045	0.044	0.045
MDD	-5.0	-5.0	-5.0	-5.0	-5.3	-5.4	-5.4

Note: This table reports the characteristics of the active strategies based on different timing of the turning point detection: in advance, right in time or late. It is an in-sample analysis. A 120/80 equity strategy and a 120/80 bond strategy are computed. Returns are monthly and annualized. Active strategies are then compared with the buy-and-hold strategy. The volatility corresponds to the annualized standard deviation. The certainty-equivalent return (CEQ) is the risk-free rate of return that the investor is willing to accept instead of undertaking the risky portfolio strategy. The Max drawdown (MDD) measures the largest single drop from peak to bottom in the value of a portfolio. * indicates the model is in the set of best models $\bar{M}_{75\%}^*$.

Table 1.7: **Equities: 120/80 investment strategies and timing of growth cycle turning points detection**

Euro area							
	Right timing	One month ahead	Two months ahead	Three months ahead	One month late	Two months late	Three months late
Average returns	6.8	6.9	7.1	7.0	5.9	4.9	4.4
Volatility	18	18.2	18.4	18.4	18	18.3	18.6
$CEQ : \gamma = 1$	0.052*	0.052*	0.054*	0.053	0.043	0.032	0.023
$CEQ : \gamma = 3$	0.019*	0.019*	0.020*	0.019*	0.010	-0.001	-0.012
$CEQ : \gamma = 5$	-0.013*	-0.014*	-0.014*	-0.015*	-0.022	-0.035	-0.046
MDD	-52.4	-53.0	-52.0	-52.5	-53.0	-55.0	-55.6
United States							
	Right timing	One month ahead	Two months ahead	Three months ahead	One month late	Two months late	Three months late
Average returns	9.3	9.5	9.3	9.2	9.2	9.1	8.7
Volatility	15.5	15.6	15.7	15.9	15.5	15.5	15.3
$CEQ : \gamma = 1$	0.081*	0.083*	0.081*	0.079	0.080	0.079	0.075
$CEQ : \gamma = 3$	0.057*	0.058*	0.056	0.054	0.056	0.055	0.052
$CEQ : \gamma = 5$	0.033*	0.034*	0.031	0.029	0.032	0.031	0.028
MDD	-43.0	-43.3	-43.0	-45.7	-43.0	-43.0	-44.0

Note: This table reports the characteristics of the active strategies based on different timing of the turning point detection: in advance, right in time or late. It is an in-sample analysis. A 120/80 equity strategy and a 120/80 bond strategy are computed. Returns are monthly and annualized. Active strategies are then compared with the buy-and-hold strategy. The volatility corresponds to the annualized standard deviation. The certainty-equivalent return (CEQ) is the risk-free rate of return that the investor is willing to accept instead of undertaking the risky portfolio strategy. The Max drawdown (MDD) measures the largest single drop from peak to bottom in the value of a portfolio. * indicates the model is in the set of best models $\widehat{M}_{75\%}^*$.

Table 1.6 highlights that bonds investors should rebalance their portfolio around the turning points or a month sooner in the United States and around the turning points or little bit later in the euro area. In the United States, the strategies "right in time" and "one month ahead" compose the best models sets $\widehat{M}_{75\%}^*$ for all the degrees of risk aversion.

In the euro area, the best models sets $\widehat{M}_{75\%}^*$ consist of many strategies and differ depending on the degree of risk aversion. Yet, the strategies: "right in time", "one month ahead", "one month late" and "two months late" belong to all best models sets. These results are in line with the fact that markets forecast a monetary tightening only when there are inflation pressures: the current growth rate of the economy has to be above its potential.

Table 1.7 paints a contrasted picture. The behaviour of the investor differ depending on the country. In the United States, for all the degrees of risk aversion, the strategies "right in time" and "one month ahead" form the best models set $\widehat{M}_{75\%}^*$. In the euro area, investors should shift weights in advance of the turning points: the strategy "three months ahead" belongs to all best models set $\widehat{M}_{75\%}^*$. This result may stem from the leading role for the United States as regards stocks markets (RAPACH et al. (2013)).

To validate this claim, the same 120/80 investment strategy is tested on European equities, but the rebalancing of investments is based on American turning points instead of European turning points. Table 1.8 highlights that the 120/80 strategy based on the American cycle achieves higher returns (8.8% against 6.8%) than the one based on euro area cycle and is the only strategy in the best models set $\widehat{M}_{75\%}^*$. Euro-based equity investors should thus focus on the American chronology instead of their own cycle.

Table 1.8: European equities: 120/80 investment strategies based on the American chronology

Euro area				
	European Growth cycle 120/80	American Growth cycle 120/80	Full period	
Average returns	6.8	8.8	4.0	
Volatility	18.0	19.1	19.5	
$CEQ : \gamma = 1$	0.052	0.070*	0.021	
$CEQ : \gamma = 3$	0.019	-0.033*	-0.017	
$CEQ : \gamma = 5$	-0.013	-0.003*	-0.055	
MDD	-52.4	-52.2	-59.9	

Note: This table reports the characteristics of active strategies based on the state of the economic cycle over the period from January 1999 to December 2013. A 120/80 equity strategy is computed. Returns are monthly and annualized. The volatility corresponds to the annualized standard deviation. The certainty-equivalent return (CEQ) is the risk-free rate of return that the investor is willing to accept instead of undertaking the risky portfolio strategy. The Max drawdown (MDD) measures the largest single drop from peak to bottom in the value of a portfolio. * indicates the model is in the set of best models $\widehat{M}_{75\%}^*$.

1.4 Strategic asset allocation and economic cycles

The primary goal of a strategic asset allocation is to create an asset mix which will provide an optimal balance between expected risk and return for a long-term investment horizon. Strategic asset allocation is the most important determinant of long run investment success (Campbell and Viceira (2002)).

The asset allocation basis is to combine assets with low correlations to reduce the variance, or riskiness, of a portfolio. It is said to be the only the "only free lunch" in investing (Markowitz (1952)). The traditional inputs required for an optimizer are expected returns, volatility and a correlation matrix. Even with more modern techniques, such as risk-based portfolio (Roncalli (2013)) or "Hierarchical Risk Parity" (López de Prado (2016a)), a correlation matrix is still needed.

1.4.1 Correlations and economic cycles

Tables 1.9 and 1.10 highlight that correlations are time varying and sensitive to economic cycles. In particular, asset classes are more correlated during bad times than during good times. We recall that bad times refer to slowdown periods (and thus recessions) and good times refer to acceleration periods. Expansions can not be considered as good times, since they are composed of both acceleration and slowdown periods.

An important implication of higher correlation is that otherwise-diversified portfolio lose some of diversification benefit during bad times, when most needed.

The correlation between stocks and bonds is arguably the most important correlation input to the asset allocation decision. The full sample average of the realized correlation is -0.14 in the euro area and -0.04 in the United States, but there has been substantial variation around this mean. From a theoretical point of view, equities should suffer during bad times and bonds should perform well. The correlation should be negative.

Good times are more complex to analyse. Equities should increase. Government bonds should benefit from the reinvestment of the coupons. Indeed, during good times, government interest rates should be high and coupons should support the performance of this asset¹⁰.

The correlation between equities and government bonds should thus be slightly positive, especially near the end of an acceleration period. Tables 1.9 and 1.10 confirm this assessment: the realized correlation during acceleration periods is 0.01 in the euro area and 0.05 in the United States.

The covariation between equities and credit bonds is also a crucial input to managing the risk of diversified portfolios. For instance, if credit appears attractive relative to equities, portfolio managers may choose to take a kind of equity risk via credit bonds.

Tables 1.9 and 1.10 emphasize that equities and credit bonds are more correlated

¹⁰Low interest rates are not the symbol of easy monetary policy, but rather an outcome of excessively tight monetary policy (see Friedman (1992) and Friedman (1997)). If monetary conditions are tight then inflation and growth expectations decrease and as a consequence bond yields will also fall.

during bad times than during good times. For instance, in the euro area, there is a marked increase from -0.04 to 0.67 in the correlation between equities and investment grade bonds during recessions. In the United States, the correlation between equities and investment grade bonds is much more significant during slowdowns (0.63) than during accelerations (0.30). From a theoretical point of view, Merton (1974) posits that equity should behave like a call option on the assets of the firm, whereas risky debt should behave like a government bond plus a short position in a put option on the firm's assets. The short option position embedded in credit bonds leads to negative convexity in their payoff profile. As a result, the relationship between credit and equity returns becomes stronger in the Merton model when firm and equity valuations fall.

As a consequence, investment-grade corporate bonds are not substitute to government bonds. Indeed, correlation to equities during bad times should be positive for investment-grade corporate bonds and negative for government bonds. The increasing importance of investment-grade corporate bonds in large institutional portfolios (Ilmanen (2011)) may partly explain the large losses during the last crisis.

Table 1.9: **Return correlations in the euro area, 1999-2013**

	Eurostoxx	Gov	IG
Eurostoxx			
Gov	-0.14		
IG	0.35	0.55	
HY	0.62	-0.08	0.62

Acceleration				Slowdown			
	Eurostoxx	Gov	IG		Eurostoxx	Gov	IG
Eurostoxx				Eurostoxx			
Gov	0.01			Gov	-0.18		
IG	0.09	0.83		IG	0.48	0.42	
HY	0.55	0.21	0.53	HY	0.66	-0.18	0.66

Expansion

	Eurostoxx	Gov	IG
Eurostoxx			
Gov	-0.12		
IG	0.08	0.73	
HY	0.55	0.04	0.56

Recession

	Eurostoxx	Gov	IG
Eurostoxx			
Gov	-0.08		
IG	0.67	0.38	
HY	0.69	-0.14	0.69

Table 1.10: Return correlations in the United States, 1988-2013

	S&P500	Gov	IG
S&P500			
Gov	-0.04		
IG	0.32	0.73	
HY	0.57	-0.02	0.48

Acceleration				Slowdown			
	S&P500	Gov	IG		S&P500	Gov	IG
S&P500				S&P500			
Gov	0.05			Gov	-0.10		
IG	0.30	0.86		IG	0.39	0.53	
HY	0.53	0.09	0.45	HY	0.63	-0.10	0.54

Expansion				Recession			
	S&P500	Gov	IG		S&P500	Gov	IG
S&P500				S&P500			
Gov	-0.02			Gov	-0.10		
IG	0.25	0.86		IG	0.51	0.34	
HY	0.51	0.06	0.42	HY	0.70	-0.20	0.61

1.4.2 A truly diversified strategic asset allocation

Since correlations vary during economic regimes, the main idea is to build a portfolio that would stay diversified when needed. Indeed, assets correlations are a key input and nothing prevents from carefully selecting the correlation matrix used to establish the asset allocation. Only a chronology of the regimes is needed.

An out-of-sample analysis in the United States illustrates this important point (there is no recession in our sample before 2007 in the euro area, thereby drastically limiting an out-of-sample analysis).

Three different strategic asset allocations are computed. Only the correlation matrix used to estimate the weights differs. The first correlation matrix is calculated over the complete period 1988-2002. The second is calculated over the months labelled as slowdown

during the period 1988-2002. The third is calculated over the months labelled as recession during the period 1988-2002.

While respecting the chosen asset allocation, the three portfolios evolve from January 2003 to December 2013.

The investment universe is composed of the following assets: cash (Fed funds), S&P500, governments bonds, investment grade bonds and high-yield bonds.

The chosen targeted volatilities equal 6%, for risk averse investors, and 10%, which is comparable to the historical volatility of the classical 60/40 asset allocation over the period 1988-2002.

Tables 1.11 and 1.12 confirm the importance of the correlation matrix for asset allocation choices. In comparison with the benchmark, allocations based on slowdown or recession matrix reduce the out-of-sample volatility and the losses in extreme negative events (MDD). This is what risk-averse investors value the most.

For instance, considering the worst times (recessions) to build the correlation matrix, allows to reduce the MDD from 18.8% to 17.5% for a targeted volatility of 6% and from 42.0% to 37.6% for a targeted volatility of 10%.

Moreover, for all the degrees of risk aversion and all the targeted volatilities, the strategies based on the slowdown matrix are selected in the best models sets $\widehat{M}_{75\%}^*$. Strategies based on the recession matrix are selected in all best models sets $\widehat{M}_{75\%}^*$ but one. These results confirm the interest of this approach. The strategy based on the full period matrix is only selected in one best models sets $\widehat{M}_{75\%}^*$, for a targeted volatility of 10% and $\gamma = 1$. This strategy achieves higher returns (8.9%) but the volatility is higher than 10% and equals 11.8%, which is not in line with the initial objective.

To sum up briefly, investors should favour asset allocations based on bad times correlation matrix, since the portfolios stay diversified when needed.

Table 1.11: Strategic asset allocation, 6% volatility

United States (2003-2013)			
	Full period	Slowdown	Recession
Average returns	6.7	6.9	6.8
Volatility	6.7	6.4	6.4
<i>CEQ</i> : $\gamma = 1$	0.064	0.067*	0.067*
<i>CEQ</i> : $\gamma = 3$	0.060	0.063*	0.062*
<i>CEQ</i> : $\gamma = 5$	0.054	0.059*	0.058*
MDD	-18.8	-17.9	-17.5

Note: This table reports out-of-sample strategic asset allocation strategies' performances in the United States. Three portfolios are build based on correlation matrix, whose elements are generated from the different economic regimes (full period, slowdown, recession) over the period 1988-2002. The targeted volatility is 6%. While respecting the chosen asset allocation, the five portfolios evolve from January 2003 to December 2013. Returns are monthly and annualized. The volatility corresponds to the annualized standard deviation. The certainty-equivalent return (*CEQ*) is the risk-free rate of return that the investor is willing to accept instead of undertaking the risky portfolio strategy. The Max drawdown (MDD) measures the largest single drop from peak to bottom in the value of a portfolio. * indicates the model is in the set of best models $\widehat{M}_{75\%}^*$.

Table 1.12: Strategic asset allocation, 10% volatility

United States (2003-2013)			
	Full period	Slowdown	Recession
Average returns	8.9	8.7	8.5
Volatility	11.8	10.4	10.2
<i>CEQ</i> : $\gamma = 1$	0.082*	0.082*	0.080
<i>CEQ</i> : $\gamma = 3$	0.068	0.071*	0.070*
<i>CEQ</i> : $\gamma = 5$	0.054	0.060*	0.059*
MDD	-42.0	-40.1	-37.6

Note: This table reports out-of-sample strategic asset allocation strategies' performances in the United States. Three portfolios are build based on correlation matrix, whose elements are generated from the different economic regimes (full period, slowdown, recession) over the period 1988-2002. The targeted volatility is 10%, which is comparable to the historical volatility of the classical 60/40 asset allocation over the period 1988-2002. While respecting the chosen asset allocation, the five portfolios evolve from January 2003 to December 2013. Returns are monthly and annualized. The volatility corresponds to the annualized standard deviation. The certainty-equivalent return (*CEQ*) is the risk-free rate of return that the investor is willing to accept instead of undertaking the risky portfolio strategy. The Max drawdown (MDD) measures the largest single drop from peak to bottom in the value of a portfolio. * indicates the model is in the set of best models $\widehat{M}_{75\%}^*$.

Conclusion

The willingness of investors to bear risk varies over time, larger in good times, and less in bad times, leading to time-varying risk premiums (Cochrane (2016)). Yet, there is still no consensus on the definition of good and bad time. This paper tries to fill this gap.

To this aim, economic cycles are first rigorously defined, namely the classical business cycle and the growth cycle, which seeks to represent the fluctuations around the trend. The growth cycle is better known as the output gap. The description of different economic

phases is refined by jointly considering both economic cycles. It improves the classical analysis of economic cycles by considering sometimes two distinct phases and sometimes four distinct phases. In particular, all recessions involve slowdowns, but not all slowdowns involve recessions. A complete chronology of the macroeconomic regimes is established.

Then, the theoretical influence of economic cycles on time-varying risk premiums is explained and exhibited. The key concept is nominal growth expectations, which is the same thing as saying that income expectations are crucial. The theory of adaptive expectations (Fisher (1911)) makes the link between the cyclical framework and nominal growth expectations. For example, when the real growth rate is above its potential, inflation pressures surge: the nominal growth rate of the economy increases. Adaptive expectations imply that nominal growth rate expectations should increase. Forward-looking investors adjust thus their portfolios according to increasing income expectations.

Dynamic investment strategies based on economic cycles turning points highlight the importance of economical cycles, especially the growth cycle, for euro and dollar-based investors. Indeed, strategies based on output gap turning points statistically outperform not only passive buy-and-hold benchmarks, but also business cycles' strategies, in the United States and in the euro area.

In theory, asset returns are thus not correlated with the business cycle but are primarily caused by the economic cycles. Assessing if the current growth rate of the economy is above or under the trend growth rate is the most crucial task for investors.

The presence of regimes with different correlations and assets' characteristics can enhance strategic asset allocation, which is the most important determinant of long run investment success (Campbell and Viceira (2002)). In particular, investment-grade corporate bonds are not substitute to government bonds and risk-averse investors should select an asset allocation based on a correlation matrix whose elements are generated from bad times periods.

Last but not least, this article opens the door for further research. An asset allocation based on real time leading indicators comes naturally to mind. It could also be very inter-

esting to study how macroeconomic regimes impact the diversifying power of alternative asset classes, such as real estate, commodities, hedge funds or private equity.

Appendix

1.A The double Hodrick-Prescott filter

A time series y_t may be decomposed into a trend or growth component g_t and a cyclical component c_t .

$$y_t = g_t + c_t, \quad t = 1, 2, \dots, T$$

The Hodrick–Prescott filter effects such a decomposition by minimizing the following:

$$\sum_{t=1}^T (y_t - g_t)^2 + \lambda \sum_{t=2}^{T-1} ((g_{t+1} - g_t) - (g_t - g_{t-1}))^2.$$

The first term above is the sum of squared cyclical components $c_t = y_t - g_t$. The second term is a multiple λ of the sum of squares of the trend component's second differences. This second term penalizes variations in the growth rate of the trend component: the larger the value of λ , the higher is the penalty and hence the smoother the trend series.

The Hodrick-Prescott filter is applied twice to become a bandpass filter. The long term trend is first filtered by setting λ to a high value. The filter is then applied again with a small λ to remove seasonal patterns.

Chapter 2

Can macroeconomists get rich nowcasting output gap turning points with a simple machine-learning algorithm?

Abstract

To nowcast output gap turning points, probabilistic indicators are created from a simple and transparent machine-learning algorithm known as Learning Vector Quantization. Those indicators need to be robust, interpretable and preserve economic consistency. The real-time ability of the indicators to quickly and accurately detect economic turning points in the United States and in the euro area is gauged. To assess the value of the indicators, profit maximization measures based on trading strategies are employed in addition to more standard criteria. When comparing predictive accuracy and profit measures, the bootstrap based model confidence set

procedure is applied to avoid data snooping. A substantial improvement in profit measures over the benchmark is found: macroeconomists can get rich nowcasting output gap turning points.

Introduction

In his seminal paper, Okun (1962) defines the potential Gross Domestic Product as the maximum amount of goods and services an economy can produce under conditions of full employment. The output gap, which is the deviation of actual output from estimated potential, has played an important role in the conduct of monetary policy and fiscal policy ever since.

A turning point of the output gap occurs when the current growth rate of the activity passes above or below the potential growth rate, thereby signalling increasing or decreasing inflation pressures. Output gap turning points provide thus extremely reliable pieces of information. For instance, if a central bank wants to tighten monetary policy, because the economy is overheating, a peak of the output gap would indicate that its strategy starts to bear fruit. Moreover, Chapter 1 emphasizes that investment strategies based on output gap turning points achieve impressive risk-adjusted returns, even better than those based on business cycle turning points ¹.

However, detecting economic turning points in real time is easier said than done. For instance, in May 2001, the Survey of Professional Forecasters, conducted by the American Statistical Association and the National Bureau of Economic Research, said there would not be a recession in 2001, even though one had already started.

One stylised fact of economic cycles is the non-linearity: the relationship between economic variables is not simply static, but instead is dynamic and fluctuating. In other words, economic dynamics during economically stressful times are potentially different

¹With a complete different approach, Cooper and Priestley (2009) point out the importance of the output gap for investment decisions. To read more on this subject, see Chapter 1.

from normal times. For example, Phillips (1958) states that the relation between unemployment and the rate of change of wage rates is likely to be highly non-linear.

Real-time regime classification requires thus methods capable of taking into account the non-linearity of the cycles. In this respect, many non-linear parametric models have been proposed, such as smooth-transition autoregressive models (see, for example, Ferrara and Guegan (2005)), non-linear probit models (see, among others, Liu and Moench (2016)) or Markov switching models (see Piger (2011) for a review). Parametric models are effective if the true data generating process (DGP) linking the observed data to the economic regime is known. Nevertheless, the assumption of a particular DGP is often too restrictive or unrealistic. It implies practical issues in estimating parametric models, especially the presence of frequent local maxima in the likelihood. Therefore, in the absence of knowledge of the true DGP, non-parametric methods, such as machine-learning algorithms, may be preferable as they do not rely on a specification of the DGP (Giusto and Piger (ming)).

The machine-learning approach treats the DGP as unknown and performs prediction and classification by generalizing from examples. Giusto and Piger (ming) introduce in economics a transparent and simple machine-learning algorithm known as Learning Vector Quantization (LVQ henceforth), which appears very competitive with commonly used alternatives. LVQ has significant computational advantages over parametric methods: the algorithm is intuitive, easily implemented and can be modified to incorporate data series that arrive with different reporting lags, as well as data series of mixed reporting frequency (Giusto and Piger (ming)). The interpretability of the resulting model makes LVQ especially attractive for complex real life applications: LVQ has been successfully applied to various supervised classification tasks such as radar data extraction or spam email detection.

To get useful signals for policy makers and for investors, this paper attempts to create indicators from LVQ, which could quickly and accurately detect output gap turning points in real time, not only in the United States but also in the euro area. Those indicators

should be robust, interpretable and preserve economic consistency. Indeed, the ability of policy makers to understand these models is indispensable, since it may be a prerequisite for trust (Ribeiro et al. (2016)).

To prove that the models built are effective, several alternative classifiers are computed. Importantly, to avoid data snooping, which occurs when the same data set is employed more than once for inference and model selection, the comparison of predictive accuracy and profit measures is assessed using the model confidence set procedure proposed by Hansen et al. (2011).

The results provide evidence that LVQ is very effective, despite its simplicity. Indeed, a substantial improvement in profit measures over the benchmark and competitive models is found: macroeconomists can get rich nowcasting output gap turning points.

2.1 Learning Vector Quantization

This section briefly introduces LVQ. Refer to Kohonen (2001) for a detailed exposition of this approach.

The LVQ is a prototype-based supervised classification algorithm. The basic idea of this algorithm is to find a natural grouping in a set of data. As supervised method, LVQ uses known target output classifications for each input pattern of the form. LVQ algorithms do not approximate density functions of class samples like Vector Quantization or Probabilistic Neural Networks² do, but directly define class boundaries based on prototypes, a nearest-neighbour rule and a winner-takes-it-all paradigm. In other words, LVQ takes both historical data and its classification as an input, which is then used to train the algorithm. Based on this training, the model can label new data that has not yet been classified.

LVQ has significant computational advantages over parametric methods: the algorithm

²LVQ can be understood as a special case of an artificial neural network: LVQ is a feedforward net with one hidden layer of neurons, fully connected with the input layer.

is intuitive, simply performed and can be easily modified to incorporate data series that arrive with different reporting lags, as well as data series of mixed reporting frequency (Giusto and Piger (ming)). Moreover, LVQ can be implemented when there is a large number of indicators with little increase in computational complexity.

The main idea of the algorithm is to cover the input space of samples with "codebook vectors" (CVs), each representing a region labelled with a class. Once these relevant points are singled out, data is classified to belong to the same class as the closest codebook vector. In this paper, the Euclidean metric is used as it is the dominant metric in the literature.

Learning means modifying the weights in accordance with adapting rules and, therefore, changing the position of a CV in the input space. Since class boundaries are built piecewise-linearly as segments of the mid-planes between CVs of neighbouring classes, the class boundaries are adjusted during the learning process. The tessellation induced by the set of CVs is optimal if all data within one cell indeed belong to the same class.

Giusto and Piger (ming) describe LVQ algorithm as follows. Let X be a collection of N observations $x_n \in \mathbb{R}^m, n = 1, \dots, N$ for which the classification in the set $\{C_k\}_{k=1}^K$ is known. Let there be $\bar{N} \in [K, N]$ codebook vectors $m_i \in \mathbb{R}^m, i = 1, 2, \dots, \bar{N}$ with given initial locations. Finally, let $g = 1, 2, \dots, G$ denote iterations of the algorithm and let α^g be a decreasing sequence of real numbers bounded between zero and one. Given the initial location of the \bar{N} codebook vectors, the LVQ algorithm makes adjustments to their location through these steps:

Step 1 Let $g = 1$ and $n = 1$

Step 2 Identify the codebook vector m_g^c closest to the data point x_n

$$c = \operatorname{argmin}_{i \in \{1, \dots, \bar{N}\}} \|x_n - m_i^g\|$$

Step 3 Adjust the location of the codebook vector with index c according to the following

rule:

$$\begin{cases} m_c^{g+1} = m_c^g + \alpha^g [x_n - m_c^g] & \text{if } x_n \text{ and } m_c^g \text{ belong to the same class} \\ m_c^{g+1} = m_c^g - \alpha^g [x_n - m_c^g] & \text{otherwise} \end{cases}$$

Step 4 If $n + 1 \leq N$ let $n = n + 1$ and repeat from step 2. Otherwise let $n = 1$ and $g = g + 1$ and if $g \leq G$ repeat from step 2; stop otherwise.

Classification after learning is based on a presented sample's vicinity to the CVs: the classifier assigns the same class label to all samples that fall into the same tessellation. This process is analogous to the in-sample parameter estimation and out-of-sample prediction steps employed with parametric statistical models.

Let x_{N+1} a new point for which the classification is unknown. Its class is predicted by first finding the codebook vector m_c that is closest to x_{N+1} in the Euclidean metric:

$$c = \operatorname{argmin}_{i \in \{1, \dots, \bar{N}\}} \|x_{N+1} - m_i\|$$

and then x_{N+1} is assigned to the same class as is assigned to codebook vector m_c .

To define parameters \bar{N} , α and G , Kohonen (2001) recommendations are followed. They are based on a survey of a large number of empirical implementations of the LVQ algorithm. \bar{N} , the total number of codebooks, is set to 70 and the same number of codebooks is assigned to each class. The parameter α equals 0.3, while the number of algorithm iterations, G , is set to 40 times the total number of codebook vectors.

Giusto and Piger (ming) test several approaches to initialize the codebook vectors (k-means, k-medoids, and self-organizing maps) and recommend to initialize the codebook vectors to a randomly drawn set of data vectors in their training sample.

Last but not least, LVQ is transparent: there is no issue on how forecasts are derived (i.e., they do not come from the magical "black box"). The interpretability of the resulting model provides further insight into the nature of the data and makes LVQ especially attractive for complex real life applications.

2.2 Empirical setup

2.2.1 Real time recursive estimation

To implement the LVQ classifier, a classification of economic regimes is needed. The turning points of the output gap separate periods of slowdowns (the growth rate of the economy is below the potential growth rate) and accelerations (the growth rate of the economy is above the potential growth rate)³. A slowdown signals a prolonged period of subdued economic growth though not necessarily an absolute decline in economic activity. This is distinct from business cycles which are defined as a succession of periods of absolute growth (expansion) and decline in economic activity (recession). As a matter of fact, all recessions involve slowdowns, but not all slowdowns involve recessions. For instance, if a slowdown is not severe enough, thanks to counter cyclical policies, then the economy does not fall into recession, but, on the contrary, starts to grow again.

Two economic phases are considered: slowdown and acceleration. Applied to the context of nowcasting, it can be summarized as follows:

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

This paper employs the turning point chronology established in Chapter 1. In the euro area, over the period from January 1999 to December 2013, the number of periods identified as slowdown equals the number of periods identified as acceleration. In the United States, 71 % of the data are classified as acceleration. Over the period from January 1999 to December 2013, there were 9 turning points in the growth cycle in the euro area and in the United States.

In the empirical analysis, a recursive estimation is computed: each month the model

³To implement the investment strategies, the sign and the magnitude of the output gap, which are subject to considerable uncertainty are not needed. The turning points estimations are more robust (Nilsson and Gyomai (2011)).

is estimated with the data and the chronology that would have been available at the time the nowcasting is done. In real time, the complete chronology is not available, but the monthly GDP introduced by Raffinot (2007)⁴ allows to quickly refine the turning point chronology.

The LVQ classifier is thus 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 known⁵. For instance, in January 2012, the chronology that would have been available to implement the LVQ classifier runs over the period from January 1988 to January 2011. The estimation windows is thus expanding as data accumulates.

The turning point identification procedure is based on 100 runs of the LVQ algorithm with different random initializations for codebook vectors, therefore providing 100 results of either 1 or 0. The mean of those results is computed and is assumed to be the probability P_t of being in the regime 1 (acceleration). For a given covariate x_n , based on the learning sample $(R_1, x_1), \dots, (R_{T-12}, x_{T-12})$:

$$P_t^x = E[R_t = 1|x_t]$$

A threshold of 0.5 appears logical to classify regimes:

$$\hat{R}_t^x = \begin{cases} 1, & \text{if } P_t^x \geq 0.5 \\ 0, & \text{otherwise} \end{cases}$$

2.2.2 Data set

The real-time detection of turning points faces the difficult issues of late release dates and data revision. As a matter of fact, key statistics are published with a long delay, are subsequently revised and are available at different frequencies. For example, official

⁴A temporal disaggregation based on business surveys of the non revised values of gross domestic product GDP is used to develop a monthly indicator of GDP

⁵In unreported results, samples through month $T - 18$ and $T - 24$ have been tested and lead to almost the same results.

estimates of economic growth are available on a quarterly basis with a time span of one to three months. To make matters worse, updates to statistical methods can cause revisions even after three years have passed.

However, monthly business and consumer surveys provide unique information such as confidence, rating of business conditions and respondents' expectations for the next months. They are released before the end of the month they relate to or just a few days after and are subject to very weak revisions. They are widely used to detect turning points in the economic cycle. In the euro area, surveys published by the European Commission have been proven to be very effective (Bengoechea et al. (2006)) (see Table 2.A.1 in Appendix 2.A for an exhaustive list).

In the United States, the surveys published by the Institute for Supply Management (ISM), the Conference Board and the National Association of Home Builders (NAHB) are often tested in the literature (Liu and Moench (2016)) (see Tables 2.A.2, 2.A.3 and 2.A.4 in Appendix 2.A for an exhaustive list). The only real economic data tested is the four-week moving average of initial claims for unemployment insurance, which is a weekly measure of the number of jobless claims filed by individuals seeking to receive state jobless benefits.

Moreover, financial series, which are not revised and often available on a daily basis, have also been considered: government bond yields, different yield curves, investment-grade yields, high-yield corporate yields, corporate spreads against government bonds, stock markets, assets volatility, the volatility of volatility of these indexes, the VIX index, the VSTOXX index, which is the VIX equivalent for the euro area and some commodity prices (see Table 2.A.5 in Appendix 2.A for an exhaustive list). This paper uses end of month values to match stock index futures and options contracts settlement prices.⁶

To detect the turning points in real-time, not only original series are screened, but also differentiated series (to underline the phases of low and high pace of growth). Because of the classical trade-off between reliability and advance, different lags of differentiation were

⁶<http://www.cmegroup.com/trading/equity-index/fairvaluefaq.html>

considered: 3 months, 6 months, 9 months, 12 months. The large dataset of predictors consists of more than 450 monthly variables in the euro area and 500 in the United States.

2.2.3 Model evaluation

Classical set of model selection

Two classical economic metrics are computed to evaluate the quality of classification of a model.

The first one is the Brier's Quadratic Probability Score (*QPS*), defined as follows:

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

where $t = 1, \dots, F$ is the number of forecasts. The best model should strive to minimize the *QPS*.

The second one is the area under the Receiver Operating Characteristics curve (AUROC). Given a classifier and an instance, there are four possible outcomes. If the instance is positive and it is classified as positive, it is counted as a true positive ($T_p(c)$). If the instance is negative and classified as negative, it is counted as a true negative ($T_n(c)$). If a negative instance is misclassified as positive, it is counted as a false positive ($F_p(c)$). If a positive instance is misclassified as negative, it is counted as a false negative ($F_n(c)$).

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. A perfect classifier would fall into the top left corner of the graph with a True Positive Rate of 1 and a False Positive Rate of 0.

Accuracy is measured by the Area Under the ROC curve (AUROC), defined by:

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

An area of 1 represents a perfect test, an area of 0.5 represents a worthless test. A general rule of thumb is that an AUROC value exceeding 0.85 indicates a useful prediction performance.

Hanley and McNeil (1982) propose a test to compare the AUROC predictive accuracy. The aim is to test the best models in the selection with another criterion, thereby further reducing the set. The t-statistic for the test of $H_0 : \text{AUROC}_1 = \text{AUROC}_2$ is given by:

$$t = \frac{\text{AUROC}_1 - \text{AUROC}_2}{\sqrt{(\sigma_1^2 + \sigma_2^2 - 2r\sigma_1 * \sigma_2)}}$$

where, AUROC1 and AUROC2 are the areas under the curve for models 1 and 2 which are being compared. Similarly, σ_1 and σ_2 refer to the variances of the AUROCs for model 1 and model 2, respectively. Finally, r is the correlation between the two AUROCs (see Hanley and McNeil (1982) or Liu and Moench (2016) for more details on the test statistic and its implementation.)

Profit maximization measures

For investors, the usefulness of a forecast depends on the rewards associated with the actions taken by the agent as a result of the forecast. In addition to more standard criteria, profit maximization measures are employed.

In order to frame the concept of active portfolio management, a specified investment strategy is required. The investment strategies are as stripped-down and simple as possible without raising concerns that the key results will not carry over to more general and intricate methods or asset classes.

We first consider an equity portfolio manager investing 100€ or 100\$ on January 1, 2007. Each month, the investor decides upon the fraction of wealth to be invested based on the current state of the economy induced by the indicator that would have been

available at the time the decision was made. If the probabilistic indicator classifies the period as acceleration, then the investor can leverage his portfolio (120% of his wealth is invested on the asset and 20% of cash is borrowed), otherwise he only invests 80% of his wealth and 20% is kept in cash.

Moreover, if different asset classes perform differently during different stages of the growth cycle, it might be reasonable to rebalance the portfolio (shifting allocation weights) based on the stage of the growth cycle. The second strategy aims at beating the classic asset allocation for an institutional portfolio, *i.e.* 60% of the portfolio allocated to equities and 40% to fixed income securities (bonds). The investor decides each month to rebalance his portfolio. If the probabilistic indicator indicates acceleration, then 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.

Pesaran and Timmermann (1994) and Han et al. (2013) demonstrate that the total cost of transactions appears to be low, less than 1% (around 50 basis points when trading in stocks while the cost for bonds is 10 basis points). To simplify, no transaction costs are considered.

There are no look-ahead bias: the reallocation takes place at the beginning of the month following the turning point. Indeed, in real time, an investor could not know at the beginning of any month whether a turning point would occur in that month.

For conventional comparison of the portfolio performances, annualized average returns, annualized standard deviation (volatility), Sharpe ratio (SR), which is the mean portfolio return in excess of the risk-free rate divided by the standard deviation of the excess portfolio return⁷ and max drawdown (MDD) are computed. The Sharpe ratio compares the expected returns of an investment to the amount of risk undertaken to capture these returns. The Max drawdown (MDD) is the largest drop from the maximum cumulative return. In brief, the MDD offers investors a worst case scenario.

⁷A risk-free interest rate of zero is assumed when calculating the SR.

Data snooping

Data snooping occurs when a the same data set is employed more than once for inference and model selection. It leads to the possibility that any successful results may be spurious because they could be due to chance (White (2000)). In other words, looking long enough and hard enough at a given data set will often reveal one or more forecasting models that look good but are in fact useless.

To avoid data snooping, the model confidence set (MCS) procedure proposed by Hansen et al. (2011) is computed. The MCS procedure is a 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.

The resulting set contains the best models with with a probability that is no less than $1 - \alpha$ with α being the size of the test.

An advantage of the test is that it not necessarily selects a single model, instead it acknowledges possible limitations in the data since the number of models in the set containing the best model will depend on how informative the data are.

More formally, define a set M_0 that contains the set of models under evaluation indexed by: $i = 0, \dots, m_0$. Let $d_{i,j,t}$ denote the loss differential between two models by

$$d_{i,j,t} = L_{i,t} - L_{j,t}, \forall i, j \in M_0$$

L is the loss calculated from some loss function for each evaluation point $t = 1, \dots, T$. The set of superior models is defined as:

$$M^* = \{i \in M_0 : E[d_{i,j,t}] \leq 0 \forall j \in M_0\}$$

The MCS uses a sequential testing procedure to determine M^* . The null hypothesis being tested is:

$$\begin{cases} H_{0,M} : E[d_{i,j,t}] = 0 \forall i, j \in M \text{ where } M \text{ is a subset of } M_0 \\ H_{A,M} : E[d_{i,j,t}] \neq 0 \text{ for some } i, j \in M \end{cases}$$

When the equivalence test rejects the null hypothesis, at least one model in the set M is considered inferior and the model that contributes the most to the rejection of the null is eliminated from the set M . This procedure is repeated until the null is accepted and the remaining models in M now equal $\widehat{M}_{1-\alpha}^*$.

According to Hansen et al. (2011), the following two statistics can be used for the sequential testing of the null hypothesis:

$$t_{i,j} = \frac{\bar{d}_{i,j}}{\sqrt{\widehat{var}(\bar{d}_{i,j})}} \text{ and } t_i = \frac{\bar{d}_i}{\sqrt{\widehat{var}(\bar{d}_i)}}$$

where m is the number of models in M , $\bar{d}_i = (m - 1)^{-1} \sum_{j \in M} \bar{d}_{i,j}$, is the simple loss of the i^{th} model relative to the averages losses across models in the set M , and $\bar{d}_{i,j} = (m)^{-1} \sum_{t=1}^m d_{i,j,t}$ measures the relative sample loss between the i^{th} and j^{th} models. Since the distribution of the test statistic depends on unknown parameters a bootstrap procedure is used to estimate the distribution.

The MCS can be applied with classical criteria loss function (Brier's Quadratic Probability Score) or with profit maximization loss function (risk-adjusted returns for example).

As regards investment strategies, it should be noted that the MCS aims at finding the best model and all models which are indistinguishable from the best, not those better than the benchmark. To determined if models are better than the benchmark, the stepwise test of multiple reality check by Romano and Wolf (2005) and the stepwise multiple superior predictive ability test by Hsu et al. (2013) should be considered. However, investors can conclude that their strategies "beat" the benchmark, if the benchmark is not selected in the best models set.

2.2.4 Model selection

Real time issue

In real time, the investor would not have known which models were selected if the classifier is evaluated over the entire period from January 1999 to December 2013. To prove that the chosen models can generate significant profits, a two-step selection process is needed.

A set of models is selected over the period from January 1999 to December 2006 based on classical criteria. Once the best models are selected, investment strategies are computed over the period from January 2007 to December 2013.

Model selection based on classical metrics

Boivin and Ng (2006) suggest that cross-correlation of regressors in large datasets might result in inaccurate forecasts and hence a smaller set is more likely to provide a smaller average forecast error. To narrow down the predictors to only those that are relevant, the identification of a first candidate set of time series of interest is done, by examining the series one by one. The best 15 series according to the QPS are retained.

All possible combinations of four variables from the selected predictors are computed (1365 models in total). Note that four is an arbitrary choice, consistent with the number of series used in Stock and Watson (1989). The best models set is selected based on the MCS procedure with QPS as loss function. Since the aim of this study is to highlight that nowcasting turning points allows to make profits, the confidence level for the MCS is set as low as possible to only select few superior models in the resulting set.

The test proposed by Hanley and McNeil (1982) to compare the Area Under the Receiver Operating Characteristics curve (AUROC) predictive accuracy is then computed. The idea is to test the selected models with another criterion, thereby potentially reducing the set.

2.2.5 Competitive models

To prove that the models are significantly better than random guessing, several alternative classifiers, which assign classes arbitrarily, are computed. The first one (*Acc*) classifies all data as "acceleration", the second one (*Slow*) classifies all data as "slowdown" and the last one (*Random*) randomly assigns classes based on the proportions found in the training data. Thousand different simulations are computed and average criteria are provided.

A comparison of LVQ against competing parametric models comes naturally to mind. Parametric models in academic studies rarely focus on output gap turning points and often employ hard data, such as industrial production or unemployment. The delay of publication and the revisions of the data make a "pseudo real-time" comparison hard to perform. Nevertheless, the term spread has been proved to be an excellent leading indicator of recession in the United States (Liu and Moench (2016)) and in the euro area (Duarte et al. (2005)). Nowcasts from probit models based on the term spread are thus computed⁸.

For a given covariate x_n , based on the learning sample $(R_1, x_1), \dots, (R_{T-12}, x_{T-12})$, the model is characterized by the simple equation:

$$P(R_t^{probit} = 1) = \Phi(\alpha_0 + \alpha_1 x_t)$$

where Φ denotes a standard Gaussian cumulative distribution function, i. e.

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{1}{2}t^2} dt$$

The probit model maximizes the following log likelihood function:

⁸Markov-switching dynamic factor models have been proven effective to identify business cycle turning points (Camacho et al. (2015)). But, as previously explained, variable selection in factor analysis is not straightforward: forecasts often improve by focusing on a limited set of highly informative series. Moreover, the factors are mixtures of all the different series which can mean that they are very difficult to interpret economically.

$$\ln L(\alpha_0, \alpha_1) = \sum_{t=1}^{T-12} (1 - R_t) \ln[1 - \Phi(\alpha_0 + \alpha_1 x_t)] + R_t \ln(\Phi(\alpha_0 + \alpha_1 x_t))$$

Moreover, some parametric models can use the information of our large data sets by throwing away series that provide little information. The most popular strategy for very high-dimensional regression is the estimation of penalized linear models, especially elastic-net (Zou and Hastie (2005)).

Elastic-net solves the following problem:

$$\min_{\beta_0, \beta} \frac{1}{N} \sum_{i=1}^N w_i l(y_i, \beta_0 + \beta^T x_i) + \lambda [(1 - \alpha) \|\beta\|_2^2 / 2 + \alpha \|\beta\|_1]$$

over a grid of values of λ covering the entire range. Here $l(y, \eta)$ is the negative log-likelihood contribution for observation i . The elastic-net penalty is controlled by α , and bridges the gap between lasso ($\alpha = 1$, the default) and ridge ($\alpha = 0$). The tuning parameter λ controls the overall strength of the penalty and is chosen by 5-fold cross validation.

2.3 Empirical results

2.3.1 United States

Models selection

The univariate analysis points out the importance of the following series⁹: Average weekly initial claims for unemployment insurance, High-yield corporate spreads (BofA Merrill Lynch US High Yield 100 Index), Investment grade corporate bonds, Current business conditions (Conference Board), S&P500, VIX, Yield curve, ISM Purchasing Managers Index (manufacturing survey), ISM new orders (manufacturing survey), ISM supplier deliveries (manufacturing survey), Jobs hard to get (Conference Board), Jobs plentiful

⁹Given the large number of realized measures, presenting summary statistics for all possible combinations is not feasible.

(Conference Board), Jobs plentiful minus jobs hard to get (Conference Board), NAHB present sales, NAHB sales Expected Next Six Months.

The MCS procedure selects two models in $\widehat{M}_{75\%}^*$. They are very close: only one component differs. The components of the first model are:

- Average weekly initial claims for unemployment insurance
- High-yield corporate spreads
- Current business conditions
- S&P500

As regards the second model, instead of the S&P500, the ISM Purchasing Managers Index (manufacturing survey) is introduced.

The choice of the predictors clearly preserves economic consistency and is in accord with other studies. In particular, Ng (2014) concludes that risky bonds and employment variables have been proved to have predictive power, when the interest rate spreads were uninformative.

Table 2.1 displays classical metrics. The performance of the models are impressive and are consistent with the results found in Berge (2015) and in Ng (2014), as regards nowcasting.

Table 2.1: **Classical evaluation criteria in the United States**

	AUROC	QPS
Model 1	0.92	0.062
Model 2	0.91	0.066

Note: This table reports two classical metrics used to evaluate the quality of classification of the best two models: the area under the ROC curve (AUROC) and the Brier's Quadratic Probability Score (QPS).

The test proposed by Hanley and McNeil (1982) highlights that no model is more effective than the other, as regards the AUROC criterion.

Investment strategies

Table 2.2 points out that active investment strategies based on the growth cycle turning points induced by the LVQ models statistically outperform the passive buy-and-hold benchmark and competitive models. Timing the stock market in real time based on economic cycles is thus difficult but possible. These results have also implications for the risk management and hedging. Especially, in the options market one can utilize the current state of the economy to hedge the portfolio against the possible price declines. For example, writing an out-of-the-money covered call or buy a put option when the stock market is expected to decrease would limit the losses. The best models set $\widehat{M}_{75\%}^*$ consists of the two LVQ models, but the first model displays better returns and a better performance to volatility ratio.

Table 2.2: **Summary of return and risk measures in the United States: 120/80 equity strategy**

	Average returns	Volatility	SR	MDD
Model 1	10.6	17.7	0.61**	-43.1
Model 2	10.3	17.4	0.59**	-43.5
<i>Prob</i>	8.0	19.5	0.41	-52.4
<i>Elastic – net</i>	9.0	18.2	0.49	-48.4
<i>Acc</i>	8.1	20.4	0.40	-57.9
<i>Slow</i>	5.4	13.6	0.40	-43.0
<i>Random</i>	7.0	17.8	0.40	-50.9
<i>Benchmark</i>	6.7	17.1	0.39	-50.9

Note: This table reports profit maximization measures for 120/80 equity strategy based on the state of the growth cycle induced by the models. Returns are monthly and annualized. The volatility corresponds to the annualized standard deviation. The performance to volatility ratio (SR) compares the expected returns of an investment to the amount of risk undertaken to capture these returns. The Max drawdown (MDD) measures the largest single drop from peak to bottom in the value of a portfolio. ** indicates the model is in the set of best models $\widehat{M}_{75\%}^*$. Model 1 refers to the first model, Model 2 refers to the first model, *Prob* refers to the probit model based on the term spread, *Elastic – net* refers to elastic-net logistic model, *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 and *Benchmark* refers to the passive buy-and-hold investment strategy.

Table 2.3 emphasizes that dynamic asset allocation delivers a substantial improvement in risk-adjusted performance as compared to static asset allocation, especially for investors who seek to avoid large losses. In real time, portfolio rebalancing based on the stage of the output gap is thus realisable.

As regards model selection, it should be noted that the strategy based on *Slow*, which is overweighted in bonds, belongs to $\widehat{M}_{75\%}^*$, in line with the thrilling and non-reproducible

performance of the bond market. Importantly, both LVQ models are still selected in the best models set. Once again, the first model exhibits better returns along with a better performance to volatility ratio.

Table 2.3: **Summary of return and risk measures in the United States: dynamic asset allocation**

	Average returns	Volatily	SR	MDD
Model 1	8.6	9.7	0.89**	-20.2
Model 2	8.3	9.8	0.84**	-18.0
<i>Prob</i>	6.7	13.1	0.51	-39.2
<i>Elastic – net</i>	7.6	10.8	0.70	-24.6
<i>Acc</i>	6.4	13.4	0.48	-41.5
<i>Slow</i>	5.7	6.5	0.87**	-18.0
<i>Random</i>	6.4	10.8	0.60	-29.8
<i>Benchmark</i>	6.1	9.8	0.61	-30.6

Note: This table reports profit maximization measures for a dynamic asset allocation between bonds and equities based on the state of the growth cycle induced by the models. Returns are monthly and annualized. The volatility corresponds to the annualized standard deviation. The performance to volatility ratio (SR) compares the expected returns of an investment to the amount of risk undertaken to capture these returns. The Max drawdown (MDD) measures the largest single drop from peak to bottom in the value of a portfolio. ** indicates the model is in the set of best models $\bar{M}_{75\%}^*$. Model 1 refers to the first model, Model 2 refers to the first model, *Prob* refers to the probit model based on the term spread, *Elastic – net* refers to elastic-net logistic model, *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 and *Benchmark* refers to the passive buy-and-hold investment strategy.

Graphic illustration

Figure 2.1 illustrates the behavior of the first model in real time. The indicator gives reliable signals. This indicator is not volatile and displays a high persistence in the regime classification. November 2012 is the only false signal.

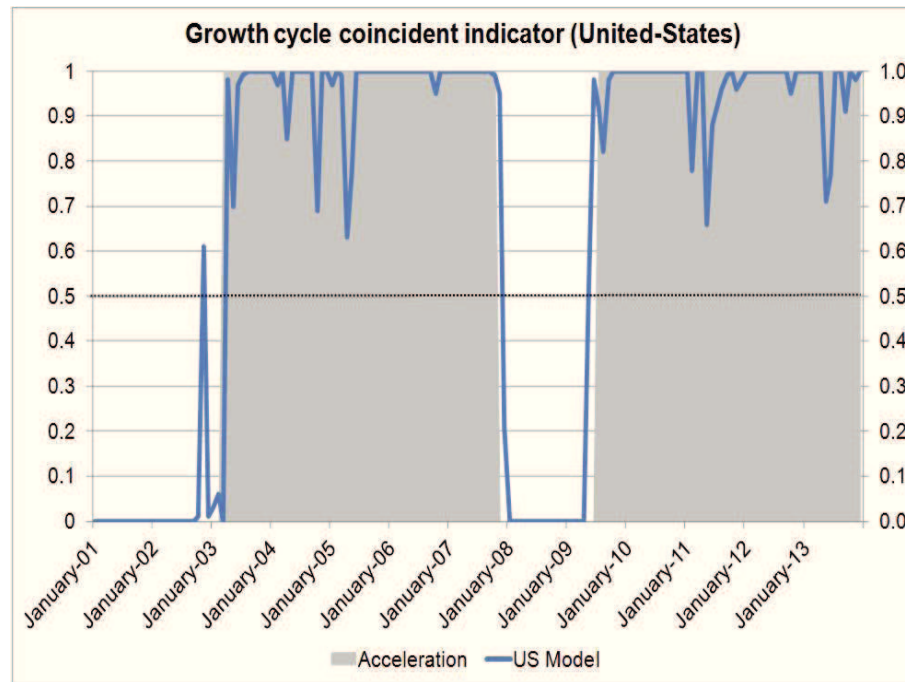


Figure 2.1: Recursive real time classification of the output gap in the United States

2.3.2 Euro area

Models selection

The univariate analysis highlights the interest of the following series: Eurostoxx, Yield curve, Economic Sentiment Indicator (ESI), Business climate indicator (BCI), Consumer confidence index, Employment expectations for the months ahead in the consumer survey, Financial situation over the last 12 months in the consumer survey, Financial situation over the next 12 months in the consumer survey, Industrial confidence indicator, Production expectations for the months ahead in the industry survey, Construction confidence indicator, Employment expectations for the months ahead in the construction survey, Price expectations for the months ahead in the construction survey, Retail confidence indicator, Future business situation in the retail survey.

Just like for the United States, $\widehat{M}_{75\%}^*$ is composed of two models. Again, these models are close and only one component differs. The components of the first model are:

- Employment expectations for the months ahead in the consumer survey
- Financial situation over the last 12 months in the consumer survey
- Production expectations for the months ahead in the industry survey
- Employment expectations for the months ahead in the construction survey

As regards the second model, the last component is the price expectations for the months ahead in the construction survey.

The importance of expectations survey questions to detect turning points should be noted. Moreover, financial variables are not introduced in the indicators. This result may stem from the preponderance of bank loans in corporate financing. About 70% of firms' external financing in the euro area comes via the banking system, compared with only 30% in the United States. It implies that the corporate bond market is still small and that corporate spreads are not necessarily representative of the businesses' financing costs.

Table 2.4 highlights classical metrics for the best models in the euro area. The choice between models is difficult to perform. The test proposed by Hanley and McNeil (1982) emphasizes that no model is more effective than the other, as regards the AUROC criterion.

Metrics in the euro area are less impressive than in the United States. Indeed, the persistence of the regimes is smaller in the euro area growth cycle, the real-time classification is thus more difficult.

Table 2.4: **Classical evaluation criteria in the euro area**

	AUROC	QPS
Model 1	0.89	0.105
Model 2	0.87	0.107

Note: This table reports two classical metrics used to evaluate the quality of classification of the best two models: the area under the ROC curve (AUROC) and the Brier's Quadratic Probability Score (QPS)..

Investment strategies

Table 2.5 emphasizes that several active investment strategies outperform the passive buy-and-hold benchmark. In real time, timing the stock market based on output gap

turning points is thus feasible. Naturally, the risk management and hedging implications described for the United States also apply for the euro area.

The probit model *Prob*, the elastic-net model and the two LVQ models are selected in the best models set $\widehat{M}_{75\%}^*$. However, the first model seems to provide more useful signals as it reduces the MDD and gets better returns associated with a better performance to volatility ratio.

Table 2.5: **Summary of return and risk measures in the euro area: 120/80 equity strategy**

	Average returns	Volatility	SR	MDD
Model 1	4.7	17.8	0.26**	-45.7
Model 2	4.5	17.7	0.25**	-45.8
<i>Prob</i>	5.0	20.9	0.24**	-48.3
<i>Elastic – net</i>	4.5	18.9	0.24**	-48.4
<i>Acc</i>	3.2	23.4	0.14	-60.8
<i>Slow</i>	2.1	15.6	0.14	-45.5
<i>Random</i>	4.0	20.6	0.20	-52.6
<i>Benchmark</i>	2.7	19.5	0.14	-53.7

Note: This table reports profit maximization measures for 120/80 equity strategy based on the state of the growth cycle induced by the models. Returns are monthly and annualized. The volatility corresponds to the annualized standard deviation. The performance to volatility ratio (SR) compares the expected returns of an investment to the amount of risk undertaken to capture these returns. The Max drawdown (MDD) measures the largest single drop from peak to bottom in the value of a portfolio. ** indicates the model is in the set of best models $\widehat{M}_{75\%}^*$. Model 1 refers to the first model, Model 2 refers to the first model, *Prob* refers to the probit model based on the term spread, *Elastic – net* refers to elastic-net logistic model, *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 and *Benchmark* refers to the passive buy-and-hold investment strategy.

Table 2.6 highlights that dynamic asset allocation delivers a substantial improvement in risk-adjusted performance as compared to static asset allocation.

The best models set $\widehat{M}_{75\%}^*$ consists of the two LVQ models, but the first model exhibits better returns along with a better performance to volatility ratio.

To sum up, Model 1 seems to be the better choice, even if it is not statistically better than Model 2.

Table 2.6: **Summary of return and risk measures in the euro area: dynamic asset allocation**

	Average returns	Volatily	SR	MDD
Model 1	5.7	10.3	0.53**	-20.9
Model 2	5.5	10.5	0.52**	-21.3
<i>Prob</i>	5.4	13.2	0.41	-25.5
<i>Elastic – net</i>	5.1	11.3	0.45	-23.5
<i>Acc</i>	3.0	15.5	0.19	-44.5
<i>Slow</i>	3.7*	7.9	0.47	-21.4
<i>Random</i>	4.4	13.0	0.34	-31.7
<i>Benchmark</i>	3.4	11.6	0.29	-33.8

Note: This table reports profit maximization measures for a dynamic asset allocation between bonds and equities based on the state of the growth cycle induced by the models. Returns are monthly and annualized. The volatility corresponds to the annualized standard deviation. The performance to volatility ratio (SR) compares the expected returns of an investment to the amount of risk undertaken to capture these returns. The Max drawdown (MDD) measures the largest single drop from peak to bottom in the value of a portfolio. ** indicates the model is in the set of best models $\widehat{M}_{75\%}^*$. Model 1 refers to the first model, Model 2 refers to the first model, *Prob* refers to the probit model based on the term spread, *Elastic – net* refers to elastic-net logistic model, *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 and *Benchmark* refers to the passive buy-and-hold investment strategy.

Graphic illustration

The figure 2.2 illustrates the behavior of the first model in real time. The euro area probabilistic indicator appears more volatile than the United States probabilistic indicator. Yet, low phases detected are not erratic and display a high persistence. There are only three false signals, which do not last for a long time: September 2003, September 2004 and June 2012.

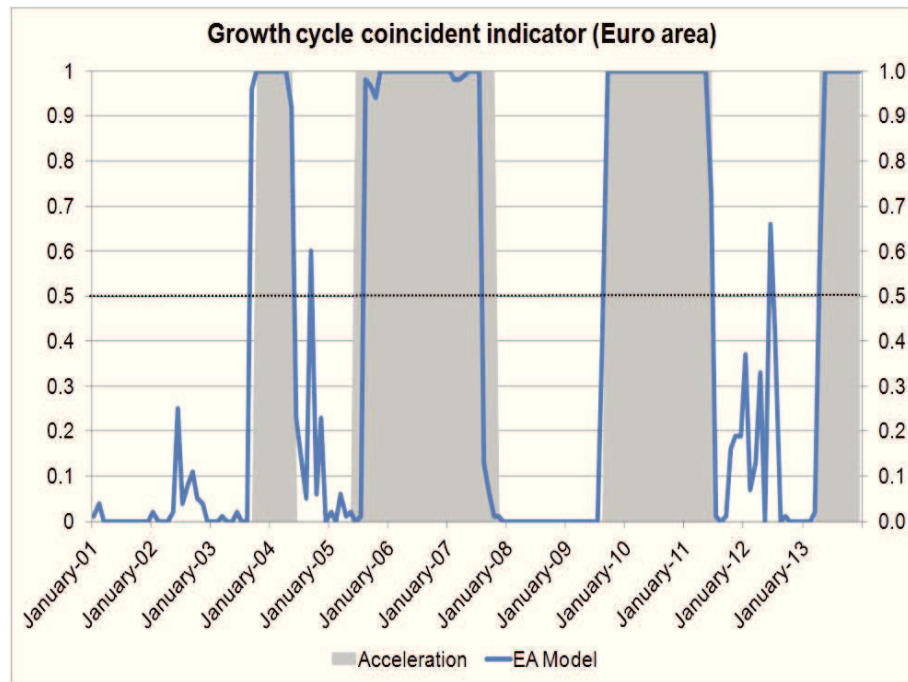


Figure 2.2: Recursive real time classification of the output gap in the euro area

Conclusion

Output gap turning points provide extremely reliable pieces of information for policy makers: a turning point of the output gap occurs when the current growth rate of the activity passes above or below the potential growth rate, thereby signalling increasing or decreasing inflation pressures. Moreover, Chapter 1 emphasizes that investment strategies based on output gap turning points achieve impressive risk-adjusted returns, even better than those based on business cycle turning points.

To quickly and accurately detect output gap turning points in real time, probabilistic indicators are created from a simple and transparent machine-learning algorithm known as Learning Vector Quantization, recently introduced in economics by Giusto and Piger (ming). Those indicators should be robust, interpretable and preserve economic consistency. Indeed, the ability of policy makers to understand these models is indispensable, since it may be a prerequisite for trust.

To select the best model, profit maximization measures based on trading strategies

are used in addition to more standard criteria, such as Quadratic Probability Score and area under the ROC curve (AUROC). When comparing predictive accuracy and profit measures, the model confidence set procedure proposed by Hansen et al. (2011) is applied to avoid data snooping.

The results provide evidence that LVQ is very effective, despite its simplicity. Indeed, a substantial improvement in profit measures over the benchmark and competitive models is found. Economic and financial indicators can thus be exploited to quickly identify turning points in real time in the United States and in the euro area. It leads to useful implications for investors practising active portfolio and risk management and for policy makers as tools to get early warning signals.

Last but not least, this article opens the door for further research. An attempt to forecast output gap turning points three to twelve months ahead comes naturally to mind.

Appendix

2.A Explanatory variables

2.A.1 Economic Surveys

Table 2.A.1: **European Commission’ surveys**

Economic Sentiment Indicator	<ul style="list-style-type: none"> Euro area Belgium Germany Ireland Greece Spain France Italy Cyprus Luxembourg Malta Netherlands Austria Portugal Finland
Industry	<ul style="list-style-type: none"> Business Climate Indicator Confidence Indicator Production trend observed in recent months Assessment of order-book levels Assessment of export order-book levels Assessment of stocks of finished products Production expectations for the months ahead Selling price expectations for the months ahead Employment expectations for the months ahead
Services	<ul style="list-style-type: none"> Confidence Indicator Business situation development over the past 3 months Evolution of the demand over the past 3 months Expectation of the demand over the next 3 months Evolution of the employment over the past 3 months Expectations of the employment over the next 3 months Expectations of the prices over the next 3 months
Retail trade	<ul style="list-style-type: none"> Confidence Indicator Business activity (sales) development over the past 3 months Volume of stock currently hold Orders expectations over the next 3 months Business activity expectations over the next 3 months Employment expectations over the next 3 months Prices expectations over the next 3 months
Construction	<ul style="list-style-type: none"> Building activity development over the past 3 months Evolution of your current overall order books Employment expectations over the next 3 months

Table 2.A.2: **Institute for Supply Management(ISM)**

Manufacturing ISM	PMI
	New Orders
	Production
	Employment
	Supplier Deliveries
	Inventories
	Customers' Inventories
	Prices
	Backlog of Orders
	New Export Orders
	Imports
Non-Manufacturing	BNI
	New Orders
	Production
	Employment
	Supplier Deliveries
	Inventories
	Customers' Inventories
	Prices
	Backlog of Orders
	New Export Orders
	Imports

Table 2.A.3: Conference Board

Consumer Confidence Index	Consumer Confidence Index
	Present Situation
	Expectations
Appraisal of Present Situation	Business Conditions
	Good
	Bad
	Normal
Employment	Jobs plentiful
	Jobs not so plentiful
	Jobs hard to get
Expectations for Six Months Hence	Business Conditions
	Better
	Worse
	Same
Employment	More jobs
	Fewer jobs
	Same
Income	Increase
	Decrease
	Same
Plans to Buy Within Six Months	Automobile
	Home
Major appliances	Total plans
	Refrigerator
	Washing machine
	TV Set
	Vacuum Cleaner
	Range
	Clothes dryer
	Air conditioner
	Carpet

Table 2.A.4: National Association of Home Builders (NAHB)

Housing Market Index (HMI)
Single Family Sales: Present
Single Family Sales: Next 6 Months
Traffic of Prospective Buyers

2.A.2 Financial series

Table 2.A.5: **Financial variables**

Government bond yields	3 months 6 months 1-year 2-year 5-year 10-year 30-year
Yield curves	6 months minus 3 months 1-year minus 3 months 1-year minus 6 months 2-year minus 3 months 2-year minus 6 months 2-year minus 1-year 5-year minus 3 months 5-year minus 6 months 5-year minus 1-year 5-year minus 2-year 10-year minus 3 months (term spread) 10-year minus 6 months 10-year minus 1-year 10-year minus 2-year 10-year minus 5-year 30-year minus 10-year
Investment-grade yields	AAA AA A BBB
High-yield corporate yields	
Corporate spreads against government bonds	AAA minus 5-year AA minus 5-year A minus 5-year BBB minus 5-year High-yield minus 5-year
Stock markets	large caps large caps sectors mid caps small caps
Asset volatility	stock markets government bonds investment-grade bonds
Implicit volatility	VIX Index (United States) VSTOXX Index (euro area)
Commodity prices (local currency)	crude oil natural gas gold silver CRB index

Chapter 3

Investing through Economic Cycles with Ensemble Machine Learning Algorithms

Abstract

Ensemble machine learning algorithms, referred to as random forest and as boosting, are applied to quickly and accurately detect economic turning points in the United States and in the euro area. The two key features of those algorithms are their abilities to entertain a large number of predictors and to perform estimation and variable selection simultaneously. The real-time ability to nowcast economic turning points is gauged. To assess the value of the models, profit maximization measures are employed in addition to more standard criteria. When comparing predictive accuracy and profit measures, the model confidence set procedure is applied to avoid data snooping. The investment strategies based on economic regimes induced by the models achieve impressive risk-adjusted returns: timing the market is thus possible.

Introduction

The potential rewards of successful market timing are more than attractive. For instance, investment strategies based on the turning points of the growth cycle, defined as the deviation of the real GDP to its long-term trend, achieve impressive risk-adjusted returns... in theory (see Chapter 1)¹.

Yet, economic turning points detection in real time is a notorious difficult task. Economists often fail to detect if a new economic phase has already begun. For example, the Survey of Professional Forecasters conducted in May 2008, by the American Statistical Association and the National Bureau of Economic Research, said there would not be a recession in 2008, even though one had already started.

One stylised fact of economic cycles is the non-linearity: the effect of a shock depends on the rest of the economic environment. For instance, small shock, such as a decrease in housing prices, can sometimes, but not always, have large effects, such as a recession. Real-time regime classification and turning points detection require thus methods capable of taking into account the non-linearity of the cycles. In this respect, many parametric models have been proposed, especially Markov switching models (see, among others, Piger (2011)) and probit models (see, for example, Liu and Moench (2016)). Parametric models are effective if the true data generating process (DGP) linking the observed data to the economic regime is known. In practice, however, one might lack such strong prior knowledge. It leads to practical issues in estimating parametric models, especially the presence of frequent local maxima in the likelihood. Therefore, in the absence of knowledge of the true DGP, non-parametric methods are advocated, such as machine-learning algorithms, as they do not rely on a specification of the DGP (Giusto and Piger (ming)).

¹If the long-term trend is considered as the estimated potential level, then the growth cycle equals the output gap. A turning point of the output gap occurs when the current growth rate of the activity is above or below the potential growth rate, thereby signalling increasing or decreasing inflation pressures. Quickly detecting growth cycle turning points provides thus extremely reliable pieces of information for the conduct of monetary policy. For instance, if a central bank wants to loosen monetary policy, because inflation is running under the target, a through of the output gap would indicate that its strategy starts to bear fruit.

The machine-learning approach assumes that the DGP is complex and unknown and attempts to learn the response by observing inputs and responses and finding dominant patterns. This places the emphasis on a model's ability to predict well and focuses on what is being predicted and how prediction success should be measured. Over the last couple of decades, researchers in the computational intelligence and machine learning community have developed complex methods, also called ensemble learning, which improve prediction performances. Ensemble methods are learning models that achieve performance by combining the opinions of multiple learners. The two most popular techniques for constructing ensembles are random forest (Breiman (2001)) and boosting (Schapire (1990)). The two features of those algorithms are their abilities to entertain a large number of predictors and to perform estimation and variable selection simultaneously. Paradoxically, both methods work by adding randomness to the data (Varian (2014)), although they have substantial differences. Random forest relies on simple averaging of models in the ensemble and boosting is an iterative process where the errors are kept being modelled.

While the random forest algorithm is usually applied in medical research and biological studies, it is largely unknown in economics and to the best of my knowledge has not been applied to economic turning point detection. Boosting is increasingly applied to empirical problems in economics. Ng (2014) and Berge (2015) apply the algorithm to the problem of identifying business cycle turning points in the United States.

In this paper, random forest and boosting algorithms are applied to create several models aiming at quickly and accurately detecting growth cycle turning points in real time, not only in the United States but also in the euro area.

The real-time ability to nowcast economic turning points is assessed. Since, for investors, the usefulness of a forecast depends on the rewards associated with the actions taken by the agent as a result of the forecast, profit maximization measures based on trading strategies are employed in addition to more standard criteria.

To avoid data snooping, which occurs when a given set of data is used more than once for purposes of inference or model selection, the comparison of predictive accuracy and

profit measures is assessed using the model confidence set procedure (Hansen et al. (2011)). It is a model selection algorithm, which filters a set of models from a given entirety of models. It aims at finding the best model and all models which are indistinguishable from the best.

The findings of the paper can be summarized as follows: ensemble machine learning algorithms detect economic turning points accurately. Investment strategies based on economic regimes induced by the models achieve thus excellent risk-adjusted returns in the United States and in the euro area. Among ensemble machine learning algorithms, there is no clear winner. Depending on the data and the objective, random forest sometimes performs better than boosting, sometimes not.

The rest of the paper proceeds as follows. Section 1 introduces ensemble machine learning algorithms, referred to as random forest and as boosting. Section 2 describes the empirical set up: the turning point chronology, the data-set, the alternative classifiers and the evaluation of the forecasts. Section 3 analyses the empirical results.

3.1 Ensemble Machine Learning Algorithms

Making decisions based on the input of multiple people or experts has been a common practice in human civilization and serves as the foundation of a democratic society. Over the last couple of decades, researchers in the computational intelligence and machine learning community have studied schemes that share such a joint decision procedure. Ensemble methods are learning models that achieve performance by combining the opinions of multiple learners.

Two of the most popular techniques for constructing ensembles are random forest (Breiman (2001)) and boosting (Schapire (1990)). The two key features of those algorithms are their abilities to entertain a large number of predictors and to perform estimation and variable selection simultaneously.

Paradoxically, both methods work by adding randomness to the data, but adding

randomness turns out to be a helpful way of dealing with the overfitting problem (Varian (2014)). Overfitting denotes the situation when a model targets particular observations rather than a general structure: the model explains the training data instead of finding patterns that generalize it. In other words, attempting to make the model conform too closely to slightly inaccurate data can infect the model with substantial errors and reduce its predictive power.

Nevertheless, those methods have substantial differences. Random forest relies on simple averaging of models in the ensemble. They derive their strength from two aspects: randomizing subsamples of the training data and randomizing the selection of features. Boosting methods are based on a different strategy of ensemble formation: boosting combines models that do not perform particularly well individually into one with much improved properties. The main idea is to add new models to the ensemble sequentially. At each particular iteration, a new weak, base-learner model is trained with respect to the error of the whole ensemble learned so far.

3.1.1 Random forest

Random forest (RF henceforth) is a non-parametric statistical method for both high-dimensional classification and regression problems, which requires no distributional assumptions on covariate relation to the response.

RF is a way of averaging multiple deep decision trees, trained on different parts of the same training set, with the goal of overcoming overfitting problem of individual decision tree. In other words, RF builds a large collection of de-correlated trees and then averages their predictions. The method is fast, robust to noise and produces surprisingly good out-of-sample fits, particularly with highly nonlinear data (Caruana and Niculescu-Mizil (2005)).

Classification and Regression Trees algorithm

Classification and regression trees (CART henceforth), introduced by Breiman et al. (1984), are machine-learning methods for constructing prediction models from data that can be used for classification or regression. The models are obtained by recursively partitioning the data space and fitting a simple prediction model within each partition. As a result, the partitioning can be represented graphically as a decision tree.

The tree is generated in a recursive binary way, resulting in nodes connected by branches. A node, which is partitioned into two new nodes, is called a parent node. The new nodes are called child nodes. A terminal node is a node that has no child nodes.

A CART procedure is generally made up of two steps. In the first step, the full tree is built using a binary split procedure. The full tree is an overgrown model, which closely describes the training set. In the second step, the model is pruned to avoid overfitting.

Given a dataset with explanatory inputs x , the CART algorithm can be summarized as follows:

Step 1 Find each predictor's best split:

Sort each predictor's entries by increasing value. Iterate over all values of the sorted predictor and find the candidate for the best split. That is the value that maximizes the splitting criterion.

Step 2 Find the node's best split:

To actually perform the split, compare all evaluated predictors from step 1 and choose the split, that maximizes the splitting criterion.

Step 3 Let s be this best split of the winning predictor. All $x \leq s$ are sent to the left node and all $x > s$ to the right node.

So constructing a CART is accomplished by finding the best split, which is just trying every possibility, calculating the "goodness" of every possible split and choose the best

one. For every split at node t a splitting criterion $\Delta i(s|t)$ is calculated. The best split s , at node t maximizes this splitting criterion $\Delta i(s|t)$, based on the Gini criterion in classification problems and measured by mean squared error in regression trees. For classification, given a node t with estimated class probabilities $p(j|t)$ with $j = 1, \dots, J$ being the class label, a measure of node impurity given t is:

$$i(s|t) = 1 - \sum_j p(j|t)^2 = \sum_{j \neq k} p(j|t)p(k|t)$$

A search is then made for the split that most reduces node, or equivalently tree, impurity.

Construction of a random forest

RF is an ensemble of tree predictors. Each decision tree is built from a bootstrapped sample of the full dataset (Efron and Tibshirani (1994)) and then, at each node, only a random sample of the available variables is used as candidate variables for split point selection. Thus, instead of determining the optimal split on a given node by evaluating all possible splits on all variables, a subset of the input variables are randomly chosen, and the best split is calculated only within this subset. Once an ensemble of K trees is built, the predicted outcome (final decision) is obtained as the average value over the K trees.

Averaging over trees, in combination with the randomisation used in growing a tree, enables random forests to approximate a rich class of functions while maintaining a low generalisation error. This enables random forests to adapt to the data, automatically fitting higher-order interactions and non-linear effects, while at the same time keeping overfitting in check (Ishwaran (2007)). As the number of trees increases, the generalization error converges to a limit (Breiman (2001)).

A RF is constructed by the following steps:

Step 1 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.

Step 2 Take a bootstrap sample Z of the N observations. So about two third of the cases are chosen. Then select randomly m features.

Step 3 Grow a CART using the bootstrap sample Z and the m randomly selected features.

Step 4 Repeat the steps 2 and 3, K times.

Step 5 Output the ensemble of trees T_1^K

For regression, to make a prediction at a new point x :

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

For classification, each tree gives a classification for x . The forest chooses the class that has the most out of n votes. Calculating the associated probability is easily done.

Since Breiman (2001) uses unpruned decision trees as base classifiers, RF has basically only one parameter to set: the number of features to randomly select at each node. Typically, for a classification problem with M features, \sqrt{M} features (rounded down) are used in each split and $M/3$ features (rounded down) with a minimum node size of 5 as the default are recommended for regression problems (Friedman et al. (2000)).

3.1.2 Boosting

Boosting is based on the idea of creating an accurate learner by combining many so-called "weak learners" (Schapire (1990)), i.e., with high bias and small variance. The main concept of boosting is to add new models to the ensemble sequentially. At each particular iteration, a new weak, base-learner model is trained with respect to the error of the whole ensemble learned so far. The final model hopefully yields greater predictive performance than the individual models. The heuristics is thus simple: an iterative process where the errors are kept being modelled.

The original boosting algorithms such as AdaBoost (Freund and Schapire (1997)), were purely algorithm-driven, which made the detailed analysis of their properties and performance rather difficult (Schapire (2003)). The gradient descent view of boosting (Friedman (2001), Friedman et al. (2000)) has connected boosting to the more common optimisation view of statistical inference. This formulation of boosting methods and the corresponding models are called the gradient boosting machines (GBM henceforth).

Using a learning sample $(y_i; \mathbf{x}_i)_{(i=1, \dots, n)}$, where the response y is continuous (regression problem) or discrete (classification problem) and $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_d)$ denotes a d -dimensional explanatory input variables, the objective is to obtain an estimate $\hat{f}(\mathbf{x})$ of the function $f(\mathbf{x})$, which maps \mathbf{x} to y . The task is thus to estimate the function $\hat{f}(\mathbf{x})$, that minimizes the expectation of some loss function, $\Psi(y, f)$, i.e.,

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

The loss function $\Psi(y, f)$ is assumed to be smooth and convex in the second argument to ensure that the gradient method works well.

An approximate solution to the minimization problem is obtained via forward stage-wise additive modeling, which approximates the solution by sequentially adding new basis functions to the expansion without adjusting the parameters and coefficients of those that have already been added.

GBM take on various forms with different programs using different loss functions, different base models, and different optimization schemes. This high flexibility makes GBM highly customizable to any particular data-driven task and introduces a lot of freedom into the model design thus making the choice of the most appropriate loss function a matter of trial and error. As a matter of fact, Friedman et al. (2000) warn that given a dataset, it is rarely known in advance which procedures and base learners should work the best, or if any of them would even provide decent results.

Loss-functions can be classified according to the type of response variable y . In the

case of categorical response, the response variable 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 $\bar{y} \in \{-1, 1\}$.

The most frequently used loss-functions for classification are the following:

-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 Binomial loss function is far more robust than the Adaboost loss function in noisy settings (mislabels, overlapping classes).

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: $\Psi(y, f(\mathbf{x})) = |y - f(\mathbf{x})|$

Several types of weak learners have been considered in the boosting literature, including decision trees (e.g., stumps, trees with two terminal nodes) (Friedman (2001)), smoothing splines (Bühlmann and Yu (2003)), wavelets (Wu et al. (2004)) and many more.

To design a particular GBM for a given task, one has to provide the choices of functional parameters $\Psi(y, f)$ and the weak learner $h(\mathbf{x}, \theta)$, characterized by a set of parameters θ . For instance, for decision trees, θ describes the axis to be split, the split points and the location parameter in terminal nodes.

The principle difference between boosting methods and conventional machine-learning techniques is that optimization is held out in the function space (Friedman (2001)). That is, the function estimate $\hat{f}(\mathbf{x})$ is parameterized in the additive functional form:

$$\hat{f}(\mathbf{x}) = \sum_{i=0}^{M_{stop}} \hat{f}_i(\mathbf{x})$$

Moreover, a common procedure is to restrict $\hat{f}(\mathbf{x})$ to take the 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 GBM algorithm can be summarized as follows:

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

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

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

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

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

The classic approach to controlling the model complexity is the introduction of the regularization through shrinkage. In the context of GBM, shrinkage is used for reducing,

or shrinking, the impact of each additional fitted base-learner. It reduces the size of incremental steps and thus penalizes the importance of each consecutive iteration. A better improvement is done by taking many small steps than by taking fewer large steps. Indeed, if one of the boosting iterations turns out to be erroneous, its negative impact can be easily corrected in subsequent steps.

The simplest form of regularization through shrinkage is the direct proportional shrinkage (Friedman (2001)). In this case the effect of shrinkage is directly defined as the parameter $\lambda \in [0, 1]$. The regularization is applied to the step 6 in the gradient boosting algorithm:

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

A crucial issue is the choice of the stopping iteration M_{stop} . Boosting algorithms should generally not be run until convergence. Otherwise, overfits resulting in a suboptimal prediction accuracy would be likely to occur (Friedman et al. (2000)).

One possible approach to choosing the number of iterations M_{stop} would be to use an information criterion like Akaike's *AIC* or some sort of minimum description length criteria. However, they have been shown to overshoot the true number of iterations (Hastie (2007)) and thus are not recommended for practical usage. Cross-validation techniques should be used to estimate the optimal M_{stop} (Hastie (2007)).

Briefly, cross-validation uses part of the available data to fit the model, and a different part to test it. K -fold cross-validation works by dividing the training data randomly into K roughly equal-sized parts. For the k^{th} part, the learning method is fit to the other $K - 1$ parts of the data, and calculate the prediction error of the fitted model when predicting the k^{th} part of the data. This is done for $k = 1, 2, \dots, K$ and the K prediction error estimates are averaged. An estimated prediction error curve as a function of the complexity parameter is obtained (Hastie et al. (2009)). Typical choices of K are 5 and 10. When it comes to time series forecasting, Bergmeir et al. (2015) demonstrate that

K-fold cross-validation performs favourably compared to both out-of-sample evaluation and other time-series-specific techniques.

In contrast to the choice of the stopping iteration, the choice of λ has been shown to be of minor importance for the predictive performance of a boosting algorithm. The only requirement is that the value of λ is small, e.g. $\lambda = 0.1$ (Friedman (2001)).

In this paper, two different approaches are tested: a combination of a binomial loss function with decision trees ("*BTB*") as in Ng (2014) and a combination of a squared error loss function with P-splines ("*SPB*") as in Berge (2015) and Taieb et al. (2015). P-splines ((Eilers and Marx (1996))) can be seen as a versatile modeling tool for non-linear effects. Examples include smooth effects, bivariate smooth effects (e.g., spatial effects), varying coefficient terms, cyclic effects and many more.

3.2 Empirical setup

3.2.1 Turning point chronology in real time

Researchers and investors focus mainly on the business cycle detection, which is meant to reproduce the cycle of the global level of activity of a country. The turning points of that cycle separate periods of recession from periods of expansion. Since absolute prolonged declines in the level of economic activity tend to be rare events, Mintz (1974) introduces the growth cycle to produce more precise information on economic fluctuations. It is defined as the deviation of the real GDP to its long-term trend and can be thus seen as the output gap². The turning points of that cycle separate periods of slowdowns and accelerations.

A slowdown signals thus a decline in the rate of growth of the economy though not necessarily an absolute decline in economic activity. If the slowdown is not severe enough,

²The sign and the magnitude of the output gap, which are subject to considerable uncertainty, are not needed for the implementation of the investment strategies. However, turning points estimations are more robust (Nilsson and Gyomai (2011)).

there will be no recession. All recessions involve thus slowdowns, but not all slowdowns involve recessions.

Chapter 1 emphasizes the importance of the growth cycle for euro and dollar-based investors. Indeed, in theory, investment strategies based on growth cycle turning points achieve better risk-adjusted returns than those based on business cycle turning points.

Two economic phases are thus considered: slowdown and acceleration. Applied to the context of nowcasting, the variable of interest y can be summarized as follows:

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

\hat{y} is thus the probability of being in the regime referred to as acceleration.

To implement the ensemble machine learning algorithms, a chronology of economic regimes is needed. This paper employs the turning point chronology established in Chapter 1.

The training sample runs over the period from January 1988 to December 2001. The performance of the models are then evaluated over the period from January 2002 to December 2013. In the euro area, 54 % of the data are classified as slowdown. In the United States, 71 % of the data are classified as acceleration. Over the period from January 2002 to December 2013, there were 7 turning points in the growth cycle in the euro area and 5 in the United States.

In the empirical analysis, a recursive estimation of the models is done: each month the model is estimated with the data and the chronology that would have been available at the time the nowcasting is done. In real time, the complete chronology is not available, but the monthly GDP introduced by Raffinot (2007)³ allows to quickly refine the turning point chronology. The models are thus trained each month on a sample that extends from the beginning of the sample through month $T - 12$, over which the turning point

³A temporal disaggregation based on business surveys of the non revised values of gross domestic product GDP is used to develop a monthly indicator of GDP.

chronology is assumed known⁴. For instance, in January 2012, the chronology that would have been available to implement the models runs over the period from January 1988 to January 2011.

Re-estimating the model at each point in time also allows the relationship between covariates and the dependent variable to change (see Ng (2014)). Since the aim of this paper is to emphasize that ensemble machine learning algorithms can provide useful signals for policymakers and for investors in real time, analysing the most frequently selected predictors is out of the scope of this study.

3.2.2 Data set

The real-time detection of turning points faces the difficult issues of late release dates and data revision. As a matter of fact, key statistics are published with a long delay, are subsequently revised and are available at different frequencies. For example, gross domestic product (GDP) is only available on a quarterly basis with a time span of one to three months, and sometimes with significant revisions. Moreover, the data available at the time are often sending a different signal from what one sees once the data are subsequently revise (Hamilton (2011)).

To partly overcome this issue, we only focus on variables published with almost no lag along with practically no revisions: surveys and financial series⁵.

As a matter of fact, many monthly economic series are released giving indications of short-term movements. Among them, business surveys provide economists for timely and reliable pieces of information on business activity. They are subject to very weak revisions and are usually less volatile than other monthly series. They are published before the end of the month they relate to or just a few days after. In the euro area, surveys published by the European Commission have been proven to be very effective (Bengoechea et al.

⁴In unreported results, samples through month $T - 18$ and $T - 24$ have been tested and lead to almost the same results.

⁵All series are provided by Datastream.

(2006)) (see Table 2.A.1 in Appendix 2.A for an exhaustive list). In the United States, the surveys published by the Institute for Supply Management (ISM), the Conference Board and the National Association of Home Builders (NAHB) are often tested in the literature (Liu and Moench (2016)) (see Tables 2.A.2, 2.A.3 and 2.A.4 in Appendix 2.A for an exhaustive list). The only real economic data tested is the four-week moving average of initial claims for unemployment insurance, which is a weekly measure of the number of jobless claims filed by individuals seeking to receive state jobless benefits.

Moreover, financial series, which are not revised and often available on a daily basis, have also been considered: government bond yields, different yield curves, investment-grade yields, high-yield corporate yields, corporate spreads against government bonds, stock markets, assets volatility, the volatility of volatility of these indexes, the VIX index, the VSTOXX index, which is the VIX equivalent for the euro area and some commodity prices (see Table 2.A.5 in Appendix 2.A for an exhaustive list). This paper uses end of month values to match stock index futures and options contracts settlement prices.⁶

To detect the turning points in real-time, not only original series are screened, but also differentiated series (to underline the phases of low and high pace of growth). Because of the classical trade-off between reliability and advance, different lags of differentiation were considered: 1 to 18 months. The large dataset of predictors consists of more than 1000 monthly variables in the euro area and in the United States.

3.2.3 Alternative classifiers

Random guessing

To prove that ensemble models are significantly better than random guessing, several alternative classifiers, which assign classes arbitrarily, are computed. The first one (*Acc*) classifies all data as "acceleration", the second one (*Slow*) classifies all data as "slowdown". The last one (*Random*) randomly assigns classes based on the proportions found in the

⁶<http://www.cmegroup.com/trading/equity-index/fairvaluefaq.html>

training data. Thousand different simulations are computed and average criteria are provided.

Parametric models

The term spread has been proved to be an excellent leading indicator of recession in the United States (Liu and Moench (2016)) and in the euro area (Duarte et al. (2005)). Nowcasts from probit models based on the term spread are thus computed⁷.

For a given covariate x_n , based on the learning sample $(R_1, x_1), \dots, (R_{T-12}, x_{T-12})$, the model is characterized by the simple equation:

$$P(R_t^{probit} = 1) = \Phi(\alpha_0 + \alpha_1 x_t)$$

where Φ denotes a standard Gaussian cumulative distribution function, i. e.

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{1}{2}t^2} dt$$

Markov-switching dynamic factor models are effective to identify economic turning points (Camacho et al. (2015)). However, variable selection in factor analysis is a challenging task, especially among more than 1000 variables. Forecasts often improve by focusing on a limited set of highly informative series: Boivin and Ng (2006) demonstrate that factor-based forecasts extracted from 40 variables perform better than those extracted from 147 variables. A proper comparison is left for future research.

⁷Since the literature on growth cycle detection is quite small, there is no classical benchmark for models comparison. Considering the euro area makes things worse, since there is no equivalent to the Chicago Fed National Activity Index (CFNAI) or the the Aruoba, Diebold and Scotti (ADS) Business Conditions Index maintained by the Federal Reserve Bank of Philadelphia. Nevertheless, the aim of the paper is to highlight that timing the market is possible in real time nowcasting growth cycle turning points, more than providing the best method to do so.

3.2.4 Model evaluation

Classical criteria

Two metrics are computed to evaluate the quality of classification of a model.

The first one is the Brier's Quadratic Probability Score (*QPS*), defined as follows:

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

where $t = 1, \dots, F$ is the number of forecasts. The best model should strive to minimize the *QPS*.

The second one is the area under the Receiver Operating Characteristics curve (AUROC). Given a classifier and an instance, there are four possible outcomes. If the instance is positive and it is classified as positive, it is counted as a true positive ($T_p(c)$). If the instance is negative and classified as negative, it is counted as a true negative ($T_n(c)$). If a negative instance is misclassified as positive, it is counted as a false positive ($F_p(c)$). If a positive instance is misclassified as negative, it is counted as a false negative ($F_n(c)$).

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. A perfect classifier would fall into the top left corner of the graph with a True Positive Rate of 1 and a False Positive Rate of 0.

Accuracy is measured by the Area Under the ROC curve (AUROC), defined by:

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

An area of 1 represents a perfect test, an area of 0.5 represents a worthless test. A general rule of thumb is that an AUROC value exceeding 0.85 indicates a useful prediction performance.

Hanley and McNeil (1982) propose a test to compare the AUROC predictive accuracy. The aim is to test the best models in the selection with another criteria, thereby further reducing the set. The t-statistic for the test of $H_0 : AUROC_1 = AUROC_2$ is given by:

$$t = \frac{AUROC_1 - AUROC_2}{\sqrt{(\sigma_1^2 + \sigma_2^2 - 2r\sigma_1 * \sigma_2)}}$$

where, AUROC1 and AUROC2 are the areas under the curve for models 1 and 2 which are being compared. Similarly, σ_1 and σ_2 refer to the variances of the AUROCs for model 1 and model 2, respectively. Finally, r is the correlation between the two AUROCs (see Hanley and McNeil (1982) or Liu and Moench (2016) for more details on the test statistic and its implementation).

In this paper, a two-step model selection is computed. The first step is to select the best set of models according to Brier's Quadratic Probability Score (*QPS*) and then the selection is refined based on the the area under the ROC curve (AUROC) and the test proposed by Hanley and McNeil (1982).

Profit maximization measures

For investors, the usefulness of a forecast depends on the rewards associated with the actions taken by the agent as a result of the forecast. Simple hypothetical trading strategies are thus created to gauge the economic magnitude of the models.

Since asset classes behave differently during different phases of the economic cycles (Chapter 1), investment strategies based on economic regimes induced by the models should generate significant profits.

In order to frame the concept of active portfolio management, a specified investment strategy is required. The investment strategies are as stripped-down and simple as possible without raising concerns that the key results will not carry over to more general and intricate methods or asset classes.

Simple dynamic trading strategies are computed⁸, which should take advantage of positive economic regimes, as well as withstand adverse economic regimes and reduce potential drawdowns.

We first consider an equity portfolio manager investing 100 € or 100\$ on January 1, 2002. Each month, the investor decides upon the fraction of wealth to be invested based on the current state of the economy induced by the model. If the model classifies the period as acceleration, then the investor can leverage his portfolio (120% of his wealth is invested on the asset and 20% of cash is borrowed), otherwise he only invests 80% of his wealth and 20% is kept in cash.

Moreover, since asset classes perform differently during different stages of the growth cycle, it might be reasonable to rebalance the portfolio (shifting allocation weights) based on the stage of the growth cycle (Chapter 1). The second strategy aims at beating the classic asset allocation for an institutional portfolio, *i.e.* 60% of the portfolio allocated to equities and 40% to fixed income securities (bonds). The investor decides each month to rebalance his portfolio. If the model indicates acceleration, then 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.

Pesaran and Timmermann (1994) and Han et al. (2013) demonstrate that the total cost of transactions appears to be low, less than 1% (around 50 basis points when trading in stocks while the cost for bonds is 10 basis points). To simplify, no transaction costs are considered.

To avoid look-ahead bias, the reallocation takes place at the beginning of the month following the turning point. As a matter of fact, an investor could not know at the beginning of any month whether a turning point would occur in that month.

For conventional comparison of the portfolio performances, annualized average returns, annualized standard deviation (volatility), Sharpe ratio (SR), which is the mean portfolio

⁸These strategies reflect the investment process in place in the asset management companies where I used to work.

return in excess of the risk-free rate divided by the standard deviation of the excess portfolio return⁹ and max drawdown (MDD) are computed. The Sharpe ratio compares the expected returns of an investment to the amount of risk undertaken to capture these returns. The Max drawdown (MDD) is the largest drop from the maximum cumulative return. In brief, the MDD offers investors a worst case scenario.

Data snooping

Data snooping, which occurs when a given set of data is used more than once for purposes of inference or model selection, leads to the possibility that any results obtained in a statistical study may simply be due to chance rather than to any merit inherent in the method yielding the results (White (2000)).

To avoid data snooping, the model confidence set (MCS) procedure (Hansen et al. (2011)) is computed.

The MCS procedure is a model selection algorithm, which filters a set of models from a given entirety of models. The resulting set contains the best models with a probability that is no less than $1 - \alpha$ with α being the size of the test.

An advantage of the test is that it not necessarily selects a single model, instead it acknowledges possible limitations in the data since the number of models in the set containing the best model will depend on how informative the data are.

More formally, define a set M_0 that contains the set of models under evaluation indexed by: $i = 0, \dots, m_0$. Let $d_{i,j,t}$ denote the loss differential between two models by

$$d_{i,j,t} = L_{i,t} - L_{j,t}, \forall i, j \in M_0$$

L is the loss calculated from some loss function for each evaluation point $t = 1, \dots, T$. The set of superior models is defined as:

⁹A risk-free interest rate of zero is assumed when calculating the SR.

$$M^* = \{i \in M_0 : E[d_{i,j,t}] \leq 0 \forall j \in M_0\}$$

The MCS uses a sequential testing procedure to determine M^* . The null hypothesis being tested is:

$$\begin{cases} H_{0,M} : E[d_{i,j,t}] = 0 \forall i, j \in M \text{ where } M \text{ is a subset of } M_0 \\ H_{A,M} : E[d_{i,j,t}] \neq 0 \text{ for some } i, j \in M \end{cases}$$

When the equivalence test rejects the null hypothesis, at least one model in the set M is considered inferior and the model that contributes the most to the rejection of the null is eliminated from the set M . This procedure is repeated until the null is accepted and the remaining models in M now equal $\widehat{M}_{1-\alpha}^*$.

According to Hansen et al. (2011), the following two statistics can be used for the sequential testing of the null hypothesis:

$$t_{i,j} = \frac{\bar{d}_{i,j}}{\sqrt{\widehat{var}(\bar{d}_{i,j})}} \text{ and } t_i = \frac{\bar{d}_i}{\sqrt{\widehat{var}(\bar{d}_i)}}$$

where m is the number of models in M , $\bar{d}_i = (m - 1)^{-1} \sum_{j \in M} \bar{d}_{i,j}$, is the simple loss of the i^{th} model relative to the averages losses across models in the set M , and $\bar{d}_{i,j} = (m)^{-1} \sum_{t=1}^m d_{i,j,t}$ measures the relative sample loss between the i^{th} and j^{th} models. Since the distribution of the test statistic depends on unknown parameters a bootstrap procedure is used to estimate the distribution.

In this paper, the MCS is applied with classical criteria loss function (Brier's Quadratic Probability Score) and with profit maximization loss function (one-year rolling Sharpe ratio). As regards investment strategies, it should be noted that the MCS aims at finding the best model and all models which are indistinguishable from the best, not those better than the benchmark. To determined if models are better than the benchmark, the stepwise test of multiple reality check by Romano and Wolf (2005) and the stepwise multiple

superior predictive ability test by Hsu et al. (2013) should be considered. However, if the benchmark is not selected in the best models set, investors can conclude that their strategies "beat" the benchmark.

3.3 Empirical results

3.3.1 United States

The two-step model selection is computed as described previously. The first step of the model selection is to find the best set of models according to the MCS procedure based on Brier's Quadratic Probability Score (*QPS*) and then the model selection is refined based on the the area under the ROC curve (*AUROC*). The *AUROC* metric is thus only computed for models included in $\widehat{M}_{75\%}^*$. Table 3.1 highlights classical metrics for the models in the United States.

Table 3.1: Classical evaluation criteria in the United States

	QPS	AUROC
<i>SPB</i>	0.13	
<i>RF</i>	0.07**	0.94
<i>BTB</i>	0.05**	0.94
<i>Prob</i>	0.22	
<i>Acc</i>	0.21	
<i>Slow</i>	0.79	
<i>Random</i>	0.25	

Note: This table reports classical metrics used to evaluate the quality of the models: the area under the ROC curve (*AUROC*) and the Brier's Quadratic Probability Score (*QPS*). ** indicates the model is in the set of best models $\widehat{M}_{75\%}^*$. *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, *Prob* refers to the probit model based on the term spread, *Acc* classifies all data as "acceleration", *Slow* classifies all data as "slowdown" and *Random* randomly assigns classes based on the proportions found in the training data.

The performance of the models are impressive and are consistent with the results found in Berge (2015). Ensemble machine learning models built are significantly and statistically better than random guessing. "*RF*" and "*BTB*" belong to $\widehat{M}_{75\%}^*$. "*SPB*" is the only ensemble model not selected in any best models set. Comparisons made with the test proposed by Hanley and McNeil (1982) between models in $\widehat{M}_{75\%}^*$ conclude that no model is better than the others.

Table 3.2 indicates that "RF" and "BTB" do not miss any turning point. They identify peaks and troughs with a maximum lag time of three months. "SPB" does not detect the 2011 slowdown. It explains why this model is not selected in the best models set.

Table 3.2: Turning point signals of the reference cycle in the United States

	<i>SPB</i>	<i>RF</i>	<i>BTB</i>
Trough: February 2003	0	-1	-2
Peak: October 2007	1	-2	-1
Trough: September 2009	1	1	0
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.

Figure 3.1 exhibits some false signals, especially for "SPB". For *RF*, November 2006 is the only false signal. As regards *BTB*, January and February 2007 are the only false signals.

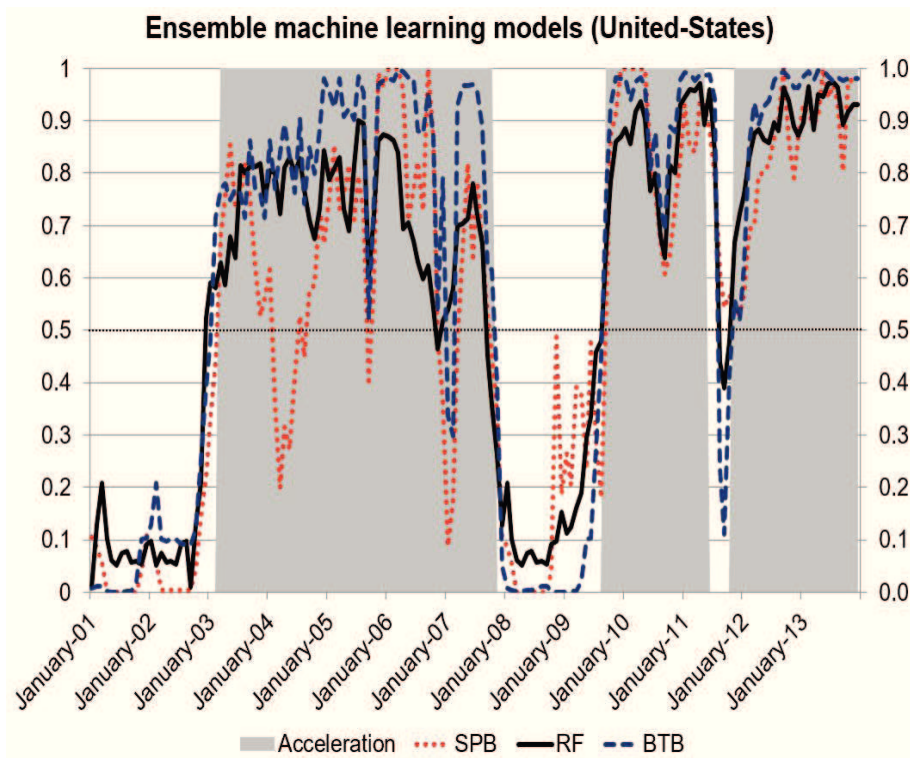


Figure 3.1: Recursive real time classification of the growth cycle in the United States

The ability to produce profits it now tested. Since models detect tuning points accurately, investment strategies should be performing well. Tables 3.3 and 3.4 emphasize that active investment strategies based on the growth cycle achieve impressive risk-adjusted returns and outperform the passive buy-and-hold benchmark.

Table 3.3: **Summary of return and risk measures in the United States: 120/80 equity strategy**

	Average returns	Volatily	SR	MDD
<i>SPB</i>	0.110	0.149	0.74**	-0.43
<i>RF</i>	0.107	0.147	0.72	-0.43
<i>BTB</i>	0.109	0.146	0.75**	-0.44
<i>Prob</i>	0.094	0.173	0.54	-0.57
<i>Acc</i>	0.099	0.177	0.56	-0.58
<i>Slow</i>	0.066	0.118	0.56	-0.43
<i>Random</i>	0.092	0.155	0.59	-0.51
<i>Benchmark</i>	0.083	0.147	0.56	-0.51

Note: This table reports profit maximization measures for 120/80 equity strategy based on the state of the growth cycle induced by the models. Returns are monthly and annualized. The volatility corresponds to the annualized standard deviation. The Sharpe ratio (SR) compares the expected returns of an investment to the amount of risk undertaken to capture these returns. The Max drawdown (MDD) measures the largest single drop from peak to bottom in the value of a portfolio. ** indicates the model is in the set of best models $\widehat{M}_{75\%}^*$. *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, *Prob* refers to the probit model based on the term spread, *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 and *Benchmark* refers to the passive buy-and-hold investment strategy.

Table 3.4: **Summary of return and risk measures in the United States: dynamic asset allocation**

	Average returns	Volatily	SR	MDD
<i>SPB</i>	0.091	0.090	1**	-0.18
<i>RF</i>	0.088	0.088	0.98	-0.18
<i>BTB</i>	0.091	0.087	1**	-0.20
<i>Prob</i>	0.074	0.113	0.66	-0.39
<i>Acc</i>	0.075	0.116	0.65	-0.42
<i>Slow</i>	0.060	0.058	1	-0.18
<i>Random</i>	0.076	0.095	0.79	-0.30
<i>Benchmark</i>	0.068	0.085	0.79	-0.31

Note: This table reports profit maximization measures for a dynamic asset allocation between bonds and equities based on the state of the growth cycle induced by the models. Returns are monthly and annualized. The volatility corresponds to the annualized standard deviation. The Sharpe ratio (SR) compares the expected returns of an investment to the amount of risk undertaken to capture these returns. The Max drawdown (MDD) measures the largest single drop from peak to bottom in the value of a portfolio. ** indicates the model is in the set of best models $\widehat{M}_{75\%}^*$. *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, *Prob* refers to the probit model based on the term spread, *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 and *Benchmark* refers to the passive buy-and-hold investment strategy.

Table 3.3 points out that several strategies outperform the benchmark: it is thus possible to time the stock market based on economic cycles in real time. These results

have naturally implications for the risk management and hedging. Especially, in the options market one can utilize the current state of the economy to hedge the portfolio against the possible price declines. For example, writing an out-of-the-money covered call or buy a put option when the stock market is expected to decrease (slowdown) would limit the losses.

Selecting the best model is still complicated. In comparison with the "classical" case, $\widehat{M}_{75\%}^*$ turns out to be different. "SPB" is not included in the best models set for economists but can be useful for investors, since it is chosen in $\widehat{M}_{75\%}^*$. Surprisingly, "RF" is not selected. Importantly, "BTB" is selected in the best models set for economists and for equity investors.

Moreover, Table 3.4 highlights that dynamic asset allocation delivers a substantial improvement in risk-adjusted performance as compared to static asset allocation, especially for investors who seek to avoid large losses. The reduction of the MDD, which focuses on the danger of permanent loss of capital as a sensible measure of risk, is what risk-averse investors value the most. Portfolio rebalancing based on the stage of the growth cycle in real time is thus realisable in the United States. One more time, *SPB* is attached to $\widehat{M}_{75\%}^*$, "RF" is not selected and "BTB" is selected in the best models set.

To sum up, ensemble machine learning algorithms detect growth cycle turning points accurately, leading to outperforming investment strategies. However, depending on the data and the objective, the model selection is quite different. Yet, "BTB" is always selected in the best models set.

3.3.2 Euro area

The same model selection methodology is applied in the euro area. The first step is to find the best set of models according to the MCS procedure based on Brier's Quadratic Probability Score (*QPS*) and then this selection is refined based on the the area under the ROC curve (*AUROC*). Table 3.5 highlights classical metrics for the models in the

euro area.

Table 3.5: Classical evaluation criteria in the euro area

	QPS	AUROC
<i>SPB</i>	0.12**	0.90
<i>RF</i>	0.11**	0.91
<i>BTB</i>	0.12**	0.90
<i>Prob</i>	0.25	
<i>Acc</i>	0.45	
<i>Slow</i>	0.54	
<i>Random</i>	0.48	

Note: This table reports classical metrics used to evaluate the quality of the models: the area under the ROC curve (AUROC) and the Brier's Quadratic Probability Score (*QPS*). ** indicates the model is in the set of best models $\widehat{M}_{75\%}^*$. *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, *Prob* refers to the probit model based on the term spread, *Acc* classifies all data as "acceleration", *Slow* classifies all data as "slowdown" and *Random* randomly assigns classes based on the proportions found in the training data.

The performance of ensemble machine learning models are notable and are significantly better than random guessing and other competitive classifiers: the three machine learning algorithms are included in the best models sets. The test proposed by Hanley and McNeil (1982) concludes that no model is better than the other.

Metrics in the euro area are less remarkable than in the United States. Indeed, the persistence of the regimes is smaller in the euro area. The real-time classification is thus harder.

Table 3.6 confirms this statement. Some turning points are difficult to detect, in particular the short slowdown of 2004-2005. Nevertheless, ensemble machine learning models detect all other turning points with a maximum lag time of three months. In real time, the models would have signalled the 2007 and 2011 slowdowns.

Table 3.6: Turning point signals of the reference cycle in the euro area

	<i>SPB</i>	<i>RF</i>	<i>BTB</i>
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.

Figure 3.2 highlights that there is no false signals.

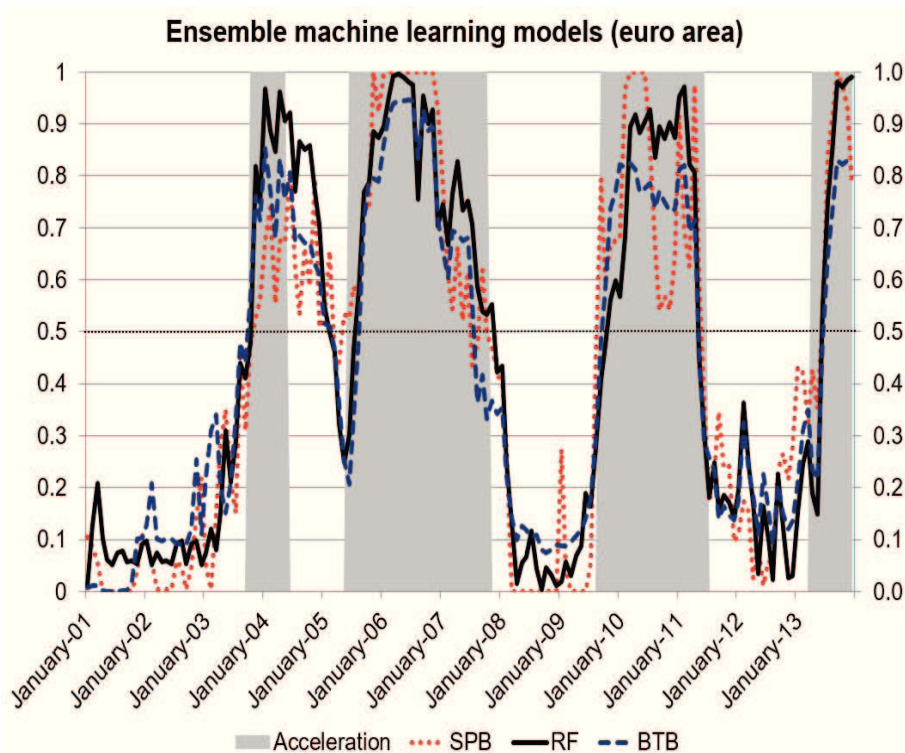


Figure 3.2: Recursive real time classification of the growth cycle in the euro area

The ability to generate profits it now analysed. Since models detect tuning points quite accurately, investment strategies should be performing well. Tables 3.7 and 3.8 highlight that active investment strategies based on the growth cycle achieve excellent risk-adjusted returns and outperform the passive buy-and-hold benchmark. Naturally,

the risk management and hedging implications described for the United States also apply in the euro area.

Table 3.7: **Summary of return and risk measures in the euro area: 120/80 equity strategy**

	Average returns	Volatility	SR	MDD
<i>SPB</i>	0.085	0.161	0.53**	-0.46
<i>RF</i>	0.083	0.160	0.52**	-0.46
<i>BTB</i>	0.079	0.158	0.50	-0.46
<i>Prob</i>	0.075	0.182	0.41	-0.48
<i>Acc</i>	0.077	0.207	0.37	-0.61
<i>Slow</i>	0.051	0.138	0.37	-0.43
<i>Random</i>	0.076	0.182	0.42	-0.53
<i>Benchmark</i>	0.064	0.173	0.37	-0.54

Note: This table reports profit maximization measures for 120/80 equity strategy based on the state of the growth cycle induced by the models. Returns are monthly and annualized. The volatility corresponds to the annualized standard deviation. The Sharpe ratio (SR) compares the expected returns of an investment to the amount of risk undertaken to capture these returns. The Max drawdown (MDD) measures the largest single drop from peak to bottom in the value of a portfolio. ** indicates the model is in the set of best models $\widehat{M}_{75\%}^*$. *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, *Prob* refers to the probit model based on the term spread, *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 and *Benchmark* refers to the passive buy-and-hold investment strategy.

Table 3.8: **Summary of return and risk measures in the euro area: dynamic asset allocation**

	Average returns	Volatility	SR	MDD
<i>SPB</i>	0.081	0.094	0.86**	-0.21
<i>RF</i>	0.080	0.093	0.86**	-0.22
<i>BTB</i>	0.075	0.091	0.83	-0.22
<i>Prob</i>	0.064	0.114	0.56	-0.25
<i>Acc</i>	0.060	0.137	0.44	-0.44
<i>Slow</i>	0.052	0.070	0.75	-0.21
<i>Random</i>	0.064	0.115	0.55	-0.32
<i>Benchmark</i>	0.06	0.10	0.55	-0.34

Note: This table reports profit maximization measures for a dynamic asset allocation between bonds and equities based on the state of the growth cycle induced by the models. Returns are monthly and annualized. The volatility corresponds to the annualized standard deviation. The Sharpe ratio (SR) compares the expected returns of an investment to the amount of risk undertaken to capture these returns. The Max drawdown (MDD) measures the largest single drop from peak to bottom in the value of a portfolio. ** indicates the model is in the set of best models $\widehat{M}_{75\%}^*$. *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, *Prob* refers to the probit model based on the term spread, *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 and *Benchmark* refers to the passive buy-and-hold investment strategy.

As regards equities, several strategies based on the growth cycle induced by several models outperform the benchmark.

Selecting the best model is still complicated. In comparison with the "classical" case, $\widehat{M}_{75\%}^*$ turns out to be different. In this case, "BTB" is less effective than the others two

ensemble models as it is not selected in the best models set. No alternative classifiers are included in the best models set.

Tables 3.8 emphasizes that dynamic asset allocation delivers a substantial improvement in risk-adjusted performance as compared to static asset allocation, especially for investors who seek to avoid large losses. It is thus possible to rebalance the portfolio based on the stage of the growth cycle in real time in the euro area.

SPB and *RF* perform well as those models belongs to $\widehat{M}_{75\%}^*$. ”*BTB*” is again less effective than the others two ensemble models.

All in all, all results found for the United States also apply for the euro area. Depending on the data and the objective the best models set can be quite different. This time, *SPB* and *RF* are always selected in the best model set.

Conclusion

Investment strategies based on the turning points of the growth cycle, better known as the output gap, achieve impressive risk-adjusted returns... in theory. But, in real time, economists often fail to detect if a new economic phase has already begun.

Over the last couple of decades, researchers in the machine learning community have developed more complex methods, also called ensemble learning, which improve prediction performances. Ensemble methods are learning models that achieve performance by combining the opinions of multiple learners. The two most popular techniques for constructing ensembles are random forests (Breiman (2001)) and boosting (Schapire (1990)). The two features of those algorithms are their abilities to entertain a large number of predictors and to perform estimation and variable selection simultaneously. Paradoxically, both methods work by adding randomness to the data (Varian (2014)), although they have substantial differences. Random forests rely on simple averaging of models in the ensemble and derive their strength from two aspects: randomizing subsamples of the training data and randomizing the selection of features. Boosting combines models that

do not perform particularly well individually into one with much improved properties. It is an iterative process where the errors are kept being modelled.

Three models based on random forest and boosting algorithms are created to quickly and accurately detect growth cycle turning points in real time, in the United States and in the euro area.

To assess the value of the models, profit maximization measures are employed in addition to more standard criteria, since, for investors, the usefulness of a signal depends on the rewards associated with the actions taken by the agent as a result of the forecast. When comparing predictive accuracy and profit measures, the model confidence set procedure (Hansen et al. (2011)) is applied to avoid data snooping.

Ensemble machine learning algorithms are effective to detect economic turning points in real time. Strategies based on the turning points of the growth cycle induced by the models achieve thus excellent risk-adjusted returns in real time: timing the market is possible. It leads to useful implications for investors practising active portfolio and risk management and for policy makers as tools to get early warning signals.

The selection of the best model is difficult. For instance, economists and investors would not always choose the same model. Moreover, depending on the data and the objective, random forest sometimes performs better than boosting, sometimes not.

Last but not least, this article opens the door for further research. An attempt to forecast growth cycle and business cycle turning points three to twelve months ahead could be very interesting. Evaluating the diversifying power of alternative asset classes in real-time may be newsworthy for investors.

Chapter 4

Hierarchical Clustering based Asset Allocation

Abstract

A hierarchical clustering based asset allocation method, which uses graph theory and machine learning techniques, is proposed. Hierarchical clustering refers to the formation of a recursive clustering, suggested by the data, not defined a priori. Several hierarchical clustering methods are presented and tested. Once the assets are hierarchically clustered, a simple and efficient capital allocation within and across clusters of assets at multiple hierarchical levels is computed. The out-of-sample performances of hierarchical clustering based portfolios and more traditional risk-based portfolios are evaluated across three disparate datasets. To avoid data snooping, the comparison of profit measures is assessed using the bootstrap based model confidence set procedure. The empirical results indicate that hierarchical clustering based portfolios are robust, truly diversified and achieve statistically better risk-adjusted performances than commonly used portfolio optimization techniques.

Introduction

Nobel Prize winner [Harry Markowitz](#) described diversification, with its ability to enhance portfolio returns while reducing risk, as the only the "only free lunch" in investing (Markowitz (1952)). Yet diversifying a portfolio in real life is easier said than done.

Investors are aware of the benefits of diversification but form portfolios without giving proper consideration to the correlations (Goetzmann and Kumar (2008)). Moreover, modern and complex portfolio optimisation methods are optimal in-sample, but often provide rather poor out-of-sample forecast performance. For instance, DeMiguel et al. (2009) demonstrate that the equal-weighted allocation, which gives the same importance to each assets, beats the entire set of commonly used portfolio optimization techniques. In fact, optimized portfolios depend on expected returns and risks, but even small estimation errors can result in large deviations from optimal allocations in an optimizer's result (Michaud (1989)).

To overcome this issue, academics and practitioners have developed risk-based portfolio optimization techniques (minimum variance, equal risk contribution, risk budgeting,...), which do not rely on return forecasts (Roncalli (2013)). However, the inversion of a positive-definite covariance matrix remains needed, which lead to errors of such magnitude that they entirely offset the benefits of diversification (López de Prado (2016b)).

Exploring a new way of capital allocation, López de Prado (2016a) introduces a portfolio diversification technique called "Hierarchical Risk Parity" (HRP), which uses graph theory and machine learning techniques. One of the main advantages of HRP is to manage to compute a portfolio on an ill-degenerated or even a singular covariance matrix.

The starting point of his analysis is that a correlation matrix is too complex to be properly analysed and understood. If you have N assets of interest, there are $\frac{1}{2}N(N - 1)$ pairwise correlations among them and that number grows quickly. For example, there are as many as 4950 correlation coefficients between stocks of the FTSE 100 and 124750 between stocks of the S&P 500. More importantly, López de Prado (2016a) points out that

correlation matrices lack the notion of hierarchy. Actually, Nobel prize laureate [Herbert Simon](#) has demonstrated that complex systems can be arranged in a natural hierarchy, comprising nested sub-structures (Simon (1962)). But, a correlation matrix makes no difference between assets. Yet, some assets seem closer substitutes of one another, while others seem complementary to one another. This lack of hierarchical structure allows weights to vary freely in unintended ways (López de Prado (2016a)).

To simplify the analysis of the relationships between this large group of relative prices, López de Prado (2016a) applies a correlation-network method known as "Minimum Spanning Tree (MST)"¹. Its main principle is easy to understand: the heart of correlation analysis is choosing which correlations really matter; in other words, choosing which links in the network are important, and removing the rest, keeping $N - 1$ links².

Graph theory is linked to unsupervised machine learning. For instance, the MST is strictly related to a hierarchical clustering algorithm, named the "Single Linkage" (Tumminello et al. (2010)) and another hierarchical clustering method, the "Average Linkage", has been shown to be associated to a slightly different version of spanning tree called Average Linkage Minimum Tree (Tumminello et al. (2007)). Hierarchical clustering refers to the formation of a recursive clustering, suggested by the data, not defined a priori. The objective is to build a binary tree of the data that successively merges similar groups of points. Hierarchical clustering is thus another way to filter correlations. Another variants of hierarchical clustering algorithm (Complete Linkage, Ward's method) are not associated to a spanning tree representation. Yet, they may provide interesting results.

Building upon López de Prado (2016a) and Simon (1962), this paper exploits the

¹Since the seminal work of Mantegna (1999), correlation-networks have been extensively used in Econophysics as tools to filter, visualise and analyse financial market data (see Baitinger and Papenbrocky (2016) for a review)

²One concrete example would be a telecommunications company which is trying to lay out cables in new neighborhood. In any case, the easiest possibility to install new cables is to bury them along roads. Some of those paths might be more expensive, because they are longer, or require the cable to be buried deeper. These paths would be represented by edges with larger weights. A spanning tree for that graph would be a subset of those paths that has no cycles but still connects to every house. There might be several spanning trees possible. A minimum spanning tree would be one with the lowest total cost would then represent the least expensive path for laying the cable.

notion of hierarchy. Different hierarchical clustering methods are presented and tested, namely Simple Linkage, Complete Linkage, Average Linkage, Ward's Method. Once the assets are hierarchically clustered, a simple and efficient capital allocation within and across clusters of assets at multiple hierarchical levels is computed.

The out-of-sample performances of hierarchical clustering based portfolios and risk-based portfolios are evaluated across three empirical datasets, which differ in terms of number of assets and composition of the universe ("S&P sectors", multi-assets and individual stocks). To avoid data snooping, which occurs when a given set of data is used more than once for purposes of inference or model selection, the comparison of profit measures is assessed using the bootstrap based model confidence set procedure proposed by Hansen et al. (2011). It prevents strategies that perform by luck to be considered as effective.

The findings of the paper can be summarized as follows: hierarchical clustering based portfolios are robust, truly diversified and achieve statistically better risk-adjusted performances than commonly used portfolio optimization techniques. Among clustering methods, there is no clear winner.

The rest of the paper proceeds as follows. Section 1 describes the risk budgeting approach. Section 2 introduces hierarchical clustering methods and their application to asset allocation. Section 3 presents the empirical set up: the datasets and the comparison criteria. Section 4 analyses the empirical results.

4.1 Risk Budgeting Approach

This section briefly describes risk budgeting portfolios. Refer to Roncalli (2013) for a detailed exposition of this approach. In a risk budgeting approach, the investor only chooses the risk repartition between assets of the portfolio, without any consideration of returns, thereby partially dealing with the issues of traditional portfolio optimization methods.

4.1.1 Notations and definitions

Consider a portfolio invested in N assets with portfolio weights vector $w = (w_1, w_2, \dots, w_N)'$. Returns are assumed to be arithmetic: $r_{t,i} = (p_{t,i} - p_{t-1,i})/p_{t-1,i} = p_{t,i}/p_{t-1,i} - 1$. The portfolio return at time t is thus:

$$r_{P,t} = \sum_{i=1}^N w_i r_{t,i}$$

Let σ_i^2 be the variance of asset i , σ_{ij} be the covariance between assets i and j and Σ be the covariance matrix.

The volatility is defined as the risk of the portfolio:

$$\mathcal{R}_w = \sigma_w = \sqrt{w' \Sigma w}$$

and μ is the expected return:

$$\mu = E(r_P) = \sum_{i=1}^N w_i E(r_i)$$

4.1.2 Risk budgeting portfolios

In a risk budgeting portfolio, the risk contribution from each components is equal to the budget of risk defined by the portfolio manager.

Since the risk measure is coherent and convex, the Euler decomposition is verified:

$$\mathcal{R}_w = \sum_{i=1}^N w_i \frac{\partial \mathcal{R}_w}{\partial w_i}$$

With the volatility as the risk measure, the risk contribution of the i^{th} asset becomes:

$$\mathcal{RC}_{w_i} = w_i \frac{(\Sigma w)_i}{\sqrt{w' \Sigma w}}$$

A long-only, full invested risk budgeting portfolio is defined as follows (Roncalli (2013)):

$$\left\{ \begin{array}{l} \mathcal{RC}_{w_i} = b_i \mathcal{RC}_w \\ b_i > 0 \\ \sum_{i=1}^N b_i = 1 \\ w_i \geq 0 \\ \sum_{i=1}^N w_i = 1 \end{array} \right.$$

Once a set of risk budgets is defined, the weights of the portfolio are computed so that the risk contributions match the risk budgets.

In this paper, four risk budgeting portfolios are considered³:

- The minimum variance (MV) portfolio is a risk budgeting portfolio where the risk budget is equal to the weight of the asset:

$$b_i = w_i$$

- The most diversified portfolio (MDP) (Choueifaty et al. (2013)) is a risk budgeting portfolio where the risk budgets are linked to the product of the weight of the asset and its volatility:

$$b_i = \frac{w_i \sigma_i}{\sum_{i=1}^N w_i \sigma_i}$$

- The equal risk contribution portfolio (ERC) (Maillard et al. (2010)) is a risk budgeting portfolio where the risk contribution from each asset is made equal:

$$b_i = \frac{1}{N}$$

³Five if the equal weighted portfolio is considered as a risk budgeting portfolio.

- The inverse-variance (IVRB) risk budgeting portfolio defines risk budgets as follows:

$$b_i = \frac{\sigma_i^{-2}}{\sum_{i=1}^N \sigma_i^{-2}}$$

The cyclical coordinate descent (CCD) algorithm for solving high dimensional risk parity problems (Griveau-Billion et al. (2013)) is employed to estimate the risk-based models.

4.2 Hierarchical clustering and asset allocation

4.2.1 Notion of hierarchy

Nobel Prize winner Herbert Simon has demonstrated that complex systems, such as financial markets, have a structure and are usually organized in a hierarchical manner, with separate and separable sub-structures (Simon (1962)). The hierarchical structure of interactions among elements strongly affects the dynamics of complex systems. The need of a quantitative description of hierarchies to model complex systems is thus straightforward (Anderson (1972)).

López de Prado (2016a) points out that correlation matrices lack the notion of hierarchy, which allows weights to vary freely in unintended ways. Moreover, he provides a concrete example to highlight the interest of the notion of hierarchy for asset allocation: "stocks could be grouped in terms of liquidity, size, industry and region, where stocks within a given group compete for allocations. In deciding the allocation to a large publicly-traded U.S. financial stock like J.P. Morgan, we will consider adding or reducing the allocation to another large publicly-traded U.S. bank like Goldman Sachs, rather than a small community bank in Switzerland, or a real estate holding in the Caribbean". To sum up, a correlation matrix makes no difference between assets. Yet, some assets seem closer substitutes of one another, while others seem complementary to one another.

4.2.2 Hierarchical clustering

The purpose of cluster analysis is to place entities into groups, or clusters, suggested by the data, not defined a priori, such that entities in a given cluster tend to be similar to each other and entities in different clusters tend to be dissimilar.

Hierarchical clustering refers to the formation of a recursive clustering. The objective is to build a binary tree of the data that successively merges similar groups of points. The tree based representation of the observations is called a dendrogram. Visualizing this tree provides a useful summary of the data.

Hierarchical clustering requires a suitable distance measure. The following distance is used (Mantegna (1999) and López de Prado (2016a)):

$$D_{i,j} = \sqrt{2(1 - \rho_{i,j})}$$

where $D_{i,j}$ is the correlation-distance index between the i^{th} and j^{th} asset and $\rho_{i,j}$ is the respective Pearson's correlation coefficients. The distance $D_{i,j}$ is a linear multiple of the Euclidean distance between the vectors i, j after z-standardization, hence it inherits the true-metric properties of the Euclidean distance (López de Prado (2016a)).

Four agglomerative clustering variants are tested in this study, namely: Single Linkage (SL), Average Linkage (AL), Complete Linkage (CL), Ward's Method (WM).

An agglomerative clustering starts with every observation representing a singleton cluster and then combines the clusters sequentially, reducing the number of clusters at each step until only one cluster is left. At each of the $N - 1$ steps the closest two (least dissimilar) clusters are merged into a single cluster, producing one less cluster at the next higher level. Therefore, a measure of dissimilarity between two clusters must be defined and different definitions of the distance between clusters can produce radically different dendrograms. The clustering variants are described below:

- Single Linkage: the distance between two clusters is the minimum of the distance between any two points in the clusters. For clusters C_i, C_j :

$$d_{C_i, C_j} = \min_{x, y} \{D(x, y) \mid x \in C_i, y \in C_j\}$$

This method is relatively simple and can handle non-elliptical shapes. Nevertheless, it is sensitive to outliers and can result in a problem called chaining whereby clusters end up being long and straggly. The SL algorithm is strictly related to the one that provides a Minimum Spanning Tree (MST). However the MST retains some information that the SL dendrogram throws away.

- Complete Linkage: the distance between two clusters is the maximum of the distance between any two points in the clusters. For clusters C_i, C_j :

$$d_{C_i, C_j} = \max_{x, y} \{D(x, y) \mid x \in C_i, y \in C_j\}$$

This method tends to produce compact clusters of similar size but, is quite sensitive to outliers.

- Average linkage: the distance between two clusters is the average of the distance between any two points in the clusters. For clusters C_i, C_j :

$$d_{C_i, C_j} = \text{mean}_{x, y} \{D(x, y) \mid x \in C_i, y \in C_j\}$$

This is considered to be a fairly robust method.

- Ward's Method (Ward (1963)): the distance between two clusters is the increase of the squared error that results when two clusters are merged. For clusters C_i, C_j with sizes m_i, m_j , respectively,

$$d_{C_i, C_j} = \frac{m_i m_j}{m_i + m_j} \|c_i - c_j\|^2.$$

where c_i, c_j are the centroids for the clusters.

This method is biased towards globular clusters, but less susceptible to noise and outliers. It is one of the most popular methods.

To determine the number of clusters, the Gap index (Tibshirani et al. (2001)) is employed. It compares the logarithm of the empirical within-cluster dissimilarity and the corresponding one for uniformly distributed data, which is a distribution with no obvious clustering.

4.2.3 Asset allocation weights

Once the clusters have been determined, the capital should be efficiently allocated both within and across groups. Indeed, a compromise between diversification across all investments and diversification across clusters of investments at multiple hierarchical levels has to be found.

Since asset allocation within and across clusters can be based on the same or different methodologies, there are countless options.

The chosen weighting scheme attempts to stay very simple and focuses not only on the clusterings, but on the entire hierarchies associated to those clusterings. The principle is to find a diversified weighting by distributing capital equally to each cluster hierarchy, so that many correlated assets receive the same total allocation as a single uncorrelated one. Then, within a cluster, an equal-weighted allocation is computed.

For example, Figure 4.1 exhibits a small dendrogram with five assets and three clusters. The first cluster is made up of assets 1 and 2, asset 5 constitutes the second cluster and the third cluster consists of assets 3 and 4. Based on the hierarchical clustering weighting, weights for cluster number one is 0.5 ($\frac{1}{2} = 0.5$) and weights for clusters 2 and 3 are 0.25 ($\frac{0.5}{2} = 0.25$). Since there are two assets in the cluster number one, final weights for assets 1 and 2 are $\frac{0.5}{2} = 0.25$. Asset 5 would have a weight of $\frac{0.25}{1} = 0.25$. At last, assets 3 and 4 would get a weight of $\frac{0.25}{2} = 0.125$.

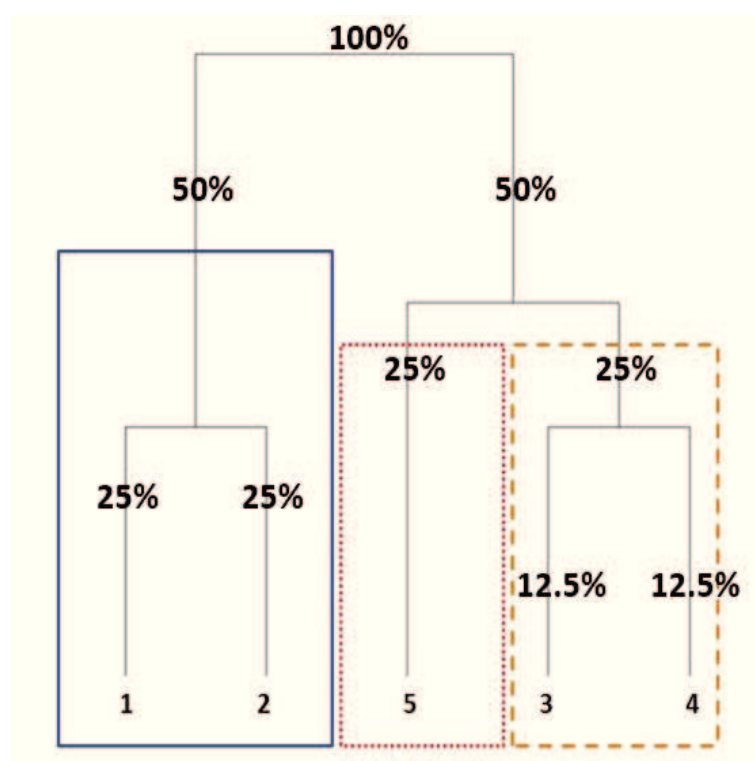


Figure 4.1: Asset allocation weights: a small example

This weighting scheme should guarantee the diversification and the robustness of the portfolio. For instance, since at least two clusters are considered, the weights are constrained: $\forall i : 0 \leq w_i \leq 0.5$. Moreover, if clusters are lasting, then weights should be very stable. At last, neither expected returns nor risk measures are required, thereby making the method more robust.

The will to exploit the nested clusters or in other words the notion of hierarchy explains why clustering methods such as K-means or K-medoids have not been tested. Indeed, these algorithms provide a single set of clusters, with no particular organisation or structure within them⁴.

⁴The results of applying K-means or K-medoids clustering algorithms depend on the choice for the number of clusters to be searched and a starting configuration assignment. In contrast, hierarchical clustering methods do not require such specifications.

4.3 Investment strategies comparison

Portfolios are updated on a daily basis via a 252 days rolling window approach, with no forward-looking biases. This approach differs from the traditional one, where portfolio are rebalanced on a more realistic monthly basis⁵. Nevertheless, the main objective of this paper is not to create a real investment strategy, but to compare asset allocation methods. The daily rebalancing framework should help highlighting the strengths and weaknesses of the different approaches, especially the robustness.

4.3.1 Datasets

The out-of-sample performances of models are evaluated across three very disparate datasets. The three considered datasets differ in term of assets' composition and number of assets⁶:

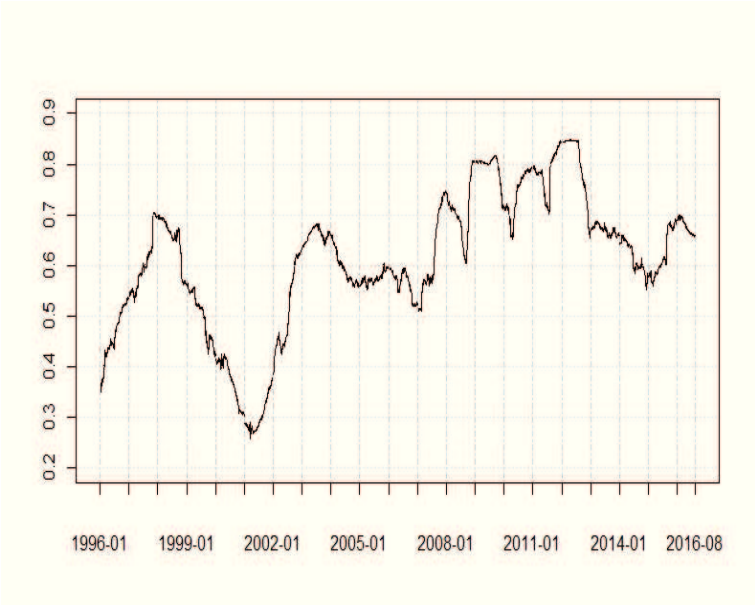
- The "S&P sectors" dataset consists of daily returns on 10 value weighted industry portfolios formed by using the Global Industry Classification Standard (GICS) developed by Standard & Poor's. The 10 industries considered are Energy, Material, Industrials, Consumer-Discretionary, Consumer-Staples, Healthcare, Financials, Information-Technology, Telecommunications, and Utilities. The data span from January 1995 to August 2016.

Figure 4.1 illustrates that the average of the absolute correlation between the "S&P sectors" is time varying and ranges between 0.26 and 0.85.

⁵For investors, the choice of the rebalancing strategy is crucial. The periodic rebalancing is not optimal and others options should be investigated (Sun et al. (2006)).

⁶Data are available from the author upon request.

Figure 4.1: S&P sectors: absolute average correlation (rolling 252 days)



Figures 4.2 and 4.3 display the distribution of the correlation matrix for the the stock market downturn of 2002 and the rebound of 2009. The "S&P sectors" correlations are always positive, yet the distribution is time varying.

Figure 4.2: Correlation matrix distribution: 2002

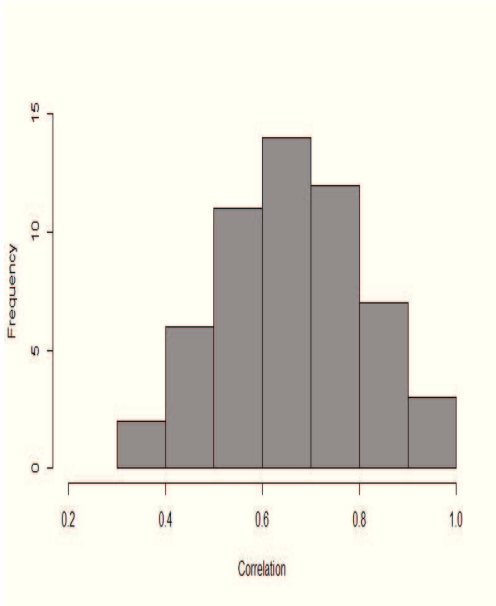
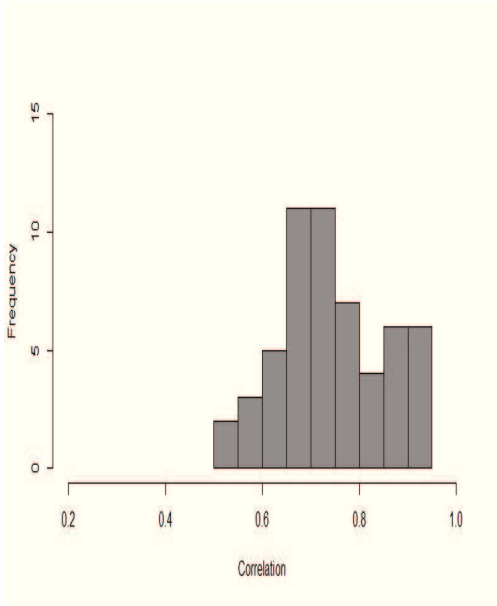


Figure 4.3: Correlation matrix distribution: 2009



- The multi-assets dataset is constituted of asset classes exhibiting different risk-return

characteristics (in local currencies): S&P 500 (US large cap), Russell 2000 (US small cap), Euro Stoxx 50 (EA large cap), Euro Stoxx Small Cap (EA small cap), FTSE 100 (UK large cap), FTSE Small Cap (UK small cap), France 2-Year bonds, France 5-Year bonds, France 10-Year bonds, France 30-Year bonds, US 2-Year bonds, US 5-Year bonds, US 10-Year bonds, US 30-Year bonds, MSCI Emerging Markets (dollars), Gold (dollars).

France has been chosen over Germany for data availability reasons. A difficult decision has been made for fixed-income indices: coupons are not reinvested. The reasoning is the following: rates are low and are expected to stay low for a long time⁷. It implies that performances in the future will not come from coupons. As the aim is to build portfolios that will perform and not that have performed, this solution has been preferred (see Appendix A for more on this subject). As a consequence, no dividends are reinvested. The data span from February 1989 to August 2016.

Figure 4.4 highlights that the average of the absolute correlation between the multi-assets changes with time and lies between 0.18 and 0.51. Asset classes are more correlated during bad times (recession for examples) than during good times. An important implication of higher correlations is that otherwise diversified portfolio lose some of diversification benefit during bad times, when most needed.

⁷Off topic: low interest rates are not the symbol of easy monetary policy, but rather an outcome of excessively tight monetary policy (see Friedman (1992) and Friedman (1997)).

Figure 4.4: Multi-assets: absolute average correlation (rolling 252 days)



Figures 4.5 and 4.6 illustrate the distribution of the correlation matrix for the years 2002 and 2009. Correlations between assets can be strong and negative, especially during bad times. It implies that hedging strategies can be implemented.

Figure 4.5: Correlation matrix distribution: 2002

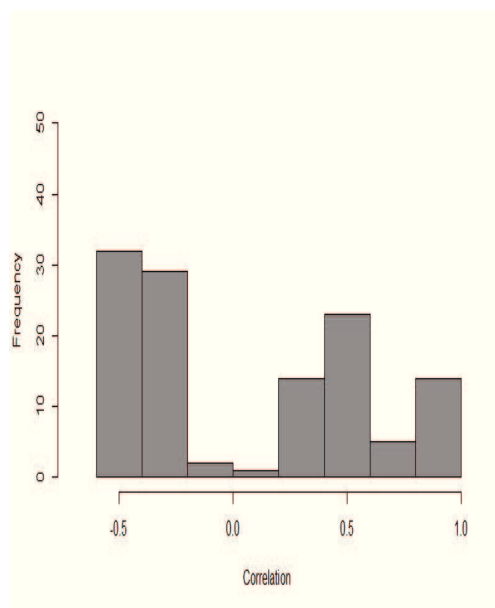
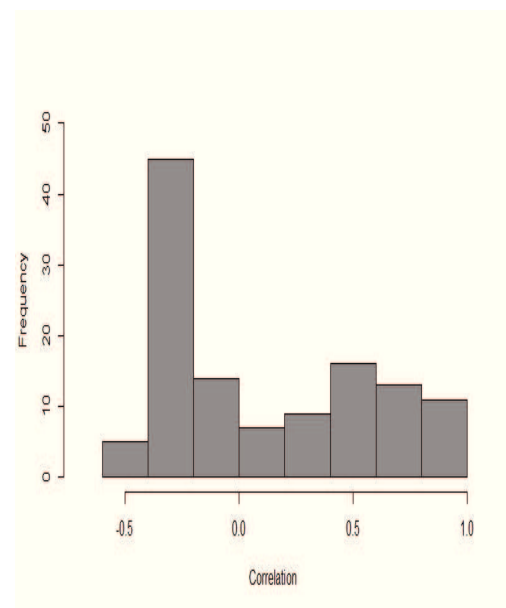


Figure 4.6: Correlation matrix distribution: 2009



- 357 individual stocks with a sufficiently long historical data from the current S&P 500

compose the last dataset. The objective is to get "real" correlations between stocks. Obviously, this dataset does not incorporate information on delistings. Since there is a strong survivor bias, comparisons with the S&P 500 are meaningless. Nevertheless, comparisons between different models are meaningful. The data span from January 1996 to August 2016.

Figure 4.7 displays that the average of the absolute correlation between the individual stocks is time varying and ranges between 0.12 and 0.59.

Figure 4.7: Individual stocks: absolute average correlation (rolling 252 days)



Figures 4.8 and 4.9 exhibit the distribution of the correlation matrix for the the stock market downturn of 2002 and the rebound of 2009. During good times, correlations are positive and higher than during bas times, which implies that the diversification effect of stock picking may be smaller during good times.

Figure 4.8: Correlation matrix distribution: 2002

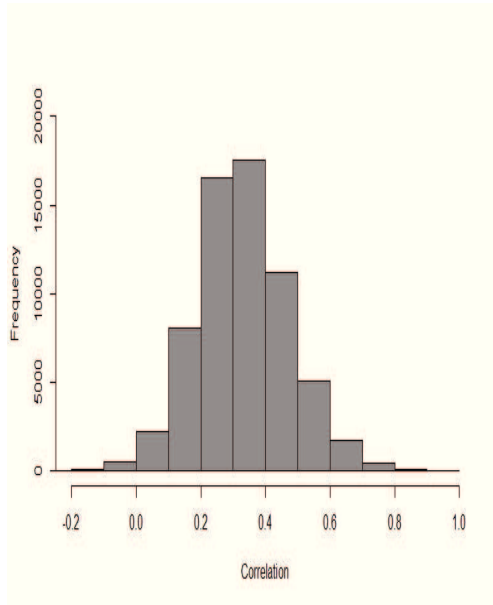
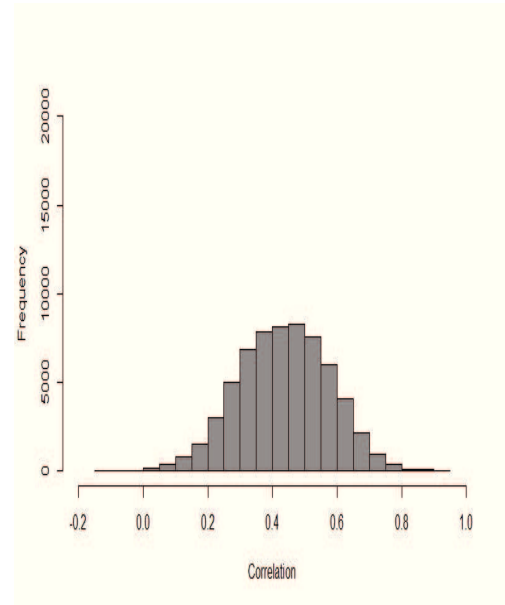


Figure 4.9: Correlation matrix distribution: 2009



Although more data history would have been desirable, the different periods cover a number of different market regimes and shocks to the financial markets and the world economy, including the "dot-com" bubble, the Great Recession and the 1994 and 1998 bond market crashes as regards the multi-asset dataset.

4.3.2 Comparison measures

Given the time series of daily out-of-sample returns generated by each strategy in each dataset, several comparison criteria are computed:

- The Adjusted Sharpe Ratio (\mathcal{ASR}) (Pezier and White (2008))⁸ explicitly adjusts for skewness and kurtosis by incorporating a penalty factor for negative skewness and excess kurtosis:

$$\mathcal{ASR} = \mathcal{SR} \left[1 + \left(\frac{\mu_3}{6} \right) \mathcal{SR} + \frac{(\mu_4 - 3)}{24} \mathcal{SR}^2 \right]$$

⁸Similar to the Adjusted Sharpe Ratio, the Modified Sharpe Ratio uses Modified VaR adjusted for skewness and kurtosis as a risk measure.

where μ_3 and μ_4 are the skewness and kurtosis of the returns distribution and \mathcal{SR} denotes the traditional Sharpe Ratio ($\mathcal{SR} = \frac{\mu - r_f}{\sigma}$, where r_f is the risk-free rate⁹).

- The certainty-equivalent return (\mathcal{CEQ}) is the risk-free rate of return that the investor is willing to accept instead of undertaking the risky portfolio strategy. DeMiguel et al. (2009) define the \mathcal{CEQ} as:

$$\mathcal{CEQ} = (\mu - r_f) - \frac{\gamma}{2}\sigma^2$$

where γ is the risk aversion. Results are reported for the case of $\gamma = 1$ but other values of the coefficient of risk aversion are also considered as a robustness check. More precisely, the employed definition of \mathcal{CEQ} captures the level of expected utility of a mean-variance investor, which is approximately equal to the certainty-equivalent return for an investor with quadratic utility (DeMiguel et al. (2009)). It is the most important number to consider to build profitable portfolios (Levy (2016)).

- The Max drawdown (\mathcal{MDD}) is an indicator of permanent loss of capital. It measures the largest single drop from peak to bottom in the value of a portfolio. In brief, the \mathcal{MDD} offers investors a worst case scenario.
- The average turnover per rebalancing (\mathcal{TO}):

$$\mathcal{TO} = \frac{1}{F} \sum_{t=2}^F |w_{i,t} - w_{i,t-1}|$$

where F is the number of out-of-sample forecasts.

- The Sum of Squared Portfolio Weights (\mathcal{SSPW}) used in Goetzmann and Kumar (2008) exhibits the underlying level of diversification in a portfolio and is defined as follows:

$$\mathcal{SSPW} = \frac{1}{F} \sum_{t=2}^F \sum_{i=1}^N w_{i,t}^2$$

⁹A risk-free interest rate of zero is assumed when calculating the \mathcal{ASR} and \mathcal{CEQ} .

$SSPW$ ranges from 0 to 1, where 1 represents the most concentrated portfolio.

No transactions costs or economic costs generated by the turnover are reported. Indeed, the study of transactions costs is difficult because investors face different fees and the same strategy can be implemented via Futures or ETF or CFD or cash. Moreover, taxes and the chosen rebalancing strategy influence costs. Nevertheless, high average turnover per rebalancing leads to expensive strategies.

Data snooping

To avoid data snooping (White (2000)), the model confidence set (MCS) procedure proposed by Hansen et al. (2011) is computed. The MCS procedure is a model selection algorithm, which filters a set of models from a given entirety of models. The resulting set contains the best models with a probability that is no less than $1 - \alpha$ with α being the size of the test (see Hansen et al. (2011)).

An advantage of the test is that it not necessarily selects a single model, instead it acknowledges possible limitations in the data since the number of models in the set containing the best model will depend on how informative the data are.

More formally, define a set M_0 that contains the set of models under evaluation indexed by: $i = 0, \dots, m_0$. Let $d_{i,j,t}$ denote the loss differential between two models by

$$d_{i,j,t} = L_{i,t} - L_{j,t}, \forall i, j \in M_0$$

L is the loss calculated from some loss function for each evaluation point $t = 1, \dots, F$. The set of superior models is defined as:

$$M^* = \{i \in M_0 : E[d_{i,j,t}] \leq 0 \forall j \in M_0\}$$

The MCS uses a sequential testing procedure to determine M^* . The null hypothesis being tested is:

$$\begin{cases} H_{0,M} : E[d_{i,j,t}] = 0 \forall i, j \in M \text{ where } M \text{ is a subset of } M_0 \\ H_{A,M} : E[d_{i,j,t}] \neq 0 \text{ for some } i, j \in M \end{cases}$$

When the equivalence test rejects the null hypothesis, at least one model in the set M is considered inferior and the model that contributes the most to the rejection of the null is eliminated from the set M . This procedure is repeated until the null is accepted and the remaining models in M now equal $\widehat{M}_{1-\alpha}^*$.

According to Hansen et al. (2011), the following two statistics can be used for the sequential testing of the null hypothesis:

$$t_{i,j} = \frac{\bar{d}_{i,j}}{\sqrt{\widehat{\text{var}}(\bar{d}_{i,j})}} \text{ and } t_i = \frac{\bar{d}_i}{\sqrt{\widehat{\text{var}}(\bar{d}_i)}}$$

where m is the number of models in M , $\bar{d}_i = (m - 1)^{-1} \sum_{j \in M} \bar{d}_{i,j}$, is the simple loss of the i^{th} model relative to the averages losses across models in the set M , and $\bar{d}_{i,j} = (m)^{-1} \sum_{t=1}^m d_{i,j,t}$ measures the relative sample loss between the i^{th} and j^{th} models. Since the distribution of the test statistic depends on unknown parameters a bootstrap procedure is used to estimate the distribution.

In this paper, the MCS is applied with profit maximization loss function (\mathcal{ASR} and \mathcal{CEQ}). It should be noted that the MCS aims at finding the best model and all models which are indistinguishable from the best, not those better than a benchmark. To determine if models are better than a benchmark, the stepwise test of multiple reality check by Romano and Wolf (2005) and the stepwise multiple superior predictive ability test by Hsu et al. (2013) should be considered. However, a small trick is possible: if the benchmark is not selected in the best models set, investors can conclude that their strategies "beat" the benchmark.

4.4 Empirical results

4.4.1 S&P sectors

Table 4.1 highlights the interest of hierarchical clustering based portfolios, especially the CL based model. It is the only model selected in the best models set $\widehat{M}_{75\%}^*$ for both ASR and CEQ . This portfolio is diversified ($SSPW=0.114$) and the average turnover per rebalancing is quite low ($TO=0.817\%$).

The MV is included in $\widehat{M}_{ASR-75\%}^*$ but its diversification ratio $SSPW$ is by far the highest of all models: the portfolio is concentrated instead of being diversified.

Table 4.1: **Investment strategies comparison: S&P 500 sectors (January 1996-August 2016)**

	ASR	CEQ	MDD	TO	$SSPW$
EW	0.422	6.81	54.2	-	0.100
MV	0.448**	5.75	37.8	1.78	0.480
MDP	0.344	5.28	57.6	1.83	0.217
ERC	0.442	6.40	51.6	0.294	0.107
IVRB	0.428	6.51	49.0	0.474	0.121
SL	0.418	6.53	53.4	0.669	0.115
CL	0.430**	6.87**	51.0	0.817	0.114
AL	0.421	6.62	52.6	0.762	0.114
WM	0.415	6.49	52.9	0.883	0.149

Note: This table reports comparison criteria used to evaluate the quality of the models: the Adjusted Sharpe Ratio (ASR), the certainty-equivalent return (CEQ) in percent, the Max drawdown (MDD) in percent, the average turnover per rebalancing (TO) in percent, the Sum of Squared Portfolio Weights ($SSPW$). ** indicates the model is in the set of best models $\widehat{M}_{75\%}^*$. EW refers to the equal weight allocation, MV refers to the minimum variance allocation, MDP refers to the most diversified portfolio allocation, ERC refers to the equal risk contribution allocation, IVRB refers to the inverse-volatility risk budget allocation, SL refers to the simple linkage based allocation, CL refers to the complete linkage based allocation, AL refers to the average linkage based allocation, WM refers to the Ward's method based allocation.

Table 4.2 exhibits that the number of clusters selected by the Gap index seems stable (the standard deviation is small for all methods). It explains why the hierarchical clustering based portfolios presents low average turnover per rebalancing.

Table 4.2: Number of clusters: S&P 500 sectors (January 1996-August 2016)

	Mean	SD	Max	Min
SL	7.0	0.23	7	6
CL	6.9	0.33	7	4
AL	6.9	0.29	7	5
WM	6.9	0.26	7	5

Note: This table reports the mean, the standard deviation (SD), the max and the min of the number of clusters. SL refers to the simple linkage based allocation, CL refers to the complete linkage based allocation, AL refers to the average linkage based allocation, WM refers to the Ward's method based allocation.

4.4.2 Multi-assets dataset

Table 4.3 paints a contrasting picture: risk-based portfolios achieve impressive ASR along with low \mathcal{CEQ} . For instance, IVRB constitutes the best models set $\widehat{M}_{ASR-75\%}^*$. Moreover, MDP and ERC display high ASR in comparison with others models: 0.717 and 0.707. That said, risk-based portfolios attain very low \mathcal{CEQ} , especially IVRB ($\mathcal{CEQ}=0.951$, while $\mathcal{CEQ}=4.71$ for AL). Above all, they do not produce diversified portfolios ($\mathcal{SSPW}=0.500$ for IVRB for example). It implies that portfolios are almost only invested in bonds, thereby being very exposed to shocks from this asset class. This is not what diversified portfolios aim at.

Hierarchical clustering based portfolios do not face the same problems. Al and SL compose $\widehat{M}_{\mathcal{CEQ}-75\%}^*$, while delivering reasonably good ASR . All portfolios are diversified and the average turnover per rebalancing is low.

Table 4.3: **Investment strategies comparison: multi-assets (February 1989-August 2016)**

	<i>ASR</i>	<i>CEQ</i>	<i>MDD</i>	<i>TO</i>	<i>SSPW</i>
EW	0.601	4.02	24.9	-	0.0625
MV	0.611	1.31	7.49	4.22	0.403
MDP	0.717	1.95	7.7	2.92	0.296
ERC	0.707	1.91	9.34	0.509	0.164
IVRB	0.581**	0.951	6.85	0.500	0.342
SL	0.597	4.67**	31.4	1.28	0.086
CL	0.586	4.51	29.9	1.15	0.085
AL	0.602	4.71**	29.7	1.19	0.085
WM	0.583	4.47	29.9	1.25	0.084

Note: This table reports comparison criteria used to evaluate the quality of the models: the Adjusted Sharpe Ratio (*ASR*), the certainty-equivalent return (*CEQ*) in percent, the Max drawdown (*MDD*) in percent, the average turnover per rebalancing (*TO*) in percent, the Sum of Squared Portfolio Weights (*SSPW*). ** indicates the model is in the set of best models $\widehat{M}_{75\%}^*$. EW refers to the equal weight allocation, MV refers to the minimum variance allocation, MDP refers to the most diversified portfolio allocation, ERC refers to the equal risk contribution allocation, IVRB refers to the inverse-volatility risk budget allocation, SL refers to the simple linkage based allocation, CL refers to the complete linkage based allocation, AL refers to the average linkage based allocation, WM refers to the Ward's method based allocation.

Table 4.4 exhibits that the number of clusters lies between 5 and 10 for CL based portfolios and between 7 and 10 for others methods. AL and SL are selected in $\widehat{M}_{CEQ-75\%}^*$ and display the lowest standard deviation. It seems thus that stable clusters lead to better performances.

Table 4.4: **Number of clusters: multi-assets (February 1989-August 2016)**

	Mean	SD	Max	Min
SL	9.8	0.53	10	7
CL	9.1	1.1	10	5
AL	9.2	0.96	10	7
WM	9.1	1.2	10	7

Note: This table reports the mean, the standard deviation (SD), the max and the min of the number of clusters. SL refers to the simple linkage based allocation, CL refers to the complete linkage based allocation, AL refers to the average linkage based allocation, WM refers to the Ward's method based allocation.

4.4.3 Individual stocks

Table 4.5 points out that hierarchical clustering based portfolios outperform risk-based portfolios. Indeed, WM is the only model selected in the best models set $\widehat{M}_{ASR-75\%}^*$ and the best models set $\widehat{M}_{CEQ-75\%}^*$ is only constituted by one model: AL. WM based portfolio seems more diversified ($SSPW=0.041$ for AL and $SSPW=0.0051$ for WM).

The main drawback is the surprising elevated average turnover per rebalancing. Table 4.6 highlight that the number of clusters is more volatile than in others datasets (the standard deviations is around 4 and the min-max ranges from 46 to 80 clusters). This point needs to be further investigated, in particular, the impact of the criteria used to select the number of clusters and the consequences of the correlation matrix "shrinkage" on the stability of the clusters.

Table 4.5: **Investment strategies comparison: Individual stocks (January 1996-August 2016)**

	<i>ASR</i>	<i>CEQ</i>	<i>MDD</i>	<i>TO</i>	<i>SSPW</i>
EW	0.595	13.3	52.2	-	0.0028
MV	0.600	13.0	51.2	0.021	0.0048
MDP	0.658	16.4	45.1	7.21	0.052
ERC	0.570	12.1	49.4	0.79	0.0036
IVRB	0.560	10.8	47.1	0.98	0.0045
SL	0.492	19.2	43.1	32.4	0.0552
CL	0.473	16.5	47.4	33.7	0.0151
AL	0.520	19.5**	46.1	33.6	0.041
WM	0.572**	14.2	51.2	33.4	0.0051

Note: This table reports comparison criteria used to evaluate the quality of the models: the Adjusted Sharpe Ratio (*ASR*), the certainty-equivalent return (*CEQ*) in percent, the Max drawdown (*MDD*) in percent, the average turnover per rebalancing (*TO*) in percent, the Sum of Squared Portfolio Weights (*SSPW*).** indicates the model is in the set of best models \widehat{M}_{75}^* . EW refers to the equal weight allocation, MV refers to the minimum variance allocation, MDP refers to the most diversified portfolio allocation, ERC refers to the equal risk contribution allocation, IVRB refers to the inverse-volatility risk budget allocation, SL refers to the simple linkage based allocation, CL refers to the complete linkage based allocation, AL refers to the average linkage based allocation, WM refers to the Ward's method based allocation.

Table 4.6: **Number of clusters: Individual stocks (January 1996-August 2016)**

	Mean	SD	Max	Min
SL	78.0	4.6	80	48
CL	76.2	4.5	80	46
AL	77.4	4.4	80	51
WM	78.2	4.4	80	50

Note: This table reports the mean, the standard deviation (SD), the max and the min of the number of clusters. SL refers to the simple linkage based allocation, CL refers to the complete linkage based allocation, AL refers to the average linkage based allocation, WM refers to the Ward's method based allocation.

4.5 Future research

The aim of this paper is to introduce hierarchical clustering based asset allocation. The previous section highlight the interest of this method. Nevertheless, there may be several

way of improving the out-of-sample performance.

- Typical machine learning issues have to be investigated, such as the choice of the distance measure and the criteria used to select the number of clusters. Different number of clusters lead to different assets weights. For instance, in the small example described before in the paper, if we consider two clusters instead of three, the weight of asset 5 decreases from 0.25 to 0.1667:

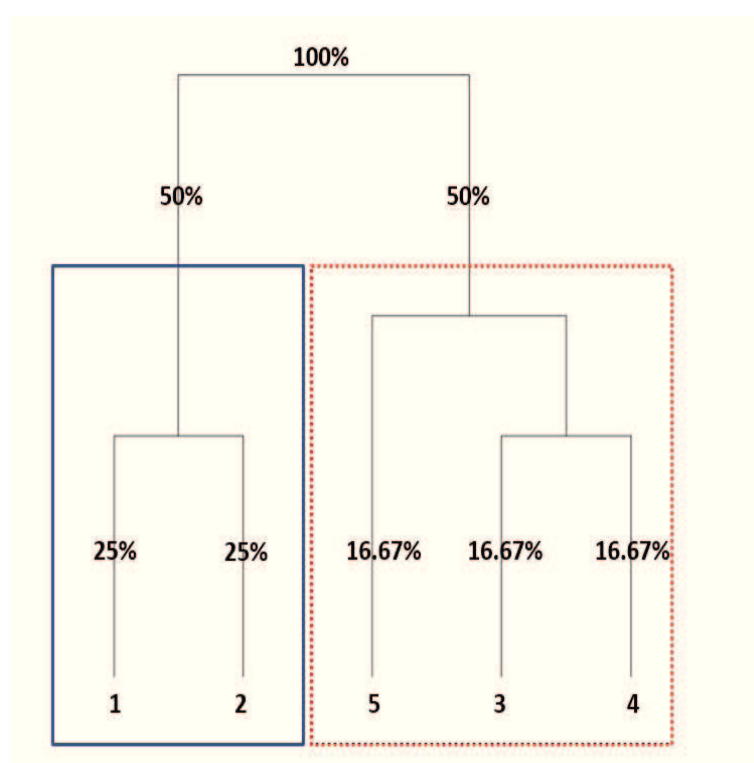


Figure 4.1: Asset allocation weights: a small example revisited

- Trying other modern clustering methods (Kernel K-means, Spectral clustering,...) may (or may not) provide impressive out-of-sample performances, even if clusters are not nested.
- Another weighting strategy is obviously possible. For instance, López de Prado (2016a) employs an inverse-variance weighting allocation with no selection of clusters. One drawback of the proposed approach is that risk management is not part of the weighting

strategy. Since assets within a clusters are correlated, a combination with risk parity or Sharpe parity could lead to interesting results.

- Even if the proposed method does not require a matrix to be inverted, the correlation matrix has still to be estimated. Historical asset returns are commonly used. But, in many cases, the length of the asset returns' time series used for estimation is not long enough compared to the number of assets considered (Jobson and Korkie (1980)). As a result, the estimated correlation matrix is unstable. One approach to improving estimation is to use "shrinkage". The general idea is that a compromise between a logical/theoretical estimator and a sample estimator will yield better results than either method (see Ledoit and Wolf (2004), Ledoit and Wolf (2014) and Gerber et al. (2015)).

Conclusion

Diversification is often spoken of as the only free lunch in investing. Yet, truly diversifying a portfolio is easier said than done. For instance, modern portfolio optimization techniques often fail to outperform a basic equal-weighted allocation (DeMiguel et al. (2009)).

Building upon the fundamental notion of hierarchy (Simon (1962)), López de Prado (2016a) introduces a new portfolio diversification technique called "Hierarchical Risk Parity", which uses graph theory and machine learning techniques.

Exploiting the same basic idea in a different way, a hierarchical clustering based asset allocation is proposed. Classical and more modern hierarchical clustering methods are tested, namely Simple Linkage, Complete Linkage, Average Linkage, Ward's Method. Once the assets are hierarchically clustered, a simple and efficient capital allocation within and across clusters of investments at multiple hierarchical levels is computed. The main principle is to find a diversified weighting by distributing capital equally to each cluster hierarchy, so that many correlated assets receive the same total allocation as a single uncorrelated one.

The out-of-sample performances of hierarchical clustering based portfolios and more traditional risk-based portfolios are evaluated across three empirical datasets, which differ in terms of number of assets and composition of the universe ("S&P sectors", multi-assets and individual stocks). To prevent strategies that perform by luck to be considered as effective, the comparison of profit measures is assessed using the bootstrap based model confidence set procedure (Hansen et al. (2011)).

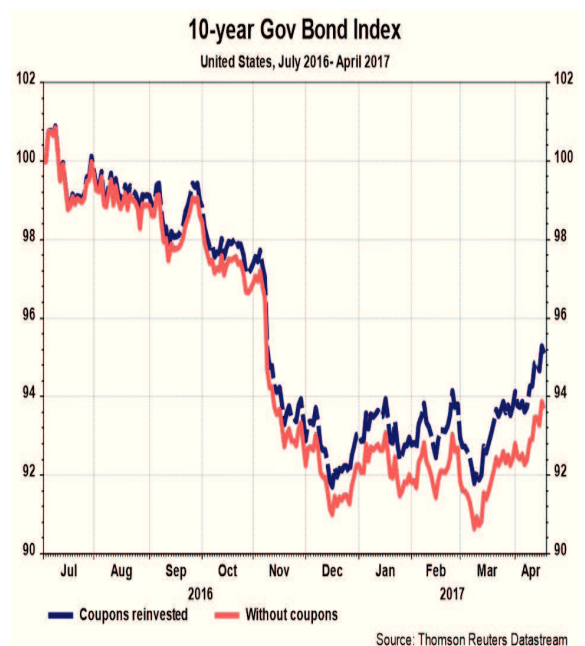
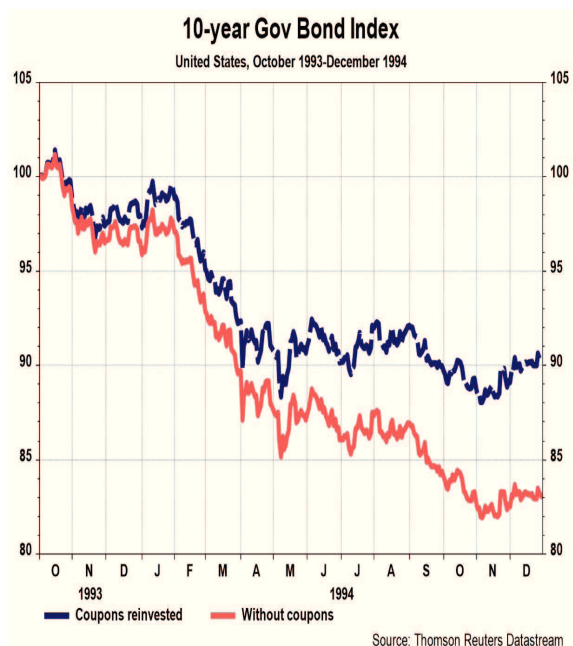
The empirical results point out that hierarchical clustering based portfolios are truly diversified and achieve statistically better risk-adjusted performances, as measured by the the Adjusted Sharpe Ratio (Pezier and White (2008)) and by the Certainty-Equivalent Return on all datasets. The only exception concerns the multi-assets dataset where risk-based portfolios produce impressive \mathcal{ASR} along with ridiculous low \mathcal{CEQ} . Among clustering methods, there is no clear winner.

Last but not least, this article opens the door for further research. Testing other clustering methods and investigating typical machine learning issues, such as the choice of the distance measure and the criteria used to select the number of clusters, come naturally to mind. Above all, improving the estimation of the correlation matrix seems to be the most important priority. Potential improvements may come from the use of "shrinkage" or/and the detection of current phase of the economic cycles.

Appendix

4.A Bond market

In 1994 and in 2016, the bond market suffered a sharp and sudden selloff. In 1994, the coupon yield was high enough to partly protect the investors. Since rates are low in 2016, the change in rates is a constant drag on performance that the initially anaemic coupon yield can not cover.



Conclusion

Diversification is the only free lunch in finance (Markowitz (1952)), yet it is easier said than done. To create portfolios that stay diversified when needed, this thesis proposes lines of thinking. The first one is to deepen the knowledge of interaction dynamics between financial markets the macroeconomy. The second one is to explore a new way of capital allocation.

The willingness of investors to bear risk varies over time, larger in good times, and less in bad times, leading to time-varying risk premiums (Cochrane (2016)). Yet, there is still no consensus on the definition of good and bad time. To fill this gap, the theoretical influence of economic cycles on time-varying risk premiums is explained based on two key economic concepts, nominal GDP and adaptive expectations. It is then exhibited over the period from January 2002 to December 2013: dynamic investment strategies based on economic cycles turning points emphasize the importance of economical cycles, especially the growth cycle, better known as the output gap, for euro and dollar-based investors. The theoretical definition of good time is thus when the current growth rate of the economy is above the trend growth rate.

To quickly and accurately detect economic turning points in real time in the United States and in the euro area, probabilistic indicators are first created from a simple and transparent machine-learning algorithm known as Learning Vector Quantization (Kohonen (2001)). Those indicators are robust, interpretable and preserve economic consistency. A more complex approach is then evaluated: ensemble machine learning algorithms, referred to as random forest (Breiman (2001)) and as boosting (Schapire (1990)), are ap-

plied. The two key features of those algorithms are their abilities to entertain a large number of predictors and to perform estimation and variable selection simultaneously. Both approaches are effective to detect economic turning points in real time over the period from January 2002 to December 2013. Strategies based on the turning points of the growth cycle induced by the models achieve thus excellent risk-adjusted returns in real time: timing the market is possible.

Modern and complex portfolio optimisation methods are optimal in-sample, but often provide rather poor out-of-sample forecast performance (DeMiguel et al. (2009)). To allocate capital differently, a new portfolio diversification technique called hierarchical clustering based asset allocation is proposed. Once the assets are hierarchically clustered, a simple and efficient capital allocation within and across clusters of investments at multiple hierarchical levels is computed. The main principle is to find a diversified weighting by distributing capital equally to each cluster hierarchy, so that many correlated assets receive the same total allocation as a single uncorrelated one. The out-of-sample performances of hierarchical clustering based portfolios and more traditional risk-based portfolios are evaluated across three disparate datasets, which differ in terms of number of assets and composition of the universe ("S&P sectors", multi-assets and individual stocks). The empirical results indicate that hierarchical clustering based portfolios are robust, truly diversified and achieve statistically better risk-adjusted performances than commonly used portfolio optimization techniques.

At last, this thesis opens doors to many areas for further research and development. An attempt to forecast business cycle turning points three to twelve months ahead could be very interesting. To this aim, the use of data generated by individuals (social media posts, product reviews, search trends, etc.), to data generated by business processes (company exhaust data, commercial transaction, credit card data, etc.) and data generated by sensors (satellite image data, foot and car traffic, ship locations, etc.) could provide new interesting explanatory variables. Moreover, deep learning methods have dramatically improved the state-of-the-art in speech recognition, visual object recognition, object

detection and many other domains such as drug discovery and genomics, but are largely unknown in economics. They could achieve better results than ensemble machine learning algorithms...or not. Finally, hierarchical clustering based asset allocation has been proven to be effective. Nevertheless, there may be several way of improving the out-of-sample performance, which have to be developed.

Bibliography

- Ahn, D.-H., Min, B.-K., and Yoon, B. (2016). Why has the size effect disappeared? Technical report.
- Anas, J., Billio, M., Ferrara, L., and Mazzi, G. L. (2008). A system for dating and detecting turning points in the euro area. *Manchester School*, 76(5):549–577.
- Anas, J. and Ferrara, L. (2004). Detecting Cyclical Turning Points: The ABCD Approach and Two Probabilistic Indicators. *Journal of Business Cycle Measurement and Analysis*, 2004(2):193–225.
- Anderson, P. W. (1972). More is different. *Science*, 177(4047):393–396.
- Baitinger, E. and Papenbrocky, J. (2016). Interconnectedness risk and active portfolio management. Technical report.
- Baltas, N. and Karyampas, D. (2016). Forecasting the equity risk premium: Predictability versus profitability. Technical report.
- Baxter, M. and King, R. (1999). Measuring business cycles: Approximate band-pass filters for economic time series. *The Review of Economics and Statistics*, 81(4):575–593.
- Bengoechea, P., Camacho, M., and Perez-Quiros, G. (2006). A useful tool for forecasting the euro-area business cycle phases. *International Journal of Forecasting*, 22(4):735 – 749.

- Berge, T. (2015). Predicting Recessions with Leading Indicators: Model Averaging and Selection over the Business Cycle. *Journal of Forecasting*, 34(6):455–471.
- Bergmeir, C., Hyndman, R. J., and Koo, B. (2015). A Note on the Validity of Cross-Validation for Evaluating Time Series Prediction. Technical report.
- Bernanke, B. (2007). Inflation expectations and inflation forecasting at the monetary economics workshop of the national bureau of economic research summer institute, cambridge, massachusetts, july 10, 2007. Speech, Board of Governors of the Federal Reserve System (U.S.).
- Boivin, J. and Ng, S. (2006). Are more data always better for factor analysis? *Journal of Econometrics*, 132(1):169–194.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45:5–32.
- Breiman, L., Friedman, J., Olshen, R., and Stone, C. (1984). *Classification and Regression Trees*. Wadsworth and Brooks, Monterey, CA.
- Brinson, G., Hood, R., and Beebower, G. (1986). Determinants of portfolio performance. *Financial Analyst Journal*.
- Brown, S. (2008). Elusive return predictability: Discussion. *International Journal of Forecasting*, 24(1):19–21.
- Bühlmann, P. and Yu, B. (2003). Boosting with the L_2 loss: Regression and classification. *Journal of the American Statistical Association*, 98:324–338.
- Burns, A. and Mitchell, W. (1946). *Measuring Business Cycles*. Number burn46-1 in NBER Books. National Bureau of Economic Research, Inc.
- Camacho, M., Perez-Quiros, G., and Poncela, P. (2015). Extracting nonlinear signals from several economic indicators. *Journal of Applied Econometrics*, 30(7):1073–1089.

- Campbell, J. and Viceira, L. (2002). *Strategic asset allocation: portfolio choice for long-term investors*. Oxford University Press, USA.
- Caruana, R. and Niculescu-Mizil, A. (2005). An empirical comparison of supervised learning algorithms using different performance metrics. In *In Proc. 23 rd Intl. Conf. Machine learning*, pages 161–168.
- 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.
- Choueifaty, Y., Froidure, T., and Reynier, J. (Spring 2013). Properties of the most diversified portfolio. *Journal of Investment Strategies*.
- Christiano, L. and Fitzgerald, T. (2003). The Band Pass Filter. *International Economic Review*, 44(2):435–465.
- Cochrane, J. H. (2016). Macro-Finance. NBER Working Papers 22485, National Bureau of Economic Research, Inc.
- Cooper, I. and Priestley, R. (2009). Time-varying risk premiums and the output gap. *Review of Financial Studies*, 22(7):2801 – 2833.
- DeMiguel, V., Garlappi, L., and Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *The Review of Financial Studies*, 22(5):1915–1953.
- Duarte, A., Venetis, I. A., and Paya, I. (2005). Predicting real growth and the probability of recession in the euro area using the yield spread. *International Journal of Forecasting*, 21(2):261–277.
- Efron, B. and Tibshirani, R. (1994). *An Introduction to the Bootstrap*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis.

- Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with b-splines and penalties. *Statist. Sci.*, 11(2):89–121.
- Ferrara, L. and Guegan, D. (2005). Detection of the industrial business cycle using setar models. *Journal of Business Cycle Measurement and Analysis*, 2005(3):353–371.
- Fisher, I. (1911). *The Purchasing Power of Money*. Macmillan.
- Freund, Y. and Schapire, R. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55:119–139.
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29:1189–1232.
- Friedman, J. H., Hastie, T., and Tibshirani, R. (2000). Additive logistic regression: A statistical view of boosting (with discussion). *The Annals of Statistics*, 28:337–407.
- Friedman, M. (1957). *A Theory of the Consumption Function*. National Bureau of Economic Research, Inc.
- Friedman, M. (1992). Do Old Fallacies Ever Die? *Journal of Economic Literature*, 30(4):2129–32.
- Friedman, M. (1997). Rx for japan: Back to the future. *Wall Street Journal*.
- Gerber, S., Markowitz, H., and Pujara, P. (2015). Enhancing Multi-Asset Portfolio Construction Under Modern Portfolio Theory with a Robust Co-Movement Measure. *Social Science Research Network Working Paper Series*.
- Giusto, A. and Piger, J. (forthcoming). Identifying business cycle turning points in real time with vector quantization. *International Journal of Forecasting*.
- Goetzmann, W. N. and Kumar, A. (2008). Equity portfolio diversification. *Review of Finance*, 12(3):433–463.

- Griveau-Billion, T., Richard, J.-C., and Roncalli, T. (2013). A fast algorithm for computing high-dimensional risk parity portfolios. Technical report.
- Hamilton, J. (2011). Calling recessions in real time. *International Journal of Forecasting*, 27(4):1006–1026.
- Han, Y., Yang, K., and Zhou, G. (2013). A new anomaly: The cross-sectional profitability of technical analysis. *Journal of Financial and Quantitative Analysis*, 48:1433–1461.
- Hanley, J. A. and McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143(1):29–36.
- Hansen, P., Lunde, A., and Nason, J. (2011). The model confidence set. *Econometrica*, 79(2):453–497.
- Harding, D. and Pagan, A. (2002). Dissecting the cycle: a methodological investigation. *Journal of Monetary Economics*, 49(2):365–381.
- Hastie, T. (2007). Comment: Boosting algorithms: Regularization, prediction and model fitting. *Statistical Science*, 22:513–515.
- Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York, 2 edition.
- Hsu, Y.-C., Kuan, C.-M., and Yen, M.-F. (2013). A generalized stepwise procedure with improved power for multiple inequalities testing. IEAS Working Paper : academic research 13-A001, Institute of Economics, Academia Sinica, Taipei, Taiwan.
- Ibbotson, R. and Kaplan, P. (2000). Does asset allocation policy explain 40, 90, 100 percent of performance? *Financial Analyst Journal*.
- Ilmanen, A. (2011). *Expected Returns : an investor's guide to harvesting market rewards*. Wiley finance.

- Ishwaran, H. (2007). Variable importance in binary regression trees and forests. *Electron. J. Statist.*, 1:519–537.
- Jobson, J. D. and Korkie, R. (1980). Estimation for Markowitz Efficient Portfolios. *Journal of the American Statistical Association*, 75.
- Keynes, J. M. (1936). *The General Theory of Employment, Interest and Money*. Macmillan.
- Kohonen, T. (2001). *Self-Organizing Maps*. Springer.
- Ledoit, O. and Wolf, M. (2004). Honey, i shrunk the sample covariance matrix. *The Journal of Portfolio Management*, 30(4):110–119.
- Ledoit, O. and Wolf, M. (2014). Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets goldilocks. *Available at SSRN 2383361*.
- Levy, M. (2016). Measuring portfolio performance: Sharpe, alpha, or the geometric mean? Technical report.
- Lin, H., Wu, C., and Zhou, G. (2016). Forecasting Corporate Bond Returns: An Iterated Combination Approach. Technical report.
- Liu, W. and Moench, E. (2016). What predicts {US} recessions? *International Journal of Forecasting*, 32(4):1138 – 1150.
- López de Prado, M. (2016a). Building diversified portfolios that outperform out of sample. *The Journal of Portfolio Management*, 42(4):59–69.
- López de Prado, M. (2016b). Mathematics and economics: A reality check. *Journal of Portfolio Management*, 43(1).
- Lustig, H. and Verdelhan, A. (2012). Business cycle variation in the risk-return trade-off. *Journal of Monetary Economics*, 59:35–49.

- Maillard, S., Roncalli, T., and Teiletche, J. (2010). The properties of equally-weighted risk contribution portfolios. *The Journal of Portfolio Management*, 36(4):60–70.
- Mantegna, R. N. (1999). Hierarchical structure in financial markets. *The European Physical Journal B-Condensed Matter and Complex Systems*, 11(1):193–197.
- Markowitz, H. (1952). Portfolio Selection. *Journal of Finance*, 7(1):77–91.
- Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates*. *The Journal of Finance*, 29(2):449–470.
- Michaud, R. (1989). The Markowitz Optimization Enigma: Is Optimized Optimal. *Financial Analysts Journal*.
- Mintz, I. (1974). Dating united states growth cycles. In *Explorations in Economic Research, Volume 1, Number 1*, pages 1–113. National Bureau of Economic Research, Inc.
- Ng, S. (2014). Viewpoint: Boosting recessions. *Canadian Journal of Economics*, 47(1):1–34.
- Nilsson, R. and Gyomai, G. (2011). Cycle Extraction: A Comparison of the Phase-Average Trend Method, the Hodrick-Prescott and Christiano-Fitzgerald Filters. OECD Statistics Working Papers 2011/4, OECD Publishing.
- Okun, A. (1962). Potential gnp: Its measurement and significance. In *Proceedings of the business and Economic Statistics Section*, pages 98–104.
- Page, S. and Taborsky, M. (2011). The myth of diversification : risk factors versus asset classes. *The journal of portfolio management*, 37(4):1–2.
- Pesaran, H. and Timmermann, A. (1994). Forecasting stock returns: an examination of stock market trading in the presence of transaction costs. *Journal of forecasting*, 13(4):335–367.

- Pezier, J. and White, A. (2008). The relative merits of alternative investments in passive portfolios. *The Journal of Alternative Investments*.
- Phillips, A. W. (1958). The relation between unemployment and the rate of change of money wage rates in the united kingdom. *Economica*, 25(100):283–299.
- Piger, J. (2011). *Econometrics: Models of Regime Changes*, pages 190–202. Springer New York, New York, NY.
- Raffinot, T. (2007). A monthly indicator of GDP for Euro-Area based on business surveys. *Applied Economics Letters*, 14(4):267–270.
- RAPACH, D. E., STRAUSS, J. K., and ZHOU, G. (2013). International stock return predictability: What is the role of the united states? *The Journal of Finance*, 68(4):1633–1662.
- Rapach, D. E., Zhou, G., et al. (2013). Forecasting stock returns. *Handbook of Economic Forecasting*, 2(Part A):328–383.
- Ribeiro, M. T., Singh, S., and Guestrin, C. (2016). Why should i trust you?: Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1135–1144. ACM.
- Roberts, J. M. (1995). New keynesian economics and the phillips curve. *Journal of Money, Credit and Banking*, 27(4):975–984.
- Romano, J. P. and Wolf, M. (2005). Stepwise multiple testing as formalized data snooping. *Econometrica*, 73(4):1237–1282.
- Roncalli, T. (2013). *Introduction to Risk Parity and Budgeting*. Chapman & Hall.
- Schapire, R. E. (1990). The strength of weak learnability. In *Machine Learning*, pages 197–227.

- Schapire, R. E. (2003). *Nonlinear Estimation and Classification*, chapter The Boosting Approach to Machine Learning: An Overview, pages 149–171. Springer New York, New York, NY.
- Simon, H. A. (1962). The architecture of complexity. In *Proceedings of the American Philosophical Society*, pages 467–482.
- Stock, J. and Watson, M. (1989). New Indexes of Coincident and Leading Economic Indicators. In *NBER Macroeconomics Annual 1989, Volume 4*, NBER Chapters, pages 351–409. National Bureau of Economic Research, Inc.
- Stock, J. and Watson, M. (2003). Has the business cycle changed and why? In *NBER Macroeconomics Annual 2002, Volume 17*, pages 159–230. National Bureau of Economic Research, Inc.
- Sumner, S. (2014). Nominal gdp targeting: A simple rule to improve fed performance. *Cato Journal*, 34(2):315–337.
- Sun, W., Fan, A., Chen, L.-W., Schouwenaars, T., and Albot, M. A. (2006). Optimal rebalancing for institutional portfolios. *The Journal of Portfolio Management*, 32(2):33–43.
- 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.
- Tibshirani, R., Walther, G., and Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):411–423.
- Tumminello, M., Coronello, C., Lillo, F., Micciche, S., and Mantegna, R. N. (2007). Spanning trees and bootstrap reliability estimation in correlation-based networks. *International Journal of Bifurcation and Chaos*, 17(07):2319–2329.

- Tumminello, M., Lillo, F., and Mantegna, R. N. (2010). Correlation, hierarchies, and networks in financial markets. *Journal of Economic Behavior & Organization*, 75(1):40–58.
- Varian, H. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives*, 28(2):3–28.
- Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301):236–244.
- White, H. (2000). A Reality Check for Data Snooping. *Econometrica*, 68(5):1097–1126.
- Wu, B., Ai, H., and Liu, R. (2004). Glasses detection by boosting simple wavelet features. In *Pattern Recognition, 2004*, volume 1, pages 292–295 Vol.1.
- Zhou, G. and Zhu, X. (2017). Bond Return Predictability and Macroeconomy: The International Link. Technical report.
- Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 67(2):301–320.

Résumé en français de la thèse : Asset Allocation, Economic Cycles and Machine Learning

Thomas Raffinot

Directrice de thèse : Anne Epaulard

Introduction

La crise économique et financière mondiale, appelée communément Grande Récession, a mis en exergue le besoin d'approfondir les liens entre le contexte macroéconomique et les marchés financiers.

la crise a mis à mal la théorie moderne du portefeuille développée en 1952 par Harry Markowitz. Markowitz (1952) avait formalisé et quantifié l'adage de ne pas mettre tous ses oeufs dans le même panier : la diversification. Ainsi, une combinaison judicieuse de nombreux actifs dans un portefeuille permet de réduire le risque total subi, mesuré par la variance de sa rentabilité, pour un taux de rentabilité espérée donné. Ces stratégies d'investissement diversifiées étaient censées pouvoir affronter des crises sans subir de lourdes pertes, la réalité a prouvé qu'il n'en était rien. La diversification a échoué au moment où elle aurait du être le plus utile (Page and Taborsky (2011)).

Cette thèse cherche à lier les cycles économiques et la gestion de portefeuille. Le premier chapitre construit un cadre théorique entre les cycles économiques et les primes de risques. Il met en évidence l'importance des points de retournement du cycle de croissance, plus connu sous le nom d'écart de production. Les deux chapitres suivants ont pour objectif de détecter en temps réel ces points de retournement. La première approche se concentre sur une méthode non paramétrique d'apprentissage automatique simple et facilement compréhensible appelée quantification vectorielle adaptative. La seconde approche utilise des méthodes plus complexes d'apprentissage automatique, dites ensemblistes : les forêts aléatoires et le boosting. Les deux démarches permettent de créer des stratégies d'investissement performantes en temps réel. Enfin, le dernier chapitre élabore une méthode d'allocation

d'actifs à partir de différents algorithmes de regroupement hiérarchique. Les résultats empiriques démontrent l'intérêt de cette tentative : les portefeuilles créés sont robustes, diversifiés et lucratifs.

Premier chapitre: Primes de risques et cycles économiques

Il est de coutume de dire que les classes d'actifs sont corrélées au cycle d'affaires (business cycle) ((Cochrane (2016))). L'objectif de ce chapitre est de démontrer que les actifs ne sont pas corrélés aux cycles économiques mais bien causés par ces derniers. Pour ce faire, il faut définir précisément les cycles économiques.

De manière simple, le cycle des affaires est le cycle du niveau de l'activité, tel que défini par Burns and Mitchell (1946). Ses points de retournement opposent les périodes de croissance négative (ou récessions) aux périodes de croissance positive (ou expansions). Suivant cette définition du cycle, les pics et les creux correspondent donc aux entrées et sorties de récession. Ce cycle des affaires est caractérisé par un mouvement commun à des nombreuses variables économiques et par la présence de fortes asymétries selon la phase du cycle. Historiquement, le cycle de croissance trouve son origine au NBER dans un article écrit par Mintz (1974) dans lequel l'auteur introduit le concept de cycle de déviation (deviation cycle) qui mesure l'écart à un instant donné entre la variable censée représenter l'évolution globale de l'économie, en général le PIB, et sa tendance de long terme. La tendance de long terme peut être vue comme la croissance tendancielle ou potentielle selon les appellations. Le pic de ce cycle correspond au moment où le taux de croissance repasse en dessous du taux de croissance tendanciel (l'output gap est alors à un maximum) et, par symétrie, le creux représente le moment où il repasse au-dessus (l'output gap atteint alors un minimum). L'OCDE a popularisé cette notion de cycle de croissance à travers la diffusion de ses indicateurs avancés composites (Composite Leading Indicators), toujours calculés par l'institution internationale depuis 1981, et visant à anticiper en temps réel les pics et les creux du cycle de croissance.

L'approche cyclique peut être raffinée en considérant à la fois le cycle d'affaire et le cycle de croissance. En effet, il existe une chronologie naturelle entre les points de retournement de ces différents cycles et l'analyse conjointe des cycles économiques selon l'approche ABCD développée par Anas and Ferrara (2004) permet une analyse plus fine du lien entre les classes d'actifs et l'économie, principalement pour les actifs risqués.

La relation théorique entre les cycles économiques et les primes de risques est clairement expliquée à partir de deux notions usuelles : le PIB nominal et les anticipations adaptatives.

Le prix d'une action ou la valeur d'une entreprise est la somme de tous les paiements de dividendes futurs actualisés à leur valeur actuelle. En d'autres termes, les investisseurs achètent une action pour son potentiel de bénéfices futurs. Puisque la somme de tous les revenus gagnés dans une économie chaque année est le PIB nominal, le prix des actions devrait donc être lié aux anticipations de croissance nominale (et non aux anticipations de croissance réelle). Lorsque le taux de croissance réel est supérieur à son potentiel, les pressions inflationnistes augmentent: le taux de croissance nominal de l'économie augmente. Les anticipations adaptatives impliquent que les anticipations de taux de croissance nominales devraient augmenter, ce qui équivaut à dire que les anticipations de revenus devraient augmenter. En théorie, les actions devraient donc bien performer lors des phases d'accélération et souffrir lors des ralentissements. Puisque les ralentissements signalent une période prolongée de croissance économique modérée, mais pas nécessairement un déclin absolu de l'activité économique, les performances des actions peuvent donc être négatives lorsque les taux de croissance réels sont positifs.

En théorie, les obligations d'État devraient bien performer pendant les ralentissements et les récessions. En effet, la théorie des anticipations de la structure par terme considère que le taux d'intérêt à long terme est une moyenne pondérée des taux d'intérêt actuels et attendus plus une prime de terme (cette dernière capte la compensation nécessaire pour supporter le risque de taux). Si le taux de croissance de l'activité est inférieur au taux de croissance potentiel, les pressions inflationnistes reculent et la banque centrale est plus susceptible de réduire les taux. En conséquence, les investisseurs devraient prévoir une trajectoire inférieure des taux d'intérêt à court terme futurs. Les taux à long terme devraient donc diminuer (et les prix des obligations augmenter).

Des stratégies dynamiques d'allocation d'actifs servent à prouver l'intérêt théorique de l'approche et mettent en évidence l'importance des points de retournement du cycle de croissance.

De plus, ce cadre théorique d'analyse des marchés permet une amélioration concrète de l'élaboration d'une allocation stratégique.

Second et troisième chapitres: Détection des points de retournement du cycle de croissance grâce à des méthodes d'apprentissage automatique

L'objectif est de détecter en temps réel les points de retournement du cycle de croissance afin de pouvoir intelligemment réallouer son portefeuille, à savoir changer la pondération de chaque actifs dans son portefeuille. Or, la principale caractéristique des cycles économiques est leur aspect non-linéaire : un choc dans un petit secteur de l'économie (l'immobilier par exemple) peut avoir des conséquences bien plus importantes pour l'économie dans son ensemble, voire même provoquer des récessions. Les méthodes statistiques envisagées doivent donc pouvoir prendre en compte ces non-linéarités. Les méthodes paramétriques traditionnellement utilisées sont les modèles à changement de régimes markoviens (Hamilton (1990)) ou les modèles de régression à seuils (Ferrara and Guegan (2005)). Giusto and Piger (ming) ont démontré que ces méthodes fonctionnaient très bien si le « Data Generating Process » était connu, autrement ils préconisent des méthodes dites non-paramétriques. Parmi ces dernières, les méthodes d'apprentissage automatique, machine learning en anglais, semblent une piste de recherche prometteuse. Certains auteurs commencent même les appliquer pour détecter les récessions aux Etats-Unis (Giusto and Piger (ming), Berge (2015), Ng (2014)), mais aucune application concerne la zone euro et encore moins le cycle de croissance.

A la maîtrise de ces techniques s'ajoute la nécessaire compréhension des interconnexions dynamiques entre les marchés financiers et la macroéconomie afin de faire un choix judicieux de variables explicatives. En effet, un adage en statistique computationnelle dit : «poubelle en entrée, poubelle en sortie », c'est-à-dire que même le meilleur des algorithmes ne pourra pas produire de résultats de qualité s'il est « nourri » par des données qui n'ont pas de sens. La difficulté est de trouver des variables économiques fiables, disponibles rapidement et peu révisées. Les données d'enquêtes et les données de marchés semblent des candidats naturels.

Le second chapitre se concentre sur une méthode non paramétrique d'apprentissage automatique simple et facilement compréhensible appelée quantification vectorielle adaptative (Kohonen (2001)). Le but est de se concentrer sur un modèle avec peu de variables explicatives, car seuls ces modèles peuvent convaincre un dirigeant ou un économiste de les suivre.

Le troisième chapitre applique des méthodes plus complexes d'apprentissage

automatique, dites ensemblistes : les forêts aléatoires (Breiman (2001)) et le boosting . Ces deux algorithmes ont la particularité d'effectuer de concert la sélection des variables et l'estimation des modèles (Schapire (1990)). Le désavantage est que ces modèles sont des boîtes noires et ne permettent pas une interprétation des modèles.

En plus des critères statistiques standards de comparaison de modèles de prévision (*QPS*, *AUROC*), des stratégies d'investissement sont mises en place. En effet, bien prévoir ne signifie pas automatiquement gagner de l'argent (Cenesizoglu and Timmermann (2012) or Brown (2008))

Enfin, pour tester le bien fondé des innovations précitées, une méthode de sélection de modèles sera nécessaire. L'emploi d'un même historique pour tester les différentes options engendre des difficultés statistiques. En effet, une avancée peut apparaître bénéfique par chance (résultat positif erroné), phénomène appelé *data snooping* dans la littérature académique (White (2000)). Pour palier ce problème, la procédure dite "Model Confidence Set" (Hansen et al. (2011)) est appliquée.

L'efficacité des deux différentes approches pour détecter les points de retournements en temps réel est démontrée. Il est alors possible de créer des stratégies d'investissement performantes à partir des modèles développés dans le second et troisième chapitre, conformément au cadre théorique développé dans le premier chapitre.

Quatrième chapitre: Allocation d'actifs à partir d'algorithmes de regroupement hiérarchique

Enfin, les études démontrent que méthodes usuelles d'allocation d'actifs, créées dans le cadre moyenne-variance de Markowitz, ne permettent pas d'obtenir des résultats satisfaisants « out-of-sample ».

Les méthodes d'optimisation de portefeuille modernes et complexes sont optimales "in-sample", mais ne permettent pas d'obtenir des résultats satisfaisants « out-of-sample ». Par exemple, DeMiguel et al. (2009) démontre que l'allocation naïve, qui donne la même importance à chaque actif, bat l'ensemble des techniques d'optimisation de portefeuille couramment utilisées.

López de Prado (2016) fait remarquer que ces méthodes font fi de la notion de hiérarchie, ce qui permet aux poids de varier librement. Or, le prix Nobel Herbert Simon a démontré que les systèmes complexes peuvent être disposés dans une hiérarchie naturelle , comprenant des sous-structures imbriquées (Simon (1962)): textit "le thème central qui traverse mes remarques est que la complexité prend souvent la forme de la hiérarchie, et que les systèmes

hiérarchiques ont des propriétés communes qui sont indépendantes de leur contenu spécifique La hiérarchie, dirais-je, est l'un des schémas structurels centraux que l'architecte de la complexité utilisée”.

Sur la base de la notion fondamentale de hiérarchie, le quatrième chapitre propose une méthode d'allocation d'actifs à partir de différents algorithmes de regroupement hiérarchique. Pour ce faire, les actifs sont dans un premier temps partitionnés, puis une pondération hiérarchique des actifs est réalisée. Une solution simple d'allocation est proposée (allocation hiérarchique naïve) dans la veine des travaux de DeMiguel et al. (2009).

Les performances des différentes méthodes (traditionnelles et hiérarchique) sont évaluées sur trois ensembles de données disparates, qui diffèrent en termes de nombre d'actifs et de composition de l'univers. Les résultats empiriques indiquent que les portefeuilles formées à partir de la classification hiérarchique sont robustes, vraiment diversifiés et atteignent des performances ajustées au risque statistiquement meilleures que les techniques d'optimisation de portefeuille communément utilisées.

References

- Anas, J. and Ferrara, L. (2004). Detecting Cyclical Turning Points: The ABCD Approach and Two Probabilistic Indicators. *Journal of Business Cycle Measurement and Analysis*, 2004(2):193–225.
- 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.
- Brown, S. (2008). Elusive return predictability: Discussion. *International Journal of Forecasting*, 24(1):19–21.
- Burns, A. and Mitchell, W. (1946). *Measuring Business Cycles*. Number burn46-1 in NBER Books. National Bureau of Economic Research, Inc.
- 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.
- Cochrane, J. H. (2016). Macro-Finance. NBER Working Papers 22485, National Bureau of Economic Research, Inc.

- DeMiguel, V., Garlappi, L., and Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the $1/n$ portfolio strategy? *The Review of Financial Studies*, 22(5):1915–1953.
- Ferrara, L. and Guegan, D. (2005). Detection of the industrial business cycle using setar models. *Journal of Business Cycle Measurement and Analysis*, 2005(3):353–371.
- Giusto, A. and Piger, J. (forthcoming). Identifying business cycle turning points in real time with vector quantization. *International Journal of Forecasting*.
- Hamilton, J. (1990). Analysis of time series subject to changes in regime. *Journal of Econometrics*, 45(1-2):39–70.
- Hansen, P., Lunde, A., and Nason, J. (2011). The model confidence set. *Econometrica*, 79(2):453–497.
- Kohonen, T. (2001). *Self-Organizing Maps*. Springer.
- López de Prado, M. (2016). Building diversified portfolios that outperform out of sample. *The Journal of Portfolio Management*, 42(4):59–69.
- Markowitz, H. (1952). Portfolio Selection. *Journal of Finance*, 7(1):77–91.
- Mintz, I. (1974). Dating united states growth cycles. In *Explorations in Economic Research, Volume 1, Number 1*, pages 1–113. National Bureau of Economic Research, Inc.
- Ng, S. (2014). Viewpoint: Boosting recessions. *Canadian Journal of Economics*, 47(1):1–34.
- Page, S. and Taborsky, M. (2011). The myth of diversification : risk factors versus asset classes. *The journal of portfolio management*, 37(4):1–2.
- Schapire, R. E. (1990). The strength of weak learnability. In *Machine Learning*, pages 197–227.
- Simon, H. A. (1962). The architecture of complexity. In *Proceedings of the American Philosophical Society*, pages 467–482.
- White, H. (2000). A Reality Check for Data Snooping. *Econometrica*, 68(5):1097–1126.

Résumé

Cette thèse cherche à lier les cycles économiques et la gestion de portefeuille. Le premier chapitre construit un cadre théorique entre les cycles économiques et les primes de risques. Il met en évidence l'importance des points de retournement du cycle de croissance, plus connu sous le nom d'écart de production. Les deux chapitres suivants ont pour objectif de détecter en temps réel ces points de retournement. La première approche se concentre sur une méthode non paramétrique d'apprentissage automatique simple et facilement compréhensible appelée quantification vectorielle adaptative. La seconde approche utilise des méthodes plus complexes d'apprentissage automatique, dites ensemblistes : les forêts aléatoires et le boosting. Les deux démarches permettent de créer des stratégies d'investissement performantes en temps réel. Enfin, le dernier chapitre élabore une méthode d'allocation d'actifs à partir de différents algorithmes de regroupement hiérarchique. Les résultats empiriques démontrent l'intérêt de cette tentative : les portefeuilles créés sont robustes, diversifiés et lucratifs.

Mots Clés

Cycles économiques, allocation d'actifs, apprentissage automatique

Abstract

A well-worked theory of macro-based investment decision is introduced. The theoretical influence of economic cycles on time-varying risk premiums is explained and exhibited. The importance of the turning points of the growth cycle, better known as the output gap, is outlined. To quickly and accurately detect economic turning points, probabilistic indicators are first created from a simple and transparent machine-learning algorithm known as Learning Vector Quantization. Those indicators are robust, interpretable and preserve economic consistency. A more complex approach is then evaluated: ensemble machine learning algorithms, referred to as random forest and as boosting, are applied. The two key features of those algorithms are their abilities to entertain a large number of predictors and to perform estimation and variable selection simultaneously. With both approaches investment strategies based on the models achieve impressive risk-adjusted returns: timing the market is thus possible. At last, exploring a new way of capital allocation, a hierarchical clustering based asset allocation method is introduced. The empirical results indicate that hierarchical clustering based portfolios are robust, truly diversified and achieve statistically better risk-adjusted performances than commonly used portfolio optimization techniques.

Keywords

Economic cycles, asset allocation, machine learning