Essays on Financial Forecasting and Risk Assessment

by

Sotiria N. Plastira

B.A. in Physics (2002) - National and Kapodistrian University of Athens
M.Sc. in Electronics & Radioelectrology (2005) - National and Kapodistrian University of Athens
M.Sc. in Actuarial Science and Risk Management (2009) - University of Piraeus

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Financial Econometrics

in the

GRADUATE DIVISION
of the
UNIVERSITY OF PIRAEUS, DEPARTMENT OF STATISTICS AND INSURANCE SCIENCE

Committee in charge:
Associate Professor Efstathios Chatzikonstadinidis, Chair
Professor Nikitas Pittis
Professor Markos Koutras

Summer 2014

Essays on Financial Forecasting and Risk Assessment

by

Sotiria N. Plastira

Abstract

Essays on Financial Forecasting and Risk Assessment

by

Sotiria N. Plastira

Doctor of Philosophy in Financial Econometrics

University of Piraeus, Department of Statistics and Insurance Science

Associate Professor Efstathios Chatzikonstadinidis, Chair

This thesis aims at investigating the performance of empirical risk factors in financial forecasting and their assessment with respect to the associated risk. The thesis consists of four essays. The first essay focuses on whether the empirical HML and SMB risk factors, along with the long-term reversal and the momentum factors exhibit both in-sample and out-of-sample forecasting ability for the U.S. stock returns, compared to the performance of the most widely used financial variables. Our findings point to the superior forecasting ability of the empirical factors. We also establish a link between financial variables and the empirical factors and find that the default and the term spread proxy for the evolution of the factors examined. The second essay extends the previous analysis by investigating the out-of-sample forecasting ability of the full set of empirical factors along with their size and value decompositions on U.S. bond and stock returns for a variety of horizons ranging from the short run (1 month) to the long run (2 years). We also examine their performance by

employing combination of the individual forecasts of the empirical factors. It turns out that these combining methods lead to particularly successful results, especially from an asset allocation perspective, with similar findings pertaining to the European and Japanese markets, as well. The third essay employs a variety of risk indices in order to quantify the embedded risk of different empirical factor portfolios, producing a relative ranking among them. The analysis also contributes to the literature by establishing a connection between size, book-to-market and stock prior-returns with risk, revealing that small size, high book-to-market and low momentum/reversal effect are related with high portfolio risk. Finally, the fourth essay provides an extensive review on traditional and more sophisticated evaluation measures focusing on premium returns adjusted for the associated risk. The implementation of these performance measures on the aforementioned empirical factors reveals that the value and momentum factor portfolios achieve the best and worst performance, respectively.

To all those who encouraged me to continue

Contents

List of Tables

Acknowledgments

Foremost, I would like to express my deep sense of gratitude to my Ph.D. supervisor (for the period 2011-2014), Assistant Professor Ekaterini Panopoulou, who has patiently and tirelessly assisted during the first three years of my dissertation and encouraged me to follow through this thesis. I would also like to thank her for her continuous intellectual clarity, assistance and encouragement, as well as, her willingness to share some of her ideas with me. Research is rarely undertaken alone. Her knowledge, guidance and support have significantly contributed to this thesis, as well as, to my overall academic development. I really appreciate her for the time and effort she invested for me and also for offering to me the chance to work as a teaching assistant for both undergraduate and postgraduate courses at the Department. Thanks to her encouragement, I had the opportunity to present parts of my thesis to various International Conferences that took place either in Greece (2nd National Conference of the F.E.B.S., 2011; 3rd National Conference of F.E.B.S., 2012; 27th Panhellenic Conference in Statistics, 2014; International Conference on Applied Economics, 2014) or abroad (5th International CFE'11, UK; 6th CSDA International Conference on CFE'12, Spain).

I am also grateful to Professor Efstathios Chatzikonstadinidis, my Ph.D. supervisor during the last year of this dissertation, for his support and guidance. I really appreciate his willingness to support my effort. Furthrmore, I would like to thank the members of my supervisory committee: Professor Nikitas Pittis and Professor Markos Koutras for careful reading and providing helpful comments on my manuscripts and also to the Department of Statistics and Insurance Science and all its faculty for providing me an appropriate research environment from the very early stages of my research. I would like to thank the Department's current Ph.D. students for the valuable discussions and support, with special thanks to Christos Argyropoulos for his support.

I am deeply grateful to my parents, Nikolaos and Ekaterini, and my sister, Stavroula, for their encouragement to follow my dreams and accomplish my goals, and also to my little Mocca for being the best company.

Last, but not least, I would like to express deepest thanks to all my friends that supported me in every difficult situation and, especially, to my beloved friend, Kiriaki Methiniti, for the constant understanding that she provided me all these years.

Financial assistance from the Research Centre of the University of Piraeus and the National Bank of Greece, under the proposition of Mr. Maligiannis George, for my attendance to International Conferences is greatly appreciated.

Chapter 1

Introduction and Overview

Evaluating the performance of an investment or being able to create reliable out-of-sample forecasts for financial time series is of utmost importance in the finance area. The widely used Capital Asset Pricing Model (CAPM) of Sharpe (1964) and Lintner (1965) leads to predictions concerning the systematic risk of an asset or a portfolio of assets with respect to the market portfolio. In other words, differences in the sensitivity of assets on the market return should explain differences in average asset returns. However, empirically the single-factor CAPM has not been very successful. This led to the development of multi-factor models aiming at explaining the cross-sectional variation of returns.

Fama and French (1992) show that size and book-to-market have explanatory power with respect to the cross-section of average returns on NYSE, Amex and NASDAQ stocks for the 1963-1990 period. Extending this analysis, Fama and French (1993) provide evidence that size and book-to-market equity proxy for the sensitivity to common risk factors in stock returns and are both related to profitability. By introducing the three-factor model, they capture the related CAPM average-return anomalies. In particular, the negative relation between size and average returns is explained by a risk factor associated with size, while the positive relation between book-to-market and average returns is attributed to another risk factor. The former is referred as the size premium or 'Small Minus Big' (SMB) and is the return on a portfolio that is long in stocks with small capitalization and short in stocks with big capitalization, while the latter is referred as the value premium or 'High Minus Low' (HML) and is the return on a portfolio that is long in stocks with high book-to-market (value stocks) and short in stocks with low book-to-market equity ratio (growth stocks).

Another market 'anomaly' was examined by Jegadeesh and Titman (1993),who examined a variety of momentum strategies and suggested that holding stocks with high returns over the previous 3 to 12 months and selling those with poor performance over the same period time leads to profits of about 1% per month for the following year. According to the authors' momentum strategies, past winners perform better than past losers by about the same magnitude as in the earlier period. In this context, Carhart (1997) proposed the four-factor model by incorporating the momentum (MOM) factor that is the difference of the monthly return between the high and low prior return portfolios.

While a vast literature exists on these anomalies with respect to the underlying explanations for these return premia and their link to systematic risk, the evidence on their forecasting ability for future returns is quite scarce. The aim of Chapters 2 and 3 is to assess, not only statistically but also from an asset allocation perspective, the ability of the famous empirical risk factors to provide efficient forecasts on stock and bond returns. The embedded risk in these empirical factors and their risk-adjusted performance is evaluated in Chapters 4 and 5. More in detail, this thesis is structured as follows.

Chapter 2 investigates whether the HML, SMB along with the long-term reversal and the momentum factors exhibit both in-sample and out-of-sample forecasting ability for U.S. stock returns for a variety of horizons ranging from the short-run to the long-run. Our analysis reveals that the empirical factors contain significantly more information for future stock market returns than the typically employed financial variables. Going one step further, we test whether the employed financial variables can proxy for the aforementioned factors and we find that the default spread and, to a lesser extent, the term spread contain important information for the evolution of the factors examined. Our analysis also sheds light on the source of this forecasting ability by investigating whether the evinced forecasting ability is attributed to either the big or the small factor components or to appropriate value counterparts. Our findings suggest that specific factor decompositions improve the forecasting ability of the model over the benchmark model.

Chapter 3 extends the analysis of Chapter 2 by investigating the predictive ability of the empirical factors not only on stock returns, but also on government bond returns. We also incorporate in the list of the empirical factors the short-term reversal factor along with its size decompositions and evaluate our findings from an asset allocation perspective. Consistent with the results of Chapter 2, the empirical factors outperform the typically employed financial variables by containing significantly more information for future U.S. bond and stock market return. This analysis also relates to the literature of combination forecasts by investigating whether combination of the empirical factors' forecasts can enhance the model's forecasting ability on stock and bond returns. Both the statistical and economic evaluation findings suggest that the empirical factors lead to significant performance fees that an investor would be willing to pay in order to have access to the information offered by our modelling approach. More importantly, similar findings pertain for the European and Japanese markets.

Chapter 4 attempts a thorough risk analysis of these empirical factor portfolios. The risk measures employed in our analysis cover not only the traditional ones, such as the standard deviation and the beta factor, but also more sophisticated ones, such the Value-at-Risk (VaR), the expected shortfall, and also measures based on downside indices or drawdown-based ones, capturing thus the kurtosis or skewness of the series distribution and other aspects of the associated risk. Implementing these risk measures on the empirical HML and SMB factors and also on the momentum and reversal ones; namely, the MOM, LT-Rev and ST-Rev, respectively, we produce a risk-based ranking for these significant predictors. Our results reveal that the MOM factor portfolios is ranked high, while the HML and LT factors are related with low risk. Both the traditional and downside indices lead to identical rankings, while drawdown-based measures reveal a different rank order. We also establish a connection between specific stock characteristics and portfolios' risk, by assessing the risk of portfolios formed according to a specific characteristic, such as size, book-to-market or momentum. The empirical findings suggest that portfolios of small stocks or high book-to-market are ranked high, while high momentum and reversal portfolios are associated with lower level of risk. This finding is robust to the measure employed revealing that an investor can decide on the most or least risky investment opportunity independently of the employed risk index.

Chapter 5 provides an extensive review on traditional and more sophisticated performance evaluation measures focusing on premium returns adjusted for the associated risk. By quantifying the portfolio risk via the plethora of risk measures presented in the previous Chapter, we implement the performance measures on the empirical portfolios, producing their rank order with respect to their risk-adjusted performance. Our results reveal that the HML and MOM factor portfolios achieve the best and worst performance, respectively, while portfolios based on stock prior-returns underperform. Similar to the prior Chapter, we investigate the connection between the specific stock characteristics (size, book-to-market and momentum) and the respective portfolio performance. Our findings indicate that portfolios of small stocks or high book-to-market are ranked high, while low momentum and high reversal portfolios are associated with lower performance.

Finally, Chapter 6 draws some conclusions and discusses some open issues that are likely to attract research interest in the future.

Chapter 2

Fama French Factors and US Stock Return Predictability

2.1 Introduction

A series of papers by Fama and French (1993, 1995, 1996) suggest that the Capital Asset Pricing Model (CAPM) fails to capture the cross-sectional variation of average stock returns. In this respect, the authors propose a three-factor model, according to which the expected return on a portfolio in excess of the risk-free rate is explained by three factors; namely, the excess return on the market portfolio, the return on a portfolio long in small stocks and short in big stocks (SMB), and the return on a portfolio long in high book-to-market stocks and short in low book-to-market stocks (HML). SMB is often referred to as the size premium, while HML as the value premium. Both the value and size premia appear to be priced risk factors omitted from the CAPM and are rather pervasive in major international equity markets. Fama and French (1993) show that size and book-to-market can proxy for the sensitivity of stock returns to common risk factors and as a result their model performs well on portfolios formed on size and book-to-market equity. Moreover, Fama and French (1995) provide evidence that size and book-to-market are related to profitability and argue that their three-factor model is consistent with Merton's (1973) Intertemporal Capital Asset-Pricing Model (ICAPM), in which size and book-to-market proxy for the sensitivity to risk factors in returns. In a subsequent paper, Fama and French (1996) show that their model captures priced default risk, and, as a result, can explain equity returns. Carhart (1997) proposes a four-factor model by adding a momentum factor in the three-factor Fama-French model. This momentum factor (MOM) is the difference between returns on portfolios of the winners and losers over the past year. These empirical factors, namely SMB, HML and MOM, are often referred as market anomalies and have motivated the use of empirical asset pricing models that incorporate their returns.[1]

The vast literature on these anomalies has generated a wide debate with respect to the underlying explanations for these return premia and their link to systematic risk. Chan *et al.* (1998), in an effort to identify the factors that capture systematic return covariation in stock returns, provide evidence that with the exception of the Fama-French factors, only the default premium and the term premium can explain return covariation. Liew and Vassalou

[1]The size anomaly reflects the empirical finding that small stocks (low market capitalization) outperform large stocks (high market capitalization), even after adjusting for market exposure (Banz, 1981; Fama and French, 1992). Likewise, the value anomaly relates to the outperformance of value stocks (stocks with high ratios of fundamental or book value to market value such as book-to-market equity, cash flow-to-price, or earnings-to-price ratios) over growth stocks, which have low book-to-price ratios (see among others DeBondt and Thaler, 1985; Fama and French, 1992; and Lakonishok *et al*, 1994). Moreover, positive momentum exists in stock returns. Stocks that have performed well relative to other stocks over the past (typically last six months to a year) continue to perform well over the future (next six months to a year), and vice versa (Jegadeesh and Titman, 1993, 2001; Fama and French, 1996).

(2000) investigate the extent to which the profitability of the HML and SMB factors can be linked to future economic growth and conclude that, indeed, the hypothesis of Fama and French (1993, 1995, 1996) is supported across various markets. Going one step further, Vassalou (2003) provides an economic interpretation and concludes that the HML and SMB factors include information related to news about future economic growth. Petkova (2006) shows that the same factors proxy for the term spread and default spread, respectively, thus establishing a link between a set of variables associated with time-series return predictability and a set of variables associated with cross-sectional return predictability. Similarly, Hahn and Lee (2006) find that changes in the default spread and the term spread capture the cross-sectional pattern of stock returns in size and book-to-market. The degree to which these factors are linked to the state variables over various time scales is examined by In and Kim (2007), who conclude that both SMB and HML play a limited role in the short run, but the opposite takes place in the long run. Recently, Vivian and Wohar (2013) employ the output gap, a key business cycle indicator and find that it cannot predict the value effect (HML) either in-sample or out-of-sample, while there is some evidence of out-of-sample predictability of the size effect (SMB).

In this paper, we assess both the in-sample and out-of-sample forecasting ability of these empirical factors for US stock returns, namely the CRSP value-weighted portfolio return. Our set of factors contains the value premium (HML), the size premium (SMB), the momentum (MOM) and the long-term reversal (LT) which is a factor related to long-term (1- 5 years) past performance and can be thought of as a value indicator (DeBondt and Thaler, 1985, 1987; Fama and French, 1996).[2] Our set of predictors is enriched with four

[2] Asness *et al* (2013) find that stock portfolios created from past five year returns display an average

financial variables that are typically employed in the return predictability literature; namely, the 1-month Treasury bill rate (Fama and Schwert, 1977; Campbell, 1991; Hodrick, 1992), the term spread (Campbell, 1987; Fama and French, 1989), the corporate bond default spread (Keim and Stambaugh, 1986; Fama and French, 1989) and the market dividend yield (Fama and French, 1988, 1989; Campbell and Shiller, 1988b). We also examine whether any of the financial variables considered can proxy for the factors at hand. These issues are addressed for a variety of horizons ranging from the short-run (1 month) to the long-run (3 years) with the aim to reveal the term structure of predictability.

Our findings suggest that all the empirical factors exhibit considerable in-sample and out-of-sample forecasting ability for the value-weighted CRSP portfolio returns at specific horizons, thus establishing the link between the time-varying investment opportunity set and the factors. With respect to the financial variables, only the term spread exhibits significant out-of-sample forecasting ability, while the remaining ones only improve in-sample forecasts. Investigating whether any financial variable can act as a proxy for the aforementioned factors, we, indeed, find that there is a link between them, with the default spread being the most important proxy. We also shed light on the source of the predictive ability of each factor by decomposing them into size and value components.[3]

The remainder of the paper is organized as follows. Section 2 describes in detail the econometric methodology employed and Section 3 presents the data and factor decomposition. Section 4 presents our in-sample predictability findings and Section 5 presents our out-of-sample results. Section 6 summarizes and concludes.

---

correlation of 86% with portfolios created from other value measures such as book-to-market.

[3]Fama and French (2012) consider the size components of international HML and MOM portfolios and find that value and momentum premiums decrease with size (except for Japan).

2.2 Econometric methodology

Following Rapach and Weber (2004), the predictive ability of factors and financial variables is evaluated by means of the following predictive AutoRegressive Distributed Lag (ARDL) model:

$$z_{t+h} = a + \sum_{i=0}^{q_1-1} \beta_i \Delta y_{t-i} + \sum_{i=0}^{q_2-1} \gamma_i x_{t-i} + \epsilon_{t+h} \tag{2.1}$$

where $z_{t+h} = \sum_{i=1}^{h} \Delta y_{t+i}$ is the return to be predicted from period t to $t+h$ with h the forecast horizon, x_t the candidate predictor variable, $\Delta y_t = y_t - y_{t-1}$ the one-period return at time t, ϵ_{t+h} the disturbance term, a the intercept, q_1 and q_2 the data-determined lag orders for Δy_t and x_t.[4] A heteroscedasticity and autocorrelation-consistent (HAC) covariance matrix should be employed when multi-step forecasts are concerned, i.e. $h > 1$, since the returns z_{t+h} overlap and this induces serial correlation to the disturbance term (Newey and West, 1987).

In order to test the in-sample forecastability of variables, we employ the whole sample and conduct a Wald test for the null hypothesis that $\gamma_0 = ... = \gamma_{q_2-1} = 0$. If the null hypothesis cannot be rejected at the desirable significance level, the variable employed does not have any forecasting ability. In order to study the out-of-sample forecasting ability, the total sample T is divided into the first R in-sample observations and the last P out-of-sample observations. In order to create the first out-of-sample forecast, we make use of the in-sample portion of the sample and estimate the OLS parameters

[4] The maximum lag value is set at 8 and is selected by means of the SIC criterion.

a, β_i and γ_i of the ARDL equation via the method of ordinary least squares (OLS) for the unrestricted form of the model, $\widehat{a}_{1,R}$, $\widehat{\beta}_{1,R,i}$, $\widehat{\gamma}_{1,R,i}$. Then, the estimated equation $\widehat{z}_{1,R+h} = \widehat{a}_{1,R} + \sum_{i=0}^{q_1-1} \widehat{\beta}_{1,R,i} \Delta y_{R-i} + \sum_{i=0}^{q_2-1} \widehat{\gamma}_{1,R,i} x_{R-i}$ creates the first out-of-sample forecast for the unrestricted form of the model, as well as, the forecast error $\widehat{u}_{1,R+h} = z_{R+h} - \widehat{z}_{1,R+h}$. Following the same procedure, we estimate the equation for the restricted form of the model: $\widehat{z}_{0,R+h} = \widehat{a}_{0,R} + \sum_{i=0}^{q_1-1} \widehat{\beta}_{0,R,i} \Delta y_{R-i}$, where $\widehat{a}_{0,R}$ and $\widehat{\beta}_{0,R,i}$ are the OLS parameter estimates, and compute the forecast error $\widehat{u}_{0,R+h} = z_{R+h} - \widehat{z}_{0,R+h}$. In order to create the next forecasts, we expand recursively the in-sample portion of the sample and repeat the whole procedure through the end of the available sample, generating totally $T-R-h+1$ out-of-sample forecast errors for the unrestricted and the restricted form of the predictive model, $\{\widehat{u}_{1,t+h}\}_{t=R}^{T-h}$ and $\{\widehat{u}_{0,t+h}\}_{t=R}^{T-h}$, respectively.

The variable x_t displays forecasting ability for the returns if the unrestricted model forecasts are superior to the restricted ones. A metric that is commonly used for this purpose is Theil's U, which is the ratio of the Mean Squared Forecast Error (MSFE) of the unrestricted model to the MSFE of the restricted one. When $U < 1$, the MSFE of the unrestricted model is less than the MSFE of the restricted model, suggesting that the candidate variable can improve forecasts. In order to statistically test the ability of a factor to improve the predictability of the ARDL model, we use the Diebold and Mariano (1995) and West (1996) t-statistic for equal MSFE, the MSE-T statistic, along with a variant of this statistic due to McCracken (2007), the MSE-F statistic. Both statistics test the null hypothesis that the unrestricted model MSFE is equal to the restricted model MSFE against the one-sided (upper-tail) alternative hypothesis that the unrestricted model MSFE is less than the restricted model MSFE.

The MSE-T and MSE-F statistics are expressed as follows:

$$MSE-T=(T-R-h+1)^{0.5}\overline{d}\widehat{S}_{dd}^{-0.5} \tag{2.2}$$

$$MSE-F=(T-R-h+1)\overline{d}/\widehat{MSFE}_1 \tag{2.3}$$

where $\overline{d}=(T-R-h+1)^{-1}\sum_{t=R}^{T-h}\widehat{d}_{t+h}=\widehat{MSFE}_0-\widehat{MSFE}_1$ is the mean loss differential, $\widehat{MSFE}_i=(T-R-h+1)^{-1}\sum_{t=R}^{T-h}\widehat{u}_{i,t+h}^2\ (i=0,1)$, $\widehat{d_{t+h}}=\widehat{u}_{0,t+h}^2-\widehat{u}_{1,t+h}^2$is the sequence of loss differentials, $\widehat{S}_{dd}=\sum_{j=-J}^{J}K(j/J)\widehat{\Gamma}_{dd}(j)$ is the long-run covariance matrix of $\widehat{d}_{t+h}$, $\widehat{\Gamma}_{dd}=(T-R-h+1)^{-1}\sum_{t=R+j}^{T-h}(\widehat{d}_{t+h}-\overline{d})(\widehat{d}_{t+h-j}-\overline{d})$ is the covariance of the loss differential $\widehat{d}_{t+h}$ at displacement j, $\widehat{\Gamma}_{dd}(-j)=\widehat{\Gamma}_{dd}(j)$, k is the number of lags. The estimator of the long-run covariance matrix of $\widehat{d}_{t+h}$, $\Omega=\lim_{j\to\infty}\sum_{-j}^{j}E(\widehat{d}_{t+h}\widehat{d}'_{t+h-j})$, is the kernel HAC estimator for Ω of the form $\widehat{S}_{dd}=\sum_{j=-J}^{J}K(j/J)\widehat{\Gamma}_{dd}(j)$. Following Clark and McCracken (2005), we use the Bartlett kernel $K(j/J)=1-[j/(J+1)]$ with bandwidth parameter $J=[1.5h]$ for $h>1$, where $[\cdot]$ is the nearest integer function.

McCracken (2007) shows that for nested models and for $h=1$, both statistics have a nonstandard asymptotic distribution, which is a function of stochastic integrals of quadratics of Brownian motion $W(\cdot)$ that depends on $\lim_{P,R\to\infty}P/R$. Moreover, Clark and McCracken (2001) show that when we focus on multi-step forecasts, i.e. $h>1$, the limiting distribution of the statistics is also nonstandard when comparing forecasts from nested models. In this case, unknown nuisance parameters exist in the limiting distribution and both the $MSE-T$ and $MSE-F$ statistics are not asymptotically pivotal. To overcome this, Clark and McCracken (2005) recommend the use of a bootstrap procedure, introduced by Kilian (1999), which enables us to calculate critical values that can yield accurate inferences, especially in the case of multi-step horizons.

An alternative way to evaluate forecasts is based on the notion of forecast encompassing. Let $\widehat{z}_{c,t+h}$ be a combination of the out-of-sample forecasts from the restricted ARDL model $\widehat{z}_{0,t+h}$ and those of the unrestricted model $\widehat{z}_{1,t+h}$ in an optimal way so that $\widehat{z}_{c,t+h} = \lambda\widehat{z}_{1,t+h} + (1-\lambda)\widehat{z}_{0,t+h}$, $0 \leq \lambda \leq 1$. If the optimal weight attached to the unrestricted model forecast is zero, $\lambda = 0$, then the restricted model forecasts encompass the competing unrestricted model forecasts. In this case we have $\widehat{z}_{c,t+h} = \widehat{z}_{0,t+h}$ from which it is obvious that only the restricted model is important. Transforming the equation $\widehat{z}_{c,t+h} = \lambda\widehat{z}_{1,t+h} + (1-\lambda)\widehat{z}_{0,t+h}$ into $\widehat{u}_{c,t+h} = \lambda(\widehat{u}_{0,t+h} - \widehat{u}_{1,t+h})$ by subtracting $\widehat{z}_{0,t+h}$ from both sides and substituting $\widehat{z}_{1,t+h} - \widehat{z}_{0,t+h} = \widehat{u}_{0,t+h} - \widehat{u}_{1,t+h}$, we conclude that when $\lambda = 1$, then the candidate variable does have predictive power and the covariance between $\widehat{u}_{0,t+h}$ and $\widehat{u}_{0,t+h} - \widehat{u}_{1,t+h}$ will be positive. If $\lambda > 0$, then not only the restricted model forecast, but also the unrestricted model forecast attributes information that is useful and important to the formation of the optimal composite forecast, and as a result the restricted model forecasts do not encompass the unrestricted model forecasts.

In order to test whether the restricted model forecasts encompass or not the unrestricted model forecasts, we employ two statistics; the ENC-T statistic proposed by Harvey *et al.* (1998) and a variant of ENC-T proposed by Clark and McCracken (2001), ENC-NEW. Both statistics test the null hypothesis of equal forecast accuracy or forecast encompassing, $\lambda = 0$, against the one-sided alternative (upper-tailed) hypothesis that $\lambda > 0$. They are calculated as follows:

$$ENC-T = (T-R-h+1)^{0.5}\bar{c}\widehat{S}_{cc}^{-0.5} \tag{2.4}$$

$$ENC-NEW = (T-R-h+1)\bar{c}/\widehat{MSFE}_1 \tag{2.5}$$

where $\bar{c} = (T-R-h+1)^{-1}\sum_{t=R}^{T-h}\widehat{c}_{t+h}$ is the mean of the sequence $\widehat{c}_{t+h}, \widehat{c}_{t+h} = \widehat{u}_{0,t+h}(\widehat{u}_{0,t+h}-\widehat{u}_{1,t+h}), \widehat{S}_{cc} = \sum_{j=-J}^{J}K(j/J)\widehat{\Gamma}_{cc}(j), \widehat{\Gamma}_{cc}(j) = (T-R-h+1)^{-1}\sum_{t=R+j}^{T-h}(\widehat{c}_{t+h}-\bar{c})(\widehat{c}_{t+h-j}-\bar{c})$, and $\widehat{\Gamma}_{cc}(-j) = \widehat{\Gamma}_{cc}(j)$. As previously, we employ the Bartlett kernel $K(j/J) = 1-[j/(J+1)]$ with a lag truncation parameter $J = [1.5h]$ for $h > 1$, where $[\cdot]$ is the nearest integer function. Clark and McCracken (2001) show that for nested models and $h = 1$, both the ENC-T and ENC-NEW statistics have a nonstandard limiting distribution, since the forecast errors for nested models are asymptotically the same and therefore perfectly correlated. Moreover, for $h > 1$ Clark and McCracken (2005) show that these statistics have a nonstandard asymptotic distribution and are not asymptotically pivotal. As in the case of the MSE-T and MSE-F statistics, Clark and McCracken (2005) recommend the use of a bootstrap procedure, which has been introduced by Kilian (1999). The bootstrapped critical values estimated seem to reflect the imprecision of the HAC variance that enters the test statistics, and according to Kilian (1999) this bootstrap method reduces the size distortions of conventional long-horizon regression tests on small samples.

Clark and McCracken (2001, 2005) show that the out-of-sample statistics have good size properties, when inference is based on a bootstrap procedure. The ENC-NEW statistic proves to be the most powerful among all with the ENC-T and MSE-F following, while the least powerful is the MSE-T statistic.

Next, we present the proposed by Killian (1999) bootstrap algorithm via which we compute the p-values to conduct properly the aforementioned test statistics.

2.2.1 Bootstrap algorithm

The procedure generates a pseudo-sample for the market returns, Δy_t, and the forecasting variables, x_t, of the same length as the original data series. We estimate by OLS the vector autoregressive equations for these series under the null hypothesis that the predictor exhibits no predictive power on the series of interest:

$$\Delta y_t = a_0 + a_1 * \Delta y_{t-1} + ... + a_{p1} * \Delta y_{t-p1} + e_{1,t} \quad (2.6)$$

$$x_t = b_0 + b_1 * \Delta y_{t-1} + ... + b_{p2} * \Delta y_{t-p2} + c_1 * x_{t-1} + ... + c_{p3} * x_{t-p3} + e_{2,t} \quad (2.7)$$

where the disturbance vector $e_t = (e_{1,t}, e_{2,t})'$ is independently and identically distributed with covariance matrix Σ.

The employed bootstrap algorithm consists of the following steps.

(1) Using the full sample of observations, we estimate the aforementioned equations via OLS, and compute the OLS residuals $\left\{\widehat{e}_t = (\widehat{e}_{1,t} - \widehat{e}_{2,t})'\right\}_{t=1}^{T}$. Lag orders are selected using the SIC criterion with maximum lag value set at 8.

(2) To generate a series of disturbances for the pseudo-sample, we randomly draw (with replacement) T+50 times from the OLS residuals, giving us a pseudo-series of disturbance terms $\{\widehat{e}_t^*\}_{t=1}^{T+50}$. We draw from the OLS residuals in tandem, preserving thus the contemporaneous correlation between the disturbances in the original data.

(3) Using the pseudo-series of disturbance terms, the OLS estimates of the aforementioned equations, and setting the initial lagged observations for the series employed

equal to zero, we create a pseudo-sample of $T+50+p$ observations, where $p = \max\{p1, p2, p3\}$:

$$\{\Delta y_t^*, x_t^*\}_{t=1}^{T+50+p} \tag{2.8}$$

The first 50 pseudo-observations are discarded as start-up transient observations, leaving us with a pseudo-sample of $T + p$ observations.

(4) The bootstrapped data are used to estimate the proposed forecasting models of our analysis (restricted and unrestricted). The resulting thus forecasts are used to calculate the test statistics employed in our analysis by incorporating the same methodology.

This procedure is repeated 500 times, giving us an empirical distribution for the employed statistics.

(5) For each statistic, we determine the p-value as the proportion of the sorted bootstrapped statistics that are greater than the statistic computed using the original sample.

2.3 Data, variables and factor decomposition

The data used in our analysis are monthly observations for the period from July 1963 to October 2009. The returns on the market portfolio (CRSP value-weighted portfolio return), the SMB (Small Minus Big), the HML (High Minus Low), the Long-Term Reversal (LT) and the Momentum (MOM) factors are taken from Kenneth French's website.[5] The SMB and HML factors are constructed from 6 value-weighted portfolios formed on size and book-to-market. Specifically, the intersections of the big/small and the value/neutral/growth portfolios form the 6 value-weighted portfolios, namely the small value (SV), small neutral

[5] This dataset can be downloaded from http://mba.tuck.dartmouth.edu/pages/faculty /ken.french/data_library.html

(SN), small growth (SG), big value (BV), big neutral (BN) and big growth (BG) portfolio. The breakpoint for year t for size is the median NYSE market equity at the end of June of year t, while for the book-to-market are the 30th and 70th NYSE percentiles. The book-to-market ratio for June of year t is the book equity for the last fiscal year end in t-1 divided by market equity for December of t-1. The portfolios for July of year t to June of t+1 include all NYSE, AMEX, and NASDAQ stocks for which we have market equity data for December of t-1 and June of t, and (positive) book equity data for t-1. The average return of the three small portfolios minus that of the three big portfolios forms the SMB portfolio, whereas the average return of the two value portfolios minus the average return of the two growth portfolios forms the HML portfolio. The LT and MOM factors are formed from 6 value-weighted portfolios formed on size and prior returns (small low, small medium, small high, big low, big medium, big high). These prior-return portfolios are constructed on prior (13-60) and (2-12) returns, respectively. The breakpoint for the equity is the median NYSE market equity, while for the prior returns are the 30th and 70th NYSE percentiles. The average return of the two low prior-return portfolios (big and small) minus the average return of the two high prior-return portfolios (big and small) forms the LT factor, while the MOM factor is the average return of the two high (big and small) prior-return portfolios, known as the winners, minus the average return of the two low (big and small) prior-return portfolios, known as the losers.

Following Fama and French (2012), who, among others, examine the effect of size on the value and momentum premium in international returns, we employ a similar decomposition of the aforementioned factors. More in detail, with the exception of SMB, we decompose all factors into their small and big counterparts. For example, the difference between the small (big) value portfolio and the small (big) growth one forms the HML_s (HML_b) portfolio as follows:

$$\begin{aligned} HML\_s &= SV - SG \\ HML\_b &= BV - BG \end{aligned} \tag{2.9}$$

In a similar manner, we form the decompositions of the LT and MOM factors as follows:

$$LT\_s = SL - SH, \quad LT\_b = BL - BH \tag{2.10}$$

$$MOM\_s = SH - SL, \quad MOM\_b = BH - BL \tag{2.11}$$

Going one step further, we test the effect of value on the size premium by decomposing the SMB factor in its three counterparts, namely value, growth and neutral. Specifically, the difference between the small value (neutral, growth) and the big value (neutral, growth) portfolio forms the SMB_v (SMB_n, SMB_g) portfolio as follows:

$$\begin{aligned} SMB\_v &= SV - BV \\ SMB\_n &= SN - BN \\ SMB\_g &= SG - BG \end{aligned} \tag{2.12}$$

With respect to financial variables, we employ four financial variables which have been typically employed in the return prediction literature and can be thought as state variables in the context of ICAPM (Petkova, 2006; Maio and Santa-Clara, 2012). These are the 1-month T-bill rate, the term spread, the default spread and the dividend yield. The term spread is the difference between the yields of the 10-year and the 1-year government bond, while the default spread is the difference between the yields of a long-term corporate Baa bond and a long-term (10-year) government bond. Data on bond yields are from the FRED database of the Federal Reserve Bank of St. Louis, while the 1-month T-bill rate is from Ibbotson and Associates Inc., available at Kenneth French's website. The data for the dividend yield come from Goyal and Welch (2008).[6] Specifically, the dividend yield ratio is the ratio of the dividends paid on the S&P 500 index and lagged stock prices (S&P 500 index).

2.4 In-sample predictive ability

2.4.1 In-sample predictive ability of the Fama French, reversal and momentum factors

Employing the predictive ARDL model (Eq. 2.1), we test the in-sample predictive ability of each of the factors on the CRSP value-weighted portfolio return for horizons of 1-36 months. Our results are reported in Table 2.1. Bold entries indicate significance at the 10% level based on the Wald test and bootstrapped critical values.

We begin our analysis with the HML factor, or else the value premium, which is

[6] This set of data can be downloaded from www.bus.emory.edu/AGoyal/Research.html

often associated with time-varying investment opportunities (see for example Petkova and Zhang, 2005; Guo *et al.*, 2009).[7] The in-sample predictive ability of this factor is mainly short-run and specifically at horizons less than a year (3-4 and 6-9 months). Decomposing the value premium into its size components, we note that this predictability almost vanishes as only the small value premium (HML_s factor) exhibits some significant in-sample predictive ability for the very short-run, namely horizons of 1 and 2 months.

Turning to the size premium, SMB factor, we find that its forecasting ability is quite limited to horizons of 1 and 5 months. On the other hand, the decomposed factors seem to explain returns at varying horizons.

More in detail, the value SMB factor (SMB_v) is useful at horizons of 3-5 and 7-9 months, the neutral one at horizons of 1, 7 and 8 months, while the growth component of SMB is significant only at the 1-month horizon.

Our findings with respect to factors related to past performance may be summarized as follows:

- The momentum factor (MOM) appears significant at horizons of 4, 5 and 22 months. More importantly, our findings suggest that only the big MOM factor exhibits in sample forecasting ability in both the short-run and the long-run. More in detail, these horizons are 3-5, 16, 19, 22, 24 and 25 months, suggesting the relative importance of the momentum of big firms over small firms for future stock market returns.

- The long-term (LT) reversal, which is constructed on the basis of 1-5 year past per-

[7]The authors argue that value stocks are riskier than growth stocks during economic recessions when the price of risk is high, resulting to an extra premium on value stocks which represents compensation for bearing systematic risk.

Table 2.1: In-sample predictability of the factors on the CRSP portfolio

Horizon	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
HML	1.773	2.600	**4.073**	**3.930**	2.798	**4.500**	**5.402**	**4.806**	**3.159**
HML_b	0.002	0.314	0.003	0.035	0.084	0.050	0.183	0.125	0.006
HML_s	**3.015**	**3.978**	1.541	0.829	0.158	0.878	2.494	1.547	0.532
SMB	**3.229**	0.015	2.030	2.133	**4.347**	2.080	1.441	3.281	3.140
SMB_g	**3.911**	0.191	0.380	0.469	1.296	0.911	0.259	1.839	1.715
SMB_n	**3.815**	0.545	0.604	0.859	2.964	1.898	**9.183**	**4.115**	2.654
SMB_v	0.185	1.479	**4.349**	**4.374**	**3.663**	2.536	**3.586**	**5.943**	**3.569**
MOM	0.202	0.015	2.571	**3.172**	**3.657**	2.318	0.864	0.385	1.044
MOM_b	0.417	0.143	**3.633**	**4.592**	**4.651**	3.282	1.768	0.941	1.858
MOM_s	0.032	0.022	1.200	1.293	1.839	0.791	0.030	0.000	0.079
LT	2.820	2.605	**4.051**	**3.587**	**4.239**	**5.520**	**6.256**	**4.095**	2.195
LT_b	0.664	0.922	1.909	1.814	**3.115**	**3.171**	**3.812**	2.190	0.938
LT_s	**6.023**	**5.281**	**6.702**	**5.395**	**4.939**	**7.946**	**8.535**	**6.502**	**4.345**
Horizon	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
HML	2.244	1.918	0.916	0.621	0.433	0.188	0.042	0.190	0.100
HML_b	0.002	0.001	0.026	0.034	0.050	0.121	0.293	0.044	0.030
HML_s	0.690	0.471	0.000	0.024	0.151	0.161	0.225	0.008	0.001
SMB	1.747	1.011	1.195	2.077	1.865	0.741	0.224	0.038	0.046
SMB_g	0.537	0.342	0.968	1.661	1.785	0.800	0.296	0.056	0.037
SMB_n	2.032	1.619	1.189	2.010	2.002	0.763	0.285	0.111	0.082
SMB_v	2.798	2.018	1.511	1.652	1.200	0.629	0.300	0.126	0.121
MOM	0.819	1.201	1.573	2.226	1.421	1.896	2.984	1.490	1.153
MOM_b	1.551	1.785	1.707	2.436	1.921	2.421	**3.654**	2.338	1.958
MOM_s	0.005	0.095	0.500	0.759	0.251	0.575	0.986	0.123	0.049
LT	1.259	1.212	1.405	1.355	1.113	1.348	0.582	0.806	1.236
LT_b	0.528	0.435	0.653	0.357	0.378	0.629	0.143	0.394	0.581
LT_s	2.374	2.545	2.570	**3.448**	2.498	2.454	1.544	1.394	2.131
Horizon	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**
HML	0.100	0.039	0.018	0.021	0.009	0.049	0.141	0.100	0.076
HML_b	0.036	0.001	0.028	0.003	0.026	0.130	0.455	0.353	0.242
HML_s	0.041	0.103	0.128	0.259	0.257	0.285	0.177	0.094	0.168
SMB	0.229	0.583	0.479	0.319	0.162	0.426	0.791	0.744	0.501
SMB_g	0.216	0.729	0.760	0.673	0.475	0.826	1.189	0.808	0.541
SMB_n	0.159	0.410	0.305	0.207	0.075	0.356	0.850	1.041	0.772
SMB_v	0.258	0.370	0.235	0.076	0.001	0.085	0.046	0.043	0.000
MOM	3.477	2.157	1.974	**3.695**	2.107	2.698	1.859	0.654	0.438
MOM_b	**4.402**	3.150	2.840	**4.626**	2.941	**3.713**	**3.653**	1.606	0.895
MOM_s	0.588	0.417	0.426	0.962	0.486	0.725	0.249	0.025	0.056
LT	1.475	0.944	0.678	0.798	0.934	0.720	0.580	0.825	0.924
LT_b	0.680	0.329	0.208	0.274	0.316	0.142	0.057	0.182	0.103
LT_s	2.593	1.970	1.418	1.519	1.733	1.948	1.970	1.988	2.658

Table 2.1 (continued)

Horizon	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**
HML	0.085	0.052	0.027	0.000	0.011	0.044	0.135	0.078	0.148
HML_b	0.269	0.435	0.214	0.027	0.021	0.001	0.049	0.000	0.025
HML_s	0.154	0.206	0.395	0.736	0.801	1.050	1.317	1.196	1.609
SMB	0.439	0.667	0.533	1.173	1.517	1.410	1.256	0.842	1.059
SMB_g	0.516	0.798	0.671	1.083	1.475	1.508	1.388	1.176	1.267
SMB_n	0.666	0.867	0.603	1.401	1.802	1.920	1.868	1.249	1.624
SMB_v	0.000	0.000	0.004	0.015	0.032	0.010	0.012	0.047	0.029
MOM	0.255	0.186	0.911	1.546	0.814	0.549	0.573	0.321	0.489
MOM_b	0.453	0.380	1.318	2.510	1.412	1.041	0.864	0.417	0.535
MOM_s	0.042	0.019	0.272	0.484	0.249	0.168	0.262	0.201	0.363
LT	1.599	1.142	1.238	1.215	1.535	1.952	1.650	1.127	1.235
LT_b	0.330	0.145	0.069	0.062	0.085	0.276	0.215	0.048	0.080
LT_s	**3.637**	**3.235**	**4.336**	**4.332**	**4.863**	**5.090**	**4.387**	**3.981**	**4.186**

Notes: (i) The table reports the in-sample predictive ability of various factors on the CRSP value-weighted portfolio return for horizons of 1-36 months.
(ii) The results given are according to the Wald test statistic and bootstrapped critical values.
(iii) Bold entries indicate significance at the 10% significance level.

formance, appears to be a significant predictor at horizons of 3-8 months. Comparing its small and big components, we find that LT_s is the most powerful one containing significant information for a wider spectrum of horizons, namely 1-9, 13 and 28-36 months. The ability of LT_b is confined to 5-7 months.

Having assessed the ability of the factors to predict returns in sample, we turn to the typically employed financial variables and compare their ability with the respective of the factors. Our findings are reported in the following subsection.

2.4.2 In-sample predictive ability of financial variables

The most commonly employed financial variables in the return predictability literature are interest rate variables, such as the short-term interest rate, the term and default spreads, and valuation ratios, such as the dividend yield. Short-term interest rates are linked to the current business cycle and monetary policy stance, as low interest rates prevail in recessions and vice versa. The term spread signals the future state of the economy, while the default spread signals credit market expectations. Fama and French (1989), among others, find that changes in the term spread and the default spread correspond to short-term and long-term business conditions. The explanation is that when business conditions are poor, income is low and expected returns on stock must be high to induce substitution from consumption to investment. With respect to valuation ratios, the dividend yield is the favorable ratio employed by researchers, since it directly reflects expectations about future returns. According to the well-known Campbell and Shiller (1988a) present-value decomposition, deviations in the dividend-price ratio from its long-term mean signal changes in expected future dividend growth rates and/or expected future stock returns; changes in the latter represent time-varying discount rates and return predictability.

Our in-sample results, with respect to the financial variables, are reported in Table 2.2. Overall, we find that all the financial variables with the exception of the term spread exhibit significant in-sample forecasting ability for the CRSP value-weighted portfolio return. Specifically, the 1-month T-bill rate emerges as a potential long-run useful predictor at horizons of 12, 14-20, 22-24, 26 and 27 months. On the other hand, the predictive ability of the default spread is rather limited at 5, 14 and 16-18 months. Quite importantly, the dividend yield emerges as a useful predictor both over the short run and the long run. More in detail, the dividend yield's predictive ability is evident at horizons of 1-15, 19, 20 and 25 months. Our results are consistent with Michou (2009) who, following a different approach, proves that the treasury bill rate along with the dividend yield exhibit rather

Table 2.2: In-sample predictability of financial variables on CRSP portfolio

Horizon	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1m Tbill	0.305	0.167	0.227	0.312	0.470	0.609	0.696	0.829	0.967
TermSpread	1.233	0.916	0.738	0.593	0.457	0.314	0.232	0.220	0.248
Def. Spr.	2.395	0.241	0.282	0.443	**7.232**	1.792	1.899	1.474	4.581
Div.Yield	**6.131**	**5.856**	**6.196**	**6.466**	**6.989**	**7.209**	**7.288**	**7.263**	**7.343**
Horizon	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
1m Tbill	1.089	1.244	**5.928**	5.101	**6.347**	**7.909**	**9.187**	**9.104**	**9.402**
TermSpread	0.338	0.400	0.436	0.490	0.558	0.689	0.887	1.056	1.331
Def. Spread	3.065	4.716	3.184	3.422	**6.061**	3.664	**5.844**	**9.117**	**9.233**
Div.Yield	**7.484**	**7.621**	**7.835**	**8.030**	**8.039**	**8.065**	8.037	8.034	8.024
Horizon	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**
1m Tbill	**7.890**	**6.323**	6.370	**6.852**	**6.599**	**6.588**	6.646	**7.054**	**6.531**
TermSpread	1.193	1.180	1.198	1.232	1.263	1.298	1.330	1.426	1.576
Def. Spread	2.498	4.704	4.222	5.145	5.124	5.672	5.620	4.004	0.065
Div.Yield	**7.986**	**7.881**	7.757	7.657	7.551	7.483	**7.397**	7.401	7.450
Horizon	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**
1m Tbill	5.979	6.070	5.825	6.147	5.818	6.202	5.669	5.823	5.936
TermSpread	1.697	1.838	1.879	1.950	2.120	2.313	2.391	2.397	2.395
Def.Spread	0.055	0.049	0.044	0.033	0.027	0.018	0.012	0.006	0.001
Div.Yield	7.561	7.697	7.832	8.004	8.201	8.394	8.629	8.937	9.267

Notes: (i) The table reports the in-sample predictive ability of various factors on the CRSP value-weighted portfolio return for horizons of 1-36 months.
(ii) The results given are according to the Wald test statistic and bootstrapped critical values.
(iii) Bold entries indicate significance at the 10% significance level.

good predictive abilities on various size and value investment strategies.

2.4.3 Can financial variables proxy in sample for the Fama French, reversal and momentum factors?

Our analysis so far has shown that the financial variables exhibit significant forecasting ability at specific horizons, while the empirical factors along with their decompositions do exactly the same, but at shorter horizons. This finding naturally leads to assessing the linkages between them. Such an assessment is not new to the literature. Hahn and Lee

(2006) examine whether the yield spread variables; namely, the term and the default spread, could proxy for the SMB and HML factors. In a simple regression framework, they prove that, indeed, the default spread proxies for the SMB factor and the term spread for the HML factor. A different approach is adopted by Petkova (2006), who tests whether these state variables could proxy for the Fama-French factors. In her analysis, the author investigates whether the SMB and HML factors proxy for the term spread, the default spread and the 1-month T-bill rate using monthly data for the post-1963 period. Her results show that the SMB factor proxies significantly for the default spread and the HML factor for the term spread, while the SMB and HML factors do not prove significant for the short-term interest rate. Employing wavelet analysis, In and Kim (2007) investigate how the SMB and HML factors interact with the innovations of state variables over various time scales. Following the approach adopted by Petkova (2006), they also prove that the SMB factor is a proxy for the default spread, while the HML factor is affected by the term spread. However, the short-term interest rate does not show any significant explanatory power for either the SMB or the HML factor. More importantly, the authors find that the predictive ability of both the SMB and HML factors is more prominent in the long run.

Employing the methodology outlined in Section 2, we investigate the link between financial variables and the factors under consideration.[8] Our results, reported in Table 2.3, indicate that there are financial variables that proxy in sample for the factors albeit at varying horizons. More in detail, the default spread emerges as a valuable in sample predictor for the value premium at any horizon exceeding 5 months. On the other hand,

[8] For brevity, we only report the horizons for which we find significant in sample predictive ability. Detailed tables are available from the authors upon request.

the term spread proxies only weakly the HML factor at the 10-month horizon. Turning to its size components, it is worth noting that small HML mimics the behavior of the full HML factor, while big HML appears completely dissected from the financial variables employed. This behavior can be explained by the fact that the small and high book-to-market firms are more vulnerable to worsening credit market conditions. Our findings are consistent with Chen *et al.* (2008), who prove that the HML factor is weakly countercyclical peaking in most of the recessions and correlates positively with the default premium. Similarly, Gulen *et al.* (2011) argue that there is a greater propensity for value firms to be exposed to bankruptcy risk during recessions than growth firms, implying that the returns of value stocks should load more heavily on the default spread than returns of growth stocks.

Our findings with respect to the SMB factor confirm those of Hahn and Lee (2006), who show that the default spread covaries positively with the SMB factor. Specifically, the default spread exhibits in-sample predictive ability for SMB at horizons of 1-22 months. More importantly, this picture is consistent for the growth, neutral and value decompositions of this factor as well. Specifically, the default spread contains significant information for the growth SMB over a variety of horizons ranging from the short run (3 months) to the long run (36 months).

With respect to SMB_n and SMB_v, the ability of the default spread is limited to horizons less than 26 and 21 months, respectively. Term spread does not emerge as a valuable in sample predictor for the aggregate SMB factor at any horizon considered. However, neutral SMB is related to the term spread at horizons of 3-4, 16-26 and 36 months and value SMB at shorter ones of 1-7 months. In the case of the growth component, some

Table 2.3: In-sample predictability of the financial variables on the factors

Variables	1m Tbill	Term Spread	Default Spread	Div. Yield
HML	-	10	5-36	-
HML_b	-	-	-	-
HML_s	-	10	3-36	-
SMB	-	-	1-22	-
SMB_g	-	10	3-36	-
SMB_n	-	3-4, 16-26, 36	1-10, 12-26	1, 3-6
SMB_v	1-5	1-7	1-9, 12-21	-
MOM	-	-	1, 20	-
MOM_b	1-4	1-7	1-9, 12-21	-
MOM_s	1-3, 6, 8-25	-	1	-
LT	-	-	-	-
LT_b	-	-	-	-
LT_s	-	27-29	-	-

Notes: (i) The table presents only the horizons for which significant in-sample predictability of various financial variables for horizons of 1-36 months exists.
(ii) The results given are according to the Wald test statistic and bootstrapped critical values at the 10% significance level.

scanty evidence of predictability is found at the horizon of 10 months. The remaining state variables, namely the 1-month T bill and the dividend yield are useful in predicting value SMB at short horizons of 1-5 months and neutral SMB at horizons of 1 and 3-6 months, respectively.

Turning to factors related to past performance, we have to note that the default spread is the only financial variable that displays significant in-sample predictive ability for the momentum factor albeit at horizons of 1 and 20 months only. However, when decomposing MOM to its big and small counterparts, both the term spread and the short-term interest rate achieve some significant explanatory power. More in detail, short-term interest rate movements, changes in the term spread and the default spread help explain future big momentum returns at horizons of 1-4, 1-7, and 1-9, 12-21 months ahead, respectively. On the other hand, small MOM can be predicted by developments in the interest rate at a wider spectrum of horizons ranging from 1 to 25 months, while the default spread can offer only short-run predictability at 1 month. Our findings are consistent with Chordia and Shivakumar (2002) who find that momentum portfolios formed on the basis of past returns vary systematically with common macroeconomic variables related to the business cycle. With respect to the long run reversal factor (LT), our results suggest that no financial variable can proxy for this factor in sample. Even after decomposing this factor, we do not find any relation of this factor with the financial variables at hand, with the exception of the small LT factor that can be proxied by the term spread at long horizons of 27-29 months.

Overall, among the state/financial variables, the default spread emerges as the most significant proxy at a variety of horizons for the factors considered (along with their decompositions) with the exception of the HML_b factor and the LT ones. Weaker in-sample predictability is found with respect to the term spread which can proxy for the HML, HML_s, the decomposed SMB factors, the MOM_b and the LT_s factors.

2.5 Out-of-sample predictive ability

While the previous Section offered a detailed analysis on the linkages between factors, state variables and future returns over a variety of horizons, their forecasting ability in a pure out-of-sample experiment has not been checked so far. In this Section, we explore the out-of-sample predictability of both the factors and the financial variables for future returns on the CRSP value weighted portfolio. Out-of-sample forecasts (for horizon h, i.e. the period $t+h$) are generated using only information available at period t. Time-varying coefficients are estimated in real-time from Eq.(2.1) using a recursive regression technique. A minimum window length of two thirds of our available sample (371 observations) is used to derive parameter estimates.[9] Thus, the 1963:07–1994:05 period provides the first coefficient estimates, 1963:07–1994:06 the second and so on. In this way the out-of-sample evaluation window contains $185 - h$ observations. The forecasting ability of the predictive ARDL model (Eq.2.1) relative to the benchmark AR specification that contains no predictors is tested through the significance of the value of the Theil's U, which is the ratio of the Mean Squared Forecast Error (MSFE) of the unrestricted model to the MSFE of the restricted one through the MSE-T, MSE-F, ENC-T, and ENC-NEW tests. Whenever at least one of the tests indicates significance (at the 10% level), the value of the Theil's U appears in boldface.[10]

2.5.1 Fama French, reversal and momentum factors vs. state variables

In Table 2.4, we present the out-of-sample forecasting ability of the HML, SMB, MOM and LT factors for subsequent market returns. Starting with the value premium (HML factor), our findings suggest that its in-sample predictive ability pertains out of sample, as well. Specifically, the aggregate value premium can improve market return forecasts at horizons less than a year, ranging from 3 to 10 months. Quite interestingly, this ability almost disappears when its size components are considered with the small value

[9] Given a total sample of T observations, the researcher must decide on how to divide the sample into the estimation part (R observations) and the out-of-sample part ($P := T - R$ observations). Obviously, there is a trade-off, since a large R improves the quality of the estimated parameters of the model but, at the same time, leaves fewer observations for the out-of-sample forecast exercise making the evaluation of the predictive ability of the model difficult. In our analysis, we keep about 1/3 of the available sample for out-of-sample forecasting. This choice gives us a sufficient number of forecasts to evaluate the estimated models, while keeping enough observations to obtain reliable estimates for the parameters of our predictive models.

[10] Detailed tables are available from the authors upon request.

premium being useful for short-run forecasts of 1 and 2 months ahead.

Turning to the size premium, our findings suggest that it can improve return forecasts at a variety of horizons, namely 3-6, 8-10, and 13-14 months. When decomposing this factor, we find that its components contain significant information for the future evolution of stock returns. The growth component displays forecasting ability both in the short run (1 and 5 months) and in the long run (25, 31, 33-36 months), while neutral and value SMB behave similar to the aggregate one.

Specifically, neutral SMB exhibits significant forecasting ability at horizons of 5, 7-11 and 13 months and value SMB at horizons of 3-11 months. Quite interestingly, the decomposed SMB factors exhibit superior forecasting ability out of sample rather than in sample. The momentum factor is associated with both short-term and medium-term predictability and specifically 4-6, 9, 12-14 and 16-19 months ahead. When decomposing the momentum factor, it becomes apparent that only the big component of the momentum factor contributes to the predictability of returns at a wide spectrum of horizons covering roughly 3 to 19 months.

Finally, the long-term reversal factor appears significant at 2-8 months and at the longer horizons of 32-34 months. Concerning its decomposition, the small component of the LT reversal factor displays similar behavior with the aggregate factor in the short run. However, it emerges as a useful predictor also in the long run and specifically at horizons exceeding 27 months. The LT_b factor does not contain more information than that included to the initial factor and the LT_s factor.

Our findings with respect to the forecasting ability of the financial variables are

Table 2.4: Out-of-sample predictability of the factors on CRSP portfolio

Horizon	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
HML	1.000	0.998	**0.996**	**0.995**	**0.996**	**0.993**	**0.990**	**0.991**	**0.994**
HML_b	1.004	1.003	1.003	1.001	1.002	1.002	1.004	1.005	1.004
HML_s	**0.997**	**0.996**	1.002	1.002	1.004	1.001	0.999	1.000	1.002
SMB	1.000	1.002	**0.997**	**0.997**	**0.994**	**0.997**	1.000	**0.998**	**0.996**
SMB_g	**0.995**	1.001	1.000	1.000	**0.999**	0.999	1.001	1.001	1.000
SMB_n	1.000	1.002	1.000	0.999	**0.997**	0.999	**0.994**	**0.997**	**0.997**
SMB_v	1.008	0.999	**0.992**	**0.992**	**0.994**	**0.996**	**0.994**	**0.988**	**0.992**
MOM	1.004	1.004	0.999	**0.998**	**0.995**	**0.997**	0.999	1.000	**0.998**
MOM_b	1.002	1.003	**0.999**	**0.996**	**0.994**	**0.996**	**0.998**	0.999	**0.997**
MOM_s	1.005	1.006	1.001	1.000	0.999	1.000	1.001	1.001	1.000
LT	1.001	**0.998**	**0.991**	**0.993**	**0.993**	**0.995**	**0.995**	**0.998**	0.999
LT_b	1.001	1.000	**0.996**	**0.996**	**0.994**	**0.996**	**0.995**	**0.997**	1.000
LT_s	1.002	**0.999**	**0.994**	**0.996**	**0.997**	1.000	1.002	1.003	1.003
Horizon	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
HML	**0.996**	0.998	1.000	1.000	1.001	1.001	1.002	1.002	1.003
HML_b	1.004	1.004	1.004	1.003	1.003	1.003	1.002	1.003	1.002
HML_s	1.002	1.004	1.005	1.004	1.003	1.002	1.003	1.004	1.003
SMB	**0.997**	0.999	0.999	**0.997**	**0.997**	0.999	1.000	1.002	1.002
SMB_g	1.000	1.000	1.000	0.998	0.998	0.998	0.999	1.002	1.002
SMB_n	**0.997**	**0.998**	0.999	**0.997**	0.998	0.999	1.000	1.001	1.001
SMB_v	**0.994**	**0.997**	0.999	0.998	0.999	1.002	1.002	1.004	1.003
MOM	1.000	1.000	**0.998**	**0.998**	**0.999**	0.998	**0.997**	**0.998**	**0.998**
MOM_b	0.999	0.999	**0.998**	**0.996**	**0.997**	**0.997**	**0.995**	**0.997**	**0.998**
MOM_s	1.001	1.002	1.001	1.001	1.002	1.001	1.000	1.000	1.000
LT	1.000	1.000	1.000	1.001	1.002	1.002	1.004	1.003	1.001
LT_b	1.001	1.001	1.001	1.002	1.002	1.002	1.004	1.003	1.001
LT_s	1.003	1.003	1.001	1.000	1.003	1.003	1.004	1.004	1.002
Horizon	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**
HML	1.004	1.003	1.002	1.003	1.004	1.003	1.003	1.003	1.003
HML_b	1.002	1.001	1.000	1.001	1.002	1.001	1.000	1.000	1.000
HML_s	1.003	1.003	1.002	1.002	1.003	1.002	1.003	1.003	1.003
SMB	1.002	1.000	1.000	1.000	1.001	1.000	0.999	0.999	1.000
SMB_g	1.002	1.000	0.999	0.999	0.999	0.999	**0.998**	0.999	1.000
SMB_n	1.001	1.000	1.000	1.000	1.001	1.000	0.999	0.999	0.999
SMB_v	1.002	1.002	1.001	1.002	1.002	1.009	1.001	1.001	1.002
MOM	**0.997**	0.999	1.001	1.000	1.000	1.000	1.000	1.001	1.000
MOM_b	**0.996**	0.999	1.002	1.001	1.001	1.000	1.001	1.002	1.001
MOM_s	1.000	1.000	1.001	1.000	1.000	1.000	1.001	1.001	1.001
LT	0.999	1.001	1.001	1.000	1.000	1.000	1.000	1.000	0.999
LT_b	1.001	1.002	1.003	1.002	1.002	1.002	1.003	1.003	1.003
LT_s	**0.999**	1.001	1.001	1.000	1.000	0.999	0.998	0.998	**0.998**

Table 2.4 (continued)

Horizon	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**
HML	1.004	1.004	1.002	1.002	1.002	1.003	1.003	1.002	1.002
HML_b	1.001	1.001	1.001	1.001	1.001	1.001	1.001	1.001	1.001
HML_s	1.003	1.003	1.001	1.000	1.000	1.001	1.000	1.001	1.001
SMB	1.000	0.999	0.999	0.997	0.998	0.998	0.998	0.999	0.999
SMB_g	1.000	0.999	0.998	**0.997**	0.997	**0.997**	**0.997**	**0.999**	**0.999**
SMB_n	0.999	0.999	1.000	0.999	0.999	0.999	0.999	0.999	0.999
SMB_v	1.001	1.001	1.001	1.000	1.000	1.000	1.000	1.001	1.001
MOM	1.001	1.001	1.001	1.000	1.000	1.001	1.001	1.001	1.001
MOM_b	1.002	1.003	1.003	1.001	1.000	1.000	1.000	1.000	1.000
MOM_s	1.001	1.001	1.001	1.001	1.002	1.003	1.003	1.003	1.004
LT	0.998	0.999	0.999	0.999	**0.999**	**0.998**	**0.999**	1.000	0.999
LT_b	1.002	1.003	1.004	1.004	1.003	1.002	1.002	1.002	1.002
LT_s	**0.997**	**0.997**	**0.995**	**0.994**	**0.993**	**0.993**	**0.995**	**0.996**	**0.995**

Notes: (i) The table reports the out-of-sample predictive ability of various factors on the CRSP value-weighted portfolio return for horizons of 1-36 months based on the significance of the value of Theil's U.
(ii) The results given are according to the MSE-T and MSE-F statistics, which test the null hypothesis of equal forecasting accuracy, and the ENC-T and ENC-NEW, which test the null hypothesis of equal forecast accuracy or forecast encompassing.
(iii) Bold indicates significance at the 10% level according to bootstrapped p-values.

Table 2.5: Out-of-sample predictability of fin.variables on CRSP portfolio

Horizon	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1m T_bill	1.001	1.003	1.004	1.005	1.005	1.004	1.003	1.002	1.000
TermSpread	1.006	1.009	1.012	1.015	1.015	1.015	1.016	1.016	1.017
Def. Spread	1.041	1.038	1.051	1.062	1.108	1.084	1.084	1.093	1.127
Div.Yield	1.007	1.012	1.017	1.021	1.026	1.032	1.040	1.047	1.052
Horizon	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
1m Tbill	0.999	0.999	1.064	1.067	1.069	1.070	1.055	1.050	1.051
TermSpread	1.016	1.014	1.011	1.007	1.004	1.000	0.997	0.994	1.005
Def. Spread	1.129	1.118	1.112	1.104	1.108	1.100	1.094	1.076	1.071
Div. Yield	1.058	1.062	1.066	1.070	1.081	1.085	1.089	1.092	1.097
Horizon	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**
1m Tbill	1.045	1.038	1.039	1.046	1.043	1.039	1.036	1.050	1.051
TermSpread	**0.990**	**0.990**	**0.989**	**0.989**	**0.988**	**0.988**	**0.988**	**0.987**	**0.987**
Def. Spread	1.051	1.044	1.039	1.038	1.035	1.033	1.033	1.030	1.021
Div.Yield	1.100	1.101	1.101	1.101	1.102	1.102	1.102	1.101	1.100
Horizon	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**
1m Tbill	1.051	1.047	1.047	1.046	1.045	1.043	1.038	1.033	1.031
TermSpread	**0.986**	**0.985**	**0.985**	**0.985**	**0.984**	**0.983**	**0.983**	**0.983**	**0.983**
Def. Spread	1.022	1.022	1.022	1.023	1.024	1.026	1.027	1.028	1.030
Div.Yield	1.100	1.100	1.100	1.099	1.097	1.094	1.091	1.088	1.084

Notes: (i) The table reports the out-of-sample predictive ability of financial variables on the CRSP value-weighted portfolio return for horizons of 1-36 months based on the significance of the value of Theil's U.
(ii) The results given are according to the MSE-T and MSE-F statistics, which test the null hypothesis of equal forecasting accuracy, and the ENC-T and ENC-NEW, which test the null hypothesis of equal forecast accuracy or forecast encompassing.
(iii) Bold indicates significance at the 10% level according to bootstrapped p-values.

reported in Table 2.5. Quite interestingly, with the exception of the term spread, all the remaining financial variables exhibit hardly any out of sample forecasting ability. More in detail, the term spread exhibits significant predictive ability in the long run, specifically at horizons greater than 1 year (19-36 months).

The well documented disparity (see, e.g. Rapach and Wohar, 2006; Campbell and Thompson, 2008; Goyal and Welch, 2008 among others) between in-sample and out-of-sample predictability is evident in our findings, as well. While our financial variables exhibit significant explanatory power in-sample at a variety of horizons, this predictability vanishes when it comes to out-of-sample forecasting. To this end, the following subsection is devoted to testing whether the same set of financial variables can be useful in predicting the Fama French, reversal and momentum factors which are market portfolios with specific characteristics, as already stated.

2.5.2 Can financial variables predict Fama French, reversal and momentum factors?

In Table 2.6, we report the horizons for which we find significant forecasting ability of the financial variables at hand for the empirical factors. In particular, our findings with respect to the HML factor suggest that only the default spread is a valuable predictor for horizons exceeding 12 months.

When decomposing the factor, the default spread retains its ability for both components, but at different horizons; 23-36 months for the big component and 11-36 for the small one. Quite importantly, the term spread offers some predictability for both components at 20, 22 months ahead and 2-6, 32, 35 and 36 months ahead for HML_b and HML_s, respectively.

Turning to the SMB factor, we have to note that with the exception of the dividend yield, the remaining variables can prove significant over the short term and specifically at horizons less than 4 months. However, this picture changes when the components are considered. The growth size premium can be forecasted by the term spread at a variety of horizons ranging from 2 to 36 months, while the default spread emerges as a consistent

Table 2.6: Out-of-sample predictability of the fin.variables on the factors

Variables	1m T_bill	Term Spread	Default Spread	Div. Yield
HML	-	-	12-36	-
HML_b	-	20, 22	23-36	-
HML_s	-	2-6, 32, 35, 36	11-36	-
SMB	2	1-4	1-3	-
SMB_g	-	2-6, 8, 13, 28, 30, 32-36	10-36	-
SMB_n	-	1-4	1-9, 11	1-12
SMB_v	1-8	1-12	1-11	-
MOM	2, 12	11, 13-22	1-12, 17-20	-
MOM_b	1-7, 9	1-12	1-11	-
MOM_s	1-21	1-3, 7, 10, 12, 13, 16	1-11	-
LT	-	31-36	3-13, 15	-
LT_b	-	30-36	2-10, 12-16	-
LT_s	-	27-29, 31-36	14, 15	-

Notes: (i) The table presents only the horizons for which significant out-of-sample forecasting ability of various financial variables for horizons of 1-36 months exists.
(ii) The results given are according to the MSE-T and MSE-F statistics, which test the null hypothesis of equal forecasting accuracy, and the ENC-T and ENC-NEW, which test the null hypothesis of equal forecast accuracy or forecast encompassing.
(iii) Bold indicates significance at the 10% level according to bootstrapped p-values.

predictor for horizons exceeding 10 months. All variables with the exception of the short-term interest rate can improve short-term (less than a year) forecasts for neutral SMB. Similarly, value SMB can be predicted over the short run by all variables with the exception of the dividend yield.

Next, we focus our attention on factors related to past performance. Among the variables considered, the term and default spreads turn out to be more powerful predictors for the aggregate momentum factor, with significant predictive ability for 11, 13-22 months and 1-12, 17-20 months for the term and default spread, respectively.

The short-term interest rate is associated with weak predictability for the aggregate factor (2 and 12 months), which is reinforced when the small and big components are considered. Specifically, the short-term interest rate improves forecasts for the big momentum portfolio at 1-7 and 9 months ahead and 1-21 months ahead for the small one. The spread variables emerge as useful predictors for both small and big momentum at a variety of horizons up to 16 months. With respect to the long term reversal factor, our findings suggest that only the term and default spread can prove valuable predictors. Specifically, there is evidence of forecasting ability of the default spread at horizons of 3-13 and 15 months and the term spread at longer horizons, ranging from 31 to 36 months. Similarly, both spreads can improve the predictive ability of both components of the long term reversal factor. In particular, LT_b is predicted by the term spread in the long run, at horizons of 30-36 months, and by the default spread at 2-10 and 12-16 months. With respect to the small LT, the same variables prove useful at horizons of 27-29 and 31-36 months for the term spread and 14-15 months for the default spread.

2.6 Conclusions

This study primarily examines whether the value premium (HML), the size premium (SMB), the momentum (MOM) and long-term reversal (LT) factors exhibit forecasting ability for US stock returns, namely the CRSP value weighted portfolio return. An important contribution of the paper is that it provides answers to the following questions i) do empirical factors exhibit stronger forecasting ability for market returns (over a wide spectrum of forecast horizons) than typically employed financial variables?, ii) what is the link between factors and these state variables? iii) can appropriate decompositions of the factors in their size and value components enhance predictability?

In order to assess the forecasting ability of candidate variables for the US returns, we employ the Autoregressive Distributed Lag (ARDL) methodology of Rapach and Weber (2004). The in-sample forecasting ability is assessed via the typical Wald test, while the out-of-sample one is assessed via tests for equal predictive ability and forecast encompassing. Our findings with respect to the in-sample predictability suggest that the HML, value SMB, big MOM and small LT emerge as useful predictors albeit at a variety of forecast horizons. Among the financial variables considered, both the short-term interest rate and the dividend yield appear as strong predictors, while the default spread and the term spread display weaker and no predictability, respectively. Investigating whether the factors are linked with the financial variables, we find that, in sample, the default spread seems to be a significant proxy for all the empirical factors at hand with the exception of LT. The term spread mainly predicts neutral and value SMB, while the short-term interest rate is associated with developments in big and small MOM, but not with the aggregate factor. Moreover, the dividend yield proxies only for neutral SMB in the short run.

More importantly, all the empirical factors can improve out-of-sample US market return forecasts over the autoregressive benchmark forecast. With respect to factor components, value and neutral SMB, big MOM and both components of LT contain more information than the aggregate factors for the evolution of market returns. Quite interestingly, while the financial variables display in-sample predictive ability, out-of-sample only the term spread improves forecasts for horizons greater than 19 months. This picture changes when we consider the forecasting ability of financial variables on the factor portfolios. Specifically, the term spread and default spread improve predictions for all the factors and their components, while the short-term interest rate helps predicting SMB, value SMB, MOM and its components. The ability of the dividend yield is confined to neutral SMB.

Overall, we provide new evidence on both the in-sample and out-of-sample predictability of factor returns for stock market returns and of typically employed financial variables on both the aggregate market return and the factor portfolios. Our findings can be particularly helpful to asset managers, who could potentially earn higher returns by incorporating them in their asset allocation decisions.

Chapter 3

Combination Forecasts of Bond and Stock Returns: An Asset Allocation Perspective

3.1 Introduction

The asset allocation decision, i.e. how much to allocate wealth in asset classes such as cash, stocks and bonds, is a key determinant of investors' portfolio performance. The importance of this decision has further been highlighted by empirical findings suggesting that stock and bond returns contain a sizeable predictable component that needs to be addressed. The degree to which bond and stock returns are predictable is a subject of ongoing debates and intensive empirical research.

The seminal contribution of Goyal and Welch (2008), who show that their long list of predictors, consisting of both macroeconomic and financial variables, can not deliver consistently superior out-of-sample performance for US stock returns, renewed the interest on stock return predictability. Contributions to this field include Campbell and Thompson (2008) who show that when imposing simple restrictions, suggested by economic theory, on predictive regressions' coefficients, the out-of-sample performance improves. The authors show that market timing strategies can deliver profits to investors (see also Ferreira and Santa-Clara (2011)). Ludvigson and Ng (2007) and Neely et al. (2013) adopt a diffusion index approach, which can conveniently track the key movements in a large set of predictors, and find evidence of improved equity premium forecasting ability.[1]

In a similar manner, various financial and macroeconomic variables are also employed to predict US government bond returns. For example, Keim and Stambaugh (1986), Fama and French (1989) and Campbell and Shiller (1991) show that yield spreads have predictive power. Cochrane and Piazzesi (2005) employed a linear combination of five forward rates and find a high degree of predictability, while Ludvigson and Ng (2009) show that the impressive predictive power, found by Cochrane and Piazzesi (2005), can be improved with five macroeconomic factors estimated from a set of 132 macroeconomic variables that measure a wide range of economic activities. More recently, Goh et al. (2013) take another route and study the predictive ability of technical indicators vis-a-vis economic variables for bond returns and find that technical indicators have both in- and out-of-sample forecasting power.

In our analysis, we also take an alternative route and investigate the forecasting ability of value, size and momentum empirical factors vis-a-vis typically employed finan-

[1] Rapach and Zhou (2012) offer a detailed review on the issue of equity return predictability.

cial variables for US bond and stock market returns. Specifically, we employ the value premium (High minus Low; HML), the size premium (Small minus Big; SMB), the momentum (Winners minus Losers over the past year; MOM), the long term reversal (Winners minus Losers over the past one to five years; LT) and the short term reversal (Winners minus Losers over the past one to one month; ST). Following Fama and French (2012), we decompose the aforementioned factors into their size and value counterparts. In this way, we can disentangle the value effect on the size premium and the size effect on the remaining factors. Our paper also relates to the broad literature of forecast combinations by considering whether combinations of individual model forecasts based on the empirical factors can further improve the predictability of bond and stock returns. Rapach, Strauss and Zhou (2010) show that combination of individual financial variables forecasts improve equity premium forecasts. The authors argue that single variable forecasts cannot generate reliable forecasts over time due to parameter instability and complexity of the real economy. To this end, they show that the success of combination forecasts is attributed mainly to their link with the real economy and their ability to stabilize forecasts. In a similar manner, we also employ a variety of combination methods applied to individual empirical factors forecasts. The performance of the proposed models is assessed not only statistically, but also economically from an asset allocation perspective.

To anticipate our key results, we find that the proposed empirical factors, aggregate and decomposed, display superior forecasting ability for bond and stock market returns compared to the financial variables, not only in the U.S. market, but also in other markets, such as Europe and Japan. From an economic perspective, the empirical factors lead to significant performance fees that an investor would be willing to pay in order to have access to the information offered by the proposed factors.

The remainder of the paper is organized as follows. Section 2 describes in detail the construction of the forecasts and the corresponding statistical significance of our results. Section 3 presents the data and the empirical results concerning the forecasting ability of the empirical factors and financial variables, when employed individually or through combining methods. The asset allocation framework along with empirical results are discussed in Section 4. Section 5 reports the results of the robustness checks and Section 6 summarizes and concludes.

3.2 Forecast methodology

3.2.1 AutoRegressive Distributed Lag (ARDL) models

Following Rapach and Weber (2004), the predictive ability of the empirical factors and financial variables is evaluated by means of the following predictive AutoRegressive Distributed Lag (ARDL) model:

$$z_{t+h} = a + \sum_{i=0}^{q_1-1} \beta_i r_{t-i} + \sum_{i=0}^{q_2-1} \gamma_i x_{t-i} + \epsilon_{t+h} \tag{3.1}$$

where $z_{t+h} = \sum_{i=1}^{h} r_{t+i}$ is the return to be predicted from period t to $t+h$ with h the forecast horizon, r_t is the one-period return at time t, x_t the candidate predictor variable, ϵ_{t+h} the disturbance term, a the intercept, q_1 and q_2 the data-determined lag orders for r_t and x_t.[2] A heteroscedasticity and autocorrelation-consistent (HAC) covariance matrix

[2]The maximum lag value is 8 and is selected by means of the SIC criterion.

should be employed when multi-step forecasts are concerned, i.e. $h > 1$, since cumulative returns z_{t+h} overlap and this induces serial correlation to the disturbance term (see e.g. Newey and West, 1987).

In order to study the out-of-sample forecasting ability, the total sample T is divided into the first R in-sample observations and the last P out-of-sample observations. In order to create the first out-of-sample forecast, we make use of the in-sample portion of the sample and get the estimated parameters $\widehat{a}_{1,R}$, $\widehat{\beta}_{1,R,i}$ and $\widehat{\gamma}_{1,R,i}$ of the ARDL equation via ordinary least squares (OLS) for the unrestricted form of the model. Then, the estimated equation: $\widehat{z}_{1,R+h} = \widehat{a}_{1,R} + \sum_{i=0}^{q_1-1} \widehat{\beta}_{1,R,i} r_{R-i} + \sum_{i=0}^{q_2-1} \widehat{\gamma}_{1,R,i} x_{R-i}$ creates the first out-of-sample forecast for the unrestricted form of the model, as well as, the forecast error: $\widehat{u}_{1,R+h} = z_{R+h} - \widehat{z}_{1,R+h}$.

Following the same procedure, we estimate the equation for the restricted form of the model: $\widehat{z}_{0,R+h} = \widehat{a}_{0,R} + \sum_{i=0}^{q_1-1} \widehat{\beta}_{0,R,i} r_{R-i}$, where $\widehat{a}_{0,R}$ and $\widehat{\beta}_{0,R,i}$ are the OLS parameter estimates and compute the forecast error: $\widehat{u}_{0,R+h} = z_{R+h} - \widehat{z}_{0,R+h}$. This restricted model forms the benchmark model in the forecast evaluation and we refer to it as the benchmark AR model. In order to create the next forecasts, we expand recursively the in-sample portion of the sample and repeat the whole procedure through the end of the available sample, generating $P = T - R - h + 1$ out-of-sample forecast errors for the unrestricted and the restricted form of the predictive model, $\{\widehat{u}_{1,t+h}\}_{t=R}^{T-h}$ and $\{\widehat{u}_{0,t+h}\}_{t=R}^{T-h}$, respectively.

3.2.2 Combination forecasts

Combination forecasts, denoted by $\widehat{z}_{CB,t+h/t}$, are linear combinations of the n individual ARDL model forecasts, $\widehat{z}_{i,t+h}$, which are constructed by employing one factor at a time at the predictive ARDL model (Equation 3.1). Specifically, combination forecasts are formed as follows:

$$\widehat{z}_{CB,t+h/t} = \sum_{i=1}^{n} w_{i,t}\widehat{z}_{i,t+h/t} \tag{3.2}$$

where $\sum_{i=1}^{n} w_{i,t} = 1$. The weights, $w_{i,t}$, allocated to each of the individual forecasts are estimated by both simple and more complicated methods.

We employ three simple combination methods, namely the mean, the median and the trimmed mean one. The mean combination forecast imposes equal weights on all individual predictive models i.e., $w_{i,t} = 1/n$ $(i = 1, ..., n)$. The median combination forecast is just the sample median of $\{\widehat{z}_{i,t+h/t}\}_{t=1}^{n}$, while the trimmed mean combination forecast sets $w_{i,t} = 1/(n-2)$ for all the individual forecasts, excluding the smallest and the largest one at time t.

We also employ the discount Mean Square Forecast Error (DMSE) combining method of Stock and Watson (2004), which assigns weights based on the historical performance of the individual ARDL models, as follows:

$$w_{i,t} = m_{i,t}^{-1} / \sum_{j=1}^{n} m_{j,t}^{-1}, m_{i,t} = \sum_{s=R}^{t-h} \psi^{t-h-s}(z_{s+h} - \widehat{z}_{i,s+h/s})^2 \tag{3.3}$$

where ψ is a discount factor that makes the recent forecasting accuracy of the individual ARDL models more important in the cases where $\psi < 1$. In particular, forecasts based on individual factors with lower MSFEs are given greater weights, and as such more accurate models are more important for the formation of this combination forecast. DMSE forecasts require a holdout out-of-sample period in order to calculate the weights attributed to each individual forecast. We employ the last P_0 observations of the in-sample period as the initial holdout window. The values of ψ we consider are 1.0 and 0.9.

Finally, we employ the cluster combining method, introduced by Aiolfi and Timmermann (2006). In order to create the cluster combining forecasts, we form K clusters of equal size based on the past MSFE performance with the first one being that with the lowest MSFE values. Then, the first combination forecast is the average of the ARDL model forecasts in the first cluster. This procedure begins over the initial holdout period and goes through the end of the available out-of-sample period using a rolling window. In our analysis, we consider $K = 2, 3$, leading to the CL(2) and CL(3) combination schemes.

3.2.3 Statistical forecast evaluation

The accuracy of forecasts is evaluated by the Campbell and Thompson (2008) out-of-sample R^2 (R^2_{os}) and the Clark and West (2007) *CW-t* statistic. The R^2_{os} statistic measures the proportional reduction in mean squared forecast error (MSFE) for the unrestricted model forecast relative to the benchmark AR specification and is defined as follows:

$$R^2_{os} = 1 - (MSFE_1/MSFE_0) \tag{3.4}$$

where $MSFE_1/MSFE_0$ is the ratio of the MSFE of either the individual unrestricted models or any of the combination schemes over the MSFE of the benchmark AR model. When $R^2_{os} > 0$, the forecast of the unrestricted model is more accurate than the AR model's forecast, suggesting that the candidate variable/combination scheme can improve forecasts.

In order to statistically test the ability of a candidate variable or combination scheme to improve forecasts over the benchmark model, we use the Clark and West (2007) statistic, *CW-t*, for equal forecasting ability. The *CW-t* is a modified Diebold and Mariano (1995) and West (1996) statistic and tests the null hypothesis that both the unrestricted model and the restricted one have equal MSFEs ($H_0 : R^2_{os} = 0$, i.e. $MSFE_1 = MSFE_0$) against the one-sided (upper-tail) alternative hypothesis that the MSFE of the unrestricted model is smaller than the restricted one ($H_A : R^2_{os} > 0$, i.e. $MSFE_i < MSFE_0$). The statistic can be easily calculated by first defining the following quantity:

$$\widehat{f_{t+h}} = (z_{t+h} - \widehat{z}_{0,t+h})^2 - [(z_{t+h} - \widehat{z}_{1,t+h})^2 - (\widehat{z}_{0,t+h} - \widehat{z}_{1,t+h})^2] \tag{3.5}$$

The first two terms in (3.5) are the sample MSFEs of the unrestricted and restricted models respectively, while the last term is an adjustment term that normalizes the bias produced in the MSFE by the nonzero parameters of the unrestricted model. The *CW-t* statistic is the t-statistic for a zero coefficient calculated by regressing $\widehat{f_{t+h}}$ on a constant and has an asymptotic distribution well approximated by the standard normal. In this respect, if the t-statistic is greater than 1.282, we reject the null hypothesis that the models have equal MSFEs at 10% level of significance (for a one-sided test). For forecast horizons greater than 1, an autocorrelation consistent standard error should be employed. (Newey and West,1987). In extensive Monte Carlo simulations, Clark and West (2007) demonstrate that the CW-t statistic performs reasonably well in terms of size and power when comparing forecasts from linear nested models.

3.3 Empirical results

3.3.1 Data

The data used in our analysis are monthly observations for the period from July 1963 to December 2010 (570 observations). The series of interest are US long-term bond returns and stock market returns. Long-term bond returns are sourced from *Ibbotson's Stocks, Bonds, Bills and Inflation Yearbook* and stock market returns are returns on the S&P 500 index sourced from the Center for Research in Security Press (CRSP).[3]

The empirical factors employed are taken from Professor Kenneth French's website.[4] The SMB and HML factors are constructed from 6 value-weighted portfolios formed on size and book/market. Specifically, the intersections of the big/small and the value/neutral/growth portfolios form the 6 value-weighted portfolios, namely the small value (SV), small neutral (SN), small growth (SG), big value (BV), big neutral (BN) and big growth (BG) portfolio.[5] The average return of the three small portfolios minus that of the three big portfolios forms the SMB portfolio, whereas the average return of the two value portfolios minus the average return of the two growth portfolios forms the HML portfolio. The ST, LT and MOM factors are formed from 6 value-weighted portfolios formed on size and prior returns (small low, small medium, small high, big low, big medium, and big high). These prior-return portfolios are constructed on prior (1-1), (13-60), and (2-12) returns, respectively.[6] The average return on the two low prior-return portfolios (big and small) minus the average return on the two high prior-return portfolios (big and small) forms the ST and LT factors, while the MOM factor is the average of the returns on the two high prior-return portfolios (big and small) minus the average return on the two low prior-return portfolios

[3]Both series are available at Prof. Goyal's website at: http://www.hec.unil.ch/agoyal/.

[4]Tha data are downloadable at: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

[5]The breakpoint for year t for size is the median NYSE market equity at the end of June of year t, while for the book/market is the 30th and 70th NYSE percentile. The book/market ratio for June of year t is the book equity for the last fiscal year end in t-1 divided by market equity for December of t-1. The portfolios for July of year t to June of t+1 include all NYSE, AMEX, and NASDAQ stocks for which we have market equity data for December of t-1 and June of t, and (positive) book equity data for t-1.

[6]The breakpoint for the equity is the median NYSE market equity, while for the prior returns is the 30th and 70th NYSE percentile.

(big and small).

Following Fama and French (2012), we decompose all the factors (except for SMB) into their small and big counterparts. For example, the difference between the small (big) value portfolio and the small (big) growth one forms the HML_s (HML_b) portfolio, as follows:

$$\begin{aligned} HML\_s &= SV - SG \\ HML\_b &= BV - BG \end{aligned} \tag{3.6}$$

Decompositions of the LT, ST and MOM factors are formed according to the following formulas:

$$\begin{aligned} LT\_s &= SL - SH, \quad LT\_b = BL - BH \\ ST\_s &= SL - SH, \quad ST\_b = BL - BH \\ MOM\_s &= SH - SL, \quad MOM\_b = BH - BL \end{aligned} \tag{3.7}$$

For the SMB factor, we construct a value decomposition. Specifically, we decompose the size premium into its value, neutral and growth components, denoted by SMB_v, SMB_n and SMB_g, respectively, and calculated as follows:

$$\begin{aligned} SMB\_v &= SV - BV \\ SMB\_n &= SN - BN \\ SMB\_g &= SG - BG \end{aligned} \tag{3.8}$$

In addition, we employ fourteen financial variables, which have been shown in the literature to exhibit predictive ability on returns. The data for the financial variables, which are used by Rapach and Zhou (2012), are described in detail by Goyal and Welch (2008)[7]. These are the dividend/price ratio (DP), dividend yield (DY), earnings/price ratio (EP), dividend/earnings ratio (DE), stock variance (SVAR), book/market ratio (BM), net equity expansion (NTIS), treasury bill rate (TBL), long-term yield (LTY), long-term government bond returns (LTR),term spread (TMS), the default yield spread (DFY), default return spread (DFR), stock market return (SP500),and the inflation rate (INF).[8]

The total sample of the 570 monthly observations is divided into the estimation period consisting of the first $R = 380$ in-sample observations (July 1963 to February 1995) and the evaluation period with the last $P = 190$ (corresponding to the 1/3 of our sample) out-of-sample observations (March 1995 to December 2010). The holdout period for the combining methods that require one is set to 7 years (84 months) prior to the start of the out-of-sample evaluation period.

3.3.2 Forecasting U.S. bond returns

We begin our analysis by evaluating the forecasting ability of the candidate predictors for US long-term government bond returns for horizons ranging from 1 to 24 months. In Table 3.1, we report the R^2_{os} associated with individual ARDL models for both the empirical factors (Panel A) and the financial/macroeconomic variables (Panel B). Bold indicates a statistically superior forecast relative to the benchmark AR(1) model on the basis of the CW-t statistic.

As is evident, the momentum and short term reversal factors display significant

[7]This set of data can be downloaded from htts://www.hec.unil.ch/AGoyal.
[8]Please refer to Goyal and Welch (2008) for details on the construction and the sources of the series.

predictive ability for a variety of horizons, while the forecasting ability of the value premium, the size premium and the long-term reversal is rather muted.

Specifically, momentum displays significant predictive ability at horizons of 1-3, 6, and 12 months, while short-term reversal for horizons less than 3 months. Examining closely the performance of the size decompositions of the momentum factor, we note that the whole information is attributed to its small component, which appears to be a significant predictor for horizons ranging from 1 month to 1 year. On the other hand, the momentum of big companies improves bond forecasts in 1- and 6-month horizons ahead. Similarly, the small component of short-term reversal emerges as a significant predictor for all the horizons considered with the exception of the 6-month one. With respect to the big component of the short term reversal, its predictability appears at horizons of 2 and 3 months ahead.

Turning to the predictive ability of the financial variables employed (Panel B), we observe that only the stock market return improves bond return forecasts, performance which is evident only in the short run, at horizons of 1-3 months. Quite interestingly, the remaining financial variables exhibit hardly any significant predictive ability.

We next examine whether combining individual forecasts can result in superior predictive ability. We consider forecast combinations of (i) the five aggregate empirical factors (HML, SMB, MOM, LT and ST), reported in Panel A of Table 3.2, (ii) the eleven decomposed factors, reported in Panel B of Table 3.2 and (iii) the 14 financial variables, reported in Panel C of Table 3.2. As already discussed in Section 2, we employ the mean, median, trimmed mean, DMSE and cluster combining methods. For the DMSE, we employ two discount factors of $\psi = 0.90$ (DMSE(0.9)) and $\psi = 1.00$ (DMSE(1)), while for

Table 3.1: Out-of-sample performance of indiv.ARDL models -US bond

Panel A. Empirical factors								
Predictor	**1**	**2**	**3**	**6**	**9**	**12**	**18**	**24**
HML	0.103	-0.710	-0.705	-1.625	-4.572	-7.880	-6.604	-2.450
SMB	-0.013	-3.076	-1.663	-4.906	-5.405	-7.139	-10.113	-19.116
MOM	**1.328**	**1.169**	**1.238**	**2.084**	0.865	**1.114**	-2.744	-3.021
LT	-0.295	-0.804	-1.460	-1.815	-0.106	-0.734	-0.430	-0.664
ST	**1.446**	**2.374**	**3.255**	-0.822	-0.818	-0.278	0.423	0.362
HML_b	-0.340	-0.577	-0.517	-0.258	-0.248	-1.775	-0.410	-0.936
HML_s	-0.377	-0.583	-0.669	-0.919	-0.613	-15.025	-10.867	-16.900
SMB_g	-0.179	-5.168	-4.604	-4.303	-5.374	**0.031**	-2.470	-9.054
SMB_n	0.091	-3.820	-4.534	-1.452	-2.533	-1.593	0.137	-0.378
SMB_v	-0.323	-0.167	-4.660	-2.279	-5.076	-2.469	-1.776	-4.679
MOM_b	**0.533**	0.144	0.444	**1.608**	-0.132	-13.215	-48.097	-8.835
MOM_ s	**1.930**	**2.351**	**1.968**	**1.835**	**1.184**	**1.796**	-0.211	0.261
LT_b	-0.131	-0.157	-0.273	-0.420	-0.494	0.178	-0.181	0.189
LT_s	-1.354	-3.479	-4.959	-4.416	-0.031	-2.565	-0.614	-3.516
ST_b	0.592	**1.433**	**1.718**	-1.259	-1.410	-1.302	-0.205	-0.233
ST_s	**1.995**	**2.694**	**4.042**	-1.516	**0.858**	**2.044**	**1.527**	**1.577**
Panel B. Financial variables								
Predictor	**1**	**2**	**3**	**6**	**9**	**12**	**18**	**24**
DP	-2.899	-1.851	-6.431	-19.903	<-20	<-20	<-20	<-20
DY	-2.749	-5.354	-7.519	-16.785	<-20	<-20	<-20	<-20
EP	-0.845	-5.853	-7.692	-3.765	-7.797	-12.603	<-20	<-20
DE	-18.583	-43.518	-2.228	-2.225	-6.869	-4.849	-4.171	-7.814
SVAR	-0.031	0.702	-1.837	-2.117	-3.302	-5.249	-8.999	<-20
BM	-0.311	-1.780	-2.310	-1.443	-1.569	-0.246	-0.110	-16.653
NTIS	-2.164	-4.310	-7.041	-10.494	-18.703	<-20	<-20	-16.078
TBL	-1.025	-2.062	-3.458	-7.604	-14.095	<-20	<-20	<-20
LTY	-0.953	-2.229	-5.124	-12.607	<-20	<-20	<-20	<-20
SP500	**2.662**	**3.186**	**2.010**	-0.413	-0.430	0.084	-1.470	-0.951
TMS	-0.153	-0.003	-0.734	-0.844	-4.128	-7.831	-14.524	<-20
DFY	-3.552	-4.332	-7.326	-16.942	<-20	<-20	<-20	<-20
DFR	-1.881	-2.088	-4.054	-10.076	-19.944	<-20	<-20	<-20
INFL	0.444	-0.526	-0.635	-0.931	-2.984	-3.227	-0.827	-0.136

Notes: (i) The table reports the out-of-sample R^2_{os} of the individual ARDL models relative to the AR benchmark.

(ii) Bold entries indicate significance at the 10% significance level according to the $CW - t$ statistic, which tests the null hypothesis: R^2_{os} =0 against the alternative: R^2_{os}>0.

the cluster combining method, we employ 2 clusters (CL(2)) and 3 clusters (CL(3)). We observe that when the five aggregate (HML, SMB, MOM, LT, and ST) factors are combined, out-of-sample predictive ability appears only short term. Specifically, at horizons of 1-3 months, the median, trimmed mean and CL(2) combining method forecasts display significant forecasting ability, while the mean, DMSE(1), DMSE(0.9) and CL(3) combining methods exhibit predictive ability at horizons of 1 and 3 months.

Our findings for the combination forecasts constructed with the decomposed factors (reported in Panel B) are quite interesting since predictability reaches the 18-month horizon. More in detail, the median and trimmed mean combining methods improve forecasts at horizons of 3 to 18 months and for 1-3, 6, 12, and 18 months, respectively. CL(2) exhibits significant forecasting ability on bond returns at horizons of 1, 2, and 9 months, while CL(3) improves forecasts for 1-3 months and 9 months ahead. Moreover, the mean and DMSE(0.9) combining schemes are associated with superior predictive ability only at the 1-month horizon, while the DMSE(1) one for horizons of 1 and 12 months. More importantly, there are no benefits associated with combination forecasts of financial variables, as suggested by Panel C of Table 3.2.

Overall, our findings so far suggest that combining empirical factors can lead to improved predictability for bond returns and that size and value decompositions of the empirical factors can further enhance it. This latter finding suggests that the disaggregated factors contain significant information for the evolution of future bond returns which is rather hidden when considering aggregate factors.[9]

[9] Unreported results suggest that combinations of both the empirical factors and the financial variables fail to improve the accuracy of forecasts relative to the performance of the AR model. This set of results are available from the authors upon request.

Table 3.2: Out-of-sample performance of comb.methods -U.S. bond

Panel A. Five empirical factors								
Method	**1**	**2**	**3**	**6**	**9**	**12**	**18**	**24**
Mean	**0.824**	0.543	**1.188**	-0.341	-0.768	-0.839	-2.030	-2.539
Median	**0.934**	**0.667**	**1.102**	-0.227	-0.273	-0.131	-1.650	-2.199
Tr. mean	**0.844**	**0.540**	**1.184**	-0.008	-0.432	-0.577	-1.911	-1.606
DMSE(1)	**0.822**	0.548	**1.190**	-0.323	-0.758	-0.840	-2.022	-2.514
DMSE(0.9)	**0.827**	0.567	**1.179**	-0.366	-0.791	-0.917	-1.973	-2.152
CL(2)	**0.974**	**1.217**	**1.729**	0.269	-0.320	-1.207	-1.548	-0.896
CL(3)	**1.053**	0.381	**1.235**	0.097	-1.931	-3.067	-2.296	-0.871
Panel B. Eleven decomposed factors								
Method	**1**	**2**	**3**	**6**	**9**	**12**	**18**	**24**
Mean	**0.554**	0.357	0.529	0.338	0.414	1.464	0.203	-0.562
Median	0.006	-0.008	**0.542**	**0.522**	**0.630**	**1.169**	**0.680**	0.076
Tr. mean	**0.353**	**0.568**	**0.759**	**0.656**	0.462	**1.338**	**1.025**	-0.336
DMSE(1)	**0.555**	0.377	0.560	0.358	0.453	**1.407**	0.300	-0.389
DMSE(0.9)	**0.560**	0.268	0.458	0.255	0.451	0.868	0.124	-0.766
CL(2)	**0.736**	**0.985**	0.358	-0.087	**1.140**	-0.057	-1.121	-1.588
CL(3)	**0.980**	**1.120**	**1.099**	-0.981	**1.017**	-2.234	-2.262	-1.227
Panel C. Financial variables								
Method	**1**	**2**	**3**	**6**	**9**	**12**	**18**	**24**
Mean	-0.654	-1.358	-0.529	-2.707	-4.788	-6.910	-10.098	<-20
Median	-0.419	-0.871	-1.458	-2.629	-3.572	-4.940	-9.041	-12.924
Tr. mean	-0.122	-0.311	-0.903	-3.071	-5.556	-7.098	-9.534	-18.198
DMSE(1)	-0.693	-1.393	-0.524	-2.524	-4.383	-6.021	-7.232	-13.543
DMSE(0.9)	-1.074	-1.780	-1.327	-4.220	-8.241	-11.206	-12.412	-18.159
CL(2)	-2.540	-1.316	-2.371	-3.714	-4.339	-4.286	-2.943	-8.036
CL(3)	-4.711	-1.473	-3.471	-6.037	-8.342	-9.982	-6.133	-12.111

Notes: (i) The table reports the out-of-sample R^2_{os} of the individual ARDL models relative to the AR benchmark.
(ii) Bold entries indicate significance at the 10% significance level according to the $CW-t$ statistic, which tests the null hypothesis: $R^2_{os}=0$ against the alternative: $R^2_{os}>0$.

3.3.3 Forecasting US stock returns

We now examine whether the forecasting ability of the candidate predictors is maintained for US stock returns (S&P500 index returns). Panel A (Table 3.3) reports the out-of-sample performance of the empirical factors and their components, while Panel B reports the related findings for the financial variables.

Among the 30 candidate predictors, the momentum factor emerges as the most powerful one, as it improves forecasts over the AR benchmark at horizons of 6, 9, 12, and 18 months. This performance is consistent with the one for bond returns and is mainly attributed to the momentum of big companies. Moreover, at a 3-month horizon, the long-term reversal factor along with both its components displays significant predictive ability. Turning to the financial variables, we have to note that their ability is rather weak and limited to horizons of 6-12 months and 18-24 months for the book to market ratio and the term spread, respectively.

Given the rather limited individual variable predictability, we do not expect combination methods to work impressively well, since they aggregate over weak predictors. Our findings, reported in Table 3.4 (Panels A to C), support this conjecture.

Specifically, when considering combination forecasts of the five empirical factors, we find improved forecasting ability only at the 3-month horizon and on the basis of the mean, DMSE and CL(3) combination schemes.

Similar findings pertain when the decomposed factors are considered (Panel B), since significant forecasting ability is evident for the trimmed mean combining method at the horizon of 3 months, as well. Finally, as expected, combination forecasts of financial

Table 3.3: Out-of-sample performance of indiv.ARDL models -US stock

Panel A. Empirical factors								
Predictor	**1**	**2**	**3**	**6**	**9**	**12**	**18**	**24**
HML	-0.168	0.098	-0.302	-0.103	-0.405	-0.716	-0.550	-0.510
SMB	-0.476	-0.161	0.373	-2.351	0.199	-0.118	-0.084	0.190
MOM	-0.777	-0821	0.247	**0.553**	**0.744**	**0.782**	**0.545**	-0.019
LT	0.025	0.906	**1.836**	1.079	-0.182	0.103	0.016	-0.039
ST	-0.768	-0.583	-0.438	-0.436	-0.084	-0.312	-0.277	-0.259
HML_b	0.016	0.078	-0.383	-1.174	-0.365	-0.311	-0.034	-0.392
HML_s	-0.720	-0.517	-0.290	-0.444	-0.294	-0.219	-0.364	-0.559
SMB_g	-1.307	-1.065	-0.762	-1.388	-0.718	-1.155	-0.301	-0.250
SMB_n	-0.518	-0.207	-0.210	-0.428	-0.260	-0.517	-0.129	-0.210
SMB_v	0.093	0.033	-0.269	-0.623	-0.667	-0.848	-0.269	-0.534
MOM_b	-0.532	-0.231	0.212	**0.816**	**0.859**	**0.899**	**0.595**	-0.056
MOM_ s	-0.976	-1.061	-0.006	-0.086	0.209	0.143	0.072	-0.138
LT_b	-0.248	0.444	**0.916**	1.032	-0.047	0.060	-0.057	-0.387
LT_s	0.300	0.657	**1.181**	0.261	-1.154	0.211	0.068	-0.080
ST_b	0.003	-0.351	-0.341	-0.479	-0.224	-0.410	-0.551	-0.602
ST_s	-0.960	-0.802	-0.465	-0.370	-0.089	-0.534	-0.471	-0.610
Panel B. Financial variables								
Predictor	**1**	**2**	**3**	**6**	**9**	**12**	**18**	**24**
DP	-2.182	-3.740	-5.791	-12.380	-19.530	-25.158	-33.246	-38.658
DY	-2.090	-3.430	-4.872	-9.531	-17.119	-21.031	-31.108	-39.367
EP	-0.935	-4.014	-9.161	-9.452	-16.894	-29.060	-11.259	-4.831
DE	-2.897	-12.300	-19.963	-26.787	-27.196	-47.349	-17.752	-15.982
SVAR	0.226	-2.549	-8.813	-4.349	-2.327	-2.477	-6.047	-4.312
BM	-0.275	-0.200	0.139	**0.603**	**1.114**	**0.837**	0.432	-0.271
NTIS	-3.299	-6.284	-10.276	-21.521	-28.734	-30.707	-33.192	-26.437
TBL	-1.178	-0.895	-1.583	-3.296	-2.747	-5.189	-5.748	-5.234
LTY	-0.973	-1.005	-0.710	-1.364	-0.835	-1.174	-0.560	**1.655**
SP500	-0.391	-1.374	-0.841	-2.344	-0.636	-0.948	-0.852	-0.720
TMS	-1.256	-1.964	-2.971	-3.861	-3.854	-1.598	**3.186**	**2.245**
DFY	-3.800	-6.382	-8.097	-8.216	-6.336	-4.923	-4.869	-12.719
DFR	0.032	-0.373	-0.444	-0.345	-0.767	-1.070	-4.318	-7.006
INFL	-2.226	-3.162	-4.528	-0.590	-0.089	0.337	-1.188	-1.814

Notes: (i) The table reports the out-of-sample R^2_{os} of the individual ARDL models relative to the AR benchmark.
(ii) Bold entries indicate significance at the 10% significance level according to the $CW-t$ statistic, which tests the null hypothesis: $R^2_{os}=0$ against the alternative: $R^2_{os}>0$.

Table 3.4: Out-of-sample performance of comb.methods -U.S. stock

Panel A. Five empirical factors								
Method	**1**	**2**	**3**	**6**	**9**	**12**	**18**	**24**
Mean	-0.133	0.244	**0.612**	0.169	0.319	0.180	0.104	0.006
Median	-0.340	-0.324	0.112	-0.228	0.159	-0.315	-0.314	0.052
Tr. mean	-0.202	-0.107	0.449	-0.201	0.000	-0.173	-0.031	0.097
DMSE(1)	-0.138	0.246	**0.613**	0.170	0.317	0.179	0.106	0.011
DMSE(0.9)	-0.162	0.236	**0.606**	0.197	0.338	0.183	0.091	-0.008
CL(2)	-0.831	-0.069	0.635	0.624	0.847	0.487	0.208	0.134
CL(3)	-1.017	-0.140	**1.260**	0.323	0.261	0.352	-0.074	-0.048
Panel B. Eleven decomposed factors								
Method	**1**	**2**	**3**	**6**	**9**	**12**	**18**	**24**
Mean	-0.140	0.043	0.232	0.130	-0.018	-0.023	0.082	-0.158
Median	0.075	-0.021	0.058	0.068	0.201	-0.056	0.110	-0.130
Tr. mean	-0.088	0.056	**0.285**	0.007	-0.043	-0.100	0.048	-0.156
DMSE(1)	-0.141	0.046	0.235	0.133	-0.018	-0.019	0.088	-0.158
DMSE(0.9)	-0.141	0.046	0.229	0.150	-0.002	-0.013	0.099	-0.164
CL(2)	-0.176	-0.270	0.191	-0.040	-0.037	0.053	0.069	-0.269
CL(3)	-0.079	-0.394	0.073	0.034	-0.121	0.098	0.170	-0.178
Panel C. Financial variables								
Method	**1**	**2**	**3**	**6**	**9**	**12**	**18**	**24**
Mean	-0.736	-2.057	-3.462	-4.770	-6.033	-6.778	-7.053	-8.032
Median	-0.182	-0.344	-0.619	-0.487	-0.430	-0.841	-2.403	-4.436
Tr. mean	-0.739	-1.475	-2.481	-2.959	-3.543	-3.637	-4.722	-6.102
DMSE(1)	-0.727	-2.021	-3.317	-4.363	-4.866	-5.396	-5.053	-5.355
DMSE(0.9)	-0.624	-1.880	-2.958	-3.624	-4.718	-6.017	-6.448	-6.946
CL(2)	-0.316	-1.570	-2.229	-2.110	-1.340	-4.130	-3.459	-3.817
CL(3)	-0.279	-2.222	-2.518	-1.737	-1.258	-2.862	-2.660	0.660

Notes: (i) The table reports the out-of-sample R^2_{os} of the individual ARDL models relative to the AR benchmark.
(ii) Bold entries indicate significance at the 10% significance level according to the $CW-t$ statistic, which tests the null hypothesis: $R^2_{os}=0$ against the alternative: $R^2_{os}>0$.

variables do not improve stock returns forecasts over the AR benchmark.

To sum up, the evidence in this section suggests that the proposed empirical factors exhibit strong forecasting ability for US bond returns and are weaker when it comes to stock returns. Their size and value decompositions further enhance their ability especially when combination of forecasts are considered. Given that statistical significance does not always imply economic significance, we next assess whether this forecasting ability can be useful from an asset allocation perspective.

3.4 Asset allocation benefits of combination forecasts

A utility-based evaluation of forecasts was first proposed by West et al. (1993) in assessing exchange rate volatility forecasts (see also Abhyankar et al. (2005), Della Corte et al. (2009) and Rime et al. (2010)). Following Fleming et al. (2001) and Della Corte et al. (2008, 2009), Thorton and Valente (2012) quantify how much a risk-averse investor is willing to pay to switch from a dynamic portfolio strategy based on a model with no predictable bond excess returns to a model that uses either forward spreads or the term structure of forward rates. Campbell and Thomson (2008), Rapach et al. (2010), Ferreira and Santa-Clara (2011), Dangl and Halling (2012) and Neely et al. (2013) provide evidence that investors who rely on equity premium forecasts based on economic variables can gain profit relative to those who just rely on the historical average forecast.

In our analysis, we investigate whether the forecasting ability of the proposed empirical factors/combination schemes can lead to significant economic gains for a mean-variance investor, who incorporates them to asset allocation decisions.

3.4.1 The framework

We consider a mean-variance investor with relative risk aversion (RRA), γ, who rebalances her portfolio every month. Her portfolio maximization problem, described in detail in Campbell and Viceira (2002), is the following:

$$\max_{w_t} \mathbf{w}_t'(E_t\mathbf{R}_{t+h} - R_{f,t\to t+h}\boldsymbol{\iota}) - \frac{\gamma}{2}\mathbf{w}_t'\boldsymbol{\Sigma}_{t+h}^{-1}\mathbf{w}_t \tag{3.9}$$

where $E_t\mathbf{R}_{t+h} - R_{f,t\to t+h}$ is the vector of expected excess returns on the risky assets over the risk-free interest rate ($R_{f,t\to t+h}$) prevailing from time t to $t+h$, $\boldsymbol{\iota}$ is a vector of ones, $\mathbf{w}_t$ is the vector of portfolio weights on risky assets, and $\mathbf{w}_t'\boldsymbol{\Sigma}_{t+h}^{-1}\mathbf{w}_t$ is the expected variance of the portfolio return. The solution to this maximization problem is:

$$w_{i,t} = \frac{1}{\gamma}\boldsymbol{\Sigma}_{t+h}^{-1}(E_t\mathbf{R}_{t+1} - R_{f,t}\boldsymbol{\iota}), i = b, s \tag{3.10}$$

where b, s stand for bond and stock returns, respectively.

The conditional expectation $E_t\mathbf{R}_{t+h}$ is given by the bond and stock return combination forecasts for each horizon and combining scheme we employed in the previous section. The expected variance/covariance matrix for bond and stock market returns, $\boldsymbol{\Sigma}_{t+h}$, is computed using a rolling window of 40 past observations.[10] The optimal weights allocated to government bonds and the stock market are winsorized to $0 < w_{i,t} < 1.5$, thus preventing short selling and extreme allocation to any of the risky assets. The investor's taste of risk, controlled by the RRA coefficient, is set equal to 3 and 5. Having estimated the optimal weights, the resulting portfolio return is equal to:

$$R_{p,t} = (1 - w_{1,t} - w_{2,t}) * R_{f,t} + w_{1,t} * R_{b,t} + w_{2,t} * R_{s,t} \tag{3.11}$$

[10] Campbell and Thomson (2008) and Goh et al. (2013) consider a 5-year rolling window of past returns.

where $R_{b,t}$ and $R_{s,t}$ are the realized bond and stock returns at each point of time, t, over the out-of-sample evaluation period (P observations). Over the forecast evaluation period the investor with initial wealth of $W_o = 1$ realizes an average utility of

$$\overline{U} = \frac{1}{P}\sum_{t=1}^{P}\left[(1 + R_{p,t}) - \frac{\gamma}{2} * \left(R_{p,t} - \overline{R}_p\right)^2\right] \tag{3.12}$$

where $\overline{R}_p$ denotes the average portfolio return over the evaluation period. In a similar way, we calculate the utility associated with the benchmark AR specification, given by the following equation:

$$\overline{U}^{AR} = \frac{1}{P}\sum_{t=1}^{P}\left[\left(1 + R_{p,t}^{AR}\right) - \frac{\gamma}{2} * \left(R_{p,t}^{AR} - \overline{R}_p^{AR}\right)^2\right] \tag{3.13}$$

where $R_{p,t}^{AR}$ refers to the portfolio returns constructed based on the benchmark model forecasts and $\overline{R}_p^{AR}$ is the respective average portfolio return over the evaluation period. The difference (ΔU) between the average utility realized from the proposed specification and the one of the benchmark specification is calculated as follows:

$$\Delta U = \overline{U} - \overline{U}^{AR} \tag{3.14}$$

It can be interpreted as the annual percentage portfolio management fee that an investor would be willing to pay to have access to our proposed forecasting methodology relative to the AR benchmark.

We also employ an alternative economic evaluation measure, which is the manipulation-proof performance measure (MPPM), proposed by Goetzmann et al. (2007). This measure takes into account the effect of non-normality, the underestimation of the performance of dynamic strategies and the choice of the utility function. It can be interpreted as the portfolio's premium return after adjusting for risk and is defined as follows:

$$MPPM = \frac{1}{1-\gamma}\ln\left[\frac{1}{P}\sum_{t=1}^{P}\left(\frac{1+R_{p,t}}{1+R_{f,t}}\right)^{1-\gamma}\right] \tag{3.15}$$

The proposed specification performs better than the benchmark one when the difference between the $MPPM$ of the proposed model and that of the benchmark one, Θ, defined as follows:

$$\Theta = \frac{1}{1-\gamma}\ln\left[\frac{1}{P}\sum_{t=1}^{P}\left(\frac{1+R_{p,t}}{1+R_{f,t}}\right)^{1-\gamma}\right] - \frac{1}{1-\gamma}\ln\left[\frac{1}{P}\sum_{t=1}^{P}\left(\frac{1+R_{p,t}^{AR}}{1+R_{f,t}}\right)^{1-\gamma}\right] \tag{3.16}$$

is positive.

3.4.2 Asset allocation: empirical results

We consider a mean-variance investor who allocates her wealth among bonds, stocks and the risk-free interest rate, and rebalances her portfolio monthly over the 1995:03 - 2010:12 out-of-sample evaluation period.[11] As already mentioned, we assume two values for the investor's RRA, $\gamma = 3$ and $\gamma = 5$, and calculate the variance covariance matrix between stocks and bond returns by employing a rolling 40-month window of past observations. Consistent with the statistical evaluation, we assess the economic value for horizons of 1, 3, 6 and 12 months.

In Table 3.5 (Panels A to D) we report the performance fees (ΔU) that a mean-variance investor would be willing to pay to have access to our models along with the

[11]The risk free interest rate considered is the 1-month US Treasury Bill.

Table 3.5: Asset allocation benefits for a US investor

Panel A. Horizon 1						
	Five Factors		**Eleven Factors**		**Fin.Variables**	
Method	**ΔU**	**Θ**	**ΔU**	**Θ**	**ΔU**	**Θ**
Mean	**1.348**	**1.630**	**1.417**	**1.652**	-1.591	-1.673
Median	**0.351**	**0.570**	**0.567**	**0.604**	-0.650	-0.748
Tr.Mean	**0.845**	**1.051**	**0.765**	**0.888**	-2.291	-2.576
DMSE(1)	**1.340**	**1.622**	**1.425**	**1.662**	-1.502	-1.580
DMSE(0.9)	**1.235**	**1.514**	**1.395**	**1.625**	-0.994	-1.005
CL(2)	**1.524**	**1.867**	**2.178**	**2.507**	**0.111**	**0.295**
CL(3)	**2.299**	**2.728**	**3.029**	**3.396**	-0.211	-0.049
Panel B. Horizon 3						
	Five Factors		**Eleven Factors**		**Fin.Variables**	
Method	**ΔU**	**Θ**	**ΔU**	**Θ**	**ΔU**	**Θ**
Mean	**2.326**	**3.309**	**2.155**	**2.432**	-4.883	-4.860
Median	**1.998**	**2.737**	-0028	**0.038**	**1.427**	**1.803**
Tr.Mean	**1.908**	**2.684**	**1.871**	**2.071**	-4.011	-3.733
DMSE(1)	**2.323**	**3.310**	**2.110**	**2.401**	-4.297	-4.267
DMSE(0.9)	**2.200**	**3.136**	**2.248**	**2.513**	-3.231	-3.047
CL(2)	**4.807**	**6.281**	**1.009**	**1.596**	-1.859	-0.767
CL(3)	**6.414**	**8.113**	**1.535**	**2.498**	-2.032	-1.041
Panel C. Horizon 6						
	Five Factors		**Eleven Factors**		**Fin.Variables**	
Method	**ΔU**	**Θ**	**ΔU**	**Θ**	**ΔU**	**Θ**
Mean	-1.440	-1.197	-0.489	-0.343	-14.087	-13.411
Median	-2.347	-2.064	**0.643**	**0.636**	-0.043	**2.418**
Tr. Mean	-1.877	-1.681	**0.292**	**0.402**	-11.103	-9.465
DMSE(1)	-1.375	-1.117	-0.504	-0.360	-11.525	-10.997
DMSE(0.9)	-1.303	-1.084	-0.496	-0.351	-8.546	-8.254
CL(2)	**1.099**	**1.509**	-0.638	-0.474	-0.853	**3.258**
CL(3)	**3.055**	**4.277**	-0.567	-0.386	-1.787	**3.265**
Panel D. Horizon 12						
	Five Factors		**Eleven Factors**		**Fin.Variables**	
Method	**ΔU**	**Θ**	**ΔU**	**Θ**	**ΔU**	**Θ**
Mean	-2.529	-2.393	-1.313	-0.948	-16.267	-14.031
Median	-1.209	-1.112	-0.877	-0.754	-1.632	-1.360
Tr. Mean	-2.108	-1.998	-0.984	-0.726	-13.491	-11.423
DMSE(1)	-2.510	-2.373	-1.310	-0.997	-12.796	-10.784
DMSE(0.9)	-2.306	-2.171	-1.480	-1.051	-12.909	-11.003
CL(2)	-2.569	-2.482	**0.979**	**0.809**	-5.939	-5.114
CL(3)	-3.491	-3.654	-0.261	**0.045**	-1.680	-1.614

Notes: (i) The table reports the average utility gain (ΔU) and the difference between the manipulation-proof performance measure (Θ) of the proposed specification relative to the benchmark AR model.
(ii) Figures are reported in annualized percentage points.
(iii) Portfolio weights are constrained to lie between 0 and 1.5.

risk-adjusted measure Θ for an investor with a risk aversion coefficient of 3. Our findings for an investment horizon of 1 month are given in Panel A. Overall, combination forecasts of both the aggregate factors and the disaggregated ones always generate positive utility gains. Utility gains range from 0.765% per year (Trimmed mean combination forecast of the disaggregate factors) to 3.029% per year (CL(3) combination method of the disaggregate factors). The best performance is achieved by the CL(3) combination method, closely followed by CL(2). However, the simplest combining method, i.e. the mean one, attains a satisfactory performance of 1.348% and 1.417% for the aggregate and disaggregated factors, respectively. Similar findings pertain when forecasts are evaluated on the basis of the risk-adjusted measure Θ. More importantly, combination forecasts of financial variables fail to generate profits to the investor in excess of the ones already contained in the benchmark AR model, with the exception of the CL(2) method.

Turning to the forecast horizon of 3 months (Panel B), our findings suggest that combination forecasts of either the five or the eleven factors can generate positive utility gains that reach 6.414% for the CL(3) method, with the exception of the median combining scheme of the disaggregate factors. When combining the aggregate factors, the cluster combining methods rank first followed by the mean and the DMSE ones. However, on the basis of the disaggregated factors, the mean and DMSE methods rank first followed by the trimmed mean and the cluster ones. Our findings with respect to Θ, are quite similar.

Moreover, similar to the 1-month forecast horizon, all the combining methods (with the exception of the median one) point to negative gains and thus greater average utility for the AR benchmark compared to the combination methods, when only financial variables are incorporated. Longer investment horizons of 6 months (Panel C) and 12 months (Panel D) do not consistently generate profits to the investor. Specifically, for the 6-month horizon an investor would be willing to have access to the forecasts generated by the cluster combinations of the five factors or the median and trimmed mean combinations of the eleven factors. The difference in MPPMs, Θ, points to benefits when a pool of the financial variables is employed. Specifically, Θ is positive at the horizon of 6 months for the median and the CL combination methods generating premium returns of up to 3.265% per year. Turning to the 12-month horizon, we note that the ability of the proposed models to generate utility gains to an investor is rather limited to the case of the CL(2) combination of the disaggregated factors.

When we allow for a more conservative investor, our findings are qualitatively similar. More in detail, Table 3.6 reports the respective findings for an investor with RRA of 5. For a short-term horizon of 1 and 3 months (Panels A and B), the investor would be willing to pay a performance fee to utilize forecasts from our combining methods on the basis of the empirical factors (both aggregate and disaggregate ones). As expected, these fees are lower compared to the ones for the less risk averse investor (Table 3.5).

On the other hand, when turning to the medium investment horizon of 6 months (Panel C), the investor can still benefit from our combination forecasts of the empirical factors and in some cases of the financial variables, as well.

The combination methods of aggregate factors are all successful and generate fees up to 1.217% (CL(2) method), whereas when disaggregate factors are considered, all but the cluster combining methods accrue benefits of up to 1.144% to the investor. More

Table 3.6: Asset allocation benefits for a US investor

Panel A. Horizon 1						
	Five Factors		**Eleven Factors**		**Fin.Variables**	
Method	**ΔU**	**Θ**	**ΔU**	**Θ**	**ΔU**	**Θ**
Mean	**1.160**	**1.374**	**0.803**	**0.979**	-2.017	-2.335
Median	**0.274**	**0.448**	**0.139**	**0.164**	-1.150	-1.534
Tr.Mean	**0.613**	**0.779**	**0.307**	**0.406**	-2.418	-2.951
DMSE(1)	**1.156**	**1.370**	**0.806**	**0.983**	-1.973	-2.286
DMSE(0.9)	**1.104**	**1.315**	**0.795**	**0.967**	-1.703	-1.949
CL(2)	**1.367**	**1.626**	**1.252**	**1.510**	-1.028	-1.062
CL(3)	**1.964**	**2.269**	**1.757**	**2.053**	-1.209	-1.312
Panel B. Horizon 3						
	Five Factors		**Eleven Factors**		**Fin. Variables**	
Method	**ΔU**	**Θ**	**ΔU**	**Θ**	**ΔU**	**Θ**
Mean	**1.927**	**2.086**	**1.299**	**1.118**	-4.669	-5.906
Median	**1.152**	**1.141**	**0.141**	**0.132**	**1.378**	**1.123**
Tr.Mean	**1.615**	**1.745**	**1.128**	**0.989**	-3.509	-4.361
DMSE(1)	**1.921**	**2.077**	**1.260**	**1.086**	-4.264	-5.553
DMSE(0.9)	**1.836**	**1.976**	**1.361**	**1.177**	-2.785	-3.942
CL(2)	**3.274**	**2.876**	**0.316**	**0.097**	-0.688	-0.939
CL(3)	**3.893**	**3.284**	**0.134**	-0.202	-1.209	-2.241
Panel C. Horizon 6						
	Five Factors		**Eleven Factors**		**Fin. Variables**	
Method	**ΔU**	**Θ**	**ΔU**	**Θ**	**ΔU**	**Θ**
Mean	**0.206**	**-0.646**	**0.127**	-0.408	-9.437	-10.347
Median	**0.334**	**2.267**	**1.144**	**1.319**	**2.842**	**5.186**
Tr. Mean	**0.179**	**1.324**	**0.347**	-0.125	-4.439	-2.333
DMSE(1)	**0.190**	-0.719	**0.121**	-0.425	-7.744	-8.478
DMSE(0.9)	**0.415**	-0.148	**0.237**	-0.163	-3.407	-1.870
CL(2)	**1.217**	-1.182	-0.298	-2.008	**2.141**	**6.996**
CL(3)	**1.134**	-5.571	-0.398	-4.087	**1.359**	**5.934**
Panel D. Horizon 12						
	Five Factors		**Eleven Factors**		**Fin. Variables**	
Method	**ΔU**	**Θ**	**ΔU**	**Θ**	**ΔU**	**Θ**
Mean	-2.657	<-20.0	-3.041	<-20.0	-11.074	**6.670**
Median	-2.848	<-20.0	-3.152	<-20.0	**5.063**	**>20.0**
Tr. Mean	-2.771	<-20.0	-2.900	<-20.0	-2.722	**>20.0**
DMSE(1)	-2.662	<-20.0	-3.061	<-20.0	-8.161	**12.412**
DMSE(0.9)	-2.501	<-20.0	-2.878	<-20.0	-10.095	**7.504**
CL(2)	-1.839	-11.166	-2.007	<-20.0	-0.626	**>20.0**
CL(3)	-2.238	<-20.0	-2.670	-15.469	**4.206**	**>20.0**

Notes: (i) The table reports the average utility gain (ΔU) and the difference between the manipulation-proof performance measure (Θ) of the proposed specification relative to the benchmark AR model.
(ii) Figures are reported in annualized percentage points.
(iii) Portfolio weights are constrained to lie between 0 and 1.5.

importantly, median and cluster combinations of the financial variables can generate positive utility gains of 2.842%. Employing Θ leads to similar findings for the financial variables pool but not for the factor ones. In some cases positive utility gains are associated with negative Θ*s*. The opposite is true for the longer horizon of 12 months and the case of the pool of financial variables. Specifically, while positive utility gains and Θ*s* are associated with the median and CL(3) methods, positive Θ*s* prevail for all the combination methods at hand.

3.5 International evidence

So far we have provided evidence for significant forecasting ability of combination forecasts of empirical factors for US bond and stock returns both in statistical and economic evaluation terms. In this section, we test whether these factors exhibit similar forecasting ability on European and Japanese stock and bond returns.

We use monthly observations of the empirical factors for the period November 1990 to April 2012.[12] European and Japanese market returns along with the aggregate factor returns and their decompositions are taken from Professor Kenneth French's website.[13] Long-term bond returns are downloaded from DataStream.[14] The total sample consists of 258 observations, 86 are reserved for the out-of-sample evaluation period.[15] The horizons examined are 1-24 months, but, for brevity, we present the results for horizons of 1-3, 6, 9,

[12] All returns are given in U.S. dollars.

[13] The European factors and portfolios include Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, and the United Kingdom.

[14] The series concerning the long-term government bond returns for Europe is the series: BOFA ML PAN EUROPE GVT 10+Y ($) - TOT RETURN IND, while for Japan is the series:BOFA ML JAPAN GVT 10+Y ($) - TOT RETURN IND.

[15] The holdout period is 3 years (36 months).

12, 18 and 24 months. As previously, we assess the out-of-sample forecasting ability of the proposed models from an asset allocation perspective, as well. In particular, we consider a mean-variance investor who allocates her wealth among bonds, stocks and the risk-free interest rate and rebalances her portfolio monthly based on information through period t over the 2005:03 - 2012:04 out-of-sample evaluation period. The investor's relative risk aversion (RRA) is set equal to $\gamma = 3$.

3.5.1 Forecasting European bond and stock returns

In Table 3.7 (Panels A to C) we report the forecasting ability of empirical aggregate and decomposed factors for European bond returns along with combinations of them. The only factor that appears valuable in forecasting bond returns is the HML factor who is significant both in the short run and in the long run. Specifically, the aggregate value premium is a significant predictor at horizons of 1-3, 18 and 24 months, while its big component is successful at horizons of 2, 3, 6 and 24 months. On the other hand, the predictive ability of the small component (HML_s) is restricted only at the horizon of 1 month. The remaining factors exhibit hardly any significant forecasting ability on bond returns.

Turning to combination forecasts of the aggregate factors (Panel B), we have to note that our findings vary with the combination method employed.

Specifically, the median and trimmed mean combining methods display significant forecasting ability at horizons of 3 and 9 months, while the DMSE combining methods

Table 3.7: Out-of-sample performance - European bond returns

Panel A. Individual ARDL models								
Predictor	**1**	**2**	**3**	**6**	**9**	**12**	**18**	**24**
HML	**3.170**	**1.745**	**3.150**	1.359	-0.287	-0.744	**3.227**	**2.793**
SMB	-0.658	-0.463	0.322	-0.293	0.140	-0.302	-1.309	-1.400
MOM	-1.756	-2.424	-5.312	-5.983	-0.371	-1.469	-3.036	-1.391
HML_b	1.001	**0.840**	**2.809**	**1.713**	-0.193	-1.354	3.568	**1.344**
HML_s	**1.998**	0.914	1.263	-0.456	-0.327	0.307	-0.688	2.104
SMB_g	-0.868	-1.321	-2.736	-1.144	-0.245	-1.494	-7.749	0.096
SMB_n	-1.791	-1.816	0.347	-0.216	-0.217	-0.366	-1.077	-4.230
SMB_v	-1.591	-0.799	0.678	-0.588	-0.040	-0.358	-1.336	-4.667
MOM_b	-0.283	-1.380	-3.196	-5.550	-1.194	-2.407	-4.693	-2.364
MOM_s	-2.764	-3.373	-7.089	-4.978	0.410	-0.333	-0.894	-0.195
Panel B. Combination forecasts - Empirical factors								
Method	**1**	**2**	**3**	**6**	**9**	**12**	**18**	**24**
Mean	0.414	-0.072	0.484	-1.106	0.202	0.167	1.975	2.424
Median	-0.095	-0.050	**1.460**	-0.370	**0.427**	0.413	0.170	0.189
Tr.mean	-0.095	-0.050	**1.460**	-0.370	**0.427**	0.413	0.170	0.189
DMSE(1)	0.423	-0.054	0.538	-1.087	0.204	0.163	**2.167**	**2.562**
DMSE(0.9)	0.441	-0.068	0.542	-1.063	0.165	0.034	1.960	**2.518**
CL(2)	**1.439**	-0.185	0.542	-0.016	-1.484	-2.208	0.088	2.197
CL(3)	**1.987**	0.753	-1.469	0.615	-2.795	-4.077	-5.596	-2.365
Panel C. Combination forecasts - Decomposed factors								
Method	**1**	**2**	**3**	**6**	**9**	**12**	**18**	**24**
Mean	-0.409	-0.576	-0.330	-1.248	0.044	-0.186	-0.279	0.491
Median	-0.261	-0.502	0.426	-1.001	0.056	-0.028	-0.431	0.141
Tr. mean	-0.402	-0.697	-0.297	-1.270	-0.122	-0.158	-0.280	0.037
DMSE(1)	-0.407	-0.726	-0.311	-1.242	0.043	-0.192	-0.227	0.582
DMSE(0.9)	-0.410	-0.877	-0.304	-1.224	0.023	-0.274	-0.430	0.554
CL(2)	-0.393	-2.088	-1.255	-0.012	-0.483	-0.861	-1.889	1.101
CL(3)	0.040	-0.810	-2.104	-0.227	-0.077	-2.951	-4.800	2.074

Notes: (i) The table reports the out-of-sample R^2_{os} of the individual ARDL models relative to the AR benchmark.
(ii) Bold entries indicate significance at the 10% significance level according to the $CW-t$ statistic, which tests the null hypothesis: $R^2_{os}=0$ against the alternative: $R^2_{os}>0$.

outperform the AR model at horizons of 18 and 24 months (DMSE(1)) and at the horizon of 2 years (DMSE (0.9)). Both cluster combining methods are associated with short run predictability of 1 month. Quite interestingly and in sharp contrast with the US market, combinations of the decomposed factors completely fail to outperform the autoregressive benchmark.

We continue by examining the level of predictability for the European stock market, which is reported in Table 3.8 (Panels A to C). Our findings suggest that the SMB factor is the dominant predictor with significant forecasting ability at horizons of 2, 6, 9 and 18 months. This performance is attributed partly to the neutral and growth component of the factor. The growth component of the size premium appears to contain useful information for 3-, 6-, 9- and 18-month future returns, while the neutral component for 1, 2, 6 and 18 months ahead. In addition, momentum along with its small and big decompositions contain useful information for the European market at the horizon of 3 months.

Similar to European bond returns, combination forecasts do not appear very successful. When we combine the individual forecasts of the aggregate factors, both the median and trimmed mean combing methods exhibit significant forecasting ability at the horizon of 3 months, while CL(3) improves forecasts at horizons of 3 and 18 months. Considering the forecasts of combinations of decomposed factors, both cluster combining methods appear significant at horizons of 3 and 18 months, while the median combining method exhibits forecasting ability at the horizon of 2 months.

Our asset allocation exercise paints a starkly different picture. Despite the anaemic

Table 3.8: Out-of-sample performance - European stock returns

Panel A. Individual ARDL models

Predictor	1	2	3	6	9	12	18	24
HML	-3.547	-1.420	-2.117	-3.340	-1.007	-1.321	-1.630	-0.811
SMB	0.320	**1.141**	1.060	**1.706**	**0.714**	0.822	**1.316**	-0.088
MOM	1.727	-0.120	**2.813**	-0.382	1.042	0.128	0.188	-0.397
HML_b	-3.926	-2.609	-2.294	-1.348	-1.652	-2.493	-2.094	-1.639
HML_s	-3.189	0.227	1.552	-1.396	0.741	0.641	-0.118	-0.110
SMB_g	-0.448	1.114	**4.132**	**1.291**	**1.123**	1.217	**1.541**	-0.381
SMB_n	**1.436**	**0.861**	0.394	**1.483**	0.529	0.781	**1.579**	-0.648
SMB_v	-2.719	-0.895	-0.125	-0.793	0.242	-0.198	-0.146	0.080
MOM_b	1.641	-0.265	**2.547**	-0.720	1.010	0.027	0.164	-1.014
MOM_s	0.836	0.070	**2.812**	-0.384	0.796	0.052	0.063	0.265

Panel B. Combination forecasts - Empirical factors

Method	1	2	3	6	9	12	18	24
Mean	0.212	0.025	0.893	-0.371	0.461	-0.017	0.045	-0.398
Median	0.553	0.392	**1.792**	0.031	-0.034	0.031	0.112	-0.694
Tr. mean	0.553	0.392	**1.792**	0.031	-0.034	0.031	0.112	-0.694
DMSE(1)	0.194	0.019	0.888	-0.363	0.451	-0.009	0.094	-0.404
DMSE(0.9)	0.192	0.028	0.909	-0.322	0.463	-0.023	0.060	-0.400
CL(2)	1.256	0.027	1.925	0.281	0.892	-0.684	0.591	-0.455
CL(3)	0.983	0.292	**1.861**	-0.004	1.658	-0.062	**1.126**	-0.340

Panel C. Combination forecasts - Decomposed factors

Method	1	2	3	6	9	12	18	24
Mean	-0.302	-0.056	1.605	-0.078	0.533	0.110	0.250	-0.409
Median	-0.195	**0.460**	2.375	-0.051	0.468	0.003	0.156	-0.100
Tr. mean	-0.352	0.329	2.044	-0.010	0.534	0.106	0.268	-0.314
DMSE(1)	-0.309	-0.061	1.611	-0.073	0.535	0.123	0.273	-0.413
DMSE(0.9)	-0.310	-0.054	1.614	-0.059	0.541	0.124	0.269	-0.409
CL(2)	0.292	0.298	**2.463**	0.325	0.893	0.406	**0.648**	-0.221
CL(3)	0.150	-0.391	**2.800**	0.277	0.790	0.351	**1.456**	-0.446

Notes: (i) The table reports the out-of-sample R^2_{os} of the individual ARDL models relative to the AR benchmark.
(ii) Bold entries indicate significance at the 10% significance level according to the $CW - t$ statistic, which tests the null hypothesis: $R^2_{os} = 0$ against the alternative: $R^2_{os} > 0$.

Table 3.9: Asset allocation benefits - European investor

	Panel A. Horizon 1				**Panel B. Horizon 3**			
	Three Factors		**Seven Factors**		**Three Factors**		**Seven Factors**	
Method	ΔU	Θ	ΔU	Θ	ΔU	Θ	ΔU	Θ
Mean	**4.622**	**4.094**	**0.730**	**0.008**	-2.355	-1.849	**2.601**	**0.266**
Median	**1.897**	**0.594**	**0.743**	-0.969	**5.070**	**1.465**	**9.144**	**2.898**
Tr.Mean	**1.897**	**0.594**	**0.011**	-1.757	**5.070**	**1.465**	**6.855**	**1.887**
DMSE(1)	**4.541**	**4.022**	**0.701**	-0.017	-2.477	-1.907	**2.618**	**0.263**
DMSE(0.9)	**4.501**	**3.954**	**0.710**	-0.057	-2.219	-1.776	**2.698**	**0.298**
CL(2)	**0.601**	-3.386	-2.975	-8.019	**7.313**	**2.238**	**9.810**	**2.582**
CL(3)	-3.644	-14.210	-2.370	-7.804	**8.865**	**3.285**	**10.707**	**2.231**

	Panel C. Horizon 6				**Panel D. Horizon 12**			
	Three Factors		**Seven Factors**		**Three Factors**		**Seven Factors**	
Method	**ΔU**	**Θ**	**ΔU**	**Θ**	**ΔU**	**Θ**	**ΔU**	**Θ**
Mean	-8.515	-1.228	-8.368	**0.092**	-7.847	**0.530**	**2.386**	**0.554**
Median	-11.396	-0.962	-2.395	-0.456	**7.431**	**0.756**	-10.422	**0.847**
Tr. Mean	-11.396	-0.962	-6.103	**0.076**	**7.431**	**0.756**	<-20.0	**0.965**
DMSE(1)	-8.485	-1.214	-8.613	**0.110**	-8.617	**0.552**	<-20.0	**0.973**
DMSE(0.9)	-8.054	-0.978	-8.386	**0.196**	-8.566	**0.520**	**0.293**	**0.027**
CL(2)	**2.997**	**1.321**	-2.902	**2.358**	-9.017	-0.980	-14.996	**1.766**
CL(3)	-0.416	-2.354	**1.718**	2.607	-11.891	**0.451**	<-20.0	**1.895**

Notes: (i) The table reports the average utility gain (ΔU) and the difference between the manipulation proof performance measure (Θ) of the proposed specification relative to the benchmark AR model.
(ii) Figures are reported in annualized percentage points.
(iii) Portfolio weights are constrained to lie between 0 and 1.5 and RRA is set equal to 3.

statistical significance of combination forecasts of both stock and bond returns, the gains for a European investor can be sizable. In Table 3.9 we report the average utility gains of a mean-variance investor who allocates her wealth between stock, bonds and the risk free interest rate along with the manipulation-proof measure of the competing models for horizons up to 1 year.

For a short term horizon of 1-month, mean and DMSE combinations of aggregate factors can lead to utility gains of up to 4.622%. Increasing the horizon to 3 months can lead to gains of 8.865% for the CL(3) method. This horizon is also associated with significant profitability of up to 10.707% on the basis of the CL(3) combinations of disaggregated factors. However, longer horizons of 6 and 12 months do not consistently generate profits to the investor.

3.5.2 Forecasting the Japanese bond and stock market

In this section, we investigate the robustness of our results employing data for the Japanese bond and stock market. In Table 3.10 and Table 3.11 we report the forecasting performance of empirical factors for bond and stock returns, respectively. With respect to bond returns our findings suggest that single factor models prove successful at short horizons. Specifically, at the 1-month horizon both the size and the value premium improve bond return forecasts, mainly stemming from the small value component and the growth and neutral size component, respectively.

Table 3.10: Out-of-sample performance - Japanese bond returns

Panel A. Individual ARDL models								
Predictor	**1**	**2**	**3**	**6**	**9**	**12**	**18**	**24**
HML	**2.031**	1.384	0.821	-0.231	0.326	-0.486	0.538	-0.720
SMB	**1.629**	-1.518	0.326	0.159	0.609	0.310	-0.492	-0.862
MOM	0.080	1.548	1.653	-0.397	0.487	-0.140	-1.112	-0.367
HML_b	0.801	-0.676	-0.021	-0.069	0.656	-0.293	0.700	-0.293
HML_s	**1.989**	**1.661**	1.088	-0.394	-0.593	-0.535	0.104	-0.271
SMB_g	**0.599**	-3.395	-0.677	-0.148	-0.110	-0.469	-0.079	-0.282
SMB_n	**1.526**	0.101	**1.769**	**0.824**	1.461	1.032	-0.391	-0.730
SMB_v	1.560	0.331	0.131	-0.325	0.559	0.584	-1.150	-1.187
MOM_b	-0.372	1.118	1.686	-0.242	0.625	-0.169	0.053	-0.180
MOM_s	0.194	0.638	0.498	-0.912	-0.374	-0.070	-2.936	-0.612
Panel B. Combination forecasts - Empirical factors								
Method	**1**	**2**	**3**	**6**	**9**	**12**	**18**	**24**
Mean	**1.357**	**0.846**	**1.294**	0.000	0.651	0.002	-0.268	-0.578
Median	**1.590**	-0.118	**0.762**	**0.264**	**0.875**	**0.470**	-0.106	-0.274
Tr. mean	**1.590**	-0.118	**0.762**	**0.264**	**0.875**	**0.470**	-0.106	-0.274
DMSE(1)	**1.356**	0.823	**1.282**	-0.006	0.651	-0.004	-0.254	-0.584
DMSE(0.9)	**1.387**	0.855	**1.309**	-0.034	**0.660**	-0.014	-0.259	-0.581
CL(2)	**1.388**	0.818	0.052	-0.446	0.348	-0.483	-0.325	-1.689
CL(3)	**1.343**	-2.148	-1.716	-1.379	**0.749**	-0.051	**1.053**	-1.163
Panel C. Combination forecasts - Decomposed factors								
Method	**1**	**2**	**3**	**6**	**9**	**12**	**18**	**24**
Mean	**1.040**	0.316	**0.981**	0.005	**0.564**	0.196	-0.387	-0.430
Median	**1.096**	0.352	**0.890**	0.051	0.118	0.245	0.257	-0.258
Tr. mean	**1.032**	0.259	**0.908**	0.052	0.302	0.261	-0.087	-0.427
DMSE(1)	**1.040**	0.294	**0.969**	-0.002	**0.559**	0.188	-0.382	-0.425
DMSE(0.9)	**1.056**	0.323	**0.989**	-0.017	**0.571**	0.174	-0.389	-0.415
CL(2)	**1.087**	-0.378	0.051	-0.596	0.570	-0.427	-0.851	-0.344
CL(3)	0.489	-2.042	-1.095	-1.873	0.170	-0.566	-1.302	-0.408

Notes: (i) The table reports the out-of-sample R^2_{os} of the individual ARDL models relative to the AR benchmark.
(ii) Bold entries indicate significance at the 10% significance level according to the $CW - t$ statistic, which tests the null hypothesis: $R^2_{os}=0$ against the alternative: $R^2_{os}>0$.

Additionally, the small value component improves forecasts for the 2-month horizon as well, while the neutral value component for the 3-month and 6-month horizons. Our findings with respect to combination forecasts are more reassuring. Specifically, with the exception of the 24-month horizon, the remaining horizons are characterized with a high degree of predictability.

The 1-month and 3-month bond returns can be predicted with almost all the methods at hand and on the basis of both the aggregate and decomposed factors. Overall, combinations of the aggregate factors perform better than the decomposed ones.

Similar findings pertain with respect to stock returns where the level of predictability is higher. The value premium is successful in improving forecasts for all the horizons up to the 9-month one. This forecasting ability is equally split between its big and small component which contains useful information for the long-run as well. Quite interestingly, the value and growth decompositions of the size premium emerge as powerful predictors for horizons greater than 18 and 24 months.

As expected, this individual forecasting ability is recorded in the success of forecast combinations. On the basis of forecast combinations of aggregate factors, the mean, DMSE and cluster combining methods improve forecasts for the majority of horizons considered. Quite interestingly, the 24-month horizon is associated with a high degree of predictability of combination methods of both aggregate and decomposed factors.

Table 3.11: Out-of-sample performance - Japanese stock returns

Panel A. Individual ARDL models								
Predictor	**1**	**2**	**3**	**6**	**9**	**12**	**18**	**24**
HML	**6.153**	**2.683**	**5.965**	**2.332**	**1.408**	-0.306	0.073	0.668
SMB	0.321	-0.195	-0.409	0.275	0.814	0.453	2.747	**2.441**
MOM	-0.239	-1.635	0.356	-0.726	-0.345	-0.156	1.197	-0.013
HML_b	**5.937**	**1.914**	**1.856**	0.206	0.167	-0.161	-0.894	-0.447
HML_s	**3.925**	0.929	4.210	**3.117**	**2.488**	-0.424	1.367	**1.204**
SMB_g	2.192	-0.456	-0.893	0.348	0.835	0.115	**1.846**	**0.976**
SMB_n	-0.373	0.003	-0.089	0.352	**1.552**	**1.329**	**3.504**	**3.195**
SMB_v	0.108	-0.376	0.150	-0.592	-0.057	-0.158	0.718	**1.092**
MOM_b	-2.637	-1.291	1.444	-0.493	-0.453	-0.379	0.458	0.591
MOM_s	-0.431	-1.288	1.884	-1.077	-1.332	-1.119	-0.030	-2.298
Panel B. Combination forecasts - Empirical factors								
Method	**1**	**2**	**3**	**6**	**9**	**12**	**18**	**24**
Mean	**3.335**	0.267	**2.343**	**0.858**	**1.035**	0.245	**1.690**	**1.164**
Median	1.499	0.290	0.752	0.358	0.012	-0.579	0.465	**0.783**
Tr. mean	1.499	0.597	0.752	0.358	0.012	-0.579	0.465	**0.783**
DMSE(1)	**3.352**	0.599	**2.371**	**0.863**	**1.028**	0.258	**1.724**	**1.168**
DMSE(0.9)	**3.362**	0.259	**2.387**	**0.871**	**1.038**	0.260	**1.712**	**1.173**
CL(2)	**2.981**	-0.175	**1.018**	**1.024**	**1.526**	**0.745**	**2.356**	**1.473**
CL(3)	**4.628**	**2.719**	0.827	0.474	**1.306**	0.327	**2.187**	**1.352**
Panel C. Combination forecasts - Decomposed factors								
Method	**1**	**2**	**3**	**6**	**9**	**12**	**18**	**24**
Mean	**2.181**	0.204	**1.618**	**0.768**	**1.056**	**0.755**	**1.422**	**0.859**
Median	1.199	**0.303**	**1.000**	0.167	**0.499**	**0.836**	**1.558**	**0.892**
Tr. mean	**1.745**	0.080	**1.139**	0.468	**0.835**	0.359	**1.387**	**0.810**
DMSE(1)	**2.171**	0.214	**1.607**	**0.756**	**1.032**	0.664	**1.437**	**0.870**
DMSE(0.9)	**2.195**	0.178	**1.602**	**0.747**	**1.099**	0.613	**1.434**	**0.879**
CL(2)	**3.398**	-0.142	0.870	0.080	**1.173**	1.052	**2.285**	**1.341**
CL(3)	**4.132**	-0.265	1.075	0.100	**2.733**	1.027	0.878	**1.792**

Notes: (i) The table reports the out-of-sample R^2_{os} of the individual ARDL models relative to the AR benchmark.
(ii) Bold entries indicate significance at the 10% significance level according to the $CW-t$ statistic, which tests the null hypothesis: $R^2_{os}=0$ against the alternative: $R^2_{os}>0$.

Table 3.12: Asset allocation benefits - Japanese investor

	Panel A. Horizon 1				Panel B. Horizon 3			
	Three Factors		Seven Factors		Three Factors		Seven Factors	
Method	ΔU	Θ	ΔU	Θ	ΔU	Θ	ΔU	Θ
Mean	2.106	1.567	1.822	1.235	5.822	2.732	4.303	1.794
Median	0.796	0.623	0.207	0.185	0.915	0.171	2.492	1.100
Tr.Mean	0.796	0.623	1.532	1.121	0.915	0.171	3.034	1.280
DMSE(1)	2.106	1.571	1.821	1.237	5.925	2.773	4.286	1.788
DMSE(0.9)	2.100	1.572	1.810	1.230	5.990	2.800	4.321	1.792
CL(2)	3.067	1.946	2.292	1.477	3.495	1.045	2.641	0.550
CL(3)	2.621	1.504	3.982	2.719	5.304	0.816	2.946	0.737

	Panel C. Horizon 6				Panel D. Horizon 12			
	Three Factors		Seven Factors		Three Factors		Seven Factors	
Method	ΔU	Θ	ΔU	Θ	ΔU	Θ	ΔU	Θ
Mean	9.012	2.068	3.083	0.546	9.748	0.805	>20.0	4.258
Median	6.860	1.626	2.247	0.517	-2.525	-0.833	18.505	2.106
Tr. Mean	6.860	1.626	2.252	0.550	-2.525	-0.833	13.006	1.482
DMSE(1)	9.085	2.084	3.146	0.567	10.479	0.887	>20.0	4.190
DMSE(0.9)	9.193	2.103	3.232	0.575	9.902	0.819	>20.0	4.067
CL(2)	16.943	3.799	10.651	2.452	9.554	0.834	>20.0	6.004
CL(3)	>20.0	3.631	16.188	3.667	6.156	-0.377	>20.0	9.978

Notes: (i) The table reports the average utility gain (ΔU) and the difference between the manipulation-proof performance measure (Θ) of the proposed specification relative to the benchmark AR model.
(ii) Figures are reported in annualized percentage points.
(iii) Portfolio weights are constrained to lie between 0 and 1.5 and RRA is set equal to 3.

Finally, the most striking result appears in Table 3.12 that reports the forecast combination benefits from an asset allocation perspective. Specifically, a mean-variance investor who employs our forecast combination methodology can always enjoy significant gains for all the horizons up to 6 months. The benefits accrued by including the information of the aggregate factors reach 3.067% at the 1-month horizon and increase to 5.990% at the horizon of 3 months, while they can even exceed 20% for the 6-month horizon. The 1-year horizon is associated with benefits reaching 10.479% for combinations of the aggregate factors and exceed 20% for combination forecasts of the decomposed ones. Our findings with respect to the MPPMs of the respective portfolios are fully consistent with the ones of utility gains.

3.6 Conclusions

In this paper, we study the forecasting ability of empirical factors; namely, the value premium (HML), the size premium (SMB) and the momentum factors (MOM, LT and ST) along with widely employed financial variables on U.S. bond and stock returns. One of our contributions to the literature consists of the decomposition of these empirical factors to their size and value components, investigating thus the size effect on the value and momentum premium and the value effect on the size premium. Our findings suggest that these empirical factors contain significantly more information for future bond and stock market returns than the typically employed financial variables, but the extent to which this forecasting ability appears differs.

To address the instability and time-variability of individual forecasts, we go one step further and combine them by employing a variety of combination methods. Specifically, we construct forecasts on the basis of three simple combining methods; namely, the mean, median and trimmed mean and two more advanced ones; the Discount Mean Square forecast Error (DMSE) combining method, which is based on the historical performance of the individual models, and the Cluster Combining method (CL), which is based on equal-sized clusters related to past forecasting performance. The forecasting ability of combination forecasts is assessed not only statistically, by means of the R^2_{os} statistic, which measures the improvement of the MSFE of the proposed model over the MSFE of the benchmark AR model forecast, but also economically by computing the performance fee that investors would be willing to pay to have access to our methodology. In addition, we calculate the risk-adjusted portfolio's premium return (manipulation-proof performance measure) in order to assess the most valuable model among the competing ones.

Our results provide evidence that combination forecasts based on decomposed factors display superior forecasting ability relative to the forecasts based on typically employed financial variables at horizons ranging from the short run to the long run. This performance is also evident from an asset allocation perspective. In particular, investors can accrue positive utility gains by employing trading strategies based on forecasts produced by the empirical factors, irrespective of the degree of relative risk aversion and borrowing constraints. Finally, the robustness of our results is assessed by conducting the same tests for markets outside the US. By employing data for the European and Japanese bond and stock market, we find that the forecasting ability of combination forecasts formed on the basis of the empirical factors is rather pervasive in these markets, as well.

Chapter 4

Measuring portfolio risk: How size, book-to-market and prior portfolio performance are related to risk?

4.1 Introduction

Measuring portfolio risk is central in the area of portfolio management. The extant literature focuses on identifying the adequate risk measure especially in the aftermath of the recent financial crisis. During the recent years, a plethora of indices able to quantify the embedded risk has been developed and, among the academics, there has been a debate on the best way to quantify the risk of a portfolio of returns.

Since the analysis of Markowitz (1959), the standard deviation along with the beta coefficient are among the most popular risk measures, with the first one quantifying the total risk and the second one measuring systematic risk, reflecting thus the sensitivity of an asset to changes in market returns. However, both standard deviation and beta give equal weight to upside and downside fluctuations, contrary to investors' appetite.

Facing this particular weakness of the traditional risk indices, alternative measures that capture only the undesirable downside volatility, have been proposed. Specifically, the introduction of the notion of downside risk, through which only the left part of a return distribution is considered, has led to a variety of different risk measures, accommodating different aspects of risk (Neil, 2001; Cheng et al., 2004). Downside risk measures have gained the preference of practitioners due to the fact that the left part of a return distribution involves risk, while the right part describes superior investment opportunities. To this direction, Estrada (2006) studied from a calculation point of view two downside risk indices, the semideviation and downside beta, which assess risk better than standard deviation, especially for skewed return distributions. Similarly, Ang, Chen and Xing (2006) proved that stocks characterized by high downside risk exhibit a premium in average returns, meaning that risk-averse investors should require a premium to hold portfolios with high sensitivities to market downturns. Alternative diversifications on downside risk measures have also been suggested, with Barndorff-Nielsen et al. (2009) proposing the realized semivariance, a measure that is based on drawdown moves using high frequency data, with important predictive qualities for future market volatility.

Another downside risk measure, used extensively by commercial banks for regulatory purposes under the Basel II Accord for years and insurance companies, as well, is the Value-at-Risk (VaR), which was first introduced in 1994 by JP Morgan Bank. VaR expresses the expected maximum loss that may incur over a defined time horizon and within a specified confidence interval. However, despite its popularity and its extensive use, this risk index ignores losses exceeding the value of the estimated VaR and also violates the diversification effect.

Shedding light to the limitations of VaR, Artzner et al. (1997, 1999), through theoretical work, showed that VaR cannot be characterized as a coherent measure of risk, since it does not satisfy the crucial in risk measurement subadditivity property, meaning that diversification has no effect and even well diversified portfolios require more regulatory capital. Apart from that, VaR leads to inadequate capital requirements and large losses due to its inability to capture large losses, especially in cases of extreme events of ruin. To alleviate the problems inherent in VaR, Artzner et al. (1999) proposed an alternative risk measure, the expected shortfall (ES) or tail conditional expectation, characterized by a series of axioms, ensuring that riskier portfolios have higher values of risk if the employed measure satisfies the proposed axioms. The ES risk index measures how much one can lose on average beyond the VaR level, specifying what happens in those bad states. Consistent with the analysis of Artzner et al. (1997, 1999), Yamai and Yoshiba (2002a) and Acerbi et al. (2001) also proved that the expected shortfall is a superior alternative to the standard Value-at-Risk, respecting the associated risk.

Similar analysis has been conducted by Acerbi (2004), who proposed the spectral risk measures by specifying the user's risk-aversion function. The ES could be thought of as a special case of spectral risk measures, assigning the same weight on all losses in the tail region, reflecting a risk-neutral investor between tail-region outcomes. However, under the assumption of a risk-averse user, neither the ES nor the VaR can be thought of as good risk measure, with the latter being an even worse risk measure as it can be regarded as a risk-loving measure by giving zero weights to losses greater than VaR. ES has also some disadvantages, such as the larger estimation errors[1], which are proved to be larger than those of VaR for fat-tailed distributions, as suggested by Yamai and Yoshiba (2005). The same authors (2002b) also showed that the expected shortfall requires a larger size of sample for its backtesting than VaR for the same level of accuracy. Comparing the properties of different risk measures based on quantiles, with the VaR and ES being among them, Dowd and Blake (2006) presented estimation methods along with various applications of the employed measures attempting to determine the best risk measure. Adding to the existing literature, Danielsson et al. (2013), standing up for VaR, suggested that VaR can also be characterized as subadditive, with the exception for the fattest tails, which are highly unlikely to be observed for the majority of assets, meaning that there is no decision-making advantages to ES over VaR in most cases.

Apart from portfolio ranking, risk measures are also used for regulatory purposes. In particular, under the accords of the Basel Committee on Banking Regulation, the banks' exposure to risk has been assessed through the 10-day, 99th percentile VaR, which was incorporated in the Basel II Capital Accord in 1999. However, a number of weaknesses identified with VaR, including its inability to capture "tail risk", encouraged recently the Committee to abandon this index. In an attempt to face the weaknesses of VaR, and under the propositions of the academics, an alternative risk index has been considered, appropriate

[1]Estimation error is the natural variability due to limited sampling size. This phenomenon becomes more intense for fat-tailed distribution, since large losses appear with high probabilities. Due to the fact that ES is affected by these realised losses, while VaR disregards loss beyond VaR level, the ES estimation varies more due to the infrequent and large losses, for more fat-tailed distributions.

to overcome the VaR's inadequacy to measure risk efficiently, especially during the unstable economic circumstances. In recognition of the inability of the 10-day VaR to capture the risk ex ante, the Committee, under the Basel III Accord[2], decided to introduce the so-called expected shortfall as a tool for assessing banks' exposure to risk. Additionally, the Committee proposed the decrease of confidence level for the new employed risk measure, from 99% to 97.5%. As mentioned previously, ES accounts for tail risk in a more comprehensive manner, as it measures the risk of a portfolio by considering both the size and likelihood of losses above a certain threshold (e.g. the 99th percentile), expressing the portfolio loss one expects to suffer, given that the portfolio loss is equal or larger than its VaR.

Attempting to cover non-normal effects, such as (negative) skewness and excess kurtosis, which are evident in financial returns, new risk measures have been proposed. In particular, Aumann and Serrano (2008) achieved to quantify the riskiness of a gamble by assigning a real number as a measure of its riskiness, independently of the specific decision-maker. The authors introduced the A/S economic index, defined as the reciprocal of the absolute risk aversion (ARA) of an investor, which looks for the critical utility regardless of wealth and, according to the definition of the authors, when riskiness increases, then less risk-averse investors are expected to take riskier gambles. Hart (2009) also provided an alternative approach that leads to the same index of riskiness, without the assumptions of Aumann and Serrano analysis, based on a different set of behavioral axioms. Following the aforementioned studies, Bali et al. (2011) proposed a generalized measure of riskiness obtained by traded options, nesting both Aumann and Serrano (2008) index and that proposed

[2]Basel Committee on Banking Supervision (2012), 'Fundamental review of the trading book'.
Basel Committee on Banking Supervision (2013), 'Fundamental review of the trading book: A revised market risk framework'.

of Foster and Hart (2009). The proposed measure incorporates the market's expectation of future return distribution, providing asset allocation implications attributed to the employed traded options, as option prices. Compared with the ES risk measure, Shalit (2013) evaluated Israeli mutual funds by employing coherent risk measures and concluded that the A/S index of riskiness adds dimensions of risk related to skewness and leptokurtosis, producing thus an unequivocal ranking of risky assets for all risk-averters, while the ES index is adequate for risk evaluation in the lower tail of distributions.

Beyond the traditional and downside risk measures, another approach, based on drawdowns occurring in stock prices, has also been proposed for risk measurement. In particular, Chekhlov et al. (2005) introduced the conditional drawdown-at-risk measure and also examined the properties of several drawdown measures. Adding to the this, Auer and Schuhmacher (2013) conducted an empirical analysis by employing well defined drawdown risk indices, introduced by Schuhmacher and Eling (2011), and constructed performance measures for portfolio evaluation.

In our analysis, we employ the most widely used risk indices and quantify the associated risk of the empirical Fama/French, reversal and momentum portfolios, revealing thus a rank order respecting the underlying risk. The empirical analysis is implemented for a 14-year period from January 2000 to December 2013 over the aforementioned factor portfolios. Additionally, we contribute to the literature by revealing the connection between specific stock characteristics, such as size, value and prior performance, and the incorporated risk of these portfolios. Despite the popularity of these portfolios among both practitioners and academics, due to their forecasting ability on stock and bond returns along with their positive mean portfolio returns, little research has been done with respect to the associated risk and the implied ranking. To this aspect, Bakshi et al. (2011) proved that equity premium, value spread, size spread, momentum spread, distress spread and excess returns of some industries are negatively related to changes in risk, suggesting that investors should reduce holdings in certain risky assets when risk increase. The same authors showed that stock portfolios with high book-to-market, small capitalization and low momentum exhibit a worse performance than portfolios with the opposite characteristics.

Our results indicate that the MOM factor portfolio appears to be the most risky portfolio, while the long-term reversal factor can be characterized as low risk, according to the majority of the risk measures employed. With respect to specific stock characteristics, small-size and value stocks are related with portfolios of high risk, while prior high performance, either momentum or reversal (long-term or short-term), is connected with low risk. The robustness of our results are also checked by implementing the same analysis for an extended sample period, from July 1963 to December 2013.

Beyond the portfolio ranking through different risk measures, this study also contributes to the existing literature by examining whether the variety of the indices employed leads to identical ranking results. From the empirical results, we conclude that portfolio rankings produced by the traditional or the downside risk measures are very similar. However, when we measure risk by employing the A/S index of risk or any of the class of the drawdown risk measures, we find a different rank order, which is attributed mainly to the exhibited relatively low (high) kurtosis and/or excess kurtosis of the series of returns.

The remainder of the paper is structured as follows. Section 2 presents the risk indices employed. The data along with the empirical ranking results are discussed in Section 3. Section 4 reports the results of the robustness checks and Section 5 concludes.

4.2 Quantifying the embedded risk for portfolios of returns

4.2.1 Traditional measures of risk

Standard deviation, denoted as $St.Dev_i$, of portfolio of returns along with the *beta* coefficient, β_i, are among the traditional risk indices in finance since the analysis of Markowitz, mainly due to the simple way of calculation along with their distribution invariance. Standard deviation measures the dispersion of a distribution and has the main advantage of being in the same units of measure as the random variable. Despite the popularity of the standard deviation, this risk index exhibits weaknesses, such as: (i) it ignores the direction of the movement of returns, as it measures the dispersion of returns around its mean by taking into account both upside and downside movements, leading thus to a false estimation of risk, especially for negatively skewed return distributions, (ii) it fails to detect only losses, (iii) it does not take into account skewness, resulting thus to misleading results and (iv) its risk estimation may be misleading in cases that the portfolio risk level has changed, due to the fact that the portfolio's risk is calculated by using past returns.

With respect to the *beta* coefficient, it depicts the correlation between portfolio returns and the market and remains an appropriate measure of risk for diversified investors although it has weaknesses similar to those of standard deviation. Computationally, the *beta* coefficient is calculated by taking the expectation of the market model:

$$R_{it} = a_i + \beta_i R_{Mt} + e_{it} \tag{4.1}$$

where R_{it} and R_{Mt} is the portfolio i and market returns, respectively, and e_{it} is a random error at time t, and then substituting for a_i in the market model. Using the basic Capital Asset Pricing Model (CAPM), developed by Sharpe (1964):

$$E(R_i) = R_f + \beta_i (E(R_M) - R_f) \tag{4.2}$$

where $E(R_i)$ and $E(R_M)$ is the average portfolio and market return, respectively, and R_f is the risk-free interest rate of return, we have the econometric model of CAPM:

$$R_{it} - R_f = \beta_i (R_{Mt} - R_f) + e_{it} \tag{4.3}$$

which assumes that the market model along with the CAPM hold every period and the *beta* coefficient is stable over time. The *beta* coefficient for a portfolio of returns is calculated by the aforementioned regression using the OLS method and shows how intensively the portfolio follows the market, with high values of *beta* indicating riskier portfolios of returns.

A portfolio with a high beta tends to go up substantially more than the market when the market grows up, even if it does not tend to fall by more than the market, when the market falls.

4.2.2 Downside risk indices

Downside risk indices, such as semi-standard deviation, $Semi - sd_i$, considered by Ogryczak and Ruszczizynski (1999), or the class of lower partial moments (LPM_{ni}) of order 1 and 3, defined by Harlow (1991), Value-at-Risk along with its Cornish-Fisher (1937) expansion, $Modified - VaR_i$[3], and expected shortfall, ES_i, have been rapidly gained the acceptance among both academics and practitioners due to the fact that they include only negative deviations, incorporating thus movements associated exclusively with losses. The aforementioned measures are computed for each of the portfolios i assumed as follows:

$$Semi - sd_i = \sqrt{E(\max(E(R_i) - R_i, 0)^2)} \tag{4.4}$$

$$LPM_{ni} = (1/N)\sum_{i=1}^{N} \max(E(R_i) - R_i, 0)^n \quad with\ order\ i = 1\ and\ 3 \tag{4.5}$$

where N represents the total number of observations. Note that the different orders considered determine the extent to which the negative deviations from the mean return of the portfolio are weighted.

With respect to the popular VaR_i and ES_i risk measures, we estimate them considering 3 different confidence levels (95%, 97.5% and 99%) consistent with the practical regulatory implications and, under the assumption of normal distributed series of returns, these risk indices are computed incorporating the estimated standard deviation of each investment, as:

$$VaR_i = z_a * \sigma_i \tag{4.6}$$

$$ES_i = (\sigma_i/(1-\alpha)\sqrt{2\pi})e^{-z_a^2/2} \tag{4.7}$$

[3]This index of risk was introduced by Zangari (1996) to estimate parametric VaR.

while the $Modified{-}VaR_i$ risk index for non-normal series of returns is calculated for the confidence levels of 95%, 97.5% and 99%, as follows:

$$M - VaR_i = \sigma_i(z_a + (z_a^2 - 1)/6 + (z_a^3 - 3z_a)E_i/24 - (2z_a^3 - 5z_a)S_i^2/36) \tag{4.8}$$

where z_a is the a-quantile of the standard normal distribution, σ_i is the standard deviation of the series assumed, S_i is the skewness and $E_i = k_i - 3$ the excess kurtosis of the series of returns.

Semi-standard deviation

With respect to the semi-standard deviation, this risk measure takes into account only negative deviations for its calculation, capturing thus the downside volatility for which investors are averse and assessing risk better than standard deviation, especially when the series distribution is skewed. Contrary to the standard deviation and the beta coefficient that both consider equal weights to upside and downside fluctuations, semi-deviation incorporates in its formula only movements associated with losses, accommodating thus different views of risk, as denoted in the following formula:

$$Semi - sd_i = \sqrt{\frac{1}{N-1} \sum_{R_{it} < E(R_i)}^{N} (R_{it} - E(R_i))^2} \tag{4.9}$$

where N is the total number of observations of the portfolio series, R_{it} is the observed value of portfolio i returns a time t, and $E(R_i)$ is the average portfolio return. The use of the specific risk index had also been mentioned at the analysis of Markowitz as an appropriate measure one could use to efficiently quantify the risk of an investment.

Value-at-Risk

The famous Value at Risk, proposed by JP Morgan in 1994, is the most frequently used downside risk index due to the fact that it can be applied to any financial series, leading thus to a rank order for various portfolios regarding the risk of extreme events. More in detail, VaR characterizes risk associated with the losses that a portfolio may suffer and can be alternatively defined as:

$$VaR_a(x) = \min(x \mid F(x) \geq a) \tag{4.10}$$

which is translated as a threshold value such that the probability of the loss on a specific portfolio over a given time horizon would exceed the probability a. Alternatively, the VaR can be thought of as the value of the worst loss not to be exceeded with a probability of at least $1 - a$ or as the maximum potential loss that a portfolio can suffer in the (e.g. 95%) a% best cases in t days. In other words, VaR is the best of the worst cases scenario underestimating thus losses associated with the specified level of probability.

From a computational aspect, VaR can be estimated either parametrically (e.g. variance-covariance approach), semi-parametrically (e.g. weighted historical simulated) or nonparametrically (e.g. historical simulation or resampled approach). The method used in our analysis for the estimation of VaR is the variance-covariance method, according to which the VaR is estimated for different levels of confidence (99%, 97.5% and 95%) by using the standard deviation of the historical data under the assumption of normal distributed series of returns, as follows:

$$VaR_{99\%} = 2.33 * \sigma_i \qquad for\ 99\%\ confidence\ leve \tag{4.11}$$

$$VaR_{97.5\%} = 1.96 * \sigma_i \qquad for\ 97.5\%\ confidence\ level \tag{4.12}$$

$$VaR_{95\%} = 1.64 * \sigma_i \qquad for\ 95\%\ confidence\ level \tag{4.13}$$

By increasing the confidence level, we expect to find higher VaR. We note that for the 95% confidence level, the standard deviation is multiplied with 1.64, corresponding to the area under the standard normal curve between -1.64 and 1.64. As noted previously, for the computation of VaR, one needs to specify the level of confidence and the time horizon, parameters, though chosen usually arbitrarily, that have an impact on the accuracy of the computed risk. In particular, longer time horizons may include observations too old to be meaningful. With respect to the confidence level, it is chosen according to the purpose of the risk estimate. For example, for capital-requirement purposes, the confidence level should be chosen at low level.

Concerning the alternative methods that could be used, the historical approach simulates the distribution of a series by generating a set of scenarios for the possible values, with no assumptions about the distribution of the series, requiring though a large database and careful selection of the sampling period, while Monte Carlo simulation estimates VaR by simulating scenarios generating them from a lognormal distribution, instead of generating the scenarios from the historical distribution. The Monte Carlo approach estimates VaR by simulating random scenarios and revaluing portfolio positions, and despite its accuracy, it is time consuming and requires a large number of simulations to get a good approximation statistically significant. On the other hand, the semi-parametric approach combines parametric and non-parametric methods by weighting differently the observations according to the changes in volatility or their age, making the newer ones more important to others.

However, despite the popularity of VaR, this risk index exhibits various limitations: (i) it tells nothing about what one loses in excess of the VaR, underestimating thus risk-return analysis, (ii) under the parametric approach, the VaR assumes that returns follow a normal distribution and the tail is well predicted, whereas, assets may have fat tails and may not follow a normal distribution, (iii) it is not subadditive, a property that plays a fundamental role in measuring risk by causing a decrease in risk for diversified portfolios, meaning that the VaR of a combined portfolio could be larger than the sum of the VaRs of its components and (iv) the underestimation of risk due to the fact that an extreme event would be included to the computation of the value-at-risk after the damage has already done.

Expected Shortfall

Taking into account that VaR fails to incorporate the severity of an incurred damage event and does not consider diversification effect, meaning that the risk of a portfolio can be larger than the sum of the stand-alone risks of its components when measured by VaR, Artzner et al. (1997, 1999) proposed the use of a risk measure which satisfies all the properties of coherence; namely, the expected shortfall (ES), which guarantees that the portfolio diversification is always positive. Specifically, the axioms of coherence are defined below.

We set as $r(\cdot)$ the employed risk measures, with x, y the portfolio positions and G the set of all risks incorporated in the portfolios of interest. Any increase in the amounts either in portfolio returns or in the estimated risk is indicated as a, while increases/decreases in risk or in returns by a factor would be indicated by λ. The proposed axioms are described as follows:

Axioms of coherence

1. *Translation invariance*: if $a \in R$, then $r(x + a) = r(x) - a$, which means that an increase in total portfolio produces a decrease in the risk total measure by the same amount.

2. *Subadditivity:* $\forall x, y \in G$, then $r(x + y) \leq r(x) + r(y)$, which covers the meaning of diversification, meaning that by adding two portfolios does not generate any additional risk[4]. The subadditivity condition suggests that a combined portfolio cannot be riskier in the aggregate than the two portfolios standing apart. The aggregation of the two portfolios should provide some diversification benefit and thereby lower overall risk.

3. *Positive homogeneity*: $\forall x \in G$ and $\lambda \in R$, then $r(\lambda x) = \lambda r(x)$, which implies that smaller positions are less risky as being more liquid. Increasing a portfolio by a factor implies corresponding increase for that risk. Increasing the size of any portfolio by a positive factor requires increase in regulatory capital by the same factor.

4. *Monotonicity*: $\forall x, y \in G$ with $x \leq y$, then $r(x) \geq r(y)$, meaning that if a portfolio offers higher returns than another portfolio in every conceivable economic state, then

[4]Diversification means that the risk associated with two combined portfolios cannot exceed the risk of the total portfolio.

the risk associated with the first portfolio cannot be higher than the second portfolio, and the regulatory capital required for the first portfolio must be less than or equal to the regulatory capital required for the second one.

The expected shortfall is the loss one expects to suffer given that the portfolio loss is equal or larger than its VaR and, for confidence level $1 - a$, is defined as:

$$ES_a(x) = E(x \mid x \geq VaR_a(x)) \tag{4.14}$$

accounting for the losses beyond the confidence interval. Alternatively, the ES could be characterized as the average loss in the worst $100 * a\%$ cases and under the assumption of normal distributed returns could be computed by Eq. 4.7.

Estimating ES under the assumption of a Normal Distribution

Let f_{SN} denote the standard normal distribution: $f_{SN}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. The scale-family of the standard normal distribution, f_N, which has a mean zero and a variance σ_t that is allowed to change over time is:

$f_N(x_t) = \frac{1}{\sigma_t} f_{SN}(\frac{x_t}{\sigma_t}) = \frac{1}{\sigma_t\sqrt{2\pi}} e^{-\frac{x_t^2}{2\sigma_t^2}}$

The derivation of ES is as follows:

$$ES_{a,t} = E[X_t \mid x_t \geq VaR_{a,t}] = \frac{E[x_t * I(x_t \geq VaR_{a,t})]}{1-a} = \frac{1}{1-a} \int_{VaR_{a,t}}^{\infty} x_t f_N(x_t) dx_t$$
$$= \frac{1}{1-a} [\frac{1}{\sigma_t\sqrt{2\pi}} \int_{VaR_{a,t}}^{\infty} x_t e^{-\frac{x_t^2}{2\sigma_t^2}} dx_t] = \frac{1}{1-a} \frac{1}{\sigma_t\sqrt{2\pi}} \left[-\sigma_t^2 e^{-\frac{x_t^2}{2\sigma_t^2}} \right]_{VaR_{a,t}}^{\infty}$$
$$= \frac{\sigma_t}{(1-a)\sqrt{2\pi}} e^{-\frac{VaR_{a,t}^2}{2\sigma_t^2}} = \frac{\sigma_t}{(1-a)\sqrt{2\pi}} e^{-\frac{z_a^2 \sigma_t^2}{2\sigma_t^2}} = \frac{\sigma_t}{(1-a)\sqrt{2\pi}} e^{-\frac{z_a^2}{2}}$$

Consistent with the analysis of Artzner et al. (1997, 1999), Acerbi et al. (2001) also suggested that the mean of the worst cases beyond the level of VaR should better distinguish different levels of risk between various portfolios and they proposed the replacement of VaR with ES for financial risk management purposes.

Taking all these into account, the Committee on Banking Regulation proposed the replacement of the VaR by the expected shortfall as a risk measurement, as it accounts for the losses beyond the confidence interval (Basel III Accord).

Despite the fact that the ES is a coherent risk measure, this measure also has limitations. In particular, it cannot be reliably backtested (crucial at Basel 2.5) in the sense that forecasts of ES cannot be verified through comparison with historical observations, while VaR can easily be backtested. So it turns out that there is not one risk measure that should be adequate in each case.

4.2.3 New approaches of risk evaluation

Another subbaditive risk index, rather popular, is the recently proposed Aumann and Serrano (2008) index of riskiness, A/S (or $Modified - A/S$), with respect to the non-normal distributions, which quantify the risk of the R_i (or the excess $R_i - R_f$) series of returns, respectively, without referring to a specific utility function or preference order. The economic interpretation of the index is straightforward; an increase in the index of riskiness leads investors from being willing to hold an asset, no longer being willing to hold it, and implies that less averse investors accept riskier gambles or less risk averse investors are expected to invest in riskier assets. The A/S risk index is the positive number that satisfies the following equation:

$$E[\exp(-\frac{R_i}{A/S\ index})] = 1 \quad (4.15)$$

In our analysis, we also estimate the risk of the employed portfolios through the $Modified - A/S$ index, which is calculated as:

$$E[\exp(-\frac{R_i - R_f}{M - A/S\ index})] = 1 \quad (4.16)$$

From the above equations it appears that it is difficult to come up with an accurate estimate of the A/S index because a distribution estimate is necessary. The equation gives a unique solution when there are negative outcomes and the expected value of the series is greater than zero.

For normal distributions, the A/S index converges to $\sigma_i^2/(2E(R_i))$, where the σ_i^2 is the variance of the portfolio, inducing thus asymptotically the same ranking with that of the standard deviation, while for series with low (high) skewness and/or relatively high (low) excess kurtosis, the portfolio is ranked lower (higher) by the A/S than by the standard deviation.

Under the suggestions of Aumann and Serrano (2008), a reasonable risk index should satisfy the following axioms:

1. *Duality*: This property states that if an agent accepts a gamble at a fixed wealth, then a uniformly less risk averse agent would accept any gamble with smaller risk index at that wealth.

2. *Positive homogeneity*[5]

[5]Described in (3) of Axioms of coherence

The A/S index of riskiness is defined as the reciprocal of the absolute risk aversion (ARA) of an individual with constant ARA who is indifferent between taking and not taking a gamble. The same authors proved that this index satisfies important properties, such as monotonicity and subadditivity. In their analysis, the same authors also discussed the relations between their risk measure and other risk measures proposed in literature, such as the Sharpe ratio, the Value-at-Risk and the coherent risk measure proposed by Artzner et al. (1999).

Compared with the traditional risk measures, which consider only the series dispersion and ignore the series' actual values, the A/S index of riskiness provides a better characterization of the underlying true risk. Additionally, contrary to downside risk measures, such as the VaR and ES, which depend on a specified confidence level, the A/S index does not require such assumptions. However, computationally, it is not easy to calculate the A/S index, unless one knows the probability distribution to calculate the expectation.

4.2.4 Drawdown-based risk indices

Finally, portfolio risk can also be quantified by employing indices based on drawdown quantities, which are rather popular, especially among commodity traders, due to the incorporated information about continually accumulated losses. Following Auer and Schuhmacher (2013) and Schuhmacher and Eling (2011), the employed risk indices are calculated on the basis of monthly excess returns:

$$z_t = R_{it} - R_{f,t} \quad for \;\; t = 1, ..., T \tag{4.17}$$

where R_{it} is the portfolio return from holding it during the period from the end of the month $t-1$ to the end of the month t and $R_{f,t}$ is the corresponding risk-free interest rate of return. In our analysis, we define drawdown as the cumulated uncompounded excess returns (CUERs), with the CUER from holding a portfolio from $t=i$ to $t=j$ with $j>i$ is given by:

$$z_{ij} = z_{i+1} + ... + z_j \tag{4.18}$$

The drawdown-based measures used to quantify risk are the maximum drawdown, defined as the largest negative cumulated uncompounded excess returns:

$$Max\ drawdown_i = \max_{ij>i}(-z_{ij}) \tag{4.19}$$

along with the mean of the K continuous drawdowns, $cdd_{i,k}$ $(k=1,...,K)$, which are the CUERs that are not interrupted by a positive excess return for each portfolio i:

$$Cum.Drawdown_i = (1/K)\sum_{k=1}^{K} cdd_{i,k} \tag{4.20}$$

Specifically, cdd_1 is the largest, cdd_2 is the second largest and cdd_K is the smallest continuous drawdown taken into consideration. The maximum drawdown risk index can also be seen as an upper bound for losses by investing on a specific portfolio during a certain period and therefore can be characterized as a rather useful tool in determining risk.

Additionally, portfolio risk is also estimated by computing the index:

$$cdd2_i = \sqrt{\sum_{k=1}^{K} cdd_{i,k}^2} \tag{4.21}$$

Note that for the calculation of the aforementioned indices K is set equal to 5, following the existing literature (Eling, 2008; Eling and Schuhmacher, 2007; Auer and Schuhmacher, 2013).

For a specific portfolio i, the risk is assessed by employing two more indices, which assign weights to drawdowns (only if there exist) from the previous peak, $ddp_{it} = \max_{1 \leq i \leq t}(-z_{it})$, and are computed using the following formulas:

$$ddp_i = (1/T) \sum_{t=1}^{T} ddp_{it} \tag{4.22}$$

$$ddp2_i = \sqrt{(1/T) \sum_{t=1}^{T} ddp_{it}^2} \tag{4.23}$$

with T being the total number of monthly observations. These risk measures can be calculated only if negative excess returns exist.

4.3 Empirical results on portfolio ranking

4.3.1 Data

The data used in the following analysis are monthly returns for the period from January 2000 to December 2013 (168 observations) on the market portfolio (CRSP value-weighted portfolio return), the risk-free interest rate of return (1-month T-bill returns from Ibbotson and Associates, Inc.), the HML and SMB factors, the momentum (MOM), Long-Term Reversal (LT) and Short-Term Reversal (ST) portfolios. Additionally, the dataset is enriched with the smallest (size1) and the biggest (size10) portfolios among 10 portfolios formed by size, the lowest (BM1) and highest (BM10) one among the 10 decile portfolios formed based on their book-to-market ratio, along with the lowest and highest ones among the 10 prior-return-based portfolios, which are constructed using NYSE prior (2-12), (13-60) and (1-1) returns, referred as mom1, mom10, LT1, LT10, ST1 and ST10, respectively. The full dataset along with details about the construction of each portfolio of returns is available at Kenneth French's website (http://mba.tuck.dartmouth.edu/pages /faculty/ken.french /data_library. html).

4.3.2 Risk assessment of the Fama/French portfolios

The present analysis contributes to the literature by assessing the popular HML, SMB, MOM, LT and ST factors with respect to the underlying risk employing the most widely used risk indices. The particular factor portfolios have gained acceptance during the last years due to their performance, as they exhibit high forecasting ability on bond and stock market returns. As depicted in Table 4.1, which reports the descriptive statistics of the particular portfolios, they all evince positive mean return and are characterized as skewed distributed with high kurtosis, with the SMB factor portfolio demonstrating the highest kurtosis, followed by the MOM factor portfolio. With respect to the evinced skewness, MOM is connected with negative, while the remaining series are positive skewed.

Turning to the ranking results produced by the different indices, Table 4.2 reports that the MOM factor portfolio, constructed based on stocks' prior performance, exhibits the highest risk, consistent with the standard deviation's estimate of risk results. Specifically, the ranking produced by the Value-at-Risk reveals the MOM factor as the most risky (suggesting that the 'winners' are relatively riskier than the 'losers'). The LT factor portfolio

Table 4.1: Descriptive statistics of the FF, reversal and momentum portfolios

Descriptive statistics	**Market**	**HML**	**SMB**	**MOM**	**LT**	**ST**
Mean	0.0045	**0.0050**	0.0042	*0.0018*	0.0026	0.0030
Median	**0.0125**	0.0030	0.0019	0.0037	0.0031	*0.0008*
Maximum	0.1134	0.1387	**0.2202**	0.1839	*0.1096*	0.1623
Minimum	-0.1715	-0.1268	-0.1639	*-0.3472*	**-0.0706**	-0.1451
St. Dev.	0.0466	0.0340	0.0355	**0.0593**	*0.0275*	0.0434
Skewness	-0.5890	0.0901	**1.0000**	*-1.5389*	0.4086	0.3624
Kurtosis	*3.6511*	6.4933	**13.4155**	11.5090	4.0422	6.0504

Notes: Bold and Italics indicate the highest and lowest value, respectively.

appears to be the least risky among the competing ones. As for the market portfolio of returns, the empirical results evince that market could be characterized as a high-risk portfolio, lying between the MOM and the ST factor portfolios. This rank order is maintained irrespective of the confidence level assumed.

Turning to the other risk measures, one can observe from Table 4.2 that this rank order is preserved for the majority of them. In particular, with the exception of the Modified Value-at-Risk, $M - VaR$, the remaining risk indices provide identical ranking for the portfolios assumed. As observed, even after penalizing excess kurtosis (or negative skewness) by applying the Aumann-Serrano index along with its modification, both measures (A/S index and M-A/S index) produce identical results for the high-risk portfolios.

With respect to the ranking deviations, the $M - VaR$ indices, irrespective of the confidence level assumed (95%, 97.5% and 95%), lead to a different rank order for the high-risk portfolios, attributed to the rather high skewness and excess kurtosis of the SMB and MOM factor, as depicted in Table 4.1.

Additionally, the risk computed by both the A/S indices reveals differences in ranking for low-risk portfolios, which is attributed to the penalty imposed for skewness and

Table 4.2: Risk indices on the FF, reversal and momentum portfolios

Risk indices	Market	HML	SMB	MOM	LT	ST
Beta	**1.0000**	-0.0762	0.2289	*-0.4696*	0.0833	0.3323
Semi-sd	0.0357	0.0236	0.0231	**0.0471**	*0.0186*	0.0292
VaR 99%	0.1085	0.0791	0.0827	**0.1383**	*0.0642*	0.1011
VaR 97.5%	0.0913	0.0665	0.0696	**0.1163**	*0.0540*	0.0850
VaR 95%	0.0764	0.0557	0.0582	**0.0973**	*0.0452*	0.0711
M-VaR 99%	0.0893	0.1092	**0.1827**	0.1367	*0.0775*	0.1417
M-VaR 97.5%	0.0780	0.0761	**0.1067**	0.0872	*0.0606*	0.1007
M-VaR 95%	0.0678	0.0540	0.0598	0.0585	*0.0477*	**0.0727**
ES 99%	0.1231	0.0897	0.0939	**0.1569**	*0.0728*	0.1147
ES 97.5%	0.1089	0.0794	0.0830	**0.1388**	*0.0644*	0.1014
ES 95%	0.0969	0.0706	0.0739	**0.1234**	*0.0573*	0.0902
A/S index	0.2473	*0.1159*	0.1427	**1.0182**	0.1394	0.3078
M-A/S index	0.3888	*0.1701*	0.2380	**10.2900**	0.3799	0.6920
LPM1	0.0181	0.0116	0.0119	**0.0189**	*0.0107*	0.0142
LPM3	0.0001	*0.0000*	*0.0000*	**0.0005**	*0.0000*	0.0001
Max drawdown	**0.6357**	0.1329	0.0637	0.2369	*0.0629*	0.2122
Cum.drawdown	**0.5930**	0.0606	*0.0412*	0.2160	0.0531	0.1323
cdd2	**1.7625**	0.0235	*0.0101*	0.2345	0.0144	0.1003
ddp	0.2018	0.1072	*0.1020*	**0.3270**	0.2074	0.1436
ddp2	0.0714	0.0207	*0.0175*	**0.1535**	0.0632	0.0374

Notes: Bold and Italics indicate the most and least risky portfolio, respectively.

excess kurtosis. Turning to the drawdown measures, Table 4.2 also depicts that they lead to a different rank order due to the incorporated information of continually accumulated losses.

Although most of the employed indices, apart from a few exceptions, imply identical rankings for the employed portfolios, some of them are proved inadequate to assess risk portfolio under certain conditions. In particular, in cases of extraordinary high returns, the popular standard deviation would be higher than the drawdown indices, overestimating thus the associated risk of the investment.

Table 4.3: Descriptive statistics of the size and B/M portfolios

Descriptive statistics	**size1**	**size10**	**BM1**	**BM10**
Mean	0.0037	*0.0030*	*0.0030*	**0.0097**
Median	0.0098	*0.0080*	0.0085	**0.0127**
Maximum	*0.1044*	0.1066	0.1121	**0.3669**
Minimum	-0.1598	**-0.1487**	-0.1620	*-0.2813*
St. Dev.	0.0446	*0.0444*	0.0496	**0.0716**
Skewness	-0.5134	-0.4530	*-0.5223*	**0.0295**
Kurtosis	3.6495	*3.5055*	3.6672	**7.1920**

Notes: Bold and Italics indicate the highest and lowest value, respectively.

4.3.3 How size, book-to-market and prior stock performance are related to risk?

Attempting to investigate whether specific stock characteristics, such as size, book-to-market or previous stock performance, can have a systematic impact on the estimated risk of specific portfolios of returns, we employ ten (10) different empirical portfolios of returns; each one characterized by either small or big size, low or high book-to-market and low or high prior stock performance. Table 4.3 presents the descriptive statistics of these portfolios, revealing positive mean return for all of them. With respect to the size-based portfolios, Table 4.3 depicts that both of them are negatively skewed with excess kurtosis, with the small-size (size1) portfolio exhibiting these characteristics intensively. With respect to the book-to-market effect, the portfolio of value stocks (BM10) evinces positive skewness and the highest kurtosis among the size and book-to-market portfolios. The implied ranking produced by the standard deviation (*ST.Dev*) indicates that stocks of high book-to-market or small size could be characterized as more risky relative to stocks of the opposite characteristics.

Table 4.4: Risk indices on size and B/M portfolios

Risk indices	**size1**	**size10**	**BM1**	**BM10**
Beta	0.9460	*0.9316*	1.0041	**1.2636**
Semi-sd	0.0338	*0.0334*	0.0375	**0.0520**
VaR 99%	0.1039	*0.1035*	0.1155	**0.1668**
VaR 97.5%	0.0874	*0.0870*	0.0972	**0.1403**
VaR 95%	0.0731	*0.0728*	0.0813	**0.1174**
M-VaR 99%	*0.0894*	0.0904	0.0991	**0.2391**
M-VaR 97.5%	*0.0768*	0.0777	0.0852	**0.1620**
M-VaR 95%	*0.0659*	0.0665	0.0731	**0.1117**
ES 99%	0.1179	*0.1174*	0.1310	**0.1893**
ES 97.5%	0.1043	*0.1038*	0.1159	**0.1674**
ES 95%	0.0928	*0.0923*	0.1031	**0.1489**
A/S index	0.2744	0.3306	**0.4111**	*0.2671*
M-A/S index	0.0500	0.0500	0.0500	0.0500
LPM1	0.0170	*0.0169*	0.0189	**0.0257**
LPM3	*0.0001*	*0.0001*	*0.0001*	**0.0004**
Max drawdown	0.7117	0.7982	**0.9309**	*0.3434*
Cum.drawdown	0.6410	0.7127	**0.8689**	*0.3105*
cdd2	2.0615	2.5501	**3.7808**	*0.4854*
ddp	0.2536	0.3193	**0.4362**	*0.1418*
ddp2	0.0962	0.1357	**0.2480**	*0.0569*

Notes: Bold and Italics indicate the most and least risky portfolio, respectively.

Turning to the risk assessment through the remaining measures, as shown in Table 4.4, our analysis leads to results consistent with those of standard deviation, revealing that low market capitalization (size1) along with value stocks (BM10) create high-risk portfolios. Similar to the results of Table 4.2, the 3 variations of $M - VaR$, both A/S indices and the drawdown-based risk measures produce alterations in rank order, attributed to either the excess kurtosis and skewness or to the considered continuous losses.

In particular, the A/S index depicts an inverse rank order among the size and value portfolios, with the growth-stock (BM1) portfolio appearing as the most risky and

Table 4.5: Descriptive statistics of the prior-performance portfolios

Descriptive statistics	**mom1**	**mom10**	**LT1**	**LT10**	**ST1**	**ST10**
Mean	0.0032	0.0066	**0.0115**	0.0049	0.0041	*0.0016*
Median	*0.0046*	**0.0180**	0.0118	0.0102	0.0116	0.0066
Maximum	**0.4577**	0.2310	0.2261	*0.1288*	0.2741	0.2148
Minimum	-0.2609	-0.2463	-0.1881	-0.2226	*-0.2971*	**-0.1758**
St. Dev.	**0.1096**	0.0648	0.0748	*0.0608*	0.0902	0.0610
Skewness	**0.7667**	-0.4284	-0.0603	*-0.6642*	-0.4371	-0.0891
Kurtosis	**6.0902**	4.5298	*3.2087*	4.0252	4.8294	3.8551

Notes: Bold and Italics indicate the highest and lowest value, respectively.

the portfolio with high book-to-market (BM10) stocks being the least risky one among the competing ones. The same index produces a different rank order for the size portfolios, as well. The $Modified-VaR$ risk measures give a different rank order only for size portfolios, revealing the big-size portfolio (size10) as the riskier compared to the small-size (size1) one. Differences are also induced by applying the drawdown risk measures, through which we conclude that high risk is connected to stocks of low book-to-market and big size.

Turning to portfolios constructed based on their, low or high, prior momentum or reversal performance, Table 4.5 presents their descriptive statistics, revealing once again the main characteristic of the employed portfolios, their positive mean return. The same table evinces that portfolios constructed based on prior-stock performance exhibit skewness and excess kurtosis. Regarding the associated risk, the portfolio of low momentum (mom1) appears to be the most risky among the 6 prior-performance employed portfolios, according to the standard deviation, while the portfolio with the high long-term reversal performance (LT10) is proved to be the least risky.

This ranking order is preserved by computing the associated risk through the remaining risk indices, as presented in Table 4.6. Specifically, the results indicate that the

Table 4.6: Risk indices on the prior-performance portfolios

Risk indices	**mom1**	**mom10**	**LT1**	**LT10**	**ST1**	**ST10**
Beta	**1.9247**	*1.0842*	1.3476	1.1957	1.6943	1.0953
Semi-sd	**0.0722**	0.0487	0.0533	0.0464	0.0675	*0.0444*
VaR 99%	**0.2553**	0.1510	0.1742	*0.1417*	0.2102	0.1422
VaR 97.5%	**0.2147**	0.1270	0.1466	*0.1192*	0.1768	0.1196
VaR 95%	**0.1797**	0.1063	0.1226	*0.0997*	0.1479	0.1001
M-VaR 99%	**0.3727**	0.1493	0.1745	*0.1164*	0.2135	0.1503
M-VaR 97.5%	**0.2684**	0.1189	0.1455	*0.1004*	0.1670	0.1205
M-VaR 95%	**0.1950**	0.0961	0.1210	*0.0866*	0.1330	0.0974
ES 99%	**0.2896**	0.1713	0.1976	*0.1608*	0.2385	0.1613
ES 97.5%	**0.2562**	0.1515	0.1748	*0.1422*	0.2109	0.1427
ES 95%	**0.2279**	0.1348	0.1555	*0.1265*	0.1876	0.1269
A/S index	**1.8381**	0.3265	*0.2432*	0.3895	1.0026	1.1486
M-A/S index	*0.0500*	0.4354	0.2846	0.5865	1.6878	**41.0100**
LPM1	**0.0375**	0.0245	0.0287	*0.0233*	0.0314	0.0235
LPM3	**0.0009**	0.0003	0.0004	0.0003	**0.0009**	*0.0002*
Max drawdown	**1.6878**	0.4917	*0.1188*	0.7972	1.1802	1.0998
Cum.drawdown	**1.4399**	0.4648	*0.0648*	0.7684	0.9419	1.0512
cdd2	**10.4481**	1.0814	*0.0170*	2.9543	4.5125	5.5285
ddp	**0.3882**	0.2484	*0.1216*	0.2986	0.3370	0.5474
ddp2	0.2834	0.1101	*0.0374*	0.1553	0.2030	**0.3562**

Notes: Bold and Italics indicate the most and least risky portfolio, respectively.

implied ranking is preserved for high-risk portfolios, with the exception of the $M - A/S$ *index* and the $ddp2$ risk measure, with the portfolio of low momentum (mom1) appearing as the most risky among the competing ones.

Concerning the high and low LT reversal portfolios, apart from the drawdown-based risk measures and both A/S indices, the remaining indices reveal that low long-term reversal (LT1) effect is related to high risk, compared to the opposite effect. Similar relative risk-based ranking is obtained for the portfolios constructed according to prior short-term reversal performance, as well. With respect to the rank order produced by the A/S risk indices and the drawdown-based ones, the exhibited differences in risk measurement induced by the associated skewness and kurtosis of the portfolios assumed and the incorporated information of the continuous losses, respectively.

To sum up, the evidence in this section suggests that high risk is attributed to specific return characteristics and this portfolio attitude is revealed by the majority of the applied risk indices.

4.4 Robustness evidence

In this section, we investigate whether our ranking results, produced by the various risk measures[6] are robust.

From Table 4.7, one can notice that the specific portfolios evince positive mean returns even after extending the sample, reinforcing thus the necessity for further analysis on these factors. Consistent with the results of Table 4.1, the MOM factor portfolio appears to be the most risky among the competing Fama/French and reversal portfolios according to the *St.Dev* risk measure, while the LT reversal portfolio seems to incorporate low risk.

In Table 4.8, we report the estimated risk of these portfolios computed by using the full set of risk measures. Apart from the market portfolio, which appears to be the most risky portfolio now, the momentum factor is ranked high in the risk order, while the LT portfolio of returns exhibits low risk compared to the other portfolios. The results indicate that the implied rank order produced by employing the different risk indices remains almost unaffected, meaning that different measures produce identical ranking, with a few exceptions, irrespective of the sample size assumed.

[6]The data used are monthly returns for the series used in the previous sections extend from July 1963 to December 2013, consisting now of 606 observations.

Table 4.7: Descriptive statistics of the FF, reversal and momentum portfolios (1963-2013)

Descriptive statistics	Market	HML	SMB	MOM	LT	ST
Mean	**0.0091**	0.0037	*0.0026*	0.0070	0.0031	0.0051
Median	**0.0126**	0.0036	*0.0009*	0.0078	0.0017	0.0035
Maximum	0.1661	*0.1387*	0.2202	**0.1839**	0.1447	0.1623
Minimum	-0.2264	-0.1268	-0.1639	*-0.3472*	**-0.0778**	-0.1451
St. Dev.	**0.0447**	0.0288	0.0311	0.0426	*0.0253*	0.0316
skewness	-0.5175	-0.0024	0.5386	*-1.4217*	**0.6340**	0.3688
kurtosis	*4.9233*	5.5695	8.5955	**14.0732**	5.6320	8.5980

Notes: Bold and Italics indicate the highest and lowest value, respectively.

Table 4.8: Risk indices on the FF, reversal and momentum portfolios (1963-2013)

Risk indices	Market	HML	SMB	MOM	LT	ST
Beta	**1.0000**	*-0.1900*	0.2168	-0.1173	-0.0090	0.2054
Semi-sd	**0.0334**	0.0203	0.0210	0.0331	*0.0167*	0.0213
VaR 99%	**0.1042**	0.0670	0.0724	0.0992	*0.0589*	0.0736
VaR 97.5%	**0.0877**	0.0564	0.0609	0.0835	*0.0496*	0.0619
VaR 95%	**0.0734**	0.0472	0.0509	0.0699	*0.0415*	0.0518
M-VaR 99%	0.1029	0.0844	0.1223	**0.1331**	*0.0826*	0.1223
M-VaR 97.5%	**0.0809**	0.0614	0.0794	0.0746	*0.0602*	0.0790
M-VaR 95%	**0.0648**	0.0456	0.0518	*0.0413*	0.0444	0.0513
ES 99%	**0.1182**	0.0761	0.0821	0.1126	*0.0668*	0.0835
ES 97.5%	**0.1046**	0.0673	0.0726	0.0996	*0.0591*	0.0739
ES 95%	**0.0930**	0.0598	0.0646	0.0886	*0.0526*	0.0657
A/S index.	0.1210	0.1122	**0.1820**	0.1546	*0.0740*	0.1018
M-A/S index	*0.2100*	5.1700	**20.5300**	0.3543	10.2900	0.5088
LPM1	**0.0170**	0.0104	0.0113	0.0141	*0.0094*	0.0105
LPM3	0.0001	*0.0000*	*0.0000*	**0.0002**	*0.0000*	*0.0000*
Max drawdown	0.2258	0.9526	**1.5893**	*0.0000*	0.9902	0.0044
Cum.drawdown	0.1324	0.8681	**1.5536**	*0.0000*	0.9703	0.0035
cdd2	0.1015	3.7801	**12.0708**	*0.0000*	4.7079	*0.0000*
ddp	*0.1342*	0.4108	**1.0489**	0.1790	0.6234	0.2840
ddp2	*0.0400*	0.2380	**1.4712**	0.0618	0.5804	0.1309

Notes: Bold and Italics indicate the most and least risky portfolio, respectively.

Table 4.9: Descriptive statistics of the size and B/M portfolios (1963-2013)

Descriptive statistics	**size1**	**size10**	**BM1**	**BM10**
Mean	0.0087	0.0085	*0.0083*	**0.0134**
Median	0.0112	0.0106	*0.0091*	**0.0152**
Maximum	*0.1812*	*0.1812*	0.2303	**0.3669**
Minimum	-0.2031	**-0.1972**	-0.2274	*-0.2813*
St. Dev.	0.0428	*0.0426*	0.0513	**0.0589**
skewness	*-0.3736*	-0.3455	-0.2091	**0.0343**
kurtosis	4.7685	4.7085	*4.4411*	**7.6915**

Notes: Bold and Italics indicate the highest and lowest value, respectively.

Table 4.10: Descriptive statistics of the prior-performance portfolios (1963-2013)

Descriptive statistics	**mom1**	**mom10**	**LT1**	**LT10**	**ST1**	**ST10**
Mean	*0.0021*	**0.0154**	0.0134	0.0088	0.0102	0.0065
Median	*0.0018*	**0.0180**	0.0115	0.0118	0.0116	0.0077
Maximum	**0.4577**	*0.2310*	0.3917	0.2551	0.3493	0.2444
Minimum	-0.2609	-0.2674	*-0.2994*	**-0.2441**	-0.2971	-0.2711
St. Dev.	**0.0806**	0.0624	0.0666	0.0596	0.0733	*0.0552*
skewness	**0.6449**	*-0.4108*	0.2676	-0.3527	-0.2665	-0.2521
kurtosis	**7.3741**	4.8135	5.9421	*4.3788*	5.9593	5.0682

Notes: Bold and Italics indicate the highest and lowest value, respectively.

With respect to the big and small-size portfolios along with the portfolios formed by value or growth stocks, Table 4.9 shows that they also exhibit positive mean returns, excess kurtosis and skewness, consistent with the results of Table 4.3. By computing the descriptive statistics for the portfolios based on stock prior-performance, presented in Table 4.10, we conclude to similar results concerning the skewness and kurtosis of the aforementioned portfolios compared with the descriptive statistics of Table 4.5.

The robustness of the ranking results for these portfolios is investigated by computing the underlying risk of the size-based, value-based and prior-performmance-based portfo-

Table 4.11: Risk indices on the size and B/M portfolios (1963-2013)

Risk indices	**size1**	**size10**	**BM1**	**BM10**
Beta	0.9434	*0.9268*	1.0654	**1.0711**
Semi-sd	0.0315	*0.0313*	0.0371	**0.0423**
VaR 99%	0.0997	*0.0992*	0.1195	**0.1373**
VaR 97.5%	0.0839	*0.0834*	0.1005	**0.1155**
VaR 95%	0.0702	*0.0698*	0.0841	**0.0966**
M-VaR 99%	*0.1035*	*0.1035*	0.1281	**0.2040**
M-VaR 97.5%	*0.0806*	0.0807	0.1002	**0.1355**
M-VaR 95%	*0.0640*	*0.0640*	0.0795	**0.0914**
ES 99%	0.1131	*0.1125*	0.1355	**0.1558**
ES 97.5%	0.1001	*0.0995*	0.1199	**0.1378**
ES 95%	0.0890	*0.0885*	0.1066	**0.1226**
A/S index.	*0.1120*	0.1140	**0.1620**	0.1381
M-A/S index	0.2090	0.2142	**0.3269**	*0.1928*
LPM1	0.0161	*0.0160*	0.0195	**0.0212**
LPM3	*0.0001*	*0.0001*	*0.0001*	**0.0002**
Max drawdown	**0.2535**	0.2519	0.1069	*0.0357*
Cum.drawdown	**0.1489**	0.1477	0.0395	*0.0357*
cdd2	**0.1273**	0.1258	0.0128	*0.0013*
ddp	0.1487	0.1713	**0.2813**	*0.1057*
ddp2	0.0463	0.0588	**0.1371**	*0.0317*

Notes: Bold and Italics indicate the most and least risky portfolio, respectively.

lios using the full set of the employed risk measures. The implied thus rank order of these portfolios is depicted in Tables 4.11 and 4.12, which show that the value-stock (BM10) portfolio still remains high in the rank order respecting the associated risk, while the least risky portfolio appears to be the portfolio of big-size (size10) stocks, consistent with the results of Table 4.4. Turning to Table 4.12, the results imply that the low momentum (mom1) portfolio is connected with high risk, similar to the realizations of Table 4.6, while differences in rank order are evident for the low risk portfolio, which now is appeared to be the portfolio based on stocks with high short-term (ST10) prior reversal performance.

Table 4.12: Risk indices on the prior-performance portfolios (1963-2013)

Risk indices	mom1	mom10	LT1	LT10	ST1	ST10
Beta	**1.4409**	1.1827	1.2064	1.2350	1.4388	*1.0638*
Semi-sd	**0.0542**	0.0462	0.0458	0.0439	0.0534	*0.0403*
VaR 99%	**0.1878**	0.1454	0.1551	0.1389	0.1708	*0.1286*
VaR 97.5%	**0.1580**	0.1223	0.1304	0.1168	0.1437	*0.1082*
VaR 95%	**0.1322**	0.1023	0.1091	0.0977	0.1202	*0.0905*
M-VaR 99%	**0.2966**	0.1491	0.2126	*0.1399*	0.2055	0.1439
M-VaR 97.5%	**0.2019**	0.1164	0.1516	0.1114	0.1485	*0.1089*
M-VaR 95%	**0.1388**	0.0925	0.1099	0.0899	0.1100	*0.0841*
ES 99%	**0.2131**	0.1649	0.1759	0.1575	0.1937	*0.1459*
ES 97.5%	**0.1885**	0.1459	0.1556	0.1393	0.1714	*0.1291*
ES 95%	**0.1676**	0.1298	0.1384	0.1239	0.1524	*0.1148*
A/S index.	**1.6100**	*0.1397*	0.1660	0.2060	0.2807	0.2320
M-A/S index	**18.5220**	*0.1859*	0.2379	0.3950	0.4655	0.6900
LPM1	**0.0283**	0.0236	0.0246	0.0227	0.0263	*0.0209*
LPM3	0.0004	0.0003	0.0003	*0.0002*	**0.0005**	*0.0002*
Max drawdown	**3.1836**	0.0001	0.0023	0.6247	*0.0000*	0.5068
Cum.drawdown	**2.9357**	0.0001	0.0023	0.4984	*0.0000*	0.4463
cdd2	**43.1726**	*0.0000*	*0.0000*	1.2627	*0.0000*	1.0017
ddp	**1.6864**	0.1451	*0.1216*	0.2796	0.1918	0.3950
ddp2	**3.4804**	0.0502	*0.0360*	0.1488	0.0876	0.2310

Notes: Bold and Italics indicate the most and least risky portfolio, respectively.

With respect to the ranking deviations, $M - VaR$ along with the A/S and the drawdown-based indices produce different ranking for the employed portfolios compared to the ranking results produced by the remaining risk measures, an attitude that is consistent to their performance depicted in our main analysis.

To sum up, we prove that the majority of the employed risk measures produce identical ranking results. The choice thus of the risk measure does not affect the implied rank order among the different portfolios, proving once again that the characterization of an investment as risky or not is not affected by the measure employed, apart from a few exceptions.

4.5 Conclusions

In this study we employ a plethora of different risk indices used extensively in literature to measure risk. The contribution of this analysis to the literature consists to ranking the empirical Fama/French, reversal and momentum factor portfolios with respect to the incorporated risk, by applying not only traditional risk measures, such as the standard deviation or the beta factor, but also more sophisticated ones, covering different aspects of risk (e.g. downside risk measures or drawdown-based ones). Although these portfolios are broadly used in forecasting stocks and bond returns due to their significant predictive ability, there has not been sufficient literature on their risk-based ranking.

Our results suggest that, among the employed factors, the MOM factor appears as a high-risk portfolio, while the HML and LT factors are related with low risk. This performance is evident not only by the traditional risk measures, but also by downside risk indices. On the other hand, risk indices based on drawdowns along with the $Modified{-}VaR$ lead to different rank order for high-risk portfolios. Differences in the implied rank order are induced by applying the A/S indices, as well, for the cases of low-risk portfolios, attributed to the considered skewness and kurtosis.

This study also contributes to the literature by establishing a connection between size, book-to-market ratio and stock prior performance with the underlying risk of portfolios with these specific characteristics. Our findings suggest that portfolios constructed by small-size or high book-to-market stocks are related with high risk, evidence which is provided by almost all the indices employed, signalling thus that the choice of the measure does not change the ranking result. Concerning the portfolios based on stocks' prior performance, ranking results reveal that portfolios of low momentum or reversal (either long-term or short-term) are related with high risk. Providing evidence for an almost identical rank order, independently of the index applied, through our empirical analysis, we suggest that an investor could use any of the traditional or downside risk measures to estimate his portfolio risk or rank different investments, respecting their associated risk.

Finally, the same risk measures are applied for an extended data sample, checking thus the robustness of our results. The analysis reveals that the rank order among the different portfolios is maintained regardless of the sample assumed and evinces that the choice of the employed measure does not produces differences in rank order among the portfolios, as well.

However, identical ranking with respect to the associated risk may lead to different portfolio performance ranking respecting their risk-adjusted returns, which is an issue that needs to be investigated.

Chapter 5

Performance Evaluation of Size, Book-to-Market and Momentum Portfolios

5.1 Introduction

In recent years, risk measurement is one of the topics of concern not only for financial institutions, due to the regulatory restrictions under the Basel II Capital Accord, but also for fund managers and the academic community. The ability of a performance measure to consistently compare different portfolios concerning their level of risk along with the fact that a performance measure should be easily understood and applied makes the choice of an appropriate measure rather important.

The construction of a performance measure demands an appropriate index that should quantify the associated risk. Apart from the traditional ones, such as the beta coefficient and the standard deviation, various specifications have been proposed to cover the evinced skewness and kurtosis of returns. Specifically, the introduction of downside risk indices, such as the Value-at-Risk, the Expected shortfall and the semi-standard deviation, through which only the left hand side of a return distribution is used to measure risk, has led to a plethora of risk-adjusted performance measures, adequate to rank investment portfolios. Moreover, a variety of new performance measures has also been developed accounting for different aspects of the incorporated risk, such as the economic performance measure of Goetzmann et al. (2007) and that of Homm and Pigorsch (2012) along with a variety of recently developed measures based on drawdown of portfolio returns.

Over the last decade, there has been a debate on the choice of the appropriate performance measure. Consistent with Phingsten et al. (2004), Eling and Schuhmacher (2005) and Eling (2008) suggested that different risk measures provide similar ranking results. Complementary to the analysis of Eling and Schuhmacher (2007), who investigated whether the choice of the risk measure affects the ranking performance of hedge funds by comparing 13 different risk measures, Auer and Schuhmacher (2013) also report similar findings about the rank order of different assets. Contrary to these analyses, Ornelas et al. (2012) and Zakamouline (2011) argued that the evaluation of investment funds is influenced by the measure employed. In particular, Zamakouline (2011) proved that the rank correlation between the Sharpe Ratio and other measures decreases for higher values of skewness.

In the present analysis, we compute an extensive set of performance measures, ranging from traditional ones to more complicated ones. Contrary to earlier studies, this analysis focuses on the performance evaluation of empirical portfolios, the Fama/French, reversal and momentum factors, used by fund traders and other practitioners to forecast and evaluate stock, bond, mutual funds and hedge funds returns. Despite the forecasting ability of these portfolios, limited literature is available concerning their performance relative to the associated risk, with the exception of Bakshi et al. (2011), who showed that changes in risk are negatively related to the equity premium, value spread, size spread and momentum spread and proved that an increase in risk is connected with an underperformance of stock portfolios with high book-to-market, small capitalization and low momentum. The ranking results of the present analysis show that the MOM factor underperforms, while the HML and SMB factors evince as high-performance portfolios. This classification is maintained when employing either the traditional or the downside measures. However, when drawdown-based performance measures are applied, the performance results reveal some differences in the rank order of the competing portfolios.

Apart from ranking the aforementioned portfolios, this study contributes to the literature by establishing a link between (small and big) size, (low and high) book-to-market ratio (B/M) and prior-return of stocks and the exhibited performance of specific portfolios. Our ranking results suggest that small and high B/M portfolios appear as high-performance investments, while high momentum and low long-term and short-term reversal portfolios are connected with lower performance. This performance is evinced not only by the traditional performance measures, but also by the downside and the more sophisticated ones providing identical ranking, with minor exceptions.

The remainder of the paper is organized as follows. Section 2 illustrates the performance measures used to evaluate portfolios. Section 3 focuses on the data used and provides the ranking results. Finally, Section 4 reports the robustness analysis and Section 5 summarizes the main results and concludes.

5.2 Performance measures for portfolio evaluation

5.2.1 Traditional performance measures

For decades, the performance of a portfolio was under evaluation through measures that quantified the embedded risk via the estimated standard deviation or the beta factor. Representative performance measures of this category are the Jensen (JR), the Treynor ratio (TR) and the Sharpe ratio (SR) (see Jensen, 1968, Treynor, 1965 and Sharpe, 1966), with the first two being calculated on the basis of the correlation between the portfolio returns and market returns; namely, the beta factor β_i, and the last one employing the standard deviation σ_i of portfolio returns, as follows:

$$JR_i = E(R_i) - E(R_f) - \beta_i[E(R_M) - E(R_f)] \tag{5.1}$$

$$TR_i = [E(R_i) - E(R_f)]/\beta_i \tag{5.2}$$

$$SR_i = [E(R_i) - E(R_f)]/\sigma_i \tag{5.3}$$

where $E(R_i)$, $E(R_M)$ and $E(R_f)$ is the average portfolio, market and risk-free interest rate return, respectively.

Additionally, a modification of the SR that employs the standard deviation of excess returns, proposed by Treynor and Black (1973) is included in the present analysis:

$$IR_i = [E(R_i) - E(R_f)]/\sigma_i^* \tag{5.4}$$

5.2.2 Downside performance measures

Downside risk indices, such as semi-standard deviation or lower partial moments (LPM) of order 1 and 3, Value-at-Risk along with its Cornish-Fisher expansion and expected shortfall, have rapidly gained acceptance among both academics and practitioners due to the fact that they include only negative deviations, incorporating thus movements associated exclusively with losses.

Based on these risk indices, a variety of performance ratios have been introduced, among of which are the Sortino Ratio, introduced by Sortino and Price (1994), which incorporates the semi-standard deviation of portfolio i of returns:

$$sd_i = \sqrt{E(\max(E(R_i) - R_i, 0)^2)}, \tag{5.5}$$

the Shadwick and Keating (2002) Omega Ratio and the Kaplan and Knowles (2004) Kappa3 Ratio, which incorporate LPM of order 1 and 3 for the full sample of returns (N represents the total number of observations), respectively, given by:

$$LPM_{ni} = (1/N)\sum_{i=1}^{N} \max(E(R_i) - R_i, 0)^n \tag{5.6}$$

The downside performance measures are given by the following formulas:

$$Sortino_i = [E(R_i) - E(R_f)]/sd_i \tag{5.7}$$

$$Omega_i = 1 + [E(R_i) - E(R_f)]/LPM_{1i} \tag{5.8}$$

$$Kappa3_i = [E(R_i) - E(R_f)]/\sqrt[3]{LPM_{3i}} \tag{5.9}$$

Additionally, this analysis employs the generalized Sharpe Ratio, G_SR, proposed by Dowd (2000), the Gregoriou and Gueyie (2003) modified Sharpe ratio, M_SR, and the Conditional Sharpe ratio, C_SR, proposed by Agarwal and Naik (2004):

$$G - SR_i = [E(R_i) - E(R_f)]/VaR_i \tag{5.10}$$

$$M - SR_i = [E(R_i) - E(R_f)]/MVaR_i \tag{5.11}$$

$$C - SR_i = [E(R_i) - E(R_f)]/ES_i \tag{5.12}$$

With respect to the employed VaR_i and ES_i, these are estimated under the assumption of normal distributed series of returns, as:

$$VaR_i = z_a \sigma_i \tag{5.13}$$

$$ES_i = (\sigma_i/(1-\alpha)\sqrt{2\pi})e^{-z_a^2/2} \quad (5.14)$$

while the $MVaR_i$ risk index for non-normal series of returns as:

$$MVaR_i = \sigma_i(z_a + (z_a^2 - 1)/6 + (z_a^3 - 3z_a)E_i/24 - (2z_a^3 - 5z_a)S_i^2/36) \quad (5.15)$$

where $z_a = 2.33$ is the $a = 99\%$-quantile of the standard normal distribution, S_i the skewness and $E_i = k_i - 3$ the excess kurtosis of the series of returns.

5.2.3 New approaches of performance measures

Two more performance measures are employed, the one proposed by Goetzmann et al. (2007), which is the portfolio's premium return after adjusting for risk for an investor with a relative risk aversion of 2 ($\gamma = 2$), known as manipulation-proof performance measure (MPPM) and the economic performance measure (EPM) proposed by Homme and Pigorsch (2012), which is a generalized form of the Sharpe Ratio, with respect to the non-normal distributions, that incorporates the Aumann and Serrano (2008) index, AS_{R_i} ($AS_{R_i - R_f}$), to quantify the risk of the R_i (or the excess $R_i - R_f$) series of returns, respectively:

$$MPPM_i = \left(\frac{1}{1-\gamma}\right) \ln\left[\frac{1}{N}\sum_{t=1}^{N}\left(\frac{1+R_i}{1+R_f}\right)^{1-\gamma}\right] \quad (5.16)$$

$$EPM_i = E(R_i)/AS_{R_i} = [E(R_i) - E(R_f)]/AS_{R_i - R_f} \quad (5.17)$$

where the risk index is the positive number that satisfies the following equation:

$$E[\exp(-R_i/AS_{R_i})] = 1 \tag{5.18}$$

For normal distributions, the EPM converges to two times the squared Sharpe Ratio, inducing thus the same ranking asymptotically, while for series with low (high) skewness and/or relatively high (low) excess kurtosis, the portfolio is ranked lower (higher) by the EPM than by the Sharpe Ratio.

5.2.4 Drawdown-based performance measures

Finally, performance measures based on drawdown quantities are rather popular, especially among commodity traders, due to the incorporated information about continually accumulated losses. Following the methodology of Auer and Schuhmacher (2013), the employed performance measures are calculated on the basis of monthly excess returns. That is, the Calmar Ratio, which quantifies risk through the largest negative cumulated uncompounded excess returns, $mdd_i = \max_{ij>i}(-z_{ij})$, two ratios that use the K largest losses, $cdd_{i,k}$; namely, the Sterling and Burke Ratio, and two more ratios that measure each portfolio' risk by assigning weights to drawdowns (only if there exist) from the previous peak, $ddp_{i,t} = \max_{1\leq i\leq t}(-z_{i,t})$; namely, the Pain Ratio along with the Martin Ratio. The aforementioned measures are given as:

$$Calmar\_Ratio_i = [E(R_i) - E(R_f)]/mdd_i \tag{5.19}$$

$$Sterling\_Ratio_i = [E(R_i) - E(R_f)]/\left[\frac{1}{K}\sum_{k=1}^{K} cdd_{i,k}\right] \tag{5.20}$$

$$Burke\_Ratio_i = [E(R_i) - E(R_f)] / \sqrt{\sum_{k=1}^{K} (cdd_{i,k})^2} \tag{5.21}$$

$$Pain\_Ratio_i = [E(R_i) - E(R_f)] / \left[\frac{1}{N}\sum_{t=1}^{N} ddp_{i,t}\right] \tag{5.22}$$

$$Martin\_Ratio_i = [E(R_i) - E(R_f)] / \sqrt{\frac{1}{N}\sum_{t=1}^{N} (ddp_{i,t})^2} \tag{5.23}$$

Note that for the calculation of the Sterling and Burke ratios K is set equal to 5 following the existing literature (Eling, 2008; Eling and Schuhmacher, 2007; Auer and Schuhmacher, 2013).

5.3 Empirical results on portfolio ranking

5.3.1 Data

The data used in the following analysis are monthly returns for the period from January 2000 to December 2013 (168 observations) on the market portfolio (CRSP value-weighted portfolio return), the risk-free interest rate return (1-month T-bill returns from Ibbotson and Associates, Inc.), the HML and SMB factors, the momentum (MOM), Long-Term Reversal (LT) and Short-Term Reversal (ST) portfolios. Additionally, the dataset is enriched with the smallest (size1) and the biggest (size10) portfolios among 10 portfolios formed by size, the lowest (BM1) and highest (BM10) one among the 10 decile portfolios formed based on their book-to-market ratio, along with the lowest and highest ones among the 10 prior-return-based portfolios, which are constructed using NYSE prior (2-

12), (13-60) and (1-1) returns, referred as mom1, mom10, LT1, LT10, ST1 and ST10, respectively. The full dataset along with details about the construction of each portfolio of returns is available at Kenneth French's website (http://mba.tuck.dartmouth.edu/pages /faculty/ken.french /data_library. html).

5.3.2 Ranking the Fama/French portfolios

The present empirical analysis aims to evaluate the performance of the popular HML, SMB, MOM, LT and ST portfolios with respect to their risk-adjusted return using the most widely used performance measures. According to the performance realizations produced by the different measures, reported in Table 5.1, portfolios constructed with respect to their previous performance are ranked low. Specifically, the ranking results produced by the Sharpe Ratio show that the MOM factor underperforms (suggesting the relative low adjusted-for-risk performance of the 'winners' over the 'losers') followed by the ST, LT, SMB and finally the HML, which achieves the best performance. As for the market portfolio of returns, the empirical results show it should be characterized as a medium-performance portfolio, lying between the LT-Rev and the SMB factor.

Similar ranking pertains when the employed portfolios are evaluated on the basis of the alternative measures. In particular, with the exception of the Treynor ratio, the Jensen and the Omega, the remaining traditional and downside performance measures along with the MPPM measure all provide identical rank order for the portfolios assumed. As observed, even after penalizing excess kurtosis (or negative skewness) by applying the EPM performance measure, the rank order is maintained and both measures lead to identical ranking.

Concerning the ranking deviations, the Treynor Ratio leads to a completely different rank order, while the Jensen measure affects only low-performance portfolios. With respect to the Omega measure, it provokes differences in rank order for low-performance portfolios; namely, the LT and ST reversal ones, mainly caused by the smaller extent to which the negative deviations from the mean return are weighed, compared to the higher-order ratios, Kappa3 and Sortino.

Turning to the drawdown-based measures, our findings point to different rankings. In particular, for the high-performance HML and SMB portfolios, the Calmar, Sterling and Burke ratios produce different rank order, while the results of the Pain and Martin ratios indicate differences in ranking only when attention is restricted to the low-performance LT-Rev and ST-Rev portfolios.

Despite the fact that the majority of the employed measures produces identical rankings, some of them face disadvantages that make them inappropriate for portfolio evaluation under certain conditions. Specifically, during periods of extraordinary high returns, the popular Sharpe ratio would appear lower than drawdown measures by incorporating both-side deviations of returns. From a practical point of view, though, the choice of performance measure does not have a crucial influence on the relative evaluation of portfolio of returns, with a few exceptions.

Table 5.1: Performance measures on the FF, reversal and mom portfolios

Perform.Measures	**Market**	**HML**	**SMB**	**MOM**	**LT**	**ST**
Treynor ratio	0.0029	*-0.0441*	0.0113	-0.0003	**0.0118**	0.0041
Jensen	*0.0000*	**0.0036**	0.0019	0.0015	0.0007	0.0004
Sharpe ratio	0.0617	**0.0991**	0.0727	*0.0021*	0.0357	0.0310
Inform. ratio	0.0613	**0.0998**	0.0725	*0.0021*	0.0357	0.0310
Sortino ratio	0.0806	**0.1426**	0.1117	*0.0027*	0.0528	0.0460
G-SR 99%	0.0265	**0.0425**	0.0312	*0.0009*	0.0153	0.0133
M-SR 99%	**0.0322**	0.0308	0.0141	*0.0009*	0.0127	0.0095
C-SR 99%	0.0233	**0.0375**	0.0275	*0.0008*	0.0135	0.0117
Omega	1.1587	**1.2911**	1.2175	*1.0067*	1.0922	1.0945
Kappa3	0.0590	**0.0966**	0.0731	*0.0016*	0.0406	0.0317
MPPM (γ=2)	0.0006	**0.0022**	0.0014	*-0.0039*	0.0002	-0.0005
EPM	0.0116	**0.0290**	0.0181	*0.0001*	0.0071	0.0044
exc_EPM	0.0074	**0.0198**	0.0109	*0.0000*	0.0026	0.0019
Calmar ratio	0.0045	0.0253	**0.0405**	*0.0005*	0.0156	0.0063
Sterling ratio	0.0048	0.0555	**0.0627**	*0.0006*	0.0185	0.0102
Burke ratio	0.0022	0.0219	**0.0257**	*0.0003*	0.0082	0.0043
Pain ratio	0.0142	**0.0314**	0.0253	*0.0004*	0.0047	0.0094
Martin ratio	0.0108	**0.0234**	0.0195	*0.0003*	0.0039	0.0070

Notes: Bold and Italics indicate the best and worst performance, respectively.

Table 5.2: Performance measures on size and B/M portfolios

Perform. Measures	size1	size10	BM1	BM10
Treynor ratio	0.0022	0.0015	*0.0014*	**0.0064**
Jensen	-0.0007	-0.0013	*-0.0015*	**0.0044**
Sharpe ratio	0.0460	0.0308	*0.0278*	**0.1128**
Inform.ratio	0.0457	0.0306	*0.0276*	**0.1126**
Sortino ratio	0.0608	0.0410	*0.0368*	**0.1552**
G-SR 99%	0.0197	0.0132	*0.0119*	**0.0484**
M-SR 99%	0.0229	0.0151	*0.0139*	**0.0338**
C-SR 99%	0.0174	0.0117	*0.0105*	**0.0427**
Omega	1.1204	1.0810	*1.0731*	**1.3136**
Kappa3	0.0443	0.0301	*0.0267*	**0.1095**
MPPM (γ=2)	0.0000	-0.0006	*-0.0012*	**0.0029**
EPM	0.0075	0.0041	*0.0034*	**0.0302**
exc_EPM	0.0410	0.0274	*0.0276*	**0.1615**
Calmar ratio	0.0029	0.0017	*0.0015*	**0.0235**
Sterling ratio	0.0032	0.0019	*0.0016*	**0.0260**
Burke ratio	0.0014	0.0009	*0.0007*	**0.0116**
Pain ratio	0.0081	0.0043	*0.0032*	**0.0569**
Martin ratio	0.0066	0.0037	*0.0028*	**0.0339**

Notes: Bold and Italics indicate the best and worst performance, respectively.

5.3.3 Ranking portfolios based on size, book-to-market and previous performance

In order to examine how specific stock characteristics, such as the size, book-to-market ratio or prior returns, are related to portfolio performance, ten (10) different empirical portfolios of returns are employed; each one characterized by either small or big size, low or high book-to-market and low or high momentum/reversal.

Our findings in Table 5.2 suggest that value-stock portfolios (high B/M) perform better closely followed by low-market-capitalization portfolios. On the other hand, growth-stock portfolios (low B/M) and big-size ones achieve low performance.

Table 5.3: Performance measures on high and low mom/reversal portfolios

Perform. Measures	**mom1**	**mom10**	**LT1**	**LT10**	**ST1**	**ST10**
Treynor ratio	0.0008	0.0046	**0.0073**	0.0027	0.0014	*0.0000*
Jensen	*-0.0040*	0.0018	**0.0060**	-0.0002	-0.0024	-0.0032
Sharpe ratio	0.0142	0.0764	**0.1318**	0.0532	0.0271	*-0.0007*
Inform. ratio	0.0142	0.0761	**0.1315**	0.0530	0.0270	*-0.0007*
Sortino ratio	0.0215	0.1016	**0.1848**	0.0698	0.0362	*-0.0009*
G-SR 99%	0.0061	0.0328	**0.0566**	0.0228	0.0116	*-0.0003*
M-SR 99%	0.0042	0.0331	**0.0565**	0.0278	0.0114	*-0.0003*
C-SR 99%	0.0054	0.0289	**0.0499**	0.0201	0.0102	*-0.0003*
Omega	1.0415	1.2016	**1.3437**	1.1389	1.0777	*0.9982*
Kappa3	0.0160	0.0730	**0.1388**	0.0499	0.0253	*-0.0007*
MPPM (γ=2)	*-0.0100*	0.0006	**0.0042**	-0.0006	-0.0062	-0.0038
EPM	0.0008	0.0152	**0.0405**	0.0083	0.0024	*0.0000*
exc_EPM	0.0311	0.0114	**0.0346**	0.0055	0.0014	*0.0000*
Calmar ratio	0.0009	0.0101	**0.0829**	0.0041	0.0021	*0.0000*
Sterling ratio	0.0011	0.0106	**0.1521**	0.0042	0.0026	*0.0000*
Burke ratio	0.0005	0.0048	**0.0756**	0.0019	0.0011	*0.0000*
Pain ratio	0.0040	0.0199	**0.0810**	0.0108	0.0072	*-0.0001*
Martin ratio	0.0029	0.0149	**0.0509**	0.0082	0.0054	*-0.0001*

Notes: Bold and Italics indicate the best and worst performance, respectively.

More importantly, identical ranking are induced by all the employed performance measures.

Turning to the performance of portfolios characterized by low/high momentum, long-term and short-term reversal, Table 5.3 provides the ranking results based on the performance measures. The results indicate almost identical ranking, with the portfolio of low momentum underperforming compared to that of high momentum, while the opposite takes place for the portfolios characterized by low long-term reversal and low short-term reversal. Deviation in the rank order is created only by the exc_EPM measure and for the ranking of the competing momentum portfolios. Overall, low long-term reversal and high short-term reversal are the best and worst performing portfolios, respectively.

5.4 Robustness evidence

So far we have produced a rank order among the empirical FF portfolios, the momentum and reversal factors along with portfolios of specific stock characteristics for the period from January 2000 to December 2013.

In this section, we investigate the robustness of our results employing data series that extend from July 1963 to December 2013. The data used are monthly returns for the series used in the previous sections, consisting now of 606 observations. The measures applied for the portfolio evaluation have been described analytically above.

The implied rank order for the Fama/French, momentum and reversal factors are reported in Table 5.4 after using the full set of performance measures. Apart from the market portfolio, which exhibits the best performance now, the momentum factor is ranked high, while the SMB portfolio of returns underperforms compared to the competing ones. Performance results indicate that, even for the extended period of time, the rank order produced by employing the same performance measures remains unaffected, meaning that different measures produce identical ranking, with a few exceptions.

Turning to portfolios consisting of either big or small stocks and those constructed by stocks of high or low book-to-market ratio, their performance ranking is reported in Table 5.5. Consistent with the ranking results of Table 5.2, one can see that portfolios of small-size stocks and those characterized by high book-to-market ratio exhibit better performance than portfolios of the same, but opposed, characteristics, meaning that their rank order remains unaffected.

We continue by examining whether the ranking results for different momentum

Table 5.4: Performance measures on FF, reversal and mom portfolios (1963-2013)

Perform. Measures	**Market**	**HML**	**SMB**	**MOM**	**LT**	**ST**
Treynor ratio	0.0050	0.0022	-0.0074	*-0.0242*	**0.1164**	0.0046
Jensen	0.0000	0.0005	*-0.0027*	**0.0034**	-0.0010	-0.0001
Sharpe ratio	**0.1114**	-0.0146	*-0.0513*	0.0666	-0.0414	0.0296
Inform. ratio	**0.1113**	-0.0146	*-0.0510*	0.0667	-0.0412	0.0296
Sortino ratio	**0.1492**	-0.0207	*-0.0758*	0.0858	-0.0629	0.0440
G-SR 99%	**0.0478**	-0.0063	*-0.0220*	0.0286	-0.0178	0.0127
M-SR 99%	**0.0485**	-0.0050	*-0.0130*	0.0213	-0.0127	0.0077
C-SR 99%	**0.0422**	-0.0055	*-0.0194*	0.0252	-0.0157	0.0112
Omega	**1.2933**	0.9595	*0.8586*	1.2013	0.8880	1.0890
Kappa3	**0.1046**	-0.0145	*-0.0535*	0.0515	-0.0468	0.0288
MPPM (γ=2)	**0.0029**	-0.0012	*-0.0025*	0.0009	-0.0017	0.0000
EPM	**0.0412**	-0.0037	-0.0088	0.0184	*-0.0142*	0.0092
exc_EPM	**0.0237**	*-0.0001*	*-0.0001*	0.0080	*-0.0001*	0.0018
Calmar ratio	0.0221	-0.0004	-0.0010	NA	*-0.0011*	**0.2127**
Sterling ratio	0.0377	-0.0005	-0.0010	NA	*-0.0011*	**0.2674**
Burke ratio	0.0156	-0.0002	*-0.0005*	NA	*-0.0005*	**0.1831**
Pain ratio	**0.0372**	-0.0010	-0.0015	0.0159	*-0.0017*	0.0033
Martin ratio	**0.0249**	-0.0009	*-0.0013*	0.0114	-0.0014	0.0026

Notes: (i) Bold and Italics indicate the best and worst performance, respectively.

(ii) NA indicates no negative returns.

Table 5.5: Performance measures on size and B/M portfolios (1963-2013)

Perform. Measures	**size1**	**size10**	**BM1**	**BM10**
Treynor ratio	0.0048	0.0047	*0.0039*	**0.0087**
Jensen	-0.0001	-0.0003	*-0.0012*	**0.0040**
Sharpe ratio	0.1065	0.1018	*0.0808*	**0.1576**
Inform. ratio	0.1064	0.1017	*0.0806*	**0.1575**
Sortino ratio	0.1445	0.1384	*0.1115*	**0.2196**
G-SR 99%	0.0457	0.0437	*0.0347*	**0.0677**
M-SR 99%	0.0440	0.0418	*0.0323*	**0.0455**
C-SR 99%	0.0403	0.0385	*0.0306*	**0.0596**
Omega	1.2830	1.2711	*1.2123*	**1.4390**
Kappa3	0.1021	0.0980	*0.0801*	**0.1516**
MPPM (γ=2)	0.0027	0.0025	*0.0015*	**0.0058**
EPM	0.0407	0.0380	*0.0256*	**0.0673**
exc_EPM	0.0218	0.0202	*0.0127*	**0.0482**
Calmar ratio	0.0180	*0.0172*	0.0388	**0.2602**
Sterling ratio	0.0306	*0.0293*	0.1048	**0.2602**
Burke ratio	0.0128	*0.0122*	0.0366	**0.2602**
Pain ratio	0.0307	0.0253	*0.0147*	**0.0879**
Martin ratio	0.0212	0.0179	*0.0112*	**0.0521**

Notes: Bold and Italics indicate the best and worst performance, respectively.

and reversal portfolios, as depicted in Table 5.6, are maintained even after extending the dataset. In particular, Table 5.6 shows that the high-mom portfolio performs better than the low-mom one, while the opposite takes place for the long-term and short-term portfolios, consistent with the ranking results of Table 5.3.

However, with respect to the total rank order of the 6 prior-return portfolios, which are reported in Table 5.6, performance results reveal mom10 as the highest-rank portfolio, followed by the ST1 and LT1 ones. This finding is in contrast to the rank order of Table 5.3, which evinces that the best-performance portfolio is the LT1, followed by the mom10

Table 5.6: Performance measures on high and low mom/reversal portfolios (1963-2013)

Perform. Measures	**mom1**	**mom10**	**LT1**	**LT10**	**ST1**	**ST10**
Treynor ratio	*-0.0014*	**0.0095**	0.0077	0.0038	0.0042	0.0022
Jensen	*-0.0092*	**0.0053**	0.0033	-0.0015	-0.0012	-0.0029
Sharpe ratio	*-0.0253*	**0.1798**	0.1396	0.0781	0.0818	0.0430
Inform. ratio	*-0.0252*	**0.1796**	0.1393	0.0780	0.0818	0.0430
Sortino ratio	*-0.0376*	**0.2426**	0.2028	0.1061	0.1124	0.0590
G-SR 99%	*-0.0108*	**0.0772**	0.0599	0.0335	0.0351	0.0185
M-SR 99%	*-0.0069*	**0.0752**	0.0437	0.0333	0.0292	0.0165
C-SR 99%	*-0.0096*	**0.0680**	0.0528	0.0295	0.0310	0.0163
Omega	*0.9281*	**1.4756**	1.3772	1.2050	1.2280	1.1138
Kappa3	*-0.0270*	**0.1712**	0.1450	0.0757	0.0771	0.0418
MPPM (γ=2)	*-0.0084*	**0.0072**	0.0049	0.0010	0.0004	-0.0007
EPM	*-0.0013*	**0.0803**	0.0560	0.0226	0.0214	0.0102
exc_EPM	*-0.0001*	**0.0604**	0.0390	0.0118	0.0129	0.0034
Calmar ratio	*-0.0006*	**112.2030**	4.0381	0.0075	NA	0.0047
Sterling ratio	*-0.0007*	**112.2030**	4.0381	0.0093	NA	0.0053
Burke ratio	*-0.0003*	**112.2030**	4.0381	0.0041	NA	0.0024
Pain ratio	*-0.0012*	**0.0773**	0.0764	0.0166	0.0313	0.0060
Martin ratio	*-0.0011*	**0.0501**	0.0489	0.0121	0.0203	0.0049

Notes: (i) Bold and Italics indicate the best and worst performance, respectively.
(ii) NA indicates no negative returns.

and LT10.

As shown, the choice of the performance measure does not alter the implied rank order among the different portfolios, proving once again that an investor could choose the best/worst investment by employing one of the traditional or any of the more sophisticated performance measures.

5.5 Conclusions

This study provides an extensive review of the most widely used performance measures for the evaluation of portfolio of returns. An important contribution of this paper is that it reveals the rank order of the popular Fama/French, reversal and momentum portfolios, which are factors mainly used by fund traders as they exhibit significant forecasting and evaluation ability. The results suggest that, among the employed factors, the MOM factor appears as a low-performance portfolio, while the SMB and HML factors perform the best. This performance is evident not only by the traditional performance measures, such as the Sharpe Ratio or the Information ratio, but also by measures based on downside risk indices. On the other hand, measures based on drawdowns lead to different rank order for middle- and high-performance portfolios, meaning that for purposes of avoiding the worst investment opportunity, the choice of the measure does not affect ranking.

In order to identify how size and book-to-market ratio are related to performance, different portfolios with these specific characteristics are evaluated. Our findings suggest that portfolios constructed by small-size stocks and portfolios of high book-to-market stocks perform better than portfolios of the opposite characteristics. This evidence is provided by almost all the measures employed, signalling that the choice of the performance measure does not change the ranking result. Additionally, this study contributes to the literature by presenting how performance is related with portfolios constructed based on prior performance, momentum or reversal. Our ranking results reveal that portfolios of high momentum or low long-term and short-term reversal exhibit high performance. This attitude is evident independently of the measure applied, revealing once again that the impact of using different measures is insignificant, and thus, from an empirical perspective, any of the employed performance measures could be used.

Finally, the robustness of our results is assessed by employing the same performance measures on specific portfolios for an extended period of time. From the analysis, we prove that the rank order is maintained regardless of the horizon examined. Additionally, we provide evidence that the choice of the employed measure does not affect the rank order among the portfolios. However, identical rankings may lead to important economic significance for investors and managers, which is an issue that reserves further research.

Chapter 6

Concluding Remarks

My main aim in this thesis was to evaluate the famous Fama and French risk factors; namely, the HML and SMB, along with the momentum and reversal factors, known in the literature as MOM, LT and ST-Rev factors, with respect to their forecasting ability and their risk-adjusted performance by assessing the associated risk. In particular, Chapter 2 focused on the in-sample and out-of-sample predictability of the aforementioned empirical factors on U.S. stock returns compared to the performance of the most widely used financial variables. By employing the Autoregressive Distributed Lag methodology of Rapach and Weber (2004), we found that the majority of the employed factors exhibit significant forecasting ability, with the in-sample one assessed statistically via the Wald test, while the out-of-sample ability of these factors to forecast the U.S. stock market is tested via statistics for equal predictive ability and forecast encompassing. This analysis also revealed that the empirical factors are related to the financial variables, as the latter proxy for the former ones.

In Chapter 3, we extended the set of our predictors by adding the short-term reversal factor with its decompositions and we focused only on their out-of-sample forecasting ability on both the U.S. bond and stock market also compared to the performance of an extended set of financial variables. In line with the existing literature, we investigated whether the combination of forecasts produced by the individual factors could lead to improved forecasting results, addressing thus the instability and time-variability of the individual forecasts. The combining methods applied include both simple methods, but also more advanced ones based on the historical performance of the individual forecasts or computed by assuming forecasts of the same past-performance cluster. As expected, combination forecasts reveal superior performance relative to that produced by employing only financial variables or individual forecasts. The exhibited improvement of the MSFE of the proposed model over the MSFE of the benchmark one is evaluated statistically via the out-of-sample R^2 statistic and also from an asset allocation by computing the performance fee that an investor would be willing to pay to accrue positive utility gains induced by the information included in the empirical factors. Checking the robustness of our results, we conducted the same analysis in markets outside U.S..

The issue of quantifying the associated risk of these empirical factors was the focus of Chapter 4. We particularly studied how these factor portfolios are ranked with respect to their embedded risk and we also established a link between specific stock characteristics and risk. We employed a plethora of different risk measures ranging from traditional to more sophisticated ones, which are based on downside deviations or drawdowns in series of returns. The analysis also employed more recent approaches, incorporating thus skewness and kurtosis of the returns distribution in the risk assessment. Our results indicate that MOM is a high-risk portfolio, while the HML and LT portfolios are ranked low. Regarding the level of risk related to size, book-to-market and prior, momentum or reversal, performance, this analysis provide evidence for first time that small-size, high book-to-market and low momentum/reversal effect exhibit high risk, which is evinced by the majority of the employed risk indices. Our results remain robust even for an extended data sample.

The last chapter incorporated a plethora of performance measures and evaluated the performance of the Fama and French factors along with the momentum and reversal factor portfolios. The embedded risk of these empirical factors has been quantified via the various risk indices of the previous Chapter, revealing thus a risk-adjusted performance ranking for these portfolios. Our analysis indicates that the portfolios based on stock prior returns exhibit low performance, while the SMB and HML factors outperform relative to the competing ones. This rank order is evinced not only by the traditional performance measures, but also by the downside ones, while the drawdown-based measures produce significant ranking deviations. Furthermore, we contributed to the literature by establishing a relation between size, book-to-market and prior stock returns and the risk-adjusted performance of portfolios with these specific characteristics. We thus provide evidence that small size, high book-to-market, high momentum and low reversal effect exhibit higher relative performance compared to portfolios with opposite characteristics.

Future research should investigate whether empirical factors, in a combination framework, can succeed in forecasting implied and/ or realised volatility. Furthermore, the issue of forecasting stock and bond market returns should also be studied using daily data, investigating whether the proposed predictable patterns suggested in the present thesis can be retained in a more high-frequency setup.

Bibliography

[1] Abhyankar A., Sarno L., and Valente G. (2005). Exchange Rates and Fundamentals: Evidence on the Economic Value of Predictability. *Journal of International Economics,* 66(2): 325-348.

[2] Acerbi C. (2004). Coherent Representations of Subjective Risk Aversion. *Risk Measures for the 21st Century* (New York: Wiley): 147-207.

[3] Acerbi C., Nordio C. and Sirtori C. (2001). Expected Shortfall as a Tool for Financial Risk Management. *Working paper. Available at http:// www.gloriamundi.org.*

[4] Agarwal V. and Naik N.Y. (2004). Risk and Portfolio Decisions Involving Hedge Funds. *Review of Financial Studies,* 17(1): 63-98.

[5] Aiolfi M., and Timmermann A. (2006). Persistence in Forecasting Performance and Conditional Combination Strategies. *Journal of Econometrics*, 135(1-2): 31-53.

[6] Ang A., Chen J. and Xing Y. (2006). Downside Risk. *Working Paper 11824 http://www.nber.org/papers/w11824.*

[7] Artzner P., Delbaen F., Eber J.M. and Heah D. (1997). Thinking Coherently. *Risk*, 10(11): 68-71.

[8] Artzner P., Delbaen F., Eber J.M. and Heah D. (1999). Coherent Measures of Risk. *Mathematical Finance*, 9(3): 203-228.

[9] Asness C., Moskowitz T. and Pedersen L. (2013). Value and Momentum Everywhere. *Journal of Finance,* 68(3): 929-985.

[10] Auer B. and Schuhmacher F. (2013). Robust Evidence on the Similarity of Sharpe Ratio and Drawdown-Based Hedge Fund Performance Rankings. *International Financial Markets, Institutions and Money*, 24: 153-165.

[11] Aumann R.J. and Serrano R. (2008). An Economic Index of Riskiness. *Journal of Political Economy*, 116(5): 810-836.

[12] Bakshi G., Chabi-Yo F. and Gao X. (2011). Riskier Times and Asset Returns. *Working paper.*

[13] Bali T., Cakici N. and Chabi-Yo F. (2011). A Generalized Measure of Riskiness. *Management Science*, 57(8): 1406-1423.

[14] Banz R. (1981). The Relationship Between Return and Market Value of Common Stocks. *Journal of Financial Economics,* 9(1): 3-18.

[15] Barndorff-Nielsen O.E., Kinnebrock S. and Shephard N. (2009). Measuring Downside Risk: Realised Semivariance. In Volatility and Time Series Econometrics: Essays in

Honor of Robert Engle, Watson MW, Bollerslev T, Russel J (eds), *Oxford University Press: Oxford:* 117-137.

[16] Campbell J.Y. (1987). Stock Returns and the Term Structure. *Journal of Financial Economics*, 18(2): 373–399.

[17] Campbell J.Y. (1991). A Variance Decomposition for Stock Returns. *Economic Journal,* 101(405): 157–179.

[18] Campbell J.Y. and Shiller R.J. (1988a). The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors. *Review of Financial Studies,* 1(3): 195–228.

[19] Campbell J.Y. and Shiller R. (1988b). Stock Prices, Earnings, and Expected Dividends. *Journal of Finance,* 43(3): 661–676.

[20] Campbell J.Y. and Thomson S.B. (2008). Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average? *Review of Financial Studies,* 21(4): 1509-1531.

[21] Campbell J.Y. and Shiller R.J. (1991). Yield Spreads and Interest Rates: A Bird's Eye View. *Review of Economic Studies*, 58(3): 495–514.

[22] Campbell J.Y. and Thomson S.B. (2008). Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average? *Review of Financial Studies*, 21(4): 1509-1531.

[23] Campbell J.Y. and Viceira L.M. (2002). Strategic Asset Allocation. *Oxford University Press.*

[24] Carhart M. (1997). On Persistence in Mutual Fund Performance. *Journal of Finance,* 52(1): 57-82.

[25] Chan K.C.L., Karceski J. and Lakonishok J. (1998). The Risk and Return from Factors. *Journal of Financial and Quantitative Analysis,* 33(2): 159-188.

[26] Chekhlov A., Uryasev S. and Zabarankin M. (2005). Drawdown Measure in Portfolio Optimization. *International Journal of Theoretical and Applied Finance*, 8(1): 13-58.

[27] Chen L., Petkova R. and Zhang L. (2008). The Expected Value Premium. *Journal of Financial Economics,* 87(2): 269-280.

[28] Cheng S., Liu Y. and Wang S. (2004). Progress in Risk Measurement. Advanced Modelling and Optimization, 6(1): 1-20.

[29] Chordia T. and Shivakumar L. (2002). Momentum, Business Cycle, and Time-Varying Expected Returns. *Journal of Finance,* 57(2): 985-1019.

[30] Clark E.T. and McCracken W.M. (2001). Tests of Equal Forecast Accuracy and Encompassing for Nested Models. *Journal of Econometrics,* 105(1): 85-110.

[31] Clark E.T. and McCracken W.M. (2005). Evaluating Direct Multi-Step Forecasts. *Econometric Reviews,* 24(4): 369-404.

[32] Clark E.T. and West D.K. (2007). Approximately Normal Tests for Equal Predictive Accuracy in Nested Models. *Journal of Econometrics*, 138(1): 291-311.

[33] Cochrane J.H. and Piazzesi M. (2005). Bond Risk Premia. *American Economic Review*, 95(1): 138–160.

[34] Cornish E. and Fisher R. (1937). Moments and Cumulants in the Specification of Distributions. *Review of the International Statistical Institute*, 5: 307-320.

[35] Dangl T. and Halling M. (2012). Predictive Regressions with Time-Varying Coefficients. *Journal of Financial Economics*, 106(1): 157-181.

[36] Danielsson J., de Vries C., Jorgensen B., Mandira S. and Samorodnitsky G. (2013). Fat Tails, VaR and Subadditivity. *Journal of Econometrics*, 172(2): 283-291.

[37] Della Corte P., Sarno L., and Thorton D.L. (2008). The Expectations Hypothesis of the Term Structure of Very Short-Term Rates: Statistical Tests and Economic Value. *Journal of Financial Economics*, 89(1): 158-174.

[38] Della Corte P., Sarno L., and Tsiakas I. (2009). An Economic Evaluation of Empirical Exchange Rate Models. *Review of Financial Studies*, 22(9): 3491-3530.

[39] DeBondt W. and Thaler R. (1985). Does the Stock Market Overreact? *Journal of Finance,* 40(3): 793-805.

[40] DeBondt W. and Thaler R. (1987). Further Evidence on Investor Overreaction and Stock Market Seasonality. *Journal of Finance,* 42(3): 557-581.

[41] Diebond X. F. and Mariano S.R. (1995). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics*, 13(3): 253-263.

[42] Dowd K. (2000). Adjusting for Risk: An Improved Sharpe Ratio. *International Review of Economics and Finance*, 9(3): 209-222.

[43] Dowd K. and Blake D. (2006). After VaR: The Theory, Estimation, and Insurance Applications of Quantile-Based Risk Measures. *The Journal of Risk and Insurance,* 73(2): 193-229.

[44] Eling M. (2008). Does the Measure Matter in the Mutual Fund Industry? *Financial Analysis Journal*, 64(3): 54-66.

[45] Eling M. and Schuhmacher F. (2005). Performance Measurement of Hedge Fund Indices-Does the Measure Matter? *Operations Research Proceedings, Berlin: Springer:* 205-210.

[46] Eling M. and Schuhmacher F. (2007). Does the Choice of Performance Measure Influence the Evaluation of Hedge Funds? *Journal of Banking and Finance*, 31(9): 2632-2647.

[47] Estrada J. (2006). Downside Risk in Practice. *Journal of Applied Corporate Finance,* 18(1): 117-125.

[48] Fama F.E. and French R.K. (1988). Dividend Yields and Expected Stock Returns. *Journal of Financial Economics,* 22(1): 3–25.

[49] Fama F.E. and French R.K. (1989). Business Conditions and Expected Returns on Stocks and Bonds. *Journal of Financial Economics*, 25(1): 23-49.

[50] Fama F.E. and French, R.K. (1992). The Cross-Section of Expected Stock Returns. *Journal of Finance,* 47(2): 427-465.

[51] Fama F.E. and French R.K. (1993). Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics,* 33(1): 3-56.

[52] Fama F.E. and French R.K. (1995). Size and Book-to-Market Factors in Earnings and Returns. *Journal of Finance,* 50(1): 131-155.

[53] Fama F.E. and French R.K. (1996). Multifactor Explanations of Asset Pricing Anomalies. *Journal of Finance,* 51(1): 55-84.

[54] Fama F.E. and French R.K. (2012). Size, Value, and Momentum in International Stock Returns. *Journal of Financial Economics,* 105(3): 457-472.

[55] Fama F.E. and Schwert G. (1977). Asset Returns and Inflation. *Journal of Financial Economics,* 5(2): 115–146.

[56] Ferreira M.I. and Santa-Clara P. (2011). Forecasting Stock Market Returns: The Sum of the Parts is More than the Whole. *Journal of Financial Economics*, 100(3): 514-537.

[57] Fleming J., Kirby C. and Ostdiek B. (2001). The Economic Value of Volatility Timing Using Realized Volatility. *Journal of Financial Economics*, 67(3): 473-509.

[58] Foster D.P. and Hart S. (2009). An Operational Measure of Riskiness. *Journal of Political Economy*, 117(5): 785-814.

[59] Goetzmann W., Ingersoll J., Spiegel M. and Welch I. (2007). Portfolio Performance Manipulation and Manipulation-proof Performance Measures. *Review of Financial Studies*, 20(5): 1503-1546.

[60] Goh J., Jiang F., Tu J. and Zhou G. (2013). Forecasting Bond Risk Premia Using Technical Indicators. *Available http://ssrn.com/abstract=1914227.*

[61] Goyal A. and Welch I. (2008). A Comprehensive Look at the Empirical Performance of Equity Premium Prediction. *Review of Financial Studies,* 21(4): 1455-1508.

[62] Gregoriou G.N. and Gueyie J.P. (2003). Risk-Adjusted Performance of Funds of Hedge Funds Using a Modified Sharpe Ratio. *Journal of Wealth Management*, 6(Winter): 77-83.

[63] Gulen H., Xing Y. and Zhang L. (2011). Value Versus Growth: Time-Varying Expected Stock Returns. *Financial Management,* 40(2): 381-407.

[64] Guo H., Savickas R., Wang Z. and Yang J. (2009). Is the Value Premium a Proxy for Time-Varying Investment Opportunities: Some Time-Series Evidence. *Journal of Financial and Quantitative Analysis,* 44(1): 133–154.

[65] Hahn J. and Lee H. (2006). Yield Spreads as Alternative Factors for Size and Book-to-Market. *Journal of Financial and Quantitative Analysis,* 41(2): 245-269.

[66] Harlow W. Van (1991). Asset Allocation in a Downside Risk Framework. *Financial Analysis Journal,* 47(5): 28-40.

[67] Hart S. (2009). A Simple Riskiness Order Leading to the Aumann-Serrano Index of Riskiness. *Discussion Paper Series dp517, The Center for the Study of Rationality, Hebrew University, Jerusalem.*

[68] Harvey I.D., Leybourne J.S. and Newbold P. (1998). Tests for Forecast Encompassing. *Journal of Business and Economic Statistics,* 16(2): 254-259.

[69] Hodrick R. (1992). Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement. *Review of Financial Studies,* 5(3): 357–386.

[70] Homm U. and Pigorsch C. (2012). Beyond the Sharpe Ratio: An Application of the Aumann-Serrano Index to Performance Measurement. *Journal of Banking & Finance*, 36(8): 2274-2284.

[71] In F. and Kim S. (2007). A Note on the Relationship Between Fama-French Risk Factors and Innovations of ICAPM State Variables. *Finance Research Letters*, 4(3): 165-171.

[72] Jegadeesh N. and Titman S. (1993). Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. *Journal of Finance,* 48(1): 65-91.

[73] Jegadeesh N. and Titman S. (2001). Profitability of Momentum Strategies: An Evaluation of Alternative Explanations. *Journal of Finance*, 56(2): 699-720.

[74] Jensen M. (1968). The Performance of Mutual Funds in the Period 1945-1968. *Journal of Finance*, 23(2): 389-416.

[75] Kaplan P.D. and Knowles J.A. (2004). Kappa: A Generalized Downside Risk-Adjusted Performance Measure. *Journal of Performance Measurement*, 8(3): 42-54.

[76] Keim D.B. and Stambaugh R.F. (1986). Predicting Returns in the Stock and Bond Markets. *Journal of Financial Economics,* 17(2): 357-390.

[77] Kilian L. (1999). Exchange Rates and Monetary Fundamentals: What Do We Learn From Long-Horizon Regressions? *Journal of Applied Econometrics*, 14(5): 491-510.

[78] Lakonishok J., Shleifer A. and Vishny W. (1994). Contrarian Investment, Extrapolation, and Risk. *Journal of Finance,* 49(5): 1541-1578.

[79] Liew J. and Vassalou M. (2000). Can Book-to-Market, Size and Momentum Be Risk Factors That Predict Economic Growth? *Journal of Financial Economics,* 57(2): 221-245.

[80] Lintner J. (1965). The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. *Review of Economics and Statistics,* 47(1): 13-37.

[81] Ludvigson S.C. and Ng S. (2007). The Empirical Risk-Return Relation: A factor Analysis Approach. *Journal of Financial Econometrics*, 83(1): 171-222.

[82] Ludvigson S.C. and Ng S. (2009). Macro Factors in Bond Risk Premia. *Review of Financial Studies*, 22(12): 5027-5067.

[83] Maio P. and Santa-Clara P. (2012). Multifactor Models and Their Consistency With the ICAPM. *Journal of Financial Economics,* 106(3): 586-613.

[84] Markowitz H. (1959). Portfolio Selection: Efficient Diversification of Investments. *Yale University Press, New Haven & London.*

[85] McCracken W.M. (2007). Asymptotics For Out of Sample Tests of Granger Causality. *Journal of Econometrics*, 140 (2): 719-752.

[86] Merton C.R. (1973). An Intertemporal Capital Asset Pricing Model. *Econometrica,* 41(5): 867-887.

[87] Michou M. (2009). Is the Value Spread a Good Predictor of Stock Returns? UK Evidence. *Journal of Business Finance & Accounting*, 36(7): 925-950.

[88] Neely C., Rapach E.D., Tu J. and Zhou G. (2013). Forecasting the Equity Risk Premium: The Role of Technical Indicators. *Management Science*, forthcoming.

[89] Neil R. (2001). A Portfolio Manager's View of Downside Risk. In the Book of Sortino Frank A. and Satchell S.E. 'Managing Downside Risk in Financial Markets'. *Butterworth-Heinemann , UK.*

[90] Newey K.W. and West D.K. (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica,* 55(3): 703-708.

[91] Ogryczak W. and Ruszczynski (1999). From Stochastic Dominance to Mean-Risk Models: Semideviation as Risk Measures. *European Journal of Operational Research*, 116(1): 33-50.

[92] Ornelas J., Junior F. and Fernandes J. (2012). Yes, the Choice of Performance Measure does Matter for Ranking of US Mutual Funds. *International Journal of Finance & Economics*, 17(1): 61-72.

[93] Petkova R. (2006). Do the Fama-French Factors Proxy for Innovations in Predictive Variables? *Journal of Finance*, 61(2): 558-612.

[94] Petkova R. and Zhang L. (2005). Is Value Riskier Than Growth? *Journal of Financial Economics,* 78(1): 187-202.

[95] Phingsten A., Wagner P. and Wolferink C. (2004). An Empirical Investigation of the Rank Correlation Between Different Risk Measures. *Journal of Risk*, 6(4): 55-74.

[96] Rapach E. D., Strauss K. J. and G. Zhou (2010). Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy. *Review of Financial Studies*, 23(2): 821-862.

[97] Rapach E.D. and Weber E.C. (2004). Financial Variables and The Simulated Out-of-Sample Forecastability of U.S. Output Growth Since 1985: An Encompassing Approach. *Economic Inquiry,* 42(4): 717-738.

[98] Rapach E.D. and Wohar E.M. (2006). In-Sample vs. Out-of-Sample Tests of Stock Return Predictability in the Context of Data Mining. *Journal of Empirical Finance,* 13(2): 231-247.

[99] Rapach E.D. and Zhou G. (2012). Forecasting Stock Returns. *Forthcoming in the Handbook of Economic Forecasting*, Volume 2.

[100] Rime D., Sarno L. and Sojli E. (2010). Exchange Rate Forecasting, Order Flow and Macroeconomic Information. *Journal of International Economics*, 80(1): 72-88.

[101] Schuhmacher F. and Eling M. (2011). Sufficient Conditions for Expected Utility to Imply Drawdown-Based Performance Rankinngs. *Journal of Banking and Finance*, 35(9): 2311-2318.

[102] Shadwick W.F. and Keating C. (2002). A Universal Performance Measure. *Journal of Performance Measurement*, 6(3): 59-84.

[103] Shalit H. (2013). Measuring Risk in Israeli Mutual Funds: Conditional Value-at-Risk vs. Aumann and Serrano Riskiness Index. *Working paper*.

[104] Sharpe W.F. (1964). Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk. *Journal of Finance*, 19(3): 425-442.

[105] Sharpe W.F. (1966). Mutual Fund Performance. *Journal of Business*, 39(1): 119-138.

[106] Sortino F. and Price L. (1994). Performance Measurement in a Downside Risk Framework. *The Journal of Investing*, 3(3): 59-65.

[107] Stock H. J. and Watson W.M. (2004). Combination Forecasts of Output Growth in a Seven-Country Data Set. *Journal of Forecasting*, 23(6): 405-430.

[108] Thorton D. L. and Valente G. (2012). Out of Sample Predictions of Bond Excess Returns and Forward Rates: An Asset Allocation Perspective. *Review of Financial Studies*, 25(10): 3141-3168.

[109] Treynor J.L. (1965). How to Rate Management of Investment Funds. *Harvard Business Review*, 43(1): 63-75.

[110] Treynor J.L. and Black F. (1973). How to Use Security Analysis to Improve Portfolio Selection. *Journal of Business*, 46(1): 66-86.

[111] Vassalou M. (2003). News Related to Future GDP Growth As A Risk Factor In Equity Returns. *Journal of Financial Economics,* 68(1): 47-73.

[112] Vivian A. and Wohar M.E. (2013). The Output Gap and Stock Returns: Do Cyclical Fluctuations Predict Portfolio Returns? *International Review of Financial Analysis,* 26: 40-50.

[113] West D.K. (1996). Asymptotic Inference About Predictive Ability. *Econometrica*, 64(5): 1067-1084.

[114] West D.K., Edison H.J. and Cho D. (1993). A Utility-Based Comparison of Some Models of Exchange Rate Volatility. *Journal of International Economics*, 35(1-2): 23-45.

[115] Yamai Y. and Yoshiba T. (2002a). On the Validity of Value-at-Risk: Comparative Analyses with Expected Shortfall. *Monetary and Economic Studies,* 20(1): 57-85.

[116] Yamai Y. and Yoshiba T. (2002b). Comparative Analysis of Expected Shortfall and Value-at-Risk: Their Estimation, Error, Decomposition, and Optimization. *Monetary and Economic Studies,* 20(1): 87-121.

[117] Yamai Y. and Yoshiba T. (2005). Value-at-Risk Versus Expected Shortfall: A Practical Perspective. *Journal of Banking and Finance,* 29(4): 997-1015.

[118] Zakamouline V. (2011). The Performance Measure You Choose Influences the Evaluation of Hedge Funds. *Journal of Performance Measurement*, 15(3): 48-64.

[119] Zangari P. (1996). Methodology for Portfolios that Include Options. *Riskmetrics Monitor, Available at: http://www.indiceperu.com/smartfolio/refer23.pdf, First Quarter*: 4-12.