Essays on the application of factor models in finance

Stéphane Matton

Université Paris Nanterre

200 av. de la République
92001 Nanterre Cedex
www.parisnanterre.fr

Membre de l'Université Paris Lumières

École Doctorale
Économie, Organisation, Société

Unité de recherche
EconomiX

Stéphane MATTON

Essais sur l'application des modèles à facteurs en finance

Thèse présentée et soutenue publiquement le 12 janvier 2024
en vue de l'obtention du doctorat de Sciences Economiques
de l'Université Paris Nanterre
sous la direction de Christophe BOUCHER (Université Paris Nanterre)

Jury :

Rapporteure :	Mme Sophie BEREAU	Professeur : Université de Lorraine
Rapporteur :	M. Yannick LE PEN	Maître de conférences : Université Paris Dauphine
Membre du jury :	M. Christophe BOUCHER	Professeur : Université Paris Nanterre
Membre du jury :	Mme Antonia LOPEZ-VILLAVICENCIO	Professeur : Université Paris Nanterre
Membre du jury :	M. Marc JOETS	Professeur : IESEG School of Management
Membre du jury :	M. Sessi TOKPAVI	Professeur : Université d'Orléans

Remerciements

Ces premiers mots sont destinés à mes directeurs de thèse, Professeurs Christophe Boucher et Sessi Tokpavi, dont la confiance, l'attention et la bienveillance à mon égard sont restées indéfectibles. Je souhaite exprimer ma sincère reconnaissance à Christophe qui m'a donné la chance de m'épanouir dans ma vie académique et professionnelle. Ses enseignements resteront des fondamentaux et ses qualités humaines un exemple. Je remercie aussi particulièrement Sessi dont l'expertise scientifique et les conseils ont été une aide précieuse dans l'avancement de ce travail.

J'adresse mes remerciements aux membres qui ont accepté de composer mon jury de thèse : Sophie Bereau et Yannick Le Pen qui ont donné leur accord pour être rapporteurs, et Antonia Lopez-Villavicencio et Marc Joëts pour leur implication en tant qu'examinateurs.

Je souhaite remercier les personnes qui ont contribué à l'avancement de ces travaux de recherche : Professeur Jean-Paul Laurent, docteur Michael Sestier et docteur Wassim Le Lann.

J'adresse également une chaleureuse pensée à tous les membres du laboratoire EconomiX, ainsi qu'à mes anciens collègues et maintenant amis, Patrick Kouontchou et Alexandre Jasinski.

Sur une note plus personnelle, les mots ne pourraient être assez forts pour exprimer ma gratitude envers mes parents, Corinne et Patrice, dont le soutien sans faille est une constante qui a forgé ma vie. Une pensée aussi pour mes beaux-parents, Estelle et Jean-Marie, mon frère, Nicolas, ma sœur, Marie, et tous les autres membres de ma famille.

A ma compagne, Inès, je lui transmets toutes mes plus tendres pensées. Ses mots savent effacer les doutes et sa présence apaiser les inquiétudes. Merci de m'avoir encouragé et soutenu sans relâche ces dernières années.

A mes amis, Benjamin, Hugo, Jules, Nicolas, Stephen, Thibaud et Thibault, malgré la distance et le temps, votre amitié est restée inébranlable. C'est une force de vous savoir à mes côtés.

Enfin, je souhaite dédier cette thèse à mes grands-pères, Georges et Jean-Claude, qui nous ont malheureusement quittés durant sa réalisation.

Abstract

Factor models, fundamental in statistical modeling, provide a versatile framework for understanding intricate relationships among variables, particularly valuable in finance where variables exhibit strong interconnections. This thesis explores the applications of factor models in finance. Chapter 1 relies on factor models for their ability to detect market inefficiencies through estimate of idiosyncratic volatility and explores the effectiveness of ESG ratings in predicting it. It reveals a negative relationship between ESG ratings and volatility. Results emphasize the need to cross-check information from multiple providers before ESG integration as divergence leads to lower informative power of ESG ratings in predicting risk. Chapter 2 delves into calibration techniques for correlation matrices. Unrestricted and restricted factor models, where some betas are set to zero, thus leading to specific region or sectoral factors, are considered. Results highlight the superiority of restricted latent factor models. We address latent factors using maximum likelihood through the Expectation-Maximization algorithm and least squares approaches, implemented with the Spectral Projection Gradient algorithm. Chapter 3 introduces an innovative approach based on meta-factors to perform stress tests on portfolios, using either historical or hypothetical economic scenarios. An application is conducted on actively managed equity portfolios. Results offer a better fit than widely used models. The research contributes to the understanding of factor model applications in finance, providing empirical insights, theoretical developments, and practical methodologies. The findings have implications for investors and researchers, guiding ESG integration, correlation estimation, and stress testing in financial decision-making. The thesis shows that factor models remain pivotal tools in addressing key challenges in finance.

ABSTRACT

Les modèles à facteurs, fondamentaux en modélisation statistique, offrent un cadre polyvalent pour comprendre les relations complexes entre plusieurs variables, caractéristique particulièrement précieuse en finance où les interconnexions sont fortes. Cette thèse présente différentes applications des modèles à facteurs en finance. Le chapitre 1 s'appuie sur les modèles à facteurs pour leur capacité à capter les inefficiences de marché, notamment au travers de l'estimation de la volatilité idiosyncratique. Le contenu informationnel des notations ESG dans la prédiction de cette volatilité est testé. Les résultats soulignent la nécessité de vérifier la qualité de la notation auprès de plusieurs fournisseurs de données avant d'intégrer l'ESG au processus d'investissement. En effet, la divergence de notation ESG pour un même titre est associé à un contenu informationnel plus faible dans la prédiction du risque. Le chapitre 2 se penche sur les techniques de calibration des matrices de corrélations générées par des modèles à facteurs. Les modèles à facteurs non restreints et restreints, où certains bêtas (sensibilités) sont fixés à zéro, et conduisant ainsi à des facteurs régionaux ou sectoriels spécifiques, sont considérés. Les résultats mettent en évidence la supériorité des modèles à facteurs latents restreints. Nous abordons les facteurs latents en utilisant la méthode du maximum de vraisemblance à travers l'algorithme Expectation-Maximization et l'approche des moindres carrés au travers de l'algorithme Spectral Projection Gradient. Le chapitre 3 présente une approche novatrice basée sur des "méta-facteurs" (facteurs de facteurs) pour réaliser des stress tests de scénarios économiques passés ou hypothétiques. Une application est réalisée sur des portefeuilles d'actions. Les résultats révèlent un comportement des actifs stressés plus proche de la réalité que certains modèles, largement utilisés dans l'industrie, peuvent suggérer. Cette thèse contribue au développement des modèles à facteurs en finance en apportant des éléments empiriques, des développements théoriques et des méthodologies pratiques. Les résultats ont des implications pour les investisseurs et les chercheurs, guidant l'intégration ESG, l'estimation des corrélations et la conduite de stress tests dans la prise de décision financière. La thèse démontre que les modèles à facteurs restent encore aujourd'hui des outils essentiels pour l'industrie financière.

Table of Contents

List of Figures

List of Tables

General Introduction

Factor models in finance

Factor models are a fundamental concept in statistical modeling. They provide a framework for understanding the complex relationships among variables. These models are particularly useful in finance where variables are strongly interrelated.

One of the key applications of factor models in finance is the extensively use in asset pricing. The Capital Asset Pricing Model – CAPM of William Sharpe (1964) and John Litner (1965) is considered as a simple form of factor model, where the returns of an asset are explained by a market factor (returns of a representative portfolio of stocks, called the market portfolio). The model witnesses the emergence of factor models in finance and remains widely used nowadays. The power of the CAPM, and factor models in general, lies in their simplicity, robustness, and ability to predict relationships (between risk and expected returns in the case of the CAPM). Extensions of this model include multi-factor models, such as the Fama-French three-factor model (Fama, 1993), which introduces additional factors like size (market capitalizations) and value (market valuations) to better capture the complexities of stock returns.

In econometrics, factor models are employed to discern underlying economic trends. For instance, a macroeconomic factor model may include variables like inflation, interest rates, and GDP growth to explain the fluctuations observed in various economic indicators. Beyond finance and economics, factor models are also widely used in other fields that exploit large and complex datasets. In psychology, for example, a factor model might be applied to understand the underlying factors influencing scores on various psychological tests. Besides, we will rely on the psychometric literature in the second chapter of this thesis.

We distinguish three major types of factor models in finance (Connor, 1995): macroeconomic, statistical, and fundamental. Macroeconomic factor models use observable economic time series as measures of pervasive factors influencing a financial variable with the assumption that this variable is assumed to respond linearly to macroeconomic or financial shocks. Statistical factor models use several statistical techniques to estimate the parameters of factor models. Principal Component Analysis – PCA (Jolliffe, 2002; Stock et al., 2002) is a common method used to identify the underlying factors by capturing the directions of maximum variance in the data. Maximum Likelihood Estimation (MLE) and Bayesian methods are also employed (Watson and Engle, 1983). Finally, fundamental factor models rely on financial assets' attributes to explain their dynamics (for instance, dividend yield, book-to-market ratio, and industry classification can explain a substantial proportion of common returns between several stocks or bonds returns).

Beyond asset pricing, factor models have numerous applications in finance. In portfolio management, factor models help to optimize asset allocations by finding sources of diversification (Meucci, 2009). In risk management, these models are crucial to quantify and manage various sources of risk. For instance, they help institutions understand the impact of different factors on portfolio value-at-risk (VaR) and expected shortfall (Jorion, 2007). In credit risk modelling and fixed income analysis, factor models can be used to assess the creditworthiness of borrowers. They help in identifying common factors that influence the credit risk of individual loans (or bonds) and can be used in modeling interest rate risk (Altman, 1968; Litterman and Scheinkman, 1991). Factor models find many other applications, for instance, in option pricing (Black and Scholes, 1973) or in performance attribution (Brinson, 1986). These applications demonstrate the versatility of factor models in addressing key challenges and providing valuable insights for the research in finance.

Factor models offer several advantages that make them still widely used. Dimension reduction, where factor models excel, is maybe the most important advantage. Indeed, factor models help in reducing the dimensionality of data by capturing commonalities (Fan and Liao, 2014). Reducing the dimensionality also leads to improving the interpretability of latent structures that drive the observed patterns or confirm the influence of observable factors. This feature makes factor models still very useful today despite the emergence of machine learning

based models. Indeed, the identified factors often have economic or theoretical meanings, making it easier to understand and explain, while some machine learning models, especially complex ones like deep neural networks, are often considered "black boxes". Factor models also tend to be more data-efficient, even when dealing with high-dimensional data (Bai and Li, 2012). They can effectively capture and summarize the essential information using a smaller set of factors while machine learning models may require massive amounts of data to perform well thus being costly for financial institutions.

Factor models come with several limitations, however. Most of these models assume linear relationships between observed variables and factors. This assumption might not hold in all cases, especially when dealing with complex, non-linear relationships in the data. The lack of robustness of some factor models is also a limitation. The performance of factor models can be sensitive to the choice of factors and model specifications. Selecting an inappropriate number of factors or using an inadequate estimation method can lead to biased results and misinterpretations. Finally, two strong limitations should be considered when dealing with factor models: the assumption that variables follow a Gaussian distribution and the assumption of homoscedasticity. Some factor models assume that the underlying factors and idiosyncratic components follow a Gaussian distribution. This assumption may not hold in all cases, particularly when dealing with financial data, where extreme events or outliers can significantly impact the results (Murray, 2013). Factor models also often assume homoscedasticity, meaning that the variance of the idiosyncratic components is constant across all observations. In real-world scenarios, this assumption may not hold, particularly during periods of high instability and ultimately leads to the inability to capture structural breaks.

In this thesis, we will use factor models to estimate: (i) idiosyncratic volatility, (ii) correlations, and (iii) market stress.

Estimating idiosyncratic volatility through factor models

Understanding and quantifying idiosyncratic volatility is crucial in various fields, from finance to economics. Idiosyncratic volatility refers to the portion of an asset's total volatility that is specific to that asset and not explained by common factors. Factor models offer a robust framework for estimating and decomposing this idiosyncratic volatility, providing insights into the unique risk associated with individual assets. In the first chapter of this thesis, we aim to measure the informative content of a variable in explaining the idiosyncratic volatility of stocks. Before doing so, we need a framework to quantify the idiosyncratic volatility. Two widely employed factor models for this purpose are the CAPM and the Fama-French factor model mentioned above. According to the joint hypothesis (Fama, 1991), if markets are efficient, the systematic risk should be the only risk for which investors require compensation. Hence, the presence of idiosyncratic volatility could indicate the existence of factors not captured by the model.

To estimate idiosyncratic volatility using factor models, we typically follow a two-step process. First, we estimate the factor loadings using historical return data and relevant factors (such as market returns, size, and value factors). Second, the estimated factor loadings are used to compute the systematic risk component, which is then subtracted from the total observed volatility to obtain the idiosyncratic volatility.

One key advantage of using factor models for estimating idiosyncratic volatility is their ability to capture the dynamic nature of relationships among assets. Traditional methods often struggle with changing market conditions, but factor models can adapt by incorporating new factors or modifying existing ones.

Several studies have explored the effectiveness of factor models in estimating idiosyncratic volatility. The research by Ang, Hodrick, Xing, and Zhang (2006) provides insights into the application of the Fama-French three-factor model in estimating idiosyncratic volatility in the stock market. They find that incorporating size and value factors significantly improves the model's explanatory power, enhancing the accuracy of idiosyncratic volatility estimates.

Moreover, the work of Bali and Cakici (2008) delves into the role of macroeconomic factors in explaining idiosyncratic volatility. They extend traditional factor models by including macroeconomic variables, demonstrating that these factors contribute significantly to the estimation of idiosyncratic volatility.

Factor models provide a versatile and powerful framework for estimating idiosyncratic volatility, allowing researchers and practitioners to disentangle the unique risk associated with individual assets from the broader market movements.

Estimating correlations through factor models

Factor models can be specifically designed to uncover underlying structures in multivariate datasets. These models acknowledge that observed correlations between variables can be influenced by common factors, offering a more nuanced perspective on the interconnections within the data.

The general form of a factor model for correlation matrices can be expressed as:

$$\Omega = \beta\beta' + \mathcal{E}$$

where Ω is the observed correlation matrix, β represents the matrix of common factors and $\mathcal{E}$ is a diagonal matrix capturing the idiosyncratic components.

Not only factor models are convenient to estimate correlation matrices, but they are also recommended by the Basel Committee on Banking Supervision (BCBS) to be used by financial institutions. Indeed, banks must deal with a large amount of very complex data. In dealing with large cross-sectional dimensions, factor models offer a parsimonious approach[1], especially relevant for managing corporate counterparties.

[1] When using raw pairwise correlations in modelling for n assets, we need to estimate: $\frac{n(n-1)}{2}$ correlations. It would only be $2n$ factor loadings if using 2-factor modelling. Hence, for a banking book of 500 names, a financial institution would need to estimate only 1000 parameters instead of 124 750 correlations.

But the Basel 3 requirements involve using a correlation matrix associated with a probabilistic factor representation, with at least "two types of systematic risk factors". It is stated that "default correlations must be based on credit spreads or on listed equity prices". While this points to latent factor models that are being dealt with by central bankers and academics, linear regression onto observable factors, typically portfolio returns, are often used to build correlation matrices compliant with regulatory requirements.

Ultimately, banks use the factor loadings (of latent factors) generating the closet correlation matrix to the historical correlation matrix to model the creditworthiness of individual names in their portfolios. Based on this modelling, banks capitalize a default risk charge. These requirements lead to practical implications of the two-factor models and correlation calibration constraints (Laurent et al., 2016).

Estimating market stress through factor models

Stress testing is a critical tool in assessing the robustness of financial systems and institutions under adverse conditions. Factor models play a key role in stress testing by providing a structured framework to evaluate how various factors impact the performance and stability of financial assets. These models help simulate extreme scenarios, enabling financial institutions to gauge potential vulnerabilities and develop strategies to mitigate risks.

In stress testing, a factor model typically incorporates a set of macroeconomic, market, and financial variables that are likely to be sources of stress. The model aims to understand how changes in these factors can affect the overall portfolio and individual assets. The factors can include variables such as interest rates, economic growth rates, exchange rates, and other relevant indicators.

Identifying the key factors that are relevant to the assets under consideration is of foremost importance. These factors should capture the major sources of risk. The factor model is then used to assess the impact of the stress scenarios on the portfolio. This assessment includes evaluating changes in asset prices, market liquidity, credit risk, or other relevant metrics.

Regulatory bodies mandate banks to undergo stress tests to ensure their ability to withstand adverse economic conditions (as mentioned above). The insights gained from stress testing with factor models help financial institutions develop effective risk mitigation strategies (Jokivuolle, 2015). Stress-testing through factor models can also help to adjust portfolio holdings according to potential market shocks.

Thesis outline

In this thesis, organized in three chapters, we exploit factor models to answer current issues in the financial industry. We touch upon various fields in the financial research: ESG, risk management, portfolio management and banking regulation. The three research articles outlined thereafter introduce innovative methods that aim to resolve the following problematics:

(i) Could we use ESG metrics to forecast idiosyncratic risk?

(ii) What are the most efficient calibration techniques to derive correlation matrices associated with factor models?

(iii) Could we propose a simple but appropriate framework to perform stress tests on equity portfolios?

In the Chapter 1, we rely on factor models for their ability to detect market inefficiencies through estimate of idiosyncratic volatility and explores the effectiveness of ESG ratings in predicting it. More precisely, we formulate a backtesting procedure to assess the effectiveness of ESG ratings in predicting a company's idiosyncratic risk. This involves extending the conditional predictive ability test proposed by Giacomini and White (2006) to a panel data context. Our approach is applied to forecast stock returns' idiosyncratic volatility, comparing two ESG rating systems — Sustainalytics and Asset4 — across three investment regions (Europe, North America, and the Asia-Pacific region).

While this first chapter provides an empirical methodology, Chapter 2 lies at the intersection of theoretical developments and empirical applications. It presents an overview of diverse specification and calibration techniques for deriving correlation matrices in factor models. We focus on both unrestricted and restricted models, where specific betas (factor loadings) are set to zero, leading to region or sectoral factors. We address latent factors through maximum likelihood using the Expectation-Maximization (EM) algorithm and least squares approaches with the Spectral Projection Gradient algorithm (SPG). Using a dataset of equity returns from major US and European credit indices, we conduct a benchmarking exercise by varying the number of factors. We evaluate fitting performance, parameter parsimony, ease of interpretation, and the ability to handle large pairwise correlations. Finally, Chapter 3 provides a robust methodology for practitioners. It introduces an innovative approach using "meta-factors" for conducting stress tests on portfolios through either historical or hypothetical economic scenarios. Stress testing in finance assesses a portfolio's resilience to adverse conditions, serving as a valuable supplement to conventional risk measures like volatility, VaR, and expected shortfall. Unlike summary statistics, stress testing provides estimates linked to specific events, adding meaningful insights. While traditional stress tests often rely on historical scenarios by reproducing past events on current holdings, they lack a forward-looking perspective. Our approach, relies in few parameters, maintains high flexibility, is easy to implement, and enhances the understanding of portfolio sensitivities to systematic factors. We illustrate the application with actively managed equity portfolios.

Chapter 1 "Are ESG ratings informative to forecast idiosyncratic risk?" [2] shows a negative relationship between ESG ratings and idiosyncratic risk, with higher ratings predicting lower levels of idiosyncratic volatility. Furthermore, the predictive accuracy gains are generally higher when assessing the environmental dimension of the ratings. Importantly, applying the test only to firms over which there is a high degree of consensus between the ESG rating agencies leads to higher predictive accuracy gains for all three universes. Beyond providing insights into the accuracy of each of the ESG rating systems, this last result suggests that information gathered from several ESG rating providers should be cross-checked before ESG is integrated into investment processes.

[2] This article is under revision in Finance.

The results have important implications for investors and researchers. For investors, our backtest procedure provides a useful and practical framework for considering ESG rating providers before integrating the ratings into the investment process. Our results suggest prudence about the information content of ESG ratings when they diverge. For researchers in asset pricing, it is crucial to check properly the quality of ESG ratings before using them, especially when the ratings are divergent. Moreover, the link between ESG ratings and idiosyncratic volatility when the ratings are convergent suggests that ESG investing is not just an issue of the preferences of investors, but that ESG ratings can also provide information about future fundamentals and risks. Chapter 2 "Correlation matrices with factor structure for credit risk exposures"[3] finds that besides the fitting performance, parameter parsimony, ease of interpretation and ability to deal with large pairwise correlations, restricted latent factor models supersede competitors in estimating correlation matrices. We find that when dealing with the naïve PCA approach, the final transformation to recover correlation matrix must be chosen carefully and restricted versions of optimization-based approaches behave better when dealing with intra-classes. It is well known that factor-model tend to smooth the correlation matrix, leading to large error for high correlations. Numerical experiments indicate that restricted calibration performed better in that aspect too. By restricting the information onto a class of issuers that should (a priori) behaves the same way, the restricted version of SPG algorithm allows a smaller error on these high empirical correlations than other methods. Chapter 3 "Stress-testing: A meta-factor approach"[4] shows that meta-factors help to translate an economic scenario into market shocks as they exert a significant influence on multiple individual factors within the financial system (style or systematic factors in our framework). Despite its simplicity, our methodology offers a better fit than other widely use models from well-known financial data providers. Beyond the risk management aspects of stress-testing, our framework offers the possibility to quickly implement forward or hypothetical stress scenarios helping managers to adjust their active portfolios in consequences.

[3] This article is under further work before submission.
[4] The methodology detailed in this article is implemented as an investment tool for a European asset manager.

1 Are ESG ratings informative to forecast idiosyncratic risk[5]?

We develop a backtesting procedure that evaluates how well ESG ratings help in predicting a company's idiosyncratic risk. Technically, the inference is based on extending the conditional predictive ability test of Giacomini and White (2006) to a panel data setting. We apply our methodology to the forecasting of stock returns idiosyncratic volatility and compare two ESG rating systems from Sustainalytics and Asset4 across three investment universes (Europe, North America, and the Asia-Pacific region).

The results show that the null hypothesis of no informational content in ESG ratings is strongly rejected for firms located in Europe, whereas results appear mixed in the other regions. In most configurations, we find a negative relationship between ESG ratings and idiosyncratic risk, with higher ratings predicting lower levels of idiosyncratic volatility. Furthermore, the predictive accuracy gains are generally higher when assessing the environmental dimension of the ratings. Importantly, applying the test only to firms over which there is a high degree of consensus between the ESG rating agencies leads to higher predictive accuracy gains for all three universes.

Beyond providing insights into the accuracy of each of the ESG rating systems, this last result suggests that information gathered from several ESG rating providers should be cross-checked before ESG is integrated into investment processes.

[5] Co-authors: Christophe Boucher, EconomiX, CNRS, University Paris Nanterre ; Wassim Le Lann, University of Orléans, LEO, Orléans, France ; Sessi Tokpavi, University of Orléans, LEO, Orléans, France.

1.1 Introduction

Sustainable investing is growing fast, and investors are increasingly integrating environmental, social, and governance (ESG) criteria. According to Bloomberg, ESG assets are on track to exceed $53 trillion by 2025, representing more than a third of projected total assets under management in North America, Europe, and Asia-Pacific capital markets[6]. To integrate extra-financial information into investment processes, investors typically rely on ESG ratings provided by rating agencies, which are scores designed to capture companies' ESG performance. However, ESG ratings are derived using heterogeneous methodologies and can be quite divergent across rating agencies (Berg et al., 2022b; Dimson et al., 2020). This issue has raised important concerns regarding the relevance of ESG ratings for guiding investment decisions[7]. Is there any informational content in the various existing ESG rating systems? Is this informational content related to companies' risk exposure? This article aims to provide a statistical methodology to answer these questions by developing a backtesting procedure allowing to assess the informational content of ESG ratings in forecasting a company's risk-related outcome. Our test evaluates the effectiveness of extra-financial metrics in predicting a company's risk exposure beyond the information conveyed by traditional financial variables.

The global craze for responsible investment has by now led to an abundant and rich literature that has tried, with mixed results, to evaluate how sustainable investment impacts market variables, and asset prices. Some studies have found that ESG has a positive impact on asset prices (Mozaffar et al., 2016; Amel-Zadeh and Serafeim, 2018; Dycket al., 2019; Hartzmark and Sussman, 2019), for instance present evidence that firms doing well on ESG issues outperform firms doing poorly on these issues. Amel-Zadeh and Serafeim (2018) reaffirm that ESG ratings have a material impact on asset prices and more specifically on the cost of capital, as investors expect higher return on equity for companies with strong ESG performance.

[6]https://www.bloomberg.com/professional/blog/esg-assets-may-hit-53-trillion-by-2025-a-third-of-global-aum

[7] See for example recent SEC discussions on ESG ratings: https://www.sec.gov/news/speech/speech-peirce-061819.

Other contributions highlight that socially responsible investors can substantially reduce the cost of capital of responsible companies by tilting their portfolio allocation towards these firms Gollier and Pouget, 2022; Zerbib, 2022. Dyck et al. (2019) also demonstrate that engagement by investors has a positive impact on ESG performance and ultimately on financial returns, especially in countries where ESG issues are important. A study of US mutual funds flows confirms that investors find value in sustainability as a positive predictor of future returns (Hartzmark and Sussman, 2019).

Arguing the other side though are some works (Riedl and Smeets, 2017; Pastor et al., 2021; Pedersen et al., 2020) based on the impact of investor preferences on the dynamics of asset prices (Fama and French, 2007), which report that ESG practices have either a negative or a positive impact on asset prices. Considering investor preferences for ESG, Riedl and Smeets (2017) notice that investors are willing to accept lower expected returns and higher management fees for holding companies with strong ESG performance. Pastor et al. (2021) model investor preferences for ESG in a mean-variance framework and show that in equilibrium, assets considered green generally have lower expected returns but provide greater utility and offer the ability to hedge against climate risk. They also introduce an ESG-factor that reacts to unexpected change in ESG, then conclude that green assets outperform when a positive shock hits this factor. Pedersen et al. (2020) extend the mean-variance-ESG framework by adding a third type of investor who is unaware of the ESG performance of firms. How the ESG ratings affect expected returns then depends on the wealth of this third investor.

Although this literature provides useful information on the link between extra-financial performance and asset price dynamics, it does not provide a formal methodology to assess whether the available rating systems are effective in measuring a company's exposure to financially material sustainability risks. This gap in the literature is even more worrying as the correlations between the ratings of the various available providers are weak.

Indeed, the divergence of ESG ratings has been widely documented (Chatterji et al., 2009; Semenova and Hassel, 2015; Chatterji et al., 2016; Berg et al., 2022b; Dimson et al., 2020), and Berg et al. (2022b) find, for instance, that correlations between the ESG ratings of providers are on 54% for a set of six different ESG providers, whereas the credit ratings from the main agencies exhibit, on average, a correlation of 99%. They further explore the source of this divergence by splitting it into three components and looking at scope, or the selection of ESG categories to be measured; measurement, or how the ESG categories are assessed; and weight, or the importance given to each category. They observed that measurement explains more than 50% of the total divergence[8]. The divergence of ESG rating systems has important implications for sustainable investing. ESG ratings disagreement can lead to completely opposite opinions on one and the same company, dispersing ESG preferences of investors (Billio et al., 2019). It also makes it difficult to empirically assess the impact of ESG performance on stock returns (Berg et al., 2022a) and can result in risk premiums for companies with high rating disagreement (Gibson Brandon et al., 2021).

Against this background, our paper introduces a statistical inferential procedure that allows testing the informational content of a given ESG rating system in forecasting a company's risk-related outcomes. Our backtest is based on a risk management approach, if informative ESG ratings should have significant power in predicting a company's specific risk exposure beyond the information conveyed by financial variables. Our procedure evaluates ESG ratings by comparing the forecasting abilities of two nested models that differ solely in the inclusion or exclusion of ESG ratings within the set of predictor variables. In this setting, our null hypothesis of a lack of informational content in ESG ratings is defined as the equality in forecast accuracy between the two nested models, implying that integrating ESG ratings among predictor variables does not help to forecast the specified target variable. Technically, our inferential procedure extends the conditional predictive ability test of Giacomini and White (2006) to a panel setting. We derive the Gaussian asymptotic distribution of the test statistic under weak assumptions.

[8] Unlike credit ratings, ESG ratings are most often created mainly from non-standardized information and are not regulated. Methodologies can be opaque and proprietary, leading to substantial rating divergence.

A notable advantage of our procedure is the robustness of the distribution of our test statistic under the null hypothesis to possible misspecifications of the econometric model used to measure the relationship between ESG ratings and the outcome variables. Monte Carlo simulations, performed under different types of model misspecification, demonstrate that our test has good small sample properties, with a good size and increasing power as the number of firms and sample length increase.

Previous research on the relationship between ESG ratings and firm-level risk outcomes has focused on different types of outcome variables. A first strand of the literature has concentrated on examining the link between ESG ratings and ESG controversies captured through news media screening. This type of outcome variable has been typically used as a proxy for corporate misbehavior or as an ex-post indicator for ESG risk materialization. Research findings in this area have been marked with mixed results. Champagne et al. (2022) constructed a controversy variable using articles published in the *Wall Street Journal* and found that ESG performance, as measured by MSCI KLD ratings, is negatively related to the likelihood of ESG controversies. On the other hand, Yang (2022) found that MSCI ratings do not help forecast future ESG controversies derived from the RepRisk database. Additionally, Bams and van der Kroft (2022) used Asset4 (Refinitiv) ESG controversies to build a measure of *realized sustainable performance* and found that ESG scores from leading rating agencies are often inversely related to their measure of sustainability performance.

Other studies have investigated the link between ESG ratings and market-based measures of firm risk, such as the idiosyncratic volatility of stock returns (IVOL) (Boutin-Dufresne and Savaria, 2004; Lee and Fa, 2009; Jo and Na, 2012; Mishra and Modi, 2013; Bouslah et al., 2013). IVOL has often been interpreted as a proxy capturing the stock market reaction to the arrival of company-specific information (Becchetti et al., 2015), or as a proxy for innovations in a missing risk factor (Chen and Petkova, 2012). Under both interpretations, ESG ratings might be linked to IVOL: Serafeim and Yoon (2022a) find that ESG ratings moderate the stock market reaction to the arrival of firm specific ESG news, while Pastor et al. (2022) show that asset prices react to global ESG news.

Multiple studies have indeed reported a link between ESG ratings and IVOL. Jo and Na (2012) find a negative link between ESG ratings and idiosyncratic risk. Mishra and Modi (2013) find that this negative relation is moderated by financial leverage, while Bouslah et al. (2013) find that different components of social performance have different effects on idiosyncratic risk. Other studies have also investigated the relationship between ESG ratings and other market-based outcomes, such as firm beta (Sodjahnin et al., 2017; Albuquerque et al., 2019), or downside risk (Diemont et al., 2016).

In this chapter, we apply our procedure to evaluate ESG ratings for forecasting a company's specific risk, as measured by the idiosyncratic volatility of its stock price. While, in practice, our testing procedure can be applied to any target variable, we choose a market-based measure of firm risk due to the numerous limitations associated with forecasting ESG controversies. First, ESG controversies are typically constructed using natural language processing methods on media sources to identify corporate adverse events. Variations in the media sources used to compute controversies and differences in the extraction methods employed can lead to significant variations among existing controversy variables, which has important implications for results replicability[9]. Second, ESG controversies built on news media screening suffer from substantial reporting bias. Barkemeyer et al. (2023) find that the location of a company's headquarters significantly influences its media coverage, with firms based in the United Kingdom and the United States being five times more likely to receive press coverage as controversies. This issue implies that ESG controversies are an inadequate proxy for corporate misbehavior. Third, ESG controversies are also a poor proxy for ESG risks, or financial risks related to ESG events, as they only capture firm-specific events. Recent findings indeed show that global events leading to a reevaluation of ESG concerns can have a significant impact on asset prices (Ardia et al., 2022; Pastor et al., 2022). These types of global events, such as a sudden risk of an increase in carbon prices that may significantly affect the valuation of carbon-intensive companies, cannot be captured through ESG controversies.

[9] For example, we find that the rank correlations between ESG incidents from *Sustainalytics* and *Asset4* in our datasets are weak: 43% for Europe, 43% for North America, and 34% for the Asia-Pacific region.

A second rationale for forecasting a market-based measure of financial risks rather than ESG controversies is related to what ESG ratings aim to capture and the reason why investors use ESG ratings in the first place. Evidence from investor surveys shows that investors primarily use ESG ratings for motives related to investment performance (Amel-Zadeh and Serafeim, 2018). Consequently, as acknowledged by the Global Reporting Initiative, most existing ESG rating systems are assessed from a financial materiality perspective, meaning that rating agencies seek to capture companies' financial exposure to ESG-related issues (GRI, 2022). On the other hand, while ESG incidents can impact stock prices, they are not necessarily followed by asset price movements (Serafeim and Yoon, 2022b). Hence, forecasting a market-based risk measure seems to be a more direct way to evaluate ESG ratings.

We conduct empirical applications to illustrate our methodology, using two leading ESG rating systems, Sustainalytics and Asset4, for Europe, North America, and the Asia-Pacific region. Our results show that the null hypothesis of no informational content in ESG ratings is strongly rejected for firms located in Europe, whereas results appear mixed in the other regions. Hence, companies headquarter location appears to be a critical moderator for the integration of ESG factors into asset prices. In most configurations, we find a negative relationship between ESG ratings and idiosyncratic risk, with higher ratings predicting lower levels of idiosyncratic volatility. Furthermore, we find that the forecasting power of ESG ratings is generally higher when assessing the environmental dimension of ESG ratings. Lastly, and importantly, we find that the predictive accuracy gains derived from ESG ratings increase with the level of consensus between rating agencies for all three universes. This final finding can be linked to that highlighted by Serafeim and Yoon (2022a), who find that the market reaction to ESG news is moderated by the consensus rating. From a practical standpoint, our results provide crucial information for portfolio managers who integrates ESG rating to assess companies' risk profile, as we show that it is necessary to cross-check the information gathered from multiple ESG rating providers before integrating ESG into the management process.

Our contribution is related to previous works investigating the link between ESG ratings and idiosyncratic risk but differs in several aspects. First, past studies do not provide a formal test to check the informational content of ESG ratings in forecasting firm risks, which is the purpose of this article. In contrast to analyses based on an in-sample setting, we test the informational content of ESG ratings using a dynamic forward-looking approach in an out-of-sample environment. This configuration integrates the possibility that the estimated relationship between ESG ratings and firm risk might not be generalizable to a future period, while recent research suggests that the integration of ESG factors into asset prices varies over time (Ardia et al., 2022; Pastor et al., 2022). Second, our approach accounts for possible misspecification of the econometric model used to measure the relationship between ESG ratings and the outcome variable. This differs from the previous literature, where the correctness of the econometric model is critical to establishing the existence of this link. Third, while previous studies identify significant correlations between ESG ratings and firm risks, they fail to quantify the improvement in model fit resulting from incorporating extra-financial information. Our method compares the predictive ability of nested models containing financial and extra-financial information, allowing for such quantification.

The rest of the chapter is organized as follows. Section 1.2 describes our backtesting procedure for ESG ratings, focusing on the formulation of the null hypothesis, the construction of the test statistic and the analysis of its asymptotic distribution. Section 1.3 simulates the small sample properties of the test statistic under various settings, and empirical applications are considered in Section 1.4. The last section concludes the paper.

1.2 The backtesting procedure

This section gives a description of the backtesting procedure for evaluating statistically the informational content in ESG ratings. In the first part, we fix the notations and clearly define the null hypothesis of interest, while in the second part we provide the test statistic and its asymptotic distribution for inference.

1.2.1 Notations and the null hypothesis

To formulate the null hypothesis of our test, we consider an investment universe with n traded firms and let $y_{i,t}$ denote the value at month t of a target variable that is intended to measure firm-specific risks. For instance, a socially motivated investor seeking to manage the environmental and social impact of their asset portfolio can use a variable $y_{i,t}$ that measures ESG incidents, such as the ones provided by well-known providers (Sustainalytics, Asset4, TrueValue Labs, etc.), to test whether ESG ratings help predict future corporate misconduct. On the other hand, investors who are interested in the materiality of ESG information on investment performance can use a target variable that measures a firm's specific exposure to financial risks, such as idiosyncratic volatility. Therefore, our framework is general, as it enables users to choose a target variable relevant to their investment objectives.

Let $x_{i,t}$ be a vector of length p in which the elements are innovations on p financial variables that measure the financial strength of firm i for the month t. Examples of such variables are dividend yield, sales over assets, debt over assets, or the quick ratio. They measure different facets of a firm's solvency including its size, returns, risk, liquidity, debt, and leverage. Innovations can be obtained through autoregressive filtering on raw financial variables, or simply as deviations from the long-term average. Finally, the value of an ESG rating is available for each firm i at month t and we denote it as $\omega_{i,t} \in \mathbb{R}$. This can be a global ESG rating measuring environmental, social and governance issues, or only one of these three components.

Now let $m_{i,t+\tau}^{(0)} = \mathrm{E}\left(\mathrm{y}_{i,\mathrm{t}+\tau} \middle| x_{i,t}\right)$ be the unknown expected value of $\mathrm{y}_{i,\mathrm{t}}$ for firm i at time $t+\tau$, conditional on its financial strength as measured by innovations $x_{i,t}$ in financial variables, with τ as a given forecast horizon. We can use a given predictive model, whether parametric, semi-parametric or non-parametric, to forecast $m_{i,t+\tau}^{(0)}$. The forecast we denote $\widehat{m}_{i,t+\tau}^{(0)}(\hat{\beta}_{t+b_t}^{(0)})$ is based on the information set available at time t for all firms, so $\mathcal{F}_t^{(0)} = \{x_{i,s}, s = t - b_t + 1, \dots, t, i = 1, \dots, n\}$ where b_t refers to the size of the estimation sample and $\hat{\beta}_{t+b_t}^{(0)}$ collects all the estimated parameters.

In a parametric model like a linear regression, $\hat{\beta}^{(0)}_{t+b_t}$ is the vector of the estimates of the unknown parameters. Otherwise, it corresponds to whatever semi-parametric or non-parametric estimators are used to forecast $m^{(0)}_{i,t+\tau}$.

Let $m^{(1)}_{i,t+\tau} = \mathrm{E}\left(\mathrm{y}_{i,\mathrm{t+\tau}} \middle| x_{i,t}, \omega_{i,t}\right)$ be defined as $m^{(0)}_{i,t+\tau}$, but with the conditional set extended to $\omega_{i,t}$, so $\mathcal{F}^{(1)}_t = \{x_{i,s}, \omega_{i,s}, s = t - b_t + 1, \dots, t, i = 1, \dots, n\}$. In other words, $m^{(1)}_{i,t+\tau}$ is the expected value of $\mathrm{y}_{i,\mathrm{t+\tau}}$ for firm i at time $t + \tau$, conditional on its financial states as given by $x_{i,t}$ and on its ESG rating as given by $\omega_{i,s}$. We denote $\widehat{m}^{(1)}_{i,t+\tau}(\hat{\beta}^{(1)}_{t+b_t})$ as the forecast value at time $t + \tau$.

Suppose that we produce T_0 out-of-sample forecasts of both the expected values $m^{(0)}_{i,t+\tau}$ and $m^{(1)}_{i,t+\tau}$ for each firm, so $\widehat{m}^{(0)}_{i,t+\tau}(\hat{\beta}^{(0)}_{t+b_t})$ and $\widehat{m}^{(1)}_{i,t+\tau}(\hat{\beta}^{(1)}_{t+b_t})$, $i = 1, \dots n, t = 1, \dots, T_0$. With a loss function at hand that we denote $\mathcal{L}(.)$, we can evaluate the predictive performance of each model, generating two panels of losses as $\mathcal{L}^{(0)}_{i,t+\tau} \equiv \mathcal{L}^{(0)}_{i,t+\tau}(\mathrm{y}_{i,\mathrm{t+\tau}}, \widehat{m}^{(0)}_{i,t+\tau}(\hat{\beta}^{(0)}_{t+b_t}))$ and $\mathcal{L}^{(0)}_{i,t+\tau} \equiv \mathcal{L}^{(1)}_{i,t+\tau}(\mathrm{y}_{i,\mathrm{t+\tau}}, \widehat{m}^{(1)}_{i,t+\tau}(\hat{\beta}^{(1)}_{t+b_t}))$, where again $\mathrm{y}_{i,\mathrm{t+\tau}}$ is the value of $\mathrm{y}_{i,\mathrm{t}}$ for firm i at time $t + \tau$. From these panels, let $\Delta\mathcal{L}_{i,t+\tau} = \mathcal{L}^{(1)}_{i,t+\tau} - \mathcal{L}^{(0)}_{i,t+\tau}$ be the panel of loss differentials, $i = 1, \dots n$, $t = 1, \dots, T_0$, and $\mu_i(\hat{\beta}^{(0)}_{t+b_t}, \hat{\beta}^{(1)}_{t+b_t})$ the expected value of the loss differentials for firm i.

Hence, our null hypothesis of overall equal predictive ability of the two forecasting models can be stated as:

$$\mathcal{H}_0 = \bar{\mu}(\hat{\beta}^{(0)}_{t+b_t}, \hat{\beta}^{(1)}_{t+b_t}) = 0 \quad (1.1)$$

with the alternative hypothesis being:

$$\mathcal{H}_1 = \bar{\mu}(\hat{\beta}^{(0)}_{t+b_t}, \hat{\beta}^{(1)}_{t+b_t}) < 0 \quad (1.2)$$

Where $\bar{\mu}(\hat{\beta}_{t+b_t}^{(0)}, \hat{\beta}_{t+b_t}^{(1)})$ is defined as:

$$\bar{\mu}(\hat{\beta}_{t+b_t}^{(0)}, \hat{\beta}_{t+b_t}^{(1)}) = \frac{1}{n}\sum_{i=1}^{n} \mu_i(\hat{\beta}_{t+b_t}^{(0)}, \hat{\beta}_{t+b_t}^{(1)}) \quad (1.3)$$

This null hypothesis calls for several remarks. First, when it holds, it means that overall (for all i and t) including the ESG rating $\omega_{i,t}$ in the information set does not help for forecasting $y_{i,t}$.

In consequence, we should conclude that the ESG rating system investigated is void of information about $y_{i,t}$. Under the alternative hypothesis, considering the ESG rating in forecasting $y_{i,t}$, overall, gives real benefit across all firms and times.

Second, in contrast to the traditional framework for comparing predictive ability in Diebold and Mariano (1995) and West (1996), we can observe that the null hypothesis involves $\mu_i(\hat{\beta}_{t+b_t}^{(0)}, \hat{\beta}_{t+b_t}^{(1)})$, which depends on $\hat{\beta}_{t+b_t}^{(0)}$ and $\hat{\beta}_{t+b_t}^{(1)}$ which are the estimated values of the parameters instead of their population values. As discussed by Giacomini and White (2006) in a pure time series context, this helps preserve the finite sample behaviour of the estimators in the evaluation procedure, hence reflecting the effect of estimation uncertainty on the relative performance of the forecasts. This estimation uncertainty allows the comparison of nested forecasting models contrary to previous tests of predictive ability. However, they underline that adopting such a framework means remembering that the null hypothesis does not check the equal predictive ability of the competing models, but rather of the forecasting methods, where these methods include the models as well as the estimation procedures and the possible choices of estimation window.

This last remark means that some care is required in applying our test procedure to check for the validity of the null hypothesis in (1.1). First, the size of the estimation window should be kept fixed in the rolling window procedure $b_t = b$ ensure that parameter uncertainty does not vanish asymptotically.

This naturally rules out an expanding window forecasting scheme but allows for iterated or fixed schemes. Second, we should retain the same forecasting model and scheme and the same estimation window length to generate the forecasts $\widehat{m}_{i,t+\tau}^{(0)}(\hat{\beta}_{t+b_t}^{(0)})$ and $\widehat{m}_{i,t+\tau}^{(1)}(\hat{\beta}_{t+b_t}^{(1)})$. This is an important requirement, as it guarantees that the two forecasts diverge only by the set of information used, $\mathcal{F}_t^{(0)}$ or $\mathcal{F}_t^{(1)}$, the first of which excludes data on the ESG ratings for all firms.

1.2.2 Test statistic and asymptotic distribution

In this section, we provide the test statistic to check for the null hypothesis of a lack of informational content in an ESG rating system as expressed in (1.1). To do this we use the literature on comparing predictive ability in panel data settings (Davies and Lahiri, 1995; Timmermann and Zhu, 2019; Akgun et al., 2020). This literature considers extending the traditional predictive accuracy test for time series to a panel framework and it provides a test for overall equal predictive ability, meaning for all cross-sectional and time units as specified in (1.1) and tests for joint equal predictive ability across cross-sectional units or time clusters.

Specifically, we draw on the framework of Akgun et al. (2020) who extend the test of Diebold and Mariano (1995) to a panel data setting, considering the following test statistic based on the sample mean of loss differentials over time and units, so

$$\bar{\mu}_{n,T_0} = (nT_0)^{-1} \sum_{i=1}^{n} \sum_{t+\tau=1}^{T_0} \Delta\mathcal{L}_{i,t+\tau} \quad (1.4)$$

and is given by

$$\mathcal{T}_{n,T_0} = \frac{\bar{\mu}_{n,T_0}}{\bar{\sigma}_{n,T_0} / \sqrt{nT_0}} \quad (1.5)$$

where

$$\bar{\sigma}^2{}_{n,T_0} = n^{-1}\sum_{i=1}^{n}\sigma^2_{i,T_0}, \quad (1.6)$$

and $\sigma^2_{i,T_0} = var(\sqrt{T_0}\bar{\mu}_{i,T_0})$ is the long run variance of the ith time series of loss differentials.

As our null hypothesis is an extension to a panel setting of the unconditional predictive ability test of Giacomini and White (2006), rather than the one of Diebold and Mariano (1995), we need here assumptions that differ from those of Akgun et al. (2020), to establish the asymptotic distribution of the test statistic in (1.5)

Assumption 1 *For a given forecast horizon* $\tau \geq 1$ *and estimation window size* $b < \infty$, *suppose that (i)* $\{(y_{i,t}, x'_{i,t}, \omega_{i,t})', t = 1, \dots, T_0\}$ *for a given* i *is mixing with* Φ *of size* $-r/(2r-2), r \geq 2$, *or* α *of size* $-r/(r-2), r > 2$; *(ii)* $E|\Delta\mathcal{L}_{i,t+\tau}|^{2r} < \infty$ *for all* t *and a given* i; *(iii)* $\sigma^2_{i,T_0} = var(\sqrt{T_0}\mu_i(\hat{\beta}^{(0)}_{t+b_t}, \hat{\beta}^{(1)}_{t+b_t})) > 0$ *for all* T_0 *sufficiently large and a given* i.

Assumption 2 $\bar{\mu}_{i,T_0} = T_0^{-1}\sum_{t+\tau=1}^{T_0}\Delta\mathcal{L}_{i,t+\tau}, i = 1, \dots n$ *are independent, and*

$$E|\bar{\mu}_{i,T_0}|^{2+\delta} < C < \infty, (1.7)$$

for some $\delta > 0$ *for all* i. $\bar{\sigma}^2{}_{n,T_0} = n^{-1}\sum_{i=1}^{n}\sigma^2_{i,T_0} > \delta' > 0$ *for all* n *sufficiently large.*

Assumption 1 includes regularity conditions for the validity of Theorem 4 in Giacomini and White (2006). These conditions ensure that the test statistic for the unconditional predictive ability applied to a fixed cross-sectional unit converges to a standard Gaussian distribution, with

$$\mathcal{T}_i = \frac{\bar{\mu}_{i,T_0}}{\sigma_{i,T_0}/\sqrt{T_0}} \xrightarrow[T_0\to\infty]{} N(0,1) \quad (1.8)$$

Assumption 2 assumes the independence between the n random variables $\bar{\mu}_{i,T_0}, i, \dots, n$, meaning the average values over time of the loss differentials for each firm. This assumption allows the Central Limit Theory (CLT) applied to independent and heterogeneous random variables (White, 2001, Theorem 5.10) to hold. Note that this assumption is not a strong one within our framework, as opposed to macroeconomic forecasting. Indeed, our focus is on target variables that are related to firm-specific risk, which is by its nature a specific measure for each firm and hence primarily driven by firm characteristics rather than common factors. The following proposition provides the asymptotic distribution of the test statistic in (1.5).

Proposition 1 *Under the null hypothesis of a lack of informational content in ESG ratings as stated in (1.1), and if Assumptions 1-2 hold, we have that*

$$\mathcal{T}_{n,T_0} = \frac{\bar{\mu}_{n,T_0}}{\bar{\sigma}_{n,T_0}/\sqrt{nT_0}} \xrightarrow[T_0,n\to\infty]{} N(0,1) \quad (1.9)$$

Thus, we reject the null hypothesis when $\mathcal{T}_{n,T_0} < \mathrm{Z}_n$ with Z_n the quantile of order η of the standard Gaussian distribution, and η the nominal significance level. The proof of Proposition 1 is straightforward following Akgun et al. (2020), as we may note that under $\mathcal{H}_0$,

$$\sqrt{nT_0}\bar{\mu}_{n,T_0} = \frac{1}{n}\sum_{i=1}^{n}\sqrt{T_0}\bar{\mu}_{i,T_0} \quad , (1.10)$$

with $\bar{\mu}_{i,T_0}$ as defined in Assumption 2. For a fixed i, if Assumption 1 holds, $\sqrt{T_0}\bar{\mu}_{i,T_0} \xrightarrow[T_0,n\to\infty]{} \psi_i$ with $\psi_i \sim N\left(0, \sigma^2_{i,T_0}\right)$ and $\sigma^2_{i,T_0} = var(\sqrt{T_0}\mu_i(\hat{\beta}^{(0)}_{t+b}, \hat{\beta}^{(1)}_{t+b}))$. See Theorem 4 in Giacomini and White (2006). Hence the rest of the proof proceeds by noting that under Assumption 2 the CLT for heterogeneous but independent variables (White, 2001, Theorem 5.10) holds and $1/\sqrt{n}\sum_{i=1}^{n}\psi_i \xrightarrow[T_0,n\to\infty]{} N(0, \bar{\sigma}^2{}_{n,T_0})$ where again $\bar{\sigma}^2{}_{n,T_0} = n^{-1}\sum_{i=1}^{n}\sigma^2_{i,T_0\ i}$.

Note that to compute our test statistic $\mathcal{T}_{n,T_0}$, we need a consistent estimate $\hat{\bar{\sigma}}^2_{n,T_0}$of $\bar{\sigma}^2{}_{n,T_0}$. Under the assumption of cross-sectional independence of loss differentials, it follows that $\hat{\bar{\sigma}}^2_{n,T_0} = n^{-1}\sum_{i=1}^{n}\hat{\sigma}^2_{i,T_0}$ where $\hat{\sigma}^2_{i,T_0}$ is a suitable HAC estimator of the long-run variance σ^2_{i,T_0} of the ith time series of loss differentials, with

$$\hat{\sigma}^2_{i,T_0} = T_0^{-1}\sum_{t+\tau=1}^{T_0}\Delta\mathcal{L}^2_{i,t+\tau} + 2\left[T_0^{-1}\sum_{j=1}^{pT_0} w_{T_0,j}\times\sum_{t+\tau=1+j}^{T_0}\Delta\mathcal{L}_{i,t+\tau}\,\Delta\mathcal{L}_{i,t+\tau-j}\right] \quad (1.11)$$

and $\{pT_0\}$ is a sequence of integers such that $pT_0 \to \infty$ as $T_0 \to \infty$, $pT_0 = o(T_0)$, and $\{w_{T_0,j}: T_0 = 1,2,\dots; j = 1,\dots,pT_0\}$ is a triangular array such that $|w_{T_0,j}| < \infty, \mathrm{T}_0 = 1,2,\dots; j = 1,\dots,pT_0, w_{T_0,j} \to 1$ as $T_0 \to \infty, \mathrm{T}_0$ for each $j = 1,\dots,pT_0$ (Andrews, 1991).

1.3 Small sample properties

In this section we use a realistic simulation framework to analyse the small sample properties of the test. We begin by describing the simulation setup and then provide results for the sizes and the powers of the test under different forms of misspecification for the forecasting method retained.

1.3.1 Simulation setup

Our simulation setup proceeds by first simulating a vector $x_{i,t}$ of length $p = 10$ of innovations in financial variables that measure the financial strength of firm i at time t, with $t = 1,\dots,T$ and $T \in \{120,180,240\}$ as the sample size corresponding to 12, 15 and 20 years of monthly data. To have a realistic setup, these p variables are generated from a multivariate Gaussian distribution with mean vector $\bar{x}$ and covariance matrix Ω calibrated using real data

(see Appendix 1 for details about the calibration). With the vector $x_{i,t}$ of length p ready at hand, we generate the logarithmic value of the target variable $y_{i,t}$ for firm i, as[10]:

$$\log(y_{i,t+1}) = c_i^* + x_i' \beta_i^* + \gamma \omega_{i,t} + u_{i,t+1}, \qquad (1.12)$$

with $u_{i,t+1}$ following a standard Gaussian distribution, c_i^* as the constant term and β_i^* as a vector of parameters of length p. Note that we allow for heterogeneity across firms with specific values for the parameters for each firm. The values of c_i^* are generated as follows:

$$c_i^* = c^* + U(-\left|\frac{c^*}{10}\right|; \left|\frac{c^*}{10}\right|, (1.13)$$

with $U(a; b)$ as a uniform random variable over the set $[a, b]$. The same perturbation principle is used to generate each component of the vector β_i^*, with:

$$\beta_{i,j}^* = \beta_j^* + U(-\left|\frac{\beta_j^*}{10}\right|; \left|\frac{\beta_j^*}{10}\right|, (1.14)$$

$j = 1, \dots, p = 1$. The reference values c^* and β^* of the parameters are calibrated using real data (see Appendix 1 for details).

In equation (1.12), $\omega_{i,j}$ is the ESG rating, which for firm i and at each date t is generated from a uniform distribution over the set [0,1], and $\gamma \in \mathbb{R}_-$ is a parameter[11]. Note that our null hypothesis holds for $\gamma = 0$, since the ESG rating does not have any predictive content for $y_{i,t}$. With γ diverging from zero, the null hypothesis does not hold, because high lagged values of the ESG rating decrease the values of $y_{i,t}$.

[10] We use the logarithm, as most of possible candidate variables for $y_{i,t}$ are positive, including ESG incident variables.

[11] Note that we also considered a setup in which the ESG ratings $\omega_{i,j}$ are generated using a persistent AR(1) process to match the stylized fact of infrequent changes in ESG ratings. Simulations results available from the authors upon request show similar small sample properties of our inferential procedure.

Based on our design and for each Monte Carlo replication, with n and T fixed, the above simulation design is run for the n firms, with $n \in \{100,250,500\}$. This leads to a pure heterogeneous panel for $y_{i,t}$ the $p = 10$ innovations in financial variables $x_{i,t}$ and the ESG rating $\omega_{i,t}$, with $i = 1, \dots, n$ and $t = 1, \dots, T$.

1.3.2 Sizes and powers under a medium level of misspecification

For each Monte Carlo replication, we use the generated variables $y_{i,t}$, $x_{i,t}$, and $\omega_{i,t}$, $i = 1, \dots, n, t = 1, \dots, T$ and a fixed forecasting method to generate the forecast of $m_{i,t+1}^{(0)} = \mathrm{E}(\mathrm{y}_{i,t+1}|x_{i,t})$ and $m_{i,t+1}^{(1)} = \mathrm{E}(\mathrm{y}_{i,t+1}|x_{i,t}, \omega_{i,t})$, so $\widehat{m}_{i,t+1}^{(0)}(\hat{\beta}_{t,b}^{(0)})$ and $\widehat{m}_{i,t+1}^{(1)}(\hat{\beta}_{t,b}^{(1)})$ with b the estimation sample that we set to $b = [0.75T]$ and $[a]$ the integer part of a. This means that we use the first 75% of the T observations for each firm as the estimation sample and generate T_0 forecasts corresponding to the last 25% of the observations, meaning $T_0 = [0.25T]$ and $T = T_0 + b$.

The forecasts for both models are obtained using pooled OLS regression models. This means that both forecasting models are misspecified because the true panel structure of the data is heterogeneous across units. Besides, there is another form of misspecification that arises because the true data generating process uses a linear form for the logarithm of $y_{i,t}$ (see Eq. 1.12), while the pooled OLS regression models are fitted for the *raw* values of the same variable. Our goal is to evaluate how robust our inferential procedure is to these two levels of misspecification, which we call medium in comparison to another more severe form of misspecification that we will consider next. It may be recalled that the asymptotic behaviour of our test statistic under the null hypothesis suggests that with $\gamma \in \mathbb{R}_-$ in (1.12) diverging from zero, the null hypothesis is more likely to be rejected for $T_0, n \to \infty$, or equivalently, $T, n \to \infty$.

Figure 1.1 displays the rejection frequencies of the null hypothesis with respect to the parameter γ for a given couple (n, T) with the nominal significance level set to 5%. The rejection frequencies are computed over 1,000 simulations.

Overall, the test exhibits very good small sample properties, and we observe that the rejection frequencies for all couples (n, T) are close to 5% for γ and increase monotonically as γ diverges from 0.

We also observe that for a fixed n and $\gamma < 0$ the powers increase with T. Indeed, for $n = 100$ and $\gamma = -0.25$ the rejection frequencies for T of 120, 180 and 240 are 39.10%, 53.30% and 61.00% respectively. The same behaviour is observed for a fixed T and $\gamma < 0$ with the powers increasing with n. For instance, with $T = 120$ and $\gamma = -0.25$ the rejection frequencies for $n = 100, 250$ and 500 are respectively 39.10%, 71.50% and 91.30%. Hence our inferential procedure exhibits very good small sample properties. Figure A2.1 in Appendix 2 displays the rejection frequencies for the same simulation setup using the absolute error loss function. We can observe similar small sample properties, offering proof that our test is robust to the loss function.

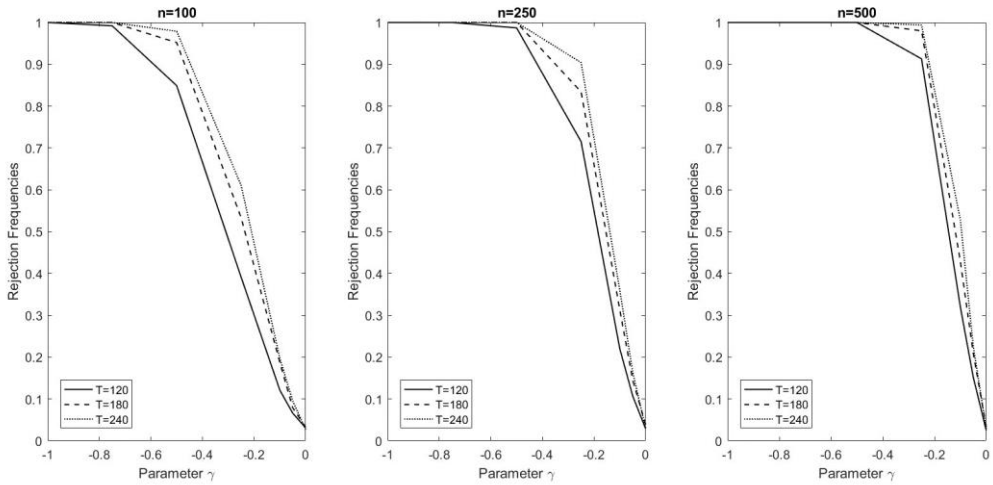


Figure 1.1 Rejection frequencies under a medium level of misspecification with the squared error loss function

1.3.3 Sizes and powers under a high level of misspecification

We now consider a configuration that will help us evaluate the properties of the test with respect to the choice of financial variables.

In the last subsection we assumed that the user of the test includes in the forecast models all the $p = 10$ innovations in the financial variables that enter the specification of the true model, but we make here the assumption that only some of these variables are retained. In each Monte Carlo replication, the following two pooled OLS models are estimated to compute out-of-sample forecasts $\widehat{m}_{i,t+1}^{(0)}(\hat{\beta}_{t,b}^{(0)})$ and $\widehat{m}_{i,t+1}^{(1)}(\hat{\beta}_{t,b}^{(1)})$ of $m_{i,t+1}^{(0)} = \mathrm{E}(\mathrm{y}_{i,t+1}|x_{i,t})$ and $m_{i,t+1}^{(1)} = \mathrm{E}(\mathrm{y}_{i,t+1}|x_{i,t}, \omega_{i,t})$:

$$y_{i,t+1} = c + \dot{x}_{i,t}'\beta + v_{i,t+1}^{(0)}, (1.15)$$

$$y_{i,t+1} = c + \dot{x}_{i,t}'\beta + \omega_{i,t}\gamma + v_{i,t+1}^{(1)}, (1.16)$$

with $v_{i,t+1}^{(0)}$ and $v_{i,t+1}^{(1)}$ as the error terms and $\dot{x}_{i,t}'$ as a vector with $p = 2$ randomly chosen financial variables from the $p = 10$ relevant ones as its elements, and $\hat{\beta}_{t,b}^{(0)} = (\hat{c}, \widehat{\beta}')'$, $\hat{\beta}_{t,b}^{(1)} = (\hat{c}, \widehat{\beta}', \hat{\gamma})'$. Assessing the small sample properties of the test with this additional form of misspecification is of great interest because such misspecification could probably arise in empirical applications where users are very likely to be wrong in their choice of the financial variables that matter.

Figure 1.2 displays the rejection frequencies over 1,000 simulations. We observe that the proposed test is robust to this form of misspecification. Indeed, the rejection frequencies are like those displayed in Figure 1.1, suggesting that making a mistake in the choice of financial variables is not harmful. Results available from the authors upon request show that the robustness holds even when the misspecification is more pronounced as only a quarter of the financial variables of interest are chosen. The robustness to the choice of the loss function can be seen in Figure A2.2 in Appendix 2.

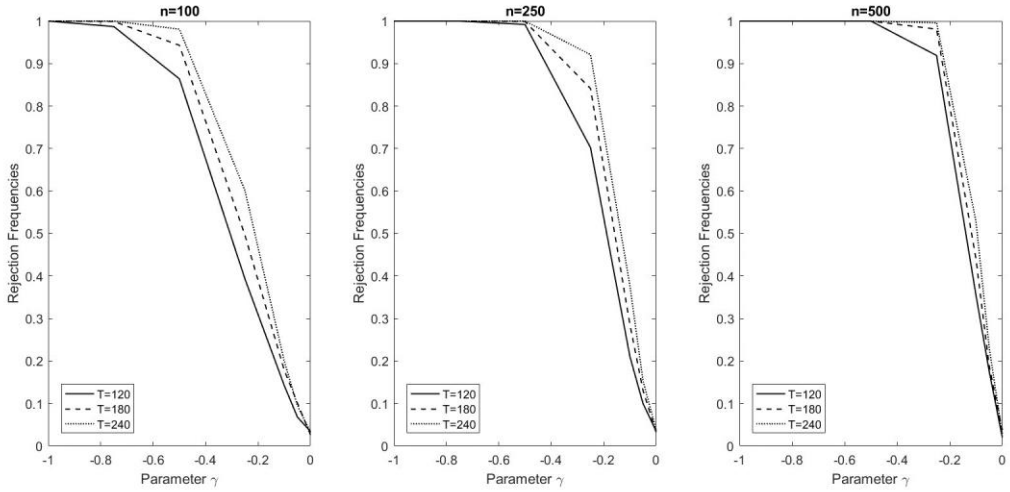


Figure 1.2 Rejection frequencies under a high level of misspecification with the squared error loss function

1.4 Empirical applications

This section illustrates our backtesting procedure using real datasets. We apply our methodology to two popular providers of ESG ratings, Sustainalytics and Asset4, over three universes from North America, Europe, and the Asia-Pacific region. We first describe our datasets and the related variables, and then conduct inferences to evaluate the informational content of each of the rating systems.

1.4.1 Description of the datasets and variables

The dataset for each of the three universes contains information for n firms at a monthly frequency over a period ranging from January 2010 to October 2018, giving a total of T=106 months. Note that we restrict our investigations to this period, as Sustainalytics has made a major change in the methodology for constructing its ratings in December 2018, with an inconsistency in the chaining of the ratings before and after this date. Precisely, before (after) this date, the ratings are performance (risk) measures, with higher (lower) ratings corresponding to best practices for environmental, social and governance issues.

Obviously, one solution would be to transform the risk-ratings into performance-ratings, but such a transformation would be arbitrary, and would not guarantee consistency in the scales of values. The North America, Europe and Asia-Pacific datasets gather information on respectively n=326, n=238 and n=217 firms. This deep panel structure ensures a high power for our backtesting methodology (see Monte Carlo simulations), with a total of 34,556, 25,228 and 23,002 pooled observations for the North America, Europe, and Asia-Pacific universes.

Table 1.1 displays pooled descriptive statistics of the ESG ratings for the two providers over the three universes. We may note that for both providers, higher values of the ESG ratings indicate higher ESG performance.

Table 1.1 Pooled descriptive statistics of the ESG ratings

	Min.	Max.	Mean	Median	Std.
			Europe		
Sustainalytics	36.0000	89.6900	66.5310	67.3000	9.6449
Asset4	5.4700	94.1500	64.4389	66.1300	15.7645
			North America		
Sustainalytics	33.0000	88.0000	59.0831	59.0000	8.6864
Asset4	2.4700	94.7700	54.4304	56.5200	18.8691
			Asia-Pacific		
Sustainalytics	32.0000	90.0900	58.5848	59.0000	8.3848
Asset4	2.3500	90.2700	53.3590	56.1900	18.2707

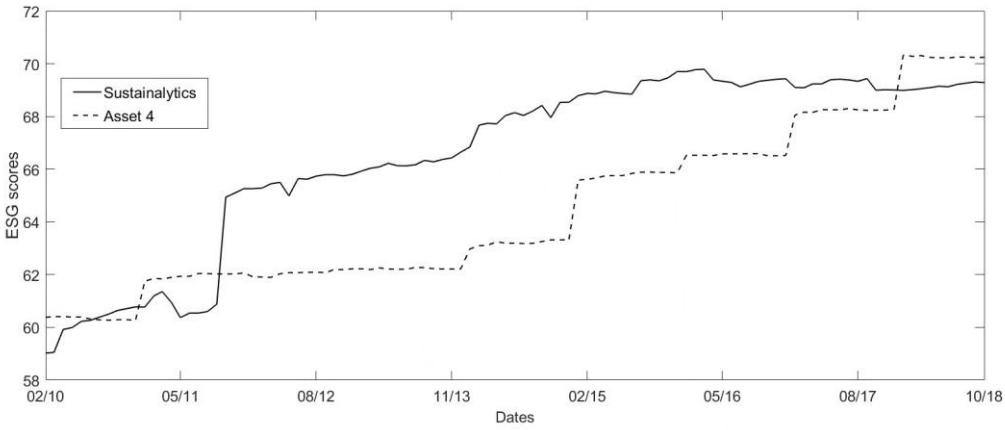
Notes: The table displays pooled descriptive statistics of the ESG ratings for the two providers (Sustainalytics and Asset4) over the three universes. The datasets contain monthly observations over the period from January 2010 to October 2018, giving a total of 106 months. The North America, Europe and Asia-Pacific datasets contain information on respectively $n = 326$, $n = 238$ and $n = 217$ firms. Min. refers to minimum, Max. to maximum, and std. to standard deviations.

The average values of the ESG ratings for the Europe universe are 66.53 for Sustainalytics and 64.43 for Asset4. This means the central statistics are similar for both providers, as is confirmed by the values of the median of 67.30 for Sustainalytics and 66.13 for Asset4 for the Europe universe. This stylised fact holds for the other two universes. However, the Asset4 ESG ratings have more variability across time and firms as given by the values of the standard deviations and ranges. The standard deviations of the Asset4 ESG ratings for instance are approximately twice as high as those for Sustainalytics.

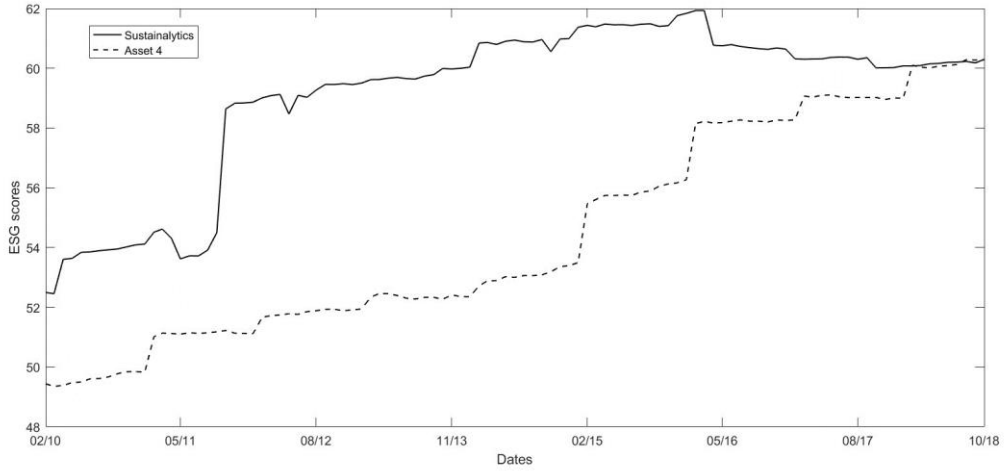
Figure 1.3 displays the evolution over time of the cross-sectional averages of the ESG ratings for the two providers in the three universes. We observe growth over time in the cross-sectional averages, which suggests a tendency towards upward revisions of the ESG ratings for firms. Assuming that ESG ratings accurately reflect ESG performance, this shows an overall improvement trend over time in the corporate behaviour of firms across the three universes regarding environmental, social, and governance best practices.

Figure 1.3 Dynamics of the cross-sectional means of the ESG ratings

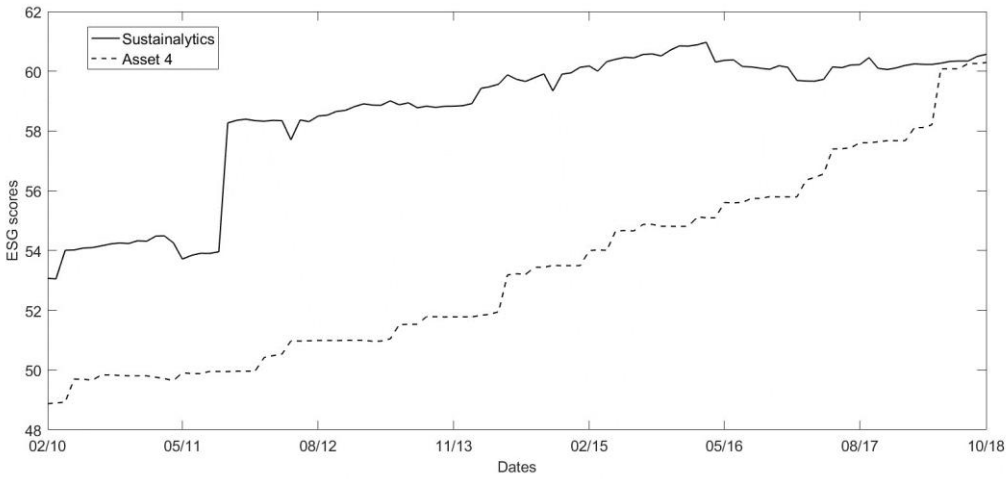
(a) Europe



(b) North America



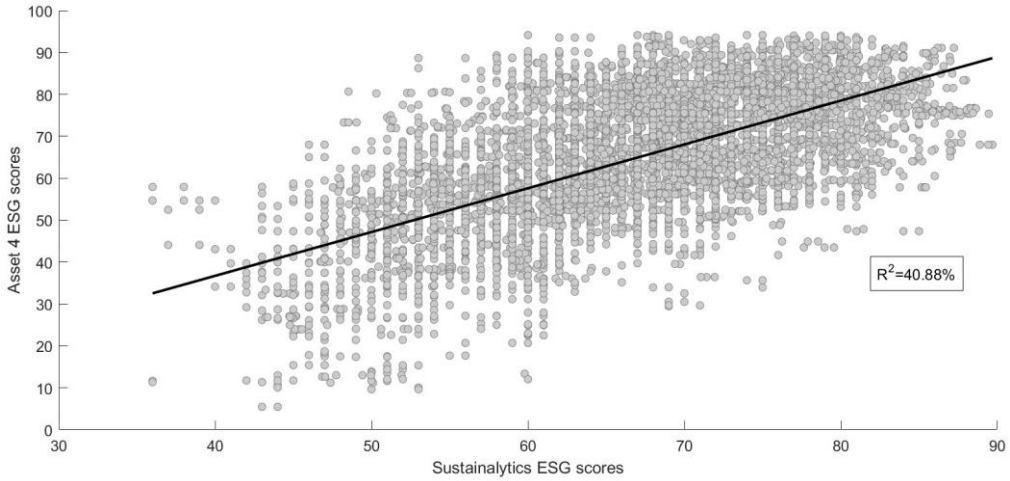
(c) Asia-Pacific



Source: The figure displays the scatter plot that shows the graphical relation between the ESG ratings for the two providers considered (Sustainalytics and Asset4). The dataset contains monthly observations for n=238 firms from January 2010 to October 2018, giving a total of 106 months.

To evaluate the link between the two rating systems, Figure 1.4 displays the scatter plot of the pooled ESG ratings from the two providers for the Europe universe. The figure also displays the fitted least square regression line, along with the adjusted R-squared, which is equal to 40.88%. Hence the link across firms and time between the two ESG ratings is weak, though it is positive. As already underlined, this has been highlighted many times in the literature and constitutes the main motivation of our paper, which proposes, in a context of limited convergence, a formal backtesting procedure for evaluating the informational content of ESG rating systems. The phenomenon is not only European and is also highlighted for the other two universes as shown by Figures A2.3 and A2.4 in Appendix 2. The trend is of the same order for the North America universe with an R-squared of 46.46%, but we observe a more pronounced divergence in the Asia-Pacific universe with an R-squared of only 32.65%.

Figure 1.4 Relation between the Sustainalytics and Asset4 ESG ratings: Europe



Source: The figure displays the scatter plot that shows the graphical relation between the ESG ratings for the two providers considered (Sustainalytics and Asset4). The dataset contains monthly observations for n=238 firms from January 2010 to October 2018, giving a total of 106 months.

Information on the target variable

In this sub-section, we provide information on the target variable. We consider the idiosyncratic volatility of stock returns as our dependent variable of interest. These variable measures idiosyncratic risk at the firm level that is not captured by traditional risk factors. ESG ratings could significantly help predict this target variable as stock markets can react to the arrival of firm specific ESG events (Serafeim and Yoon, 2022a) or global news corresponding to innovations in an ESG factor (Pastor et al., 2021; Ardia et al., 2022).

Another choice could be a variable or score measuring ESG incidents from leading providers. However, they seem divergent across providers, as the rank correlations between ESG incidents from Sustainalytics and Asset4 for instance are weak at 43% for Europe, 43% for North America and 34% for the Asia-Pacific region. Moreover, in this paper, we adopt the perspective of an investor using ESG information for its materiality on investment performance because this is the primary reason why investors use ESG information and many

rating agencies adopt this perspective (Amel-Zadeh and Serafeim, 2018). Since ESG news is not always financially relevant for investors (Serafeim and Yoon, 2022b), using a direct measure of financial risks, such as idiosyncratic volatility, seems more appropriate in our context.

To compute idiosyncratic volatility for each firm i we collect daily stock returns $r_{i,s}$ over our period of investigation from January 2010 to October 2018, with a total of 2,304 observations. For each universe, we also collect the daily returns $r_{m,s}$ of the MSCI stock index over the same period, using MSCI Europe, MSCI USA and MSCI Pacific for the Europe, North America, and Asia-Pacific universes. Residual returns are thus extracted assuming that the Capital Asset Pricing Model (CAPM) holds, with:

$$r_{i,s} = \alpha_i + \beta_i r_{m,s} + \epsilon_{i,s}, \quad (1.17)$$

where α_i is the alpha of the stock, β_i is the beta or exposure of the stock to the market, and $\epsilon_{i,s}$ is the innovation or residual return for stock i at day s. With the daily residual returns, we compute monthly idiosyncratic realized volatility as follows:

$$IRV_{i,t} = \sum_{s_k=1}^{v_t} \hat{\epsilon}_{i,s_k}^2, \quad (1.18)$$

with t the index of the month, v_t the number of daily observations in month t, and $\hat{\epsilon}_{i,s_k}$ the $s_k th$ fitted residual returns within month t. For each firm i in a given universe, we obtain a time series of monthly idiosyncratic realized volatility of length 106, which thus matches the monthly frequency and the length of the ESG data analysed in the previous sub-section. The backtesting procedure is then applied using the logarithmic transform of the idiosyncratic realized volatility as the target variable.

Remark 1 *The CAPM model in (1.17) is likely to be misspecified. In this case, our target variable* $y_{i,t} \equiv log\,(IRV_{i,t})$ *would be correlated across firms. However, recall that Proposition 1 does not require cross-sectional independence between* $y_{i,t}$*, but rather between loss differentials averaged over time. Besides, we further use a multi-factorial model to check the sensitivity of our results to the choice of the factor model.*

Table 1.2 Pooled descriptive statistics of idiosyncratic realized volatility

	Min (%)	Max (%)	Mean (%)	Median (%)	Std (%)
Europe	0.0133	24.6450	0.4634	0.2927	0.7064
North America	0.0099	50.6339	0.4970	0.2632	0.9217
Asia-Pacific	0.0236	33.3244	0.5416	0.3697	0.7303

Notes: The table displays pooled descriptive statistics of monthly idiosyncratic realised volatilities for the three universes. Idiosyncratic realised volatilities are computed from residual asset returns from the CAPM. The datasets contain monthly observations from January 2010 to October 2018, giving a total of 106 months. The North America, Europe and Asia-Pacific datasets contain information on respectively $n = 326$, $n = 238$ and $n = 217$ firms. Min. refers to minimum, Max. to maximum, and std. to standard deviations.

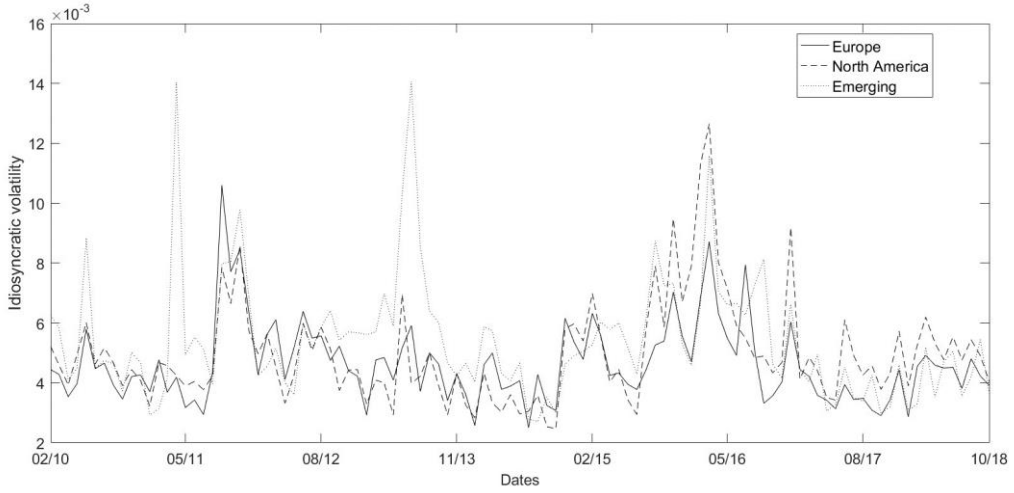
Table 2.2 displays the pooled descriptive statistics of monthly idiosyncratic volatilities for the three universes. The Asia-Pacific universe appears as the one where firms have on average the highest levels of idiosyncratic volatility. In terms of dispersion, the North America universe has more variability in the measure of the volatility of residual returns, as given by the values for the standard deviation and the range.

To get an overhead view of the monthly series of idiosyncratic realised volatilities, Figure 1.5 displays the evolution over time of the cross-sectional means of monthly idiosyncratic realised volatilities. We observe the typical dynamics, with volatility clusters that nevertheless seem less pronounced because we are dealing with idiosyncratic volatility, and not total volatility which includes the systematic part.

It may be recalled that our backtesting procedure is designed to test the informational content of the ESG ratings by checking whether they have predictive power for future idiosyncratic risk, as measured here by increased idiosyncratic volatility of stock returns. Hence, the relationship that the test aims to validate is that high ESG ratings lead to low idiosyncratic volatilities and low ratings lead to high volatilities.

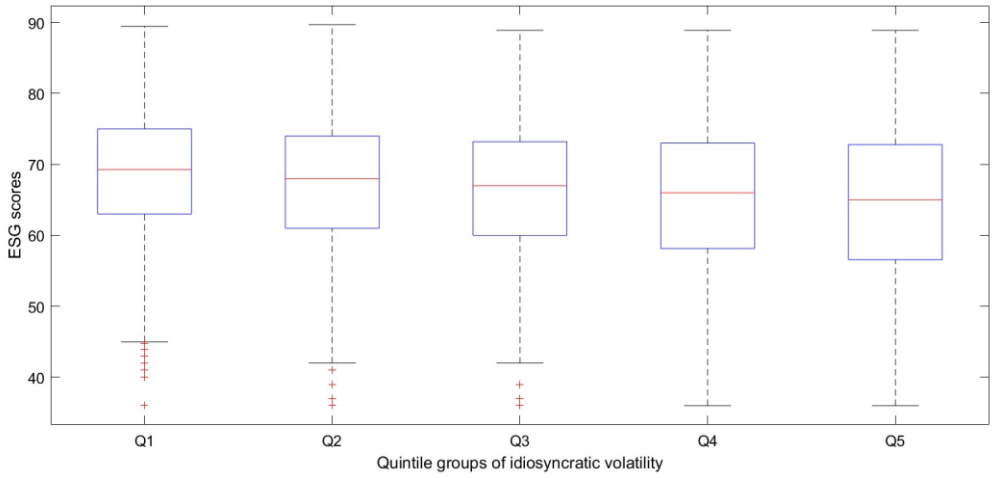
So, before we apply the backtesting procedure formally, Figures 1.6 and 1.7 try to illustrate whether there is such a relationship in the Europe universe. These figures report the distribution of the lagged values of the ESG ratings (Figure 1.6 for Sustainalytics and Figure 1.7 for Asset4) by idiosyncratic volatility quintiles. Overall, we observe that a negative relation arises, with high values of lagged ESG ratings associated with low idiosyncratic volatilities, while the median values of the lagged ESG ratings decrease with the order of the quintiles. Robustness across the universes is confirmed in Appendix 2, with Figures A2.5 and A2.6 for the North America universe, and A2.7 and A2.8 for the Asia-Pacific universe.

Figure 1.5 Dynamics of the cross-sectional means of idiosyncratic realised volatility



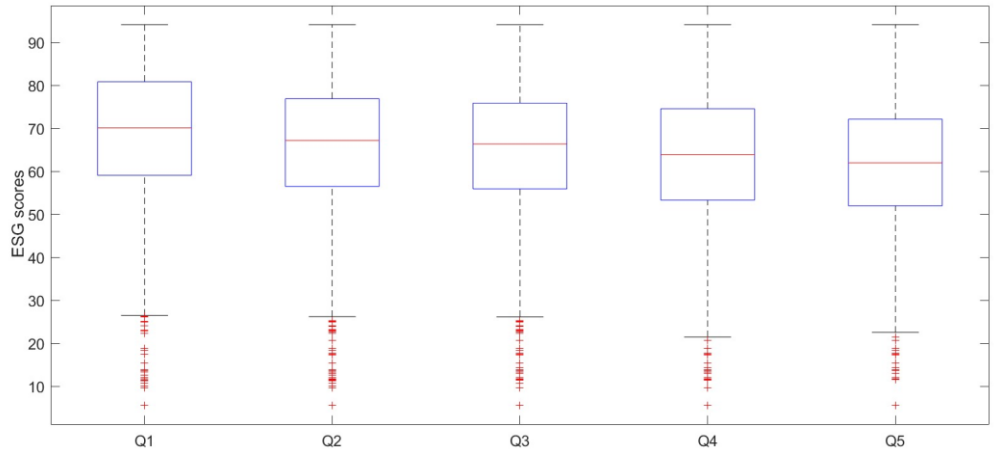
Source: The figure displays the evolution over time of the cross-sectional means of monthly idiosyncratic realised volatilities. Idiosyncratic realised volatilities are computed from residual stock returns from the CAPM. The datasets contain monthly observations from January 2010 to October 2018, giving a total of 106 months. The North America, Europe and Asia-Pacific datasets contain information on respectively $=326, n=238 and n=217 firms.

Figure 1.6 ESG ratings by idiosyncratic volatility quintiles: Sustainalytics (Europe)



Source: For the Europe universe, the figure displays the means of Sustainalytics ESG ratings within the five groups defined by the quintiles of idiosyncratic volatility computed with residual asset returns from the CAPM. The dataset contains monthly observations for n=238 firms from January 2010 to October 2018, giving a total of 106 months.

Figure 1.7 ESG ratings by idiosyncratic volatility quintiles: Asset4 (Europe)



Source: For the Europe universe, the figure displays the means of Asset4 ESG ratings within the five groups defined by the quintiles of idiosyncratic volatility computed with residual asset returns from the CAPM. The dataset contains monthly observations for n=238 firms from January 2010 to October 2018, giving a total of 106 months.

To control for potential confounding factors of the link between ESG ratings and idiosyncratic volatility, retain $p = 10$ financial variables for which the monthly observations are available for all firms over the three universes and the timespan considered. These variables are tax burden, interest burden, operating margin, asset turnover, leverage, current ratio, net debt to earnings before interest, taxes, depreciation, and amortisation (EBITDA), capital expenditure (Capex) to depreciation, current assets, and current liabilities (see Table 1.3 for a complete description of these variables). Innovations are extracted for each of these financial variables and for each firm by centering the raw values on the time average.

Table 1.3 Description of financial variables

Variables	Ratios	Description
Tax Burden	Net Income/Pretax Income	Profits retained after taxes
Interest Burden	Pretax Income/EBIT	Profits retained after interest paid
Operating Margin	EBIT/Revenue	Return on sales
Asset Turnover	Revenue/Total Assets	Revenue generated by own resources
Leverage	Total Assets/Total Equity	Measure of financial leverage
Current Ratio	Current Assets/Current Liab.	Measure of short-term resources
Net Debt to EBITDA	Net Debt/EBITDA	Capacity to finance debt
Capex to Dep.	Capex/Depreciation	Rate at which assets are renewed
Current Assets	Current Assets/Total Assets	Measure of short-term resources
Current Liab.	Current Liab./Total Liab.	Measure of short-term liabilities

Notes: The table gives the description of the financial variables retained. Innovations in these variables are used to control for the impact of financial factors when assessing the predictive contents of ESG ratings on the idiosyncratic volatility of a firm's assets.

1.4.2 Backtest results

Using the three categories of variables defined above as ESG ratings, idiosyncratic volatility and innovations in financial variables, we compute our test statistics and make inference for the predictive content of the two ESG rating systems considered. To predict the target idiosyncratic volatility variable, we consider a pooled OLS regression for the two models needed to run our backtesting procedure, which are the model that contains only innovations in the $p = 10$ financial variables, and the model that extends this set to include the lagged values of the ESG ratings. Recall that our procedure compares the predictive performance of the two models:

$$\log(IRV_{i,t+1}) = \alpha_0 + \beta_0 X_{i,t} + \epsilon^{(0)}_{i,t+1} \quad (1.19)$$

$$\log(IRV_{i,t+1}) = \alpha_1 + \beta_1 X_{i,t} + \gamma ESG_{i,t} + \epsilon^{(1)}_{i,t+1} \quad (1.20)$$

where $X_{i,t}$ denotes the vector of innovations in financial variables.

In line with our out-of-sample testing environment, we consider two different forecasting schemes: (i) a fixed forecasting scheme where the first 75% of the total T=106 months for each firm are used to estimate both models, and the forecasts are computed over the last 25% of observations, which are considered as the test sample; (ii) a rolling-window forecasting scheme with the forecasts computed by moving the estimation sample forward by including one more month and excluding the first, giving different estimation samples with the same fixed size of b=[0.75T].

Table 1.4 displays the outcome of the test for each provider across the three panel datasets. The test statistics are computed using the squared error loss. To gain more insights on the predictive power of the ratings, we perform inference on the aggregate ESG ratings of each provider and also on the specific dimensions of the ratings (environmental, social and governance). The values displayed represent the MSE variation in percentage when the ESG rating (in column) is added to the information set containing only innovations in financial variables. This presentation allows us to test our null hypothesis and to measure the magnitude of the predictive accuracy gain. Negative values are associated to MSE reductions with respect to the model excluding information about the ESG rating (or rating component), and hence to gains in predictive ability. We also report the sign of the regression coefficient associated with the ratings in parentheses. For the rolling window forecasting scheme, the coefficient is averaged across the estimation windows.

Table 1.4 Backtest of ESG ratings: results for squared error loss and idiosyncratic returns from CAPM

Sustainalytics					
		ESG	E	S	G
Rolling Window	EU	**−3.0%⋆⋆⋆** **(−0.01)**	**−3.8%⋆⋆⋆** **(−0.009)**	**−1.9%⋆⋆⋆** **(−0.008)**	**−0.54%⋆⋆⋆** **(−0.005)**
	NA	0.12% (−0.01)	**−0.53%⋆⋆⋆** **(−0.008)**	0.51% (−0.006)	0.21% (−0.006)
	AP	**−0.21%⋆⋆** **(−0.004)**	**−0.56%⋆⋆⋆** **(−0.004)**	−0.030% (−0.001)	0.019% (0.0001)
Fixed Window	EU	**−4.0%⋆⋆⋆** **(−0.01)**	**−4.7%⋆⋆⋆** **(−0.009)**	**−2.6%⋆⋆⋆** **(−0.008)**	**−1.1%⋆⋆⋆** **(−0.007)**
	NA	0.81% (−0.01)	−0.42% (−0.01)	1.3% (−0.008)	0.54% (−0.01)
	AP	**−0.55%⋆** **(−0.005)**	**−0.66%⋆⋆** **(−0.004)**	−0.17% (−0.003)	**−0.30%⋆** **(−0.003)**
Asset 4					
		ESG	E	S	G
Rolling Window	EU	**−3.1%⋆⋆⋆** **(−0.008)**	**−3.1%⋆⋆⋆** **(−0.006)**	**−3.5%⋆⋆⋆** **(−0.007)**	−0.063% (−0.002)
	NA	−0.0069% (−0.006)	**−0.30%⋆** **(−0.003)**	−0.10% (−0.005)	0.39% (−0.002)
	AP	**−0.028%⋆⋆** **(3e−05)**	**−0.12%⋆⋆⋆** **(−0.0003)**	**−0.36%⋆⋆⋆** **(−0.0008)**	−0.061% (0.0006)
Fixed Window	EU	**−4.2%⋆⋆⋆** **(−0.008)**	**−3.3%⋆⋆⋆** **(−0.005)**	**−4.5%⋆⋆⋆** **(−0.006)**	**−0.55%⋆** **(−0.003)**
	NA	−0.67% (−0.008)	−0.52% (−0.004)	−0.71% (−0.007)	0.51% (−0.003)
	AP	**−0.17%⋆⋆⋆** **(−0.0005)**	**−0.30%⋆⋆⋆** **(−0.0006)**	**−0.50%⋆⋆⋆** **(−0.0009)**	0.081% (−0.0003)

Notes: This table displays the variation in mean squared error (MSE) when ESG information is included in the model. Idiosyncratic volatilities are computed using the residual asset returns from CAPM. The datasets contain monthly observations from January 2010 to October 2018, giving a total of 106 months. The North America, Europe and Asia-Pacific datasets include information on respectively $n = 326$, $n = 238$ and $n = 217$ firms. $\star$, $\star\star$ and $\star\star\star$ indicate rejection of the null hypothesis of lack of informational content in ESG ratings at the 10%, 5% and 1% nominal risk levels respectively. Regression coefficients associated to the ESG rating are reported in parentheses.

For the Europe (EU) universe, the inclusion of ESG information significantly improves the model's predictive accuracy in all configurations except one. Among Sustainalytics ratings, the environmental rating is the strongest predictor of idiosyncratic volatility, with an MSE reduction of 3.8% and 4.7% for the rolling and fixed forecasting schemes, respectively. Among Asset4 ratings, the social rating provides more information, resulting in a 3.5% (4.5%) reduction in MSE using a rolling (fixed) forecasting scheme. Overall, the governance rating appears to be less informative in predicting stock return idiosyncratic volatility, as it is associated with the lowest predictive accuracy gains in all configurations. The results are mixed for the North America (NA) and Asia-Pacific (AP) universes. For the NA universe, and for both ESG rating systems, only the inclusion of the environmental rating in the rolling window forecasting scheme leads to significant predictive accuracy gains. The predictive accuracy gains are also lower than those for the EU universe. For the AP universe, we reject our null hypothesis in several configurations, but predictive accuracy gains remain modest compared to those for the EU universe. Furthermore, for most rejections of our null hypothesis, we find a negative association between ESG ratings and idiosyncratic volatility, indicating that higher ESG ratings are, on average, associated with lower stock return idiosyncratic volatility.

1.4.3 Robustness to factor models

Here we evaluate the sensitivity of our results to the choice of factor model used to compute the target idiosyncratic realized volatility variable. We thus extend the CAPM model and consider a multifactorial model. This extension is anchored to the findings of academic research into the existence of common risk factors beyond the market index. This strand of the literature, which can be dated back to the seminal work of Fama and French (1992), has discovered many market variables or factors that may be able to explain the cross-sectional variations of stock returns. These include the size and value factors in Fama and French (1992) and the momentum factor in Jegadeesh and Titman (1993).

To consider the multifactorial model, we extend the CAPM model in (1.17) by adding investable factors identified in the literature to drive the cross-sectional variations of the stock's returns. For the Europe and the North America universes these are the MSCI Small/Large Capitalisation factor, which approximates the size anomaly, the MSCI Value/Growth factor associated with the value premium, the MSCI Momentum factor, the MSCI quality factor, and the MSCI Minimum Volatility factor. The lack of data for the Asia-Pacific universe means we consider three factors beyond the market, these being the MSCI Small/Large Capitalisation factor, the MSCI Value/Growth factor, and the MSCI Minimum Volatility factor. Table 1.5 displays the tests results for the idiosyncratic volatility computed using a multifactorial model and the squared error loss function. Overall, we reach qualitatively similar conclusions, suggesting that our results are robust to the choice of the factor model. We also evaluate the sensitivity of the test to the choice of the loss function used to evaluate forecasts. Table A2.1 and Table A2.2 replicates the results of Table 1.4 and Table 1.5 using the absolute error loss to assess models' predictive performance. In comparison with the squared error loss function, the absolute error loss function is more robust to outliers. We find that results are highly similar for the two loss functions, suggesting that our conclusions are robust to the choice of the loss function.

Table 1.5 Backtest of ESG ratings: results for squared error loss and idiosyncratic returns from multifactorial model

Sustainalytics					
		ESG	E	S	G
Rolling Window	EU	**−3.3%***** **(−0.01)**	**−4.2%***** **(−0.01)**	**−2.1%***** **(−0.009)**	**−0.67%***** **(−0.006)**
	NA	−0.076% (−0.01)	**−0.63%***** **(−0.009)**	0.39% (−0.009)	0.088% (−0.009)
	AP	**−0.14%*** **(−0.003)**	**−0.59%***** **(−0.005)**	**−0.030%*** **(0.0002)**	0.028% (0.002)
Fixed Window	EU	**−4.4%***** **(−0.01)**	**−5.0%***** **(−0.01)**	**−2.9%***** **(−0.009)**	**−1.3%***** **(−0.008)**
	NA	0.56% (−0.02)	−0.43% (−0.01)	1.2% (−0.01)	0.27% (−0.01)
	AP	**−0.42%*** **(−0.005)**	**−0.65%**** **(−0.005)**	−0.084% (−0.001)	−0.13% (−0.002)
Asset 4					
		ESG	E	S	G
Rolling Window	EU	**−3.4%***** **(−0.009)**	**−3.4%***** **(−0.006)**	**−3.6%***** **(−0.007)**	−0.054% (−0.003)
	NA	−0.33% (−0.008)	**−0.53%**** **(−0.004)**	−0.26% (−0.007)	0.42% (−0.003)
	AP	0.15% (0.0008)	**−0.043%***** **(−9e−05)**	**−0.089%***** **(−0.0001)**	**−0.23%***** **(0.001)**
Fixed Window	EU	**−4.5%***** **(−0.009)**	**−3.6%***** **(−0.006)**	**−4.6%***** **(−0.007)**	**−0.64%**** **(−0.003)**
	NA	**−1.2%*** **(−0.01)**	**−0.86%*** **(−0.005)**	−0.88% (−0.008)	0.38% (−0.005)
	AP	0.15% (0.0004)	**−0.17%***** **(−0.0003)**	**−0.042%***** **(−8e−05)**	**−0.11%**** **(0.0004)**

Notes: This table displays the variation in mean squared error (MSE) when ESG information is included in the model. Idiosyncratic volatilities are computed using the residual asset returns from a multifactorial model. The datasets contain monthly observations from January 2010 to October 2018, giving a total of 106 months. The North America, Europe and Asia-Pacific datasets include information on respectively $n = 326$, $n = 238$ and $n = 217$ firms. $\star$, $\star\star$ and $\star\star\star$ indicate rejection of the null hypothesis of lack of informational content in ESG ratings at the 10%, 5% and 1% nominal risk levels respectively. Regression coefficients associated to the ESG rating are reported in parentheses.

So far, our results show that the predictive power of ESG ratings varies depending on the universe considered. We find strong evidence that higher ESG ratings are associated with lower future stock return idiosyncratic volatility for the European universe, and to a lesser extent for the North America and Asia-Pacific universes. This finding can be explained by the fact that European regulation on ESG issues is more stringent, with the establishment of a high-level expert group on sustainable finance (HLEG) in 2016 and the subsequent introduction of the EU taxonomy for sustainable activities[12]. As a result, European investors are more likely to consider ESG information valuable for their investment decisions compared to US investors (Amel-Zadeh and Serafeim, 2018). Regarding the rating dimensions, the environmental rating appears to carry the most information, followed by the social rating, while predictive accuracy gains are consistently lower for the governance rating. This is consistent with the findings of Berg et al. (2022b), who reported that the noise in ratings is higher for the governance component, followed by the social component, with the environmental component being the least noisy. In the next subsection, we conduct additional empirical investigations to check the robustness of our results.

1.4.4 Disagreement between raters and the informational content of the ESG ratings

Our results suggest that both rating systems are informative for forecasting idiosyncratic volatility in Europe, where regulation on ESG is more stringent, and to a lesser extent in other regions. Another factor that could affect the link between ESG ratings and return volatility is ESG ratings disagreement. Serafeim and Yoon (2022a) analysed the link between ESG ratings and ESG risks as measured by ESG-related events and showed that the consensus rating predicts future news, but its predictive ability diminishes for firms where there is a large disagreement between raters. They also found that the consensus rating moderates the stock market reaction to ESG risks. Therefore, the forecasting power of ESG ratings could be moderated by ESG rating disagreement as it affects both the likelihood of ESG events and the stock market reaction to ESG risk materialization.

[12] https://finance.ec.europa.eu/publications/high-level-expert-group-sustainable-finance-hleg_en

In our sample, we find that ESG ratings are quite divergent across the three universes. The R-squared for the linear regression between the two rating agencies is equal to 40.88% for the EU universe, 46.46% for the NA universe, and 32.65% for the AP universe (see Figures 1.4, A2.3, and A2.4).

To check for this stylised fact, we replicate the results of Table 1.5 but partition each panel into consensus and disagreement groups, based on the firm level correlation between the ratings of the two providers. For each universe, the consensus group contains firms belonging to the top 25% of highest correlations, while the disagreement group contains firms belonging to the top 25% lowest correlations. Among the consensus group, the average correlation between ESG ratings of the two providers are equal to 75%, $72% and $70% for the EU, NA and AP universes, respectively. Among the disagreement group, these figures are equal to -35%, -46% and -44%, meaning that there is considerable divergence between ratings. Table 1.6 shows the sector distribution in the consensus and disagreement samples. Since the consensus rating is driven by the different methodologies used by rating agencies rather than by firm characteristics (Berg et al., 2022b), we observe a similar distribution of sectors across the two groups.

Table 1.6 Distribution of sectors in consensus and disagreement samples

Sector	Consensus	Disagreement
Consumer Discretionary	20.3%	19.5%
Industrials	20.3%	21.1%
Information Technology	14.7%	13.5%
Energy	10.2%	8.4%
Materials	9.1%	12.0%
Consumer Staples	8.1%	8.0%
Healthcare	6.1%	6.0%
Communication Services	5.6%	5.2%
Utilities	4.6%	5.6%
Financials	1.0%	0.8%

Notes: The table displays distribution of sectors in the consensus and disagreement samples.

Table 1.7 displays the backtest results for the consensus and disagreement groups using a rolling window forecasting scheme. The results using a fixed window are displayed in Appendix 2 (Table A2.3). We observe significant differences in terms of the rejection of the null hypothesis between the two groups. Among the consensus group, we observe 17 rejections out of 24 tests at the 1% nominal risk level, while this figure drops to 5 rejections for the disagreement group. Moreover, the forecasting power of ESG ratings is consistently greater for consensus firms across the three universes. For example, considering the EU universe, the MSE reduction due to the inclusion of Sustainalytics environmental rating is equal to 7.7% in the consensus sample, but only 2.3% in the disagreement sample. For the NA universe, these figures are 2.6% for the consensus sample and 0.14% for the disagreement sample. Similar conclusions hold for most configurations and universes.

To assess the sensitivity of the previous results to the threshold used to define the consensus firms, we repeated the analysis for alternative levels of ESG consensus. We started with the full sample and excluded the top x% of firms with the highest level of disagreement before applying our inferential procedure. Figure 1.8 displays the results for values of x ranging between 0% and 75% using a rolling window forecasting scheme. Results obtained using a fixed window forecasting scheme are displayed in the appendix (Figure A2.9). We find that the forecasting power of ESG ratings increases with the level of ESG consensus. This result is consistent for both rating agencies and across the three universes. Overall, predictive accuracy gains due to the inclusion of ESG information increase with the level of ESG consensus.

Table 1.7 Consensus vs disagreement between providers using a rolling window forecasting scheme (MSE)

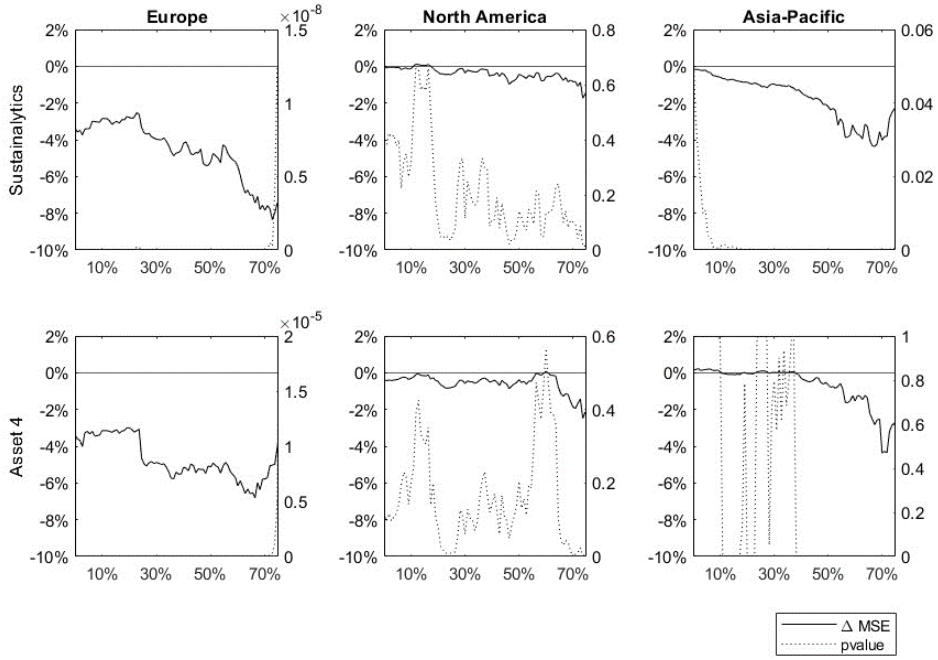
Sustainalytics					
		ESG	E	S	G
Consensus	EU	**−7.7%***** **(−0.02)**	**−7.1%***** **(−0.01)**	**−5.0%***** **(−0.01)**	**−3.7%***** **(−0.01)**
	NA	**−1.6%**** **(−0.02)**	**−2.6%***** **(−0.01)**	**−1.0%*** **(−0.01)**	0.97% (−0.009)
	AP	**−2.4%***** **(−0.01)**	**−3.5%***** **(−0.01)**	**−0.64%***** **(−0.003)**	**−1.1%**** **(−0.008)**
Disagreement	EU	**−2.9%***** **(−0.01)**	**−2.3%***** **(−0.007)**	**−2.8%***** **(−0.01)**	−0.032% (−0.003)
	NA	−0.091% (−0.01)	−0.14% (−0.01)	0.035% (−0.009)	0.027% (0.0003)
	AP	**−1.3%***** **(0.006)**	0.29% (−0.001)	**−1.1%*** **(0.009)**	**−1.4%**** **(0.008)**
Asset 4					
		ESG	E	S	G
Consensus	EU	**−4.0%***** **(−0.008)**	**−4.8%***** **(−0.007)**	**−4.7%***** **(−0.007)**	**−0.056%*** **(−0.0002)**
	NA	**−2.3%***** **(−0.009)**	**−5.1%***** **(−0.008)**	**−2.1%***** **(−0.006)**	0.46% (−0.002)
	AP	**−2.5%***** **(−0.006)**	**−3.2%***** **(−0.005)**	**−0.78%**** **(−0.003)**	**−0.33%***** **(−0.001)**
Disagreement	EU	**−1.2%*** **(−0.008)**	**−3.4%***** **(−0.005)**	**−1.4%*** **(−0.008)**	0.74% (−0.001)
	NA	0.64% (−0.008)	0.21% (−0.003)	0.55% (−0.005)	0.48% (−0.005)
	AP	−0.30% (0.002)	**−0.92%**** **(0.002)**	0.28% (0.001)	**−0.24%*** **(0.0007)**

Notes: This table displays the variation in mean squared error (MSE) when ESG information is included in the model. Idiosyncratic volatilities are computed using the residual asset returns from a multifactorial model. Results are computed using a rolling window forecasting scheme. For a given universe, the consensus group contains firms with the 25% highest correlations between the ratings of the two providers. The disagreement group contains firms with the 25% lowest correlations. $\star$, $\star\star$ and $\star\star\star$ indicate rejection of the null hypothesis of lack of informational content in ESG ratings at the 10%, 5% and 1% nominal risk levels respectively. Regression coefficients associated to the ESG rating are reported in parentheses.

We next test if the predictive ability of ESG ratings is the same for consensus firms with high and low ratings. To do so, we apply our test separately to consensus firms with a high ESG rating (above the median) and a low ESG rating (below the median). Results using a rolling (fixed) window forecasting scheme are displayed in Table 1.8 (Table A2.4). We find that for both rating agencies, the predictive ability is greater for consensus firms with a low rating in the NA universe, but that the predictive accuracy gains depend on the rating agency considered for the other universes.

From a practical point of view, our results provide crucial information for portfolio managers who integrate ESG information into their investment decisions. We show that it is necessary to cross-check the information gathered from multiple ESG rating providers before integrating ESG into the management process. The focal point of our results is that consensus about the ESG ratings is informative about idiosyncratic risk, while ESG ratings with disagreement are less valuable from this viewpoint.

Figure 1.8 Decrease in forecast error in function of ESG consensus (rolling window)



Source: This figure displays the variation in MSE when ESG information is included in the model as a function of the level of consensus between ESG providers. The x-axis represents the level of consensus between rating agencies. For a level of consensus x, only the firms with

the 1-x highest correlations between the ratings of the two providers were included in the sample.

Table 1.8 Consensus firms: high vs low ESG ratings (rolling window)

		Sustainalytics			
		ESG	E	S	G
High ESG	EU	**−2.9%***** **(−0.01)**	**−5.8%***** **(−0.01)**	−0.16% (−0.005)	0.19% (0.002)
	NA	4.0% (0.01)	0.67% (0.002)	6.8% (0.009)	−0.37% (−0.003)
	AP	**−0.68%**** **(−0.005)**	−0.47% (−0.005)	**−0.75%**** **(−0.003)**	0.037% (6e−05)
Low ESG	EU	**−4.8%***** **(−0.01)**	**−4.3%***** **(−0.009)**	**−2.1%***** **(−0.008)**	**−3.7%***** **(−0.01)**
	NA	**−1.7%**** **(−0.02)**	**−2.6%***** **(−0.01)**	−0.60% (−0.01)	0.12% (−0.008)
	AP	**−1.9%***** **(−0.005)**	**−0.63%**** **(−0.0007)**	**−1.4%***** **(−0.005)**	**−2.1%***** **(−0.008)**
		Asset 4			
		ESG	E	S	G
High ESG	EU	**−8.5%***** **(−0.02)**	**−2.5%***** **(−0.005)**	**−6.6%***** **(−0.01)**	**−1.2%***** **(−0.003)**
	NA	0.23% (0.003)	−0.75% (−0.004)	0.66% (0.008)	0.88% (−0.002)
	AP	**−3.6%***** **(−0.009)**	**−0.97%**** **(−0.004)**	**−0.86%**** **(−0.003)**	**−0.54%**** **(−0.0009)**
Low ESG	EU	0.25% (0.0007)	**−2.8%***** **(−0.004)**	**−0.71%***** **(−0.001)**	**−1.7%***** **(0.005)**
	NA	**−4.6%***** **(−0.01)**	**−7.0%***** **(−0.009)**	**−4.3%***** **(−0.008)**	**−0.20%*** **(−0.0003)**
	AP	−0.14% (−0.0006)	−0.074% (−0.0003)	−0.11% (0.0008)	0.47% (−0.001)

Notes: This table displays the variation in mean squared error (MSE) when ESG information is included in the model for consensus firms. Idiosyncratic volatilities are computed using the residual asset returns from a multifactorial model. Results are computed using a rolling window forecasting scheme. High (low) ESG sample represents firms above (below) the median ESG rating. $\star$, $\star\star$ and $\star\star\star$ indicate rejection of the null hypothesis of lack of informational content in ESG ratings at the 10%, 5% and 1% nominal risk levels respectively. Regression coefficients associated to the ESG rating are reported in parentheses.

1.5 Conclusion

The contribution of this article is to propose a formal statistical procedure for assessing the informational content in ESG ratings. The test proceeds by evaluating how well these extra-financial metrics help in predicting a given target variable intended to measure firm-specific risks. Our framework allows users to choose a target variable related to their investment objectives. Technically, our inferential procedure for checking the informational content in ESG ratings is based on extending the conditional predictive ability test of Giacomini and White (2006) to a panel setting. Under weak assumptions, including cross-sectional dependencies among loss functions for firms, we derive the Gaussian asymptotic distribution of the test statistic. Monte Carlo simulations conducted under different types of model misspecification show that the test has good small sample properties.

Empirical applications are conducted using the idiosyncratic volatility of stock returns, a measure of firm-specific risk, as our target variable. We apply our procedure to evaluate two leading ESG rating systems (Sustainalytics and Asset4) in three investment universes (Europe, North America, and the Asia-Pacific region). The results show that the null hypothesis of a lack of informational content in ESG ratings is strongly rejected for Europe, while the results are mixed, and predictive accuracy gains are lower for the other regions. Furthermore, we find that the predictive accuracy gains are higher for the environmental dimension of the ESG ratings. Importantly, we find that the predictive accuracy gains derived from ESG ratings increase with the level of consensus between rating agencies in all three universes, while they are low for firms over which there is a high level of disagreement.

The results have important implications for investors and researchers. For investors, our backtest procedure provides a useful and practical framework for considering ESG rating providers before integrating the ratings into the investment process. Our results suggest prudence about the information content of ESG ratings when they diverge. For researchers in asset pricing, it is crucial to check properly the quality of ESG ratings before using them, especially when the ratings are divergent. Moreover, the link between ESG ratings and idiosyncratic volatility when the ratings are convergent suggests that ESG investing is not just

an issue of the preferences of investors, but that ESG ratings can also provide information about future fundamentals and risks. A future application for investors could be to compare the ratings of competing ESG rating agencies since our inferential procedure can be easily adapted to compare the informational content in the ESG ratings. This would help investors in selecting one agency among several competing ones in non-nested comparisons, or in considering additional competing agencies to combine with their already existing ratings in nested comparisons.

1.6 Appendices

Appendix 1.1: Details on the Monte Carlo simulations

In this Appendix we provide details about the simulations of innovations in the financial variables for generating the small sample properties of the test (see Section 1.3). These variables are generated via a multivariate Gaussian distribution with mean vector $\bar{x}$ and covariance matrix Ω calibrated using real data. The dataset we use contains historical monthly values of $p = 10$ innovations in the financial variables for 238 European firms from January 2010 to October 2018, giving a total of 106 months.

Innovations are computed as deviations from the overall means. The financial variables are, in order: tax burden ratio, interest burden ratio, operating margin ratio, asset turnover ratio, leverage as measured by the ratio of total assets to total equity, current ratio as measured by the ratio of current assets to current liabilities, debt ratio, capex as measured by the ratio of capital expenditures to depreciation, current assets as measured by the ratio of current assets to total assets, current liabilities as measured by the ratio of current liabilities to total liabilities.

The mean vector is thus equal to

$$\bar{x} = [0.8137; 0.8333; 0.1391; 0.8265; 3.8713; 1.4031; 1.7466;\ 1.2779; 0.3634; 0.2880]$$

and the covariance matrix Ω equal to

$$\Omega = \begin{pmatrix} 4.098 & -0.061 & -0.007 & -0.003 & -0.003 & 0.008 & 0.209 & -0.023 & -0.001 & -0.001 \\ -0.061 & 17.732 & -0.008 & 0.037 & 5.704 & 0.057 & -0.136 & -0.021 & 0.017 & 0.007 \\ -0.007 & -0.008 & 0.012 & -0.025 & -0.369 & 0.010 & -0.018 & 0.025 & -0.005 & -0.006 \\ -0.003 & 0.037 & -0.025 & 0.284 & -0.559 & -0.025 & -0.358 & -0.067 & 0.042 & 0.037 \\ -0.003 & 5.704 & -0.369 & -0.559 & 5012.291 & -0.521 & 30.792 & 4.391 & -0.160 & -0.092 \\ 0.008 & 0.057 & 0.010 & -0.025 & -0.521 & 0.642 & -0.420 & 0.042 & 0.053 & -0.040 \\ 0.209 & -0.136 & -0.018 & -0.358 & 30.792 & -0.420 & 32.172 & 0.550 & -0.154 & -0.058 \\ -0.023 & -0.021 & 0.025 & -0.067 & 4.391 & 0.042 & 0.550 & 1.841 & -0.023 & -0.022 \\ -0.001 & 0.017 & -0.005 & 0.042 & -0.160 & 0.053 & -0.154 & -0.023 & 0.029 & 0.014 \\ -0.001 & 0.007 & -0.006 & 0.037 & -0.092 & -0.040 & -0.058 & -0.022 & 0.014 & 0.018 \end{pmatrix}.$$

For the simulation of the target variable of idiosyncratic volatility, we run a pooled OLS regression with the dependent variable being the logarithm of the monthly time series of idiosyncratic realised volatility over the same period (January 2010 to October 2018) for the 238 European firms. The explanatory variables are the innovations in the 10 financial variables as described above.

$c^\star$	$\beta_1^\star$	$\beta_2^\star$	$\beta_3^\star$	$\beta_4^\star$	$\beta_5^\star$	$\beta_6^\star$	$\beta_7^\star$	$\beta_8^\star$	$\beta_9^\star$	$\beta_{10}^\star$
-5.9165	0.0070	-0.0015	-0.8739	-0.0679	0.0048×10^{-2}	0.0941	0.0044	0.0605	0.1869	0.0824

For the $p = 10$ financial variables, the estimated coefficients are displayed above. These estimates are used to generate data for simulating the logarithm of idiosyncratic realised volatility, and applying the exponential function leads to the target variable.

Appendix 1.2: Additional Tables and Figures

Figure A2.1: Rejection frequencies under a medium level of misspecification with the absolute error loss function

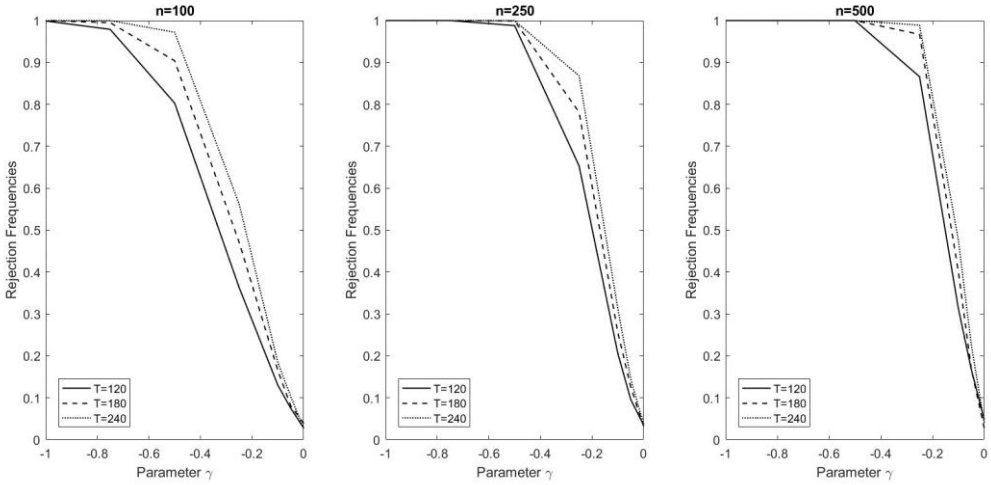


Figure A2.2: Rejection frequencies under a high level of misspecification with the absolute error loss function

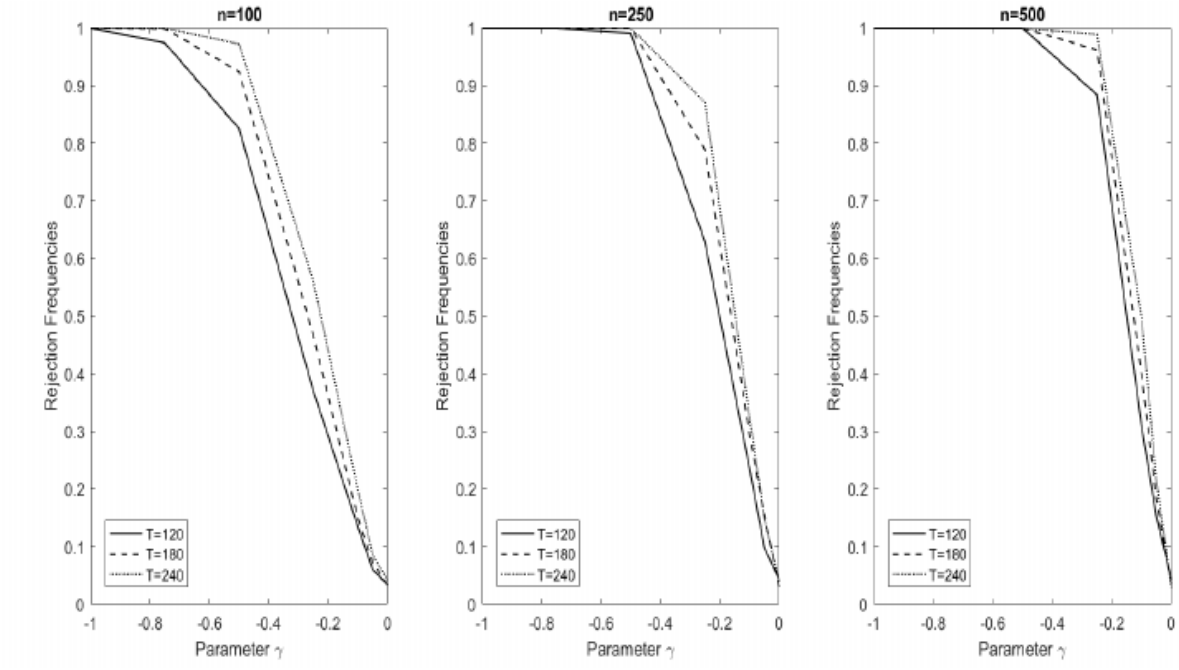


Figure A2.3: Relation between the Sustainalytics and Asset4 ESG ratings: North America

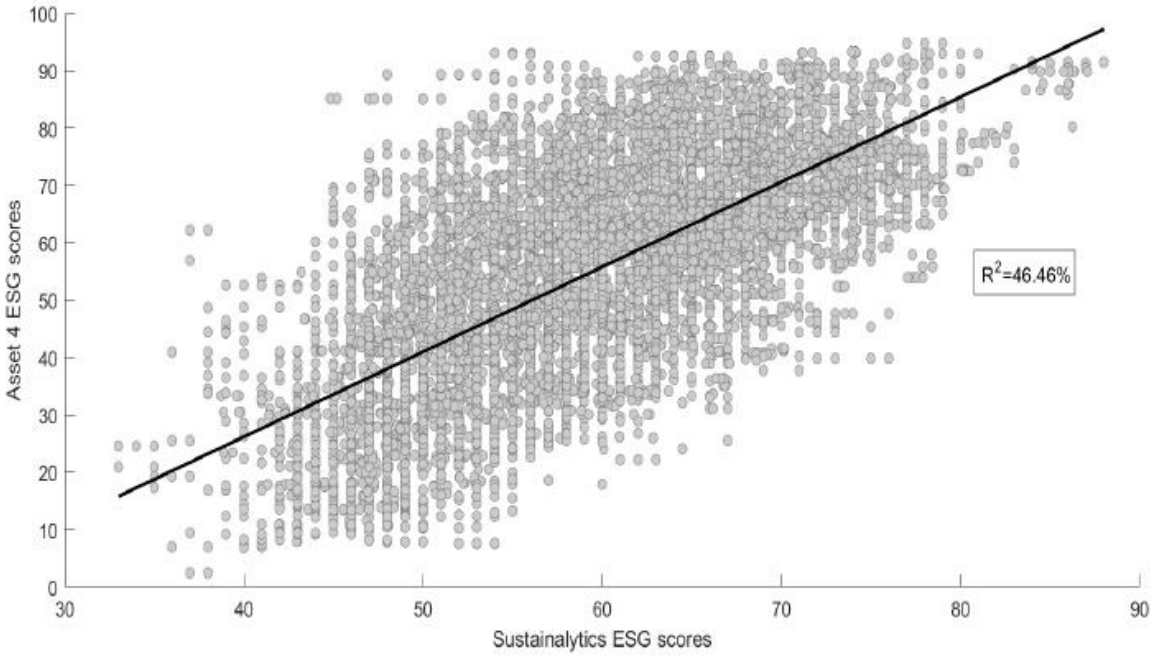


Figure A2.4: Relation between the Sustainalytics and Asset4 ESG ratings: Asia-Pacific

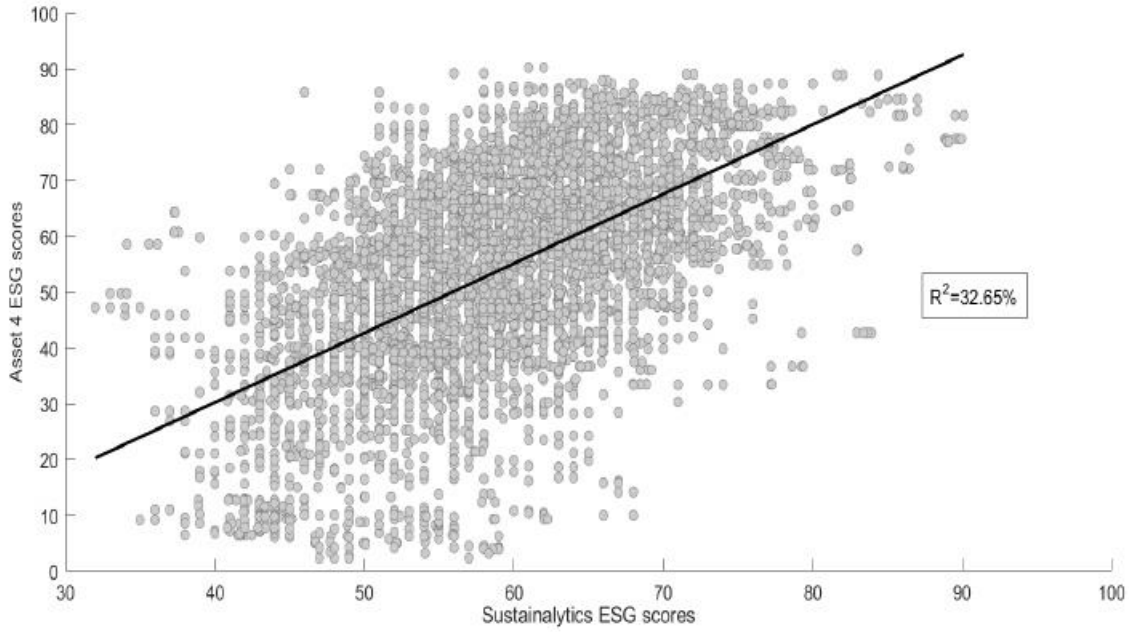


Figure A2.5: ESG ratings by idiosyncratic volatility quintiles: Sustainalytics (North America)

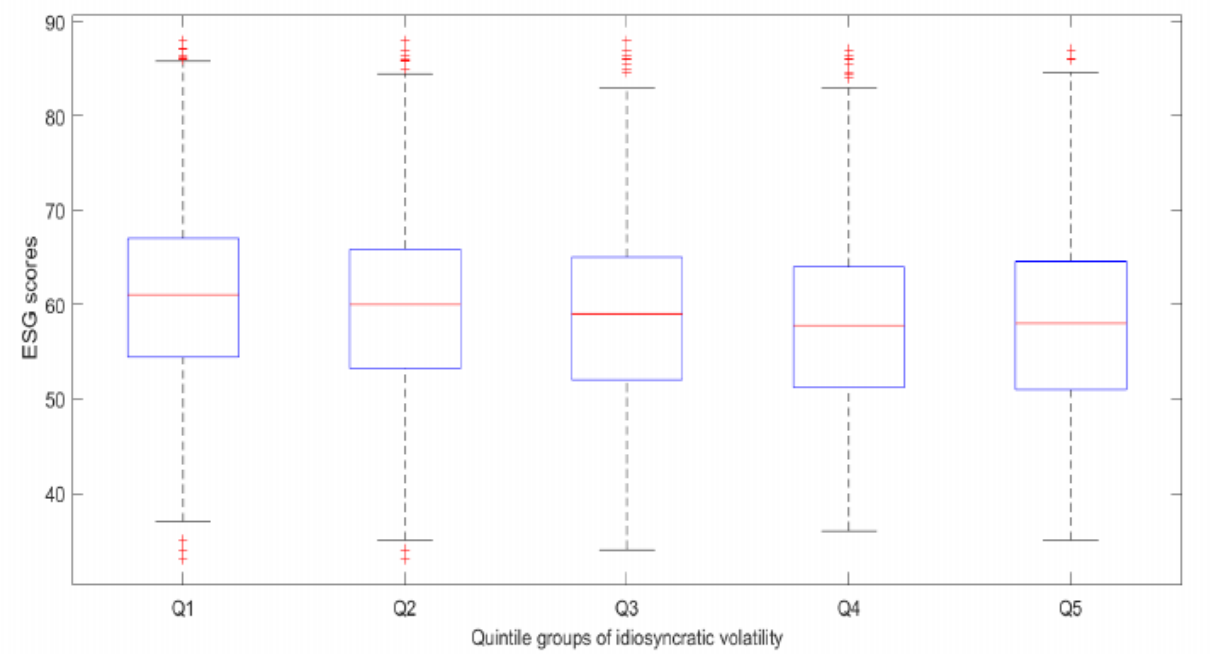


Figure A2.6: ESG ratings by idiosyncratic volatility quintiles: Asset4 (North America)

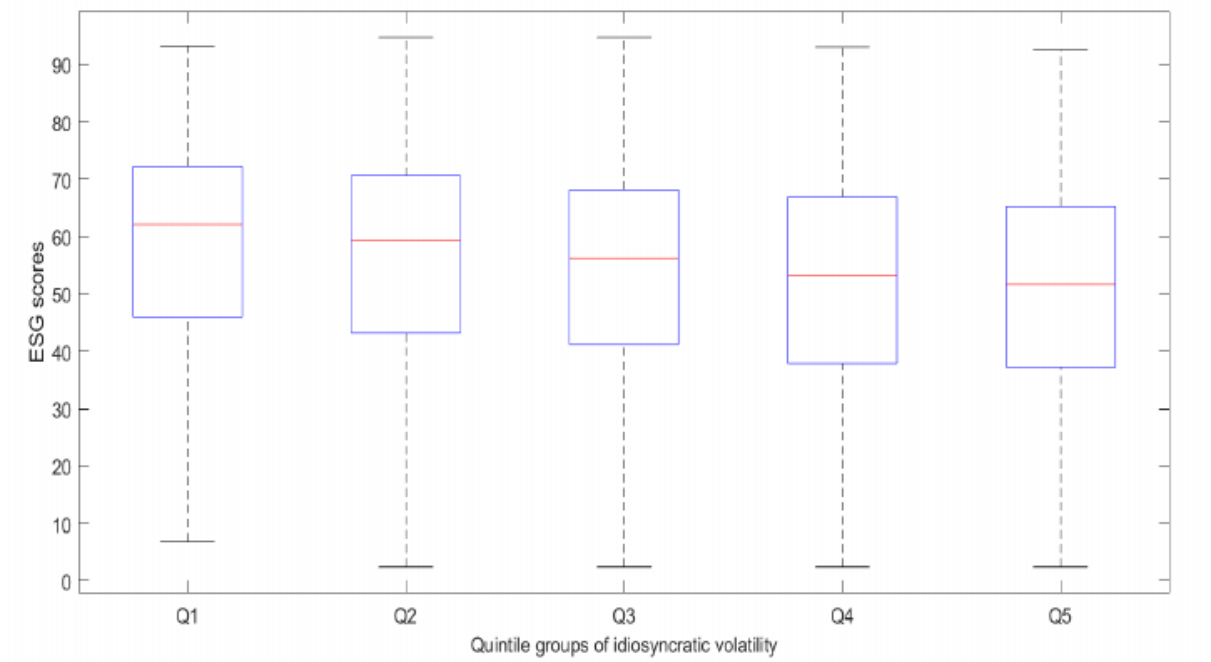


Figure A2.7: ESG ratings by idiosyncratic volatility quintiles: Sustainalytics (Asia-Pacific)

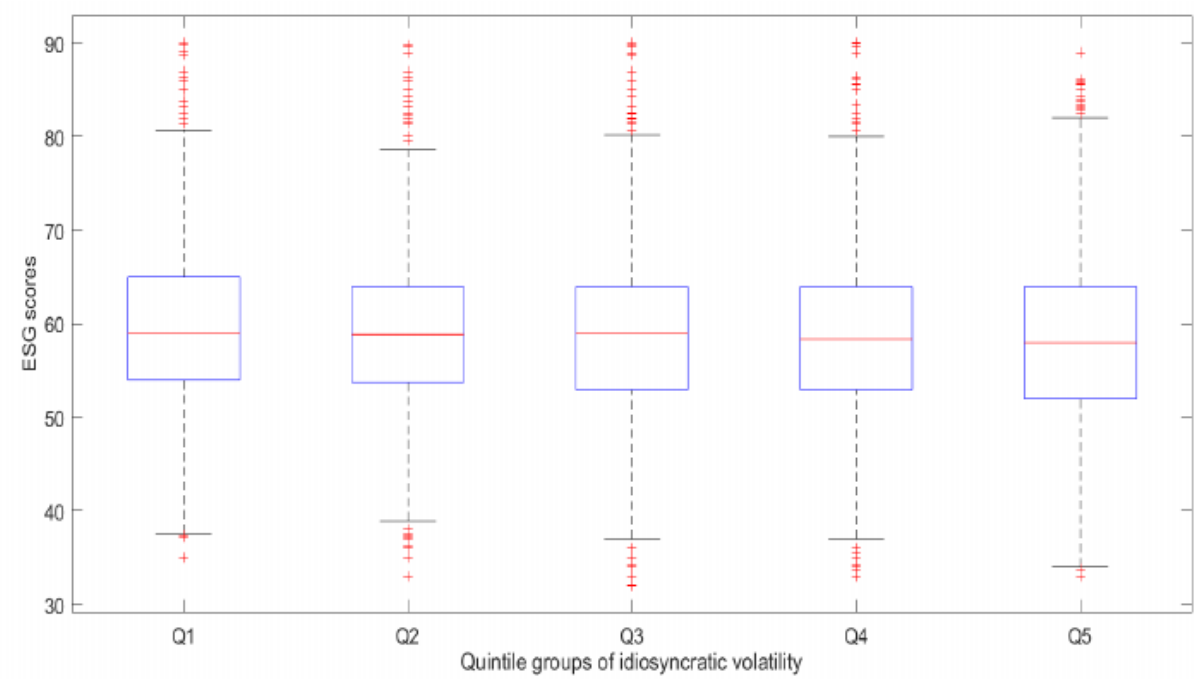


Figure A2.8: ESG ratings by idiosyncratic volatility quintiles: Asset4 (Asia-Pacific)

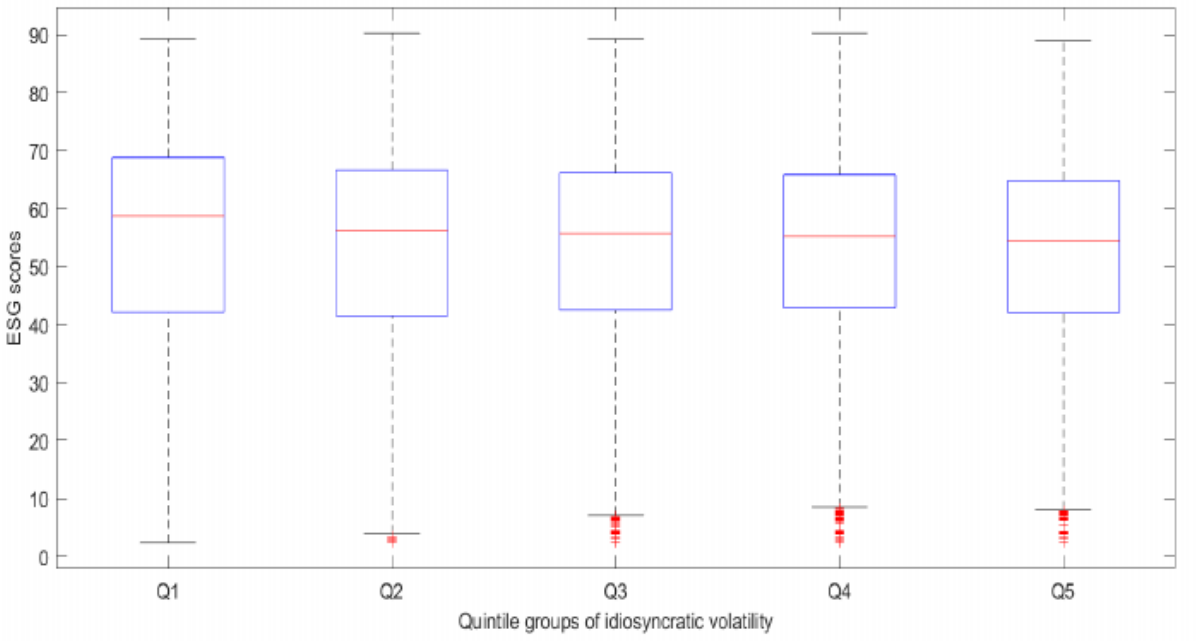


Table A2.1: Backtest of ESG ratings: results for absolute error loss and idiosyncratic returns from CAPM

Sustainalytics					
		ESG	E	S	G
Rolling Window	EU	**−1.6%***** **(−0.01)**	**−2.1%***** **(−0.009)**	**−0.98%***** **(−0.008)**	**−0.38%***** **(−0.005)**
	NA	−0.075% (−0.01)	**−0.43%***** **(−0.008)**	0.17% (−0.006)	0.060% (−0.006)
	AP	**−0.14%**** **(−0.004)**	**−0.39%***** **(−0.004)**	−0.0024% (−0.001)	0.0079% (0.0001)
Fixed Window	EU	**−2.3%***** **(−0.01)**	**−2.7%***** **(−0.009)**	**−1.4%***** **(−0.008)**	**−0.74%***** **(−0.007)**
	NA	0.14% (−0.01)	**−0.56%**** **(−0.01)**	0.52% (−0.008)	0.22% (−0.01)
	AP	**−0.37%**** **(−0.005)**	**−0.44%**** **(−0.004)**	−0.12% (−0.003)	**−0.19%**** **(−0.003)**
Asset 4					
		ESG	E	S	G
Rolling Window	EU	**−2.0%***** **(−0.008)**	**−1.6%***** **(−0.006)**	**−2.3%***** **(−0.007)**	**−0.19%***** **(−0.002)**
	NA	**−0.23%*** **(−0.006)**	**−0.26%**** **(−0.003)**	**−0.29%**** **(−0.005)**	0.10% (−0.002)
	AP	$2.9e-05\%$ $(3e{-}05)$	**−0.069%***** **(−0.0003)**	**−0.17%***** **(−0.0008)**	**−0.062%**** **(0.0006)**
Fixed Window	EU	**−2.5%***** **(−0.008)**	**−1.8%***** **(−0.005)**	**−2.8%***** **(−0.006)**	**−0.46%***** **(−0.003)**
	NA	**−0.81%**** **(−0.008)**	**−0.51%**** **(−0.004)**	**−0.76%**** **(−0.007)**	0.12% (−0.003)
	AP	**−0.089%***** **(−0.0005)**	**−0.18%***** **(−0.0006)**	**−0.25%***** **(−0.0009)**	0.055% (−0.0003)

Notes: This table displays the variation in mean absolute error (MAE) when ESG information is included in the model. Idiosyncratic volatilities are computed using the residual asset returns from CAPM. The datasets contain monthly observations from January 2010 to October 2018, giving a total of 106 months. The North America, Europe and Asia-Pacific datasets include information on respectively $n = 326$, $n = 238$ and $n = 217$ firms. $\star$, $\star\star$ and $\star\star\star$ indicate rejection of the null hypothesis of lack of informational content in ESG ratings at the 10%, 5% and 1% nominal risk levels respectively. Regression coefficients associated to the ESG rating are reported in parentheses.

Table A2.2: Backtest of ESG ratings: results for absolute error loss and idiosyncratic returns from multifactorial model

Sustainalytics					
		ESG	E	S	G
Rolling Window	EU	**−1.8%⋆⋆⋆** **(−0.01)**	**−2.3%⋆⋆⋆** **(−0.01)**	**−1.0%⋆⋆⋆** **(−0.009)**	**−0.44%⋆⋆⋆** **(−0.006)**
	NA	−0.16% (−0.01)	**−0.52%⋆⋆⋆** **(−0.009)**	0.16% (−0.009)	0.037% (−0.009)
	AP	**−0.11%⋆⋆** **(−0.003)**	**−0.45%⋆⋆⋆** **(−0.005)**	0.0095% (0.0002)	0.033% (0.002)
Fixed Window	EU	**−2.3%⋆⋆⋆** **(−0.01)**	**−2.7%⋆⋆⋆** **(−0.01)**	**−1.4%⋆⋆⋆** **(−0.009)**	**−0.81%⋆⋆⋆** **(−0.008)**
	NA	0.034% (−0.02)	**−0.56%⋆⋆** **(−0.01)**	0.49% (−0.01)	0.028% (−0.01)
	AP	**−0.34%⋆⋆** **(−0.005)**	**−0.49%⋆⋆** **(−0.005)**	−0.066% (−0.001)	**−0.090%⋆** **(−0.002)**
Asset 4					
		ESG	E	S	G
Rolling Window	EU	**−2.0%⋆⋆⋆** **(−0.009)**	**−1.7%⋆⋆⋆** **(−0.006)**	**−2.2%⋆⋆⋆** **(−0.007)**	**−0.20%⋆⋆** **(−0.003)**
	NA	**−0.31%⋆** **(−0.008)**	**−0.34%⋆⋆** **(−0.004)**	**−0.29%⋆** **(−0.007)**	0.17% (−0.003)
	AP	0.097% (0.0008)	**−0.022%⋆⋆** **(−9e−05)**	**−0.026%⋆⋆** **(−0.0001)**	**−0.14%⋆⋆⋆** **(0.001)**
Fixed Window	EU	**−2.5%⋆⋆⋆** **(−0.009)**	**−1.8%⋆⋆⋆** **(−0.006)**	**−2.7%⋆⋆⋆** **(−0.007)**	**−0.52%⋆⋆⋆** **(−0.003)**
	NA	**−0.86%⋆⋆** **(−0.01)**	**−0.57%⋆⋆** **(−0.005)**	**−0.68%⋆** **(−0.008)**	0.064% (−0.005)
	AP	0.084% (0.0004)	**−0.11%⋆⋆⋆** **(−0.0003)**	**−0.024%⋆⋆⋆** **(−8e−05)**	**−0.065%⋆⋆** **(0.0004)**

Notes: This table displays the variation in mean absolute error (MAE) when ESG information is included in the model. Idiosyncratic volatilities are computed using the residual asset returns from a multifactorial model. The datasets contain monthly observations from January 2010 to October 2018, giving a total of 106 months. The North America, Europe and Asia-Pacific datasets include information on respectively $n = 326$, $n = 238$ and $n = 217$ firms. $\star$, $\star\star$ and $\star\star\star$ indicate rejection of the null hypothesis of lack of informational content in ESG ratings at the 10%, 5% and 1% nominal risk levels respectively. Regression coefficients associated to the ESG rating are reported in parentheses.

Table A2.3: Consensus vs disagreement between providers using a fixed wind forecasting scheme (MSE)

		Sustainalytics			
		ESG	E	S	G
Consensus	EU	**−8.9%⋆⋆⋆ (−0.02)**	**−6.8%⋆⋆⋆ (−0.009)**	**−6.3%⋆⋆⋆ (−0.01)**	**−5.9%⋆⋆⋆ (−0.01)**
	NA	**−3.0%⋆⋆ (−0.02)**	**−4.8%⋆⋆⋆ (−0.01)**	−0.93% (−0.01)	0.70% (−0.01)
	AP	**−3.0%⋆⋆ (−0.01)**	**−2.8%⋆⋆ (−0.008)**	−0.49% (−0.002)	**−4.4%⋆⋆⋆ (−0.01)**
Disagreement	EU	**−1.8%⋆⋆ (−0.007)**	**−1.8%⋆⋆ (−0.005)**	**−1.4%⋆⋆ (−0.005)**	−0.28% (−0.004)
	NA	−4.2% (−0.02)	−4.7% (−0.01)	−2.5% (−0.01)	−0.97% (−0.004)
	AP	**−0.63%⋆ (0.004)**	0.47% (−0.002)	−0.43% (0.006)	**−0.76%⋆ (0.003)**
		Asset 4			
Consensus	EU	**−4.8%⋆⋆⋆ (−0.007)**	**−5.1%⋆⋆⋆ (−0.006)**	**−5.8%⋆⋆⋆ (−0.006)**	−0.052% (−0.0007)
	NA	**−6.2%⋆⋆⋆ (−0.01)**	**−8.3%⋆⋆⋆ (−0.009)**	**−4.7%⋆⋆⋆ (−0.007)**	−0.24% (−0.004)
	AP	**−3.1%⋆ (−0.007)**	**−3.9%⋆⋆⋆ (−0.005)**	−1.2% (−0.004)	−0.51% (−0.002)
Disagreement	EU	**−1.8%⋆ (−0.005)**	**−1.8%⋆⋆ (−0.002)**	**−2.5%⋆ (−0.006)**	0.84% (−0.003)
	NA	−5.0% (−0.01)	−3.0% (−0.004)	−3.6% (−0.008)	−3.3% (−0.007)
	AP	0.24% (0.002)	−0.38% (0.002)	0.58% (0.001)	**−0.059%⋆ (0.0001)**

Notes: This table displays the variation in mean squared error (MSE) when ESG information is included in the model. Idiosyncratic volatilities are computed using the residual asset returns from a multifactorial model. Results are computed using a fixed window forecasting scheme. For a given universe, the consensus group contains firms with the 25% highest correlations between the ratings of the two providers. The disagreement group contains firms with the 25% lowest correlations. ⋆, ⋆⋆ and ⋆⋆⋆ indicate rejection of the null hypothesis of lack of informational content in ESG ratings at the 10%, 5% and 1% nominal risk levels respectively. Regression coefficients associated to the ESG rating are reported in parentheses.

Table A2.4: Consensus firms: high vs low ESG rating (fixed window)

		Sustainalytics			
		ESG	E	S	G
High ESG	EU	−3.2% (−0.01)	**−7.1%*** (−0.009)**	−1.7% (−0.01)	3.4% (−0.007)
	NA	2.7% (−0.003)	0.66% (−0.0008)	3.8% (−0.003)	0.91% (−0.003)
	AP	**−1.2%** (−0.005)**	**−2.6%** (−0.009)**	1.2% (0.006)	**−2.4%*** (−0.01)**
Low ESG	EU	**−3.6%*** (−0.008)**	**−1.6%*** (−0.002)**	**−1.7%** (−0.005)**	**−4.9%*** (−0.01)**
	NA	**−3.1%** (−0.02)**	**−5.7%*** (−0.01)**	−0.59% (−0.02)	0.70% (−0.01)
	AP	−1.8% (−0.01)	−1.4% (−0.006)	0.71% (−0.007)	**−5.2%*** (−0.01)**
		Asset 4			
		ESG	E	S	G
High ESG	EU	−3.3% (−0.01)	**−1.4%*** (−0.004)**	**−8.6%*** (−0.01)**	0.044% (−0.0002)
	NA	15.% (−0.009)	18.% (−0.01)	**−1.5%*** (0.002)**	13.% (−0.004)
	AP	**−9.4%*** (−0.01)**	**−6.6%*** (−0.01)**	**−2.3%*** (−0.003)**	0.14% (0.0002)
Low ESG	EU	1.0% (0.002)	**−1.6%*** (−0.003)**	0.29% (0.0003)	−1.4% (0.004)
	NA	**−7.6%*** (−0.01)**	**−9.7%*** (−0.009)**	**−6.3%*** (−0.008)**	−0.75% (−0.002)
	AP	0.50% (−0.004)	−1.2% (−0.002)	0.45% (−0.001)	1.0% (−0.002)

Notes: This table displays the variation in mean squared error (MSE) when ESG information is included in the model for consensus firms. Idiosyncratic volatilities are computed using the residual asset returns from a multifactorial model. Results are computed using a fixed window forecasting scheme. High (low) ESG sample represents firms above (below) the median ESG rating. $\star$, $\star\star$ and $\star\star\star$ indicate rejection of the null hypothesis of lack of informational content in ESG ratings at the 10%, 5% and 1% nominal risk levels respectively. Regression coefficients associated to the ESG rating are reported in parentheses.

2 Correlation matrices with factor structure for credit risk exposures[13]

We provide an overview of various specification and calibration techniques to derive correlation matrices associated with factor models. We consider unrestricted and restricted models, where some betas are set to zero, thus leading to specific region or sectoral factors. We address latent factors using maximum likelihood through the EM algorithm and least squares approaches, implemented with the Spectral Projection Gradient algorithm.

The case of endogenous observable factors is dealt with algebraic techniques and OLS regression. PCA approaches are a bridge between these two approaches. PCA appear as optimal regressors on the one hand. On the other, ML estimators in the isotropic model can be derived from modified PCA. Simpler specifications such as block correlation matrices, associated with constant inner and inter-bucket correlations, are also considered. Based on a dataset of equity returns of corporate names included in the major US and European credit indices, we conduct a benchmarking exercise, varying the number of factors.

Besides the fitting performance, parameter parsimony, ease of interpretation and ability to deal with large pairwise correlations are considered. In our experiment, restricted latent factor models supersede competitors.

[13] Co-authors: Jean-Paul Laurent, Sorbonne & Labex RéFi, University Paris 1 Panthéon ; Michael Sestier, Sorbonne & Labex RéFi, University Paris 1 Panthéon.

2.1 Introduction

We consider the estimation of correlation matrices in a credit risk measurement context. The Basel 3 (FRTB DRC) requirements involve using a correlation matrix associated with a probabilistic factor representation, with at least "two types of systematic risk factors". It is stated that "default correlations must be based on credit spreads or on listed equity prices". While this points to latent factor models, that are being dealt with by central bankers and academics, linear regression onto observable factors, typically portfolio returns, are often used to build correlation matrices compliant with regulatory requirements. By benchmarking with various specifications, including factor analysis, linear regression, and Principal Component Analysis (PCA), we conclude that restricted latent models approach supersedes other techniques.

Dealing with covariance/correlation matrices is a standard issue in finance, that we briefly review to help the understanding the specificities of the addressed problem. In the case of multivariate Gaussian distributions, it is well-known that the Maximum Likelihood (ML) estimate of a covariance matrix is closely related to the sample covariance matrix. Even though, it is equally well-known to be related to numerous difficulties. For instance, pairwise deletion usually leads to non-PSD empirical correlation matrices. This issue also arises in a stress-testing context as outlined by Finger (1997). It can be dealt with by a variety of techniques, including computing the nearest correlation matrix with respect to the Frobenius norm through alternating projections, Newton, and hypersphere algorithms (see for example Higham (2002), Rebonato and Jäckel (2000), Qi and Sun (2006, 2010) and the references therein)[14]. While of interest, these methods are not designed to cope with the constraints associated with factor models.

[14] If X is a $n \times m$ matrix, the Frobenius norm of X is such that: $\|X\|_F^2 = \text{trace}(XX^*) = \sum_{i=1}^{n}\sum_{j=1}^{m} X_{ij}^2$. It corresponds to the Euclidian norm of the columns of X stacked in a single vector. Higham (1988) provides the nearest symmetric PSD matrix to a given real matrix with respect to the Frobenius norm: When the initial matrix is symmetric, the nearest symmetric PSD matrix is derived from the spectral decomposition by replacing the negative eigenvalues by zero. However, the resulting diagonal

When the cross-sectional dimension is large, factor models are one way to parsimony: banks typically deal with several thousand corporate counterparties. Our work aims at estimating constrained correlation matrices, not explaining cross-sectional expected returns, even though factor models have been used in these contexts. For instance, expanding upon the single-index model of Sharpe (1963), Fama & French (1993) have considered multivariate linear regressions on portfolio returns, a technique also used in assessing the validity of the CAPM (see MacKinlay (1987), Gibbons et al (1989), Shanken (1992) and the textbook of Campbell, Lo & MacKinlay (1997)). On the other hand, Ross (1976) has considered a (latent) unrestricted factor model in a pricing context. King (1966) is an early example of applying factor analysis technique to security returns (see also Elton et al (2014)). Later, financial economists were faced with the correlation of idiosyncratic terms in the returns, the so-called "approximate factor model" framework. If n denotes the cross-sectional dimension and K the number of common factors, Chamberlain & Rothschild (1983) assume that the largest eigenvalue of the covariance matrix of the returns is bounded as $n \to \infty$ or equivalently that the $K+1$ largest eigenvalue of the covariance matrix is bounded as $n \to \infty$. This can be understood as specific variances becoming negligible compared to communalities, a framework under which principal component analysis, factor analysis and low rank approximations coincide; see Connor & Korajczyk (1986, 1988, 1993), Bai & Ng (2002) and the review paper of Fan, Liao & Liu (2016) for further work on the asymptotic principal component technique.

Factor models are also being used in the context of pricing derivatives in an arbitrage-free framework, especially in Libor Market Models, as above with parsimony in mind. Contrasting the previous framework, the cross-sectional dimension n is fixed, say the number of observable forward rates, the number of latent factors K, such that $K < n$. The rank of the fitted (instantaneous) correlation matrix needs to be equal or lower than K. We are then led to find the closest rank K correlation matrix to a given correlation matrix. While Eckart-Young-

elements may differ from one. Finding the nearest correlation matrix is thus involved and requires optimisation techniques in most cases.

Mirsky theorem provides simple answers for covariance matrices, for a number of matrix forms, the problem is more involved when considering correlation matrices. A number of approaches, geometric (Grubišić and Pietersz (2007), Rapisarda, Brigo & Mercurio (2007), Kercheval (2008) expanding on earlier ideas of Rebonato and Jäckel (2000)), based on Lagrange multipliers method (Zhang and Wu (2003), Wu (2002), on majorization (Pietersz and Groenen (2004)) have been considered. However, this no-arbitrage pricing and rank reduction framework is somehow too parsimonious. It is at odds with sound risk management, since it involves setting specific variances to zero, ironing out basis risks.

Fan, Fan & Lv (2008), Fan, Liao, & Mincheva (2011, 2013) consider observable factor models and control for correlation of residuals by using sparsity constraints on the covariance matrix of residuals. Sparsity constraints can also be put on the precision matrix (the inverse of the covariance matrix) since it appears that a large number of partial correlations are close to zero (Dempster (1972)). Sparse estimation can be achieved by penalized ML with the graphical lasso (Friedman et al (2008)) or by using MTP2 constraints (Agrawal et al (2019)). The use of shrinkage techniques is another approach to large covariance/correlation matrices (see for instance Ledoit & Wolf (2004, 2020). The sample covariance matrix might be shrunk to a single index (one factor) matrix. However, this is at odds with sound credit risk analysis since single factor models typically understate the risk of long-short exposures. In a typical credit risk model, the loss involves default indicators, thus is highly non-linear in the underlying returns. Our purpose is to provide a proper input for correlation matrices of a multivariate Probit model. While assuming latent variables to be normally distributed is standard for modelling corporate credit risk, this readily extends to elliptical distributions. When validating internal risk models, supervisors favour multivariate Student t distributions. As regulatory prescribed, such models are then used to compute values-at-risk over a one-year time horizon. Since the Basel Committee prescribes that “correlations must be based on data covering a period of 10 years” and given that we use weekly stock returns for which the autocorrelations are usually low, our framework will be static and we will consider the marginal covariance/correlation matrix, leaving aside potential dynamics of factors.

Putting equality constraints on factor loadings, such as some being equal to zero and equality among some factor loadings due to homogeneity within buckets of names leads more parsimonious models and complements the penalisation techniques for dealing with scarcity. This corresponds to the confirmatory factor analysis framework put forward by Joreskög (1969a), (see also the textbooks of Bartholomew et al (2011) or Mulaik (2009)). Such restricted factor models, typically involving regional and sectoral factors, are associated with constrained factor loading structures. They have been considered in the context of international dynamic factor models by Banbura, Giannone & Reichlin (2010), Kose, Otrok, & Prasad (2012), Miranda-Agrippino & Rey (2015), among others. In a credit risk context, one can refer to Glasserman and Suchintabandid (2007). Further theoretical examples are provided by Joreskög (1969b)[15]. We further assess and benchmark the usefulness of such restrictions in our context and show quite significant improvement over existing methods.

In the latent factor analysis, one needs a discrepancy criterium between the sample correlation matrix and the factor constrained matrix. This is the case with Gaussian likelihood or (weighted) least-squares (distance associated with Frobenius norms). In most cases, minimization relies on numerical algorithms. Maximum likelihood estimates can be computed through the EM algorithm or based on first-order conditions. As for the least squares criteria, Laurent et al (2016) derived the nearest (from a given input correlation matrix, say the empirical correlation matrix) correlation matrix with two (latent) factors structure. They considered the iterative principal factor[16] and the Spectral Projection Gradient (SPG) respectively considered in a credit risk context by Andersen et al (2003), Jäckel (2005) and Borsdorf et al (2010).

[15] As a side remark Jöreskog (1970) has investigated more general covariance structures. These can be seen as nested factor models or "Russian dolls" as illustrated in a credit risk context by Gregory and Laurent (2004) or Kakushadze (2015).

[16] The iterated principal factor method that relies on iterations of PCA, each one based on the optimality result of Eckart-Young theorem. While PCA betas do not provide the optimal factor loadings, unless the residual principal components are uncorrelated, there are usually not far from optimality, when the number of factors is small and the cross-dimension is large. This usually results in a quick convergence of the algorithm.

In Ross (1976), in the above-mentioned papers and in Pykhtin (2004), Düllmann and Masschelein (2007), Puzanova and Düllmann (2013), there are no constraints on the factor loadings. This corresponds to the unrestricted model, often used in explanatory factor analysis. A common and useful feature of unrestricted and restricted models is that the gradient of the discrepancy function can be easily written. Convergence issues may arise, for instance due to non-convexity or to some binding constraints; for instance, specific variances being equal to zero (so-called Heywood cases). Optimisation can seldomly be achieved through algebraic techniques. For instance, in the isotropic model, where all specific variances are equal, Tipping and Bishop (1999) show how ML estimates of latent factor models can be computed from the spectral decomposition of the correlation matrix.

Lastly, a widely used approach is to derive correlation matrices estimates through linear regressions of returns on a set of portfolio returns (independent variables). If error terms are uncorrelated, one may use ordinary least squares estimation to reconstruct covariance matrices from the estimated betas and the covariance matrix of independent variables. When the error terms are cross-sectionally correlated, ordinary least squares (OLS) do not anymore coincide with maximum likelihood (ML) estimators unless the set of independent variables (i.e., portfolio returns) is the same for all dependent variables (individual name returns) as outlined in the SUR approach of Zellner (1962). This may be the case for unrestricted models, but not anymore for restricted models. Then, generalized least squares (GLS) algorithms should provide better estimates of the betas. In our case, this approach is flawed since independent variables are linear combinations of dependent variables with fixed weights. For instance, CDX or iTraxx credit indices (composing our dataset) are equally weighted. Then, the matrix of residuals is degenerate, GLS and ML approaches cannot be considered[17]. Still

[17] In a linear regression model, specification of error terms is as important as the choice of regressors. In our case, the covariance of error terms is fully determined by the covariance matrix and the portfolio weights.

OLS can be implemented, corresponding to the minimisation of the Frobenius norm of the difference between the design matrix and the matrix of "predicted" returns[18].

In section 2.2, we present restricted and unrestricted latent factor models. In section 2.3, we present the design matrix, related concepts and discuss algorithms to compute the nearest correlation matrix with latent factor structure in restricted and unrestricted cases. Section 2.4 is dedicated to analysing and comparing outcomes of the different approaches.

2.2 Factoring variables driving default

2.2.1 Latent unrestricted factor models

When it comes to sovereigns or large corporates in trading book portfolios, typically involving long and short exposures, credit models predominantly rely on the multivariate probit approach with latent factors.

For instance, Pykhtin (2004) has considered the standard $K-$ factor model, where X_i is the latent variable associated with name $i \in \{1, \dots, n\}$. $X = (X_1, \dots, X_n)^*$ can be written as follows, using the matrix representation of Borsdorf et al. (2010):

$$X = \beta Z + \sigma(\beta)\varepsilon$$

- (Z, ε) is a $K + n -$ standardized Gaussian vector. Z correspond to $K \leq n$ common factors, while ε is associated with residual risks.
- β is a $n \times K$ matrix of factor loadings.
- $\sigma(\beta)$ is a $(n \times n)$ diagonal matrix, such that: $\sigma^2(\beta) = Id_{n\times n} - \text{diag}(\beta\beta^*)$ where β^* denotes the transpose of β and $Id_{n\times n}$ is the identity matrix.

[18] Let us remark that we consider Frobenius norms related to the design matrix on the one hand and Frobenius norms related to the correlation matrix on the other.

This results in a covariance/correlation matrix of X equal to $C(\beta) = \beta\beta^* + \sigma^2(\beta)$. The (diagonal) terms in $\sigma^2(\beta)$ are the specific variances, while the diagonal terms in $\beta\beta^*$ are the communalities. The non-diagonal terms of $\beta\beta^*$ are the pairwise correlations.

Reiersøl (1950), Th. 3.2 or Anderson and Rubin (1956), Th. 4.1 provide a necessary and sufficient condition for the covariance matrix of the X_i to be consistent with a $K-$ factor model.

On top of this, the factor loadings must fulfil the inequalities: $\sum_{k=1}^{K} \beta_{ik}^2 \leq 1$ for $i = 1, \dots, n$. The vector of factor loadings associated with each name must stay within the unit $K-$ dimensional sphere[19]. The set of admissible factor loadings is closed and convex.

The above model can also be written in expanded form:

- $X_i = \sum_{k=1}^{K} \beta_{ik} Z_k + \sigma_i \varepsilon_i$
- $\sigma_i^2 = 1 - \sum_{k=1}^{K} \beta_{ik}^2$

Pykhtin (2004), Düllmann and Masschelein (2007), Puzanova and Düllmann (2013) use the composite factor representation. Let us define ρ_i by $\rho_i = \sqrt{1 - \sigma_i^2}$. Then, we can write the latent factors X_i as:

- $X_i = \rho_i F_i + \sqrt{1 - \rho_i^2}\varepsilon_i$,
- $F_i = \sum_{k=1}^{K} \alpha_{ik} Z_k$
- $\alpha_{ik} = \beta_{ik}/\rho_i$, $k = 1, \dots, K$, $i = 1, \dots, n$ since $\rho_i F_i = \sum_{k=1}^{K} \beta_{ik} Z_k$

[19] If as in Laurent et al (2016), we define $\beta_i = (\beta_{i1} \dots \beta_{iK})$, the row matrix associated with factor loadings of name i, then the constraints can be written as $\beta_i \beta_i^* \leq 1$, $i \in \{1, \dots, n\}$

Let us remark that since $\sum_{k=1}^{K} \alpha_{ik}^2 = 1$, the composite factors F_i have unit variances and are also uncorrelated to the idiosyncratic terms ε_i. Since $X_i = \sum_{k=1}^{K}(\rho_i \times \alpha_{k,i})Z_k + \sqrt{1-\rho_i^2}\varepsilon_i$ for $i \in \{1, \ldots, n\}$, the correlation between the latent variables X_i and X_j is provided by: $\rho_{ij} = \rho_i \times \rho_j \times \left(\sum_{k=1}^{K} \alpha_{ik}\alpha_{jk}\right) = \sum_{k=1}^{K} \beta_{ik}\beta_{jk}$.

2.2.2 Latent factor models, constraints on idiosyncratic risks

The special case where $\sigma^2(\varepsilon)$ is a scalar (or spherical) matrix, i.e., the identity matrix up to a multiplicative constant: $\sigma^2(\varepsilon) = \sigma^2 \times \mathrm{Id}_{n\times n}$ is known as the isotropic model. There is no economic rationale for such a specification. However, it is associated with a simple derivation of the Maximum Likelihood estimator.

2.2.3 Latent restricted factor models, constraints on factor loadings

What has been described so far is the unrestricted factor model. To ease interpretation of factors and minimize the determination and management of unreasonable number of parameters, it is useful to consider restricted factor models (Lawley and Maxwell (1963)), where some factor loadings are set to zero. This approach is part of so-called confirmatory factor analysis.

As an illustration, consistently with standard credit analysis, we can introduce region and sector specific factors in the following way. The latent variable associated with name i in region $C(i)$ and sector $I(i)$ is written as:

$$X_i = \beta_{i,G} \times Z_G + \beta_{i,C(i)} \times Z_{C(i)} + \beta_{i,I(i)} \times Z_{I(i)} + \sqrt{1-\rho_i^2} \times \varepsilon_i$$

Where the global factor Z_G, the country specific factors $Z_{C(i)}$ and the industry specific factors $Z_{I(i)}$ form a set of independent standardized Gaussian variables. As an example, if a name is

bucketed to the financial sector, the factor loadings associated with the latent variables associated with other sectors, say corporate non-financials and sovereigns will be set to zero. Such specification was introduced by Joreskög (1969b) and used by Glasserman and Suchintabandid (2007) in a credit risk context.

For X_i to have unit variance, the following constraints $\beta_{i,G}^2 + \beta_{i,C(i)}^2 + \beta_{i,I(i)}^2 = \rho_i^2$ must be met. As a consequence, the composite factor F_i is such that: $\rho_i \times F_i = \beta_{i,G} \times Z_G + \beta_{i,C(i)} \times Z_{C(i)} + \beta_{i,I(i)} \times Z_{I(i)}$.

The correlation structure of the latent factors X_i and X_j, assuming a single global factor and that any name is bucketed to a single sector and region is then given by:

$$\rho_{i,j} = \beta_{i,G}\beta_{j,G} + 1_{C(i)=C(j)}\beta_{i,C(i)}\beta_{j,C(j)} + 1_{I(i)=I(j)}\beta_{i,I(i)}\beta_{j,I(j)}$$

for $i \neq j$ and subject to constraints: $\beta_{i,G}^2 + \beta_{i,C(i)}^2 + \beta_{i,I(i)}^2 \leq 1$ for all $i \in \{1, \dots, n\}$.

This is to be compared with the unrestricted three factor model:

$$X_i = \beta_{i,1} \times Z_1 + \beta_{i,2} \times Z_2 + \beta_{i,3} \times Z_3 + \sqrt{1 - \rho_i^2} \times \varepsilon_i$$

with $\rho_i^2 = \beta_{i,1}^2 + \beta_{i,2}^2 + \beta_{i,3}^2$ and (for $i \neq j$):

$$\rho_{i,j} = \beta_{i,1}\beta_{j,1} + \beta_{i,2}\beta_{j,2} + \beta_{i,3}\beta_{j,3}$$

The restricted and unrestricted factor models above involve the same number of parameters, i.e., $3 \times n$.

2.2.4 Latent factor models, constraints on the correlation matrix

Practitioners often postulates that correlation matrices can be partitioned by blocks, with constant inner and outer pairwise correlations, denoted as block CS (compound symmetry) or block homogeneous matrices (see below with three buckets):

1	0.4	0.4	0.4	0.1	0.1	0.1	0.1	0.1
0.4	1	0.4	0.4	0.1	0.1	0.1	0.1	0.1
0.4	0.4	1	0.4	0.1	0.1	0.1	0.1	0.1
0.4	0.4	0.4	1	0.1	0.1	0.1	0.1	0.1
0.1	0.1	0.1	0.1	1	0.3	0.3	0.15	0.15
0.1	0.1	0.1	0.1	0.3	1	0.3	0.15	0.15
0.1	0.1	0.1	0.1	0.3	0.3	1	0.15	0.15
0.1	0.1	0.1	0.1	0.15	0.15	0.15	1	0.2
0.1	0.1	0.1	0.1	0.15	0.15	0.15	0.2	1

Figure 2.1 Example of block correlation matrix

Whether such matrices are PSD can simply be answered. Let us consider the matrix with diagonal terms equal to the inner bucket correlations (here 0.4 0.3 0.2) and the non-diagonal terms to the inter bucket correlations (inter bucket correlations are also known as peripherical correlations). This results in a (3×3) matrix: $\begin{pmatrix} 0.4 & 0.1 & 0.1 \\ 0.1 & 0.3 & 0.15 \\ 0.1 & 0.15 & 0.2 \end{pmatrix}$. Such a transformation is known as the block average map.

Roustant and Deville (2017) have shown that the block CS matrix is PSD if and only if the matrix resulting from the block average map (also known as the group level correlation matrix) is PSD. As a side consequence, it caps the inter bucket correlations to the geometric means of the inter bucket correlations.

Demey et al (2004), Huang & Yang (2010) provide a factor representation for Gaussian distributions with the above block structure. Let us denote by O the $(K \times K)$ group level correlation matrix. Let us assume that it is PSD. Let us denote by G the $(K \times K)$ matrix such

that $GG^* = \mathrm{O}$. G can be derived from Cholewski decomposition or spectral analysis of O. Let $Z_m^{(k)}$ be the $m-$th random variable from bucket k. Let us consider independent unitary Gaussian variables $W_1, \dots, W_K, \varepsilon_1^{(1)}, \dots, \varepsilon_{m(1)}^{(1)}, \dots, \varepsilon_1^{(K)}, \dots, \varepsilon_{m(K)}^{(K)}$, where $m(k)$ is the number of names in bucket k. Then,

$$Z_m^{(k)} = \sum_{j=1}^{K} G_{kj} W_j + \sqrt{1-\rho_k} \times \varepsilon_m^{(k)}$$

where ρ_k is the intra bucket k correlation, has the stated block correlation structure. Let us remark that the group level correlation matrix is the covariance matrix of the common factors. Let $\bar{Z}_k$ be such that $\sqrt{\rho_k}\bar{Z}_k = \sum_{j=1}^{K} G_{kj} W_j$. Then, $Z_m^{(k)} = \sqrt{\rho_k}\bar{Z}_k + \sqrt{1-\rho_k} \times \varepsilon_m^{(k)}$. $\bar{Z}_k$ can be seen as a composite – bucket specific factor.

2.2.5 Endogenous observable factors, unrestricted and restricted factor models

In unrestricted and restricted latent factor models, the Gaussian vector (X, Z) is non degenerated. By contrast, we consider models where factors Z are linear functions of the X. These so-called composite variables are weighted averages of the components of X, for instance principal components or equally weighted averages of components of X. Since X is associated with financial returns, the factors Z are both observable and endogenous. As above, X denotes a $n-$dimensional Gaussian vector with standardized returns. The covariance/correlation matrix of X is denoted by C is and assumed to be of full rank. We define a $K-$ vector of factors, $K < n$, as follows:

$$Z = Q^* X$$

where Q is a $n \times K$ matrix of "portfolio weights"[20], with $\operatorname{rank} Q = K$. Q may be based on financial ideas: for instance, credit indices are usually equally weighted and associated to industry, region and/or credit quality. We can also consider principal components of X.

Appendix 2 adapted from Rao (1964, 1973), see also Basilevsky (2009), Theorem 3.9 (*i*) or Jolliffe (2010, Sect. 2.1, Property A.5) discusses that point from a correlation matrix calibration point of view.

We then write:

$$X = \beta Z + \varepsilon$$

where β is a $n \times K$ matrix of Betas (or factor loadings). As for latent variable models, the factors Z and the residual or idiosyncratic risks ε are assumed to be independent. Under the assumption that X is a Gaussian vector, $(Z, \varepsilon) = (Q^* X, X - \beta Q^* X)$ is a Gaussian vector too. From $E[\varepsilon Z^*] = 0$, $\beta = E[XZ^*](E[ZZ^*])^{-1}$ as in the standard linear regression model. We end-up with:

$$\beta = CQ(Q^* CQ)^{-1}$$

Where C is the covariance/correlation matrix of X. While the derivation of factor loadings under latent factor model is usually involved, using endogenous observable factors calls on simple linear algebra.

Let us remark that we cannot deal with a strict factor model. Since $\varepsilon = X - \beta Z$, $Q^* \varepsilon = Q^* X - Q^* \beta Q^* X$. Since $Q^* \beta Q^* = Q^*$, then $Q^* \varepsilon = 0$. Therefore, the residual terms cannot be independent.

[20] The columns of Q might not sum-up to one, thus the Z need to be rescaled to lead to portfolio returns. This does not alter projections on subspaces.

Still, the approximate factor model framework can be achieved. Let us for instance consider the case of constant pairwise correlations $\rho, -\frac{1}{n-1} < \rho < 1$[21]. The vector $e \in \mathbb{R}^n$, such that $e^* = (1, \dots, 1)$ is the first eigenvector of C. The first eigenvalue equals to $1 + (n-1)\rho$, while the other eigenvalues are equal to $1 - \rho$. Since they are bounded, we are indeed the framework of an approximate single factor model.

On the other hand, for $\rho \geq 0$, the constant correlation model is associated with the probabilistic latent factor representation, $X_i = \sqrt{\rho}Z + \sqrt{1-\rho}\varepsilon_i, \quad i = 1, \dots, n,$ with $(Z, \varepsilon_1, \dots, \varepsilon_n) \sim N\big(0, Id_{(n+1)\times(n+1)}\big)$, which corresponds to a strict factor model. This shows that the same correlation matrix might be associated either with a latent factor or an endogenous factor structure.

What has been described so far is an unrestricted factor model: factors Z are common to all $X_i, \ i = 1, \dots, n$ or equivalently no β_{ik} is set to zero. To ease the determination and management of an unreasonable number of parameters, it is useful to consider restricted factor models. This is indeed common practice, where the factors are typically sector or region specific. To parallel the exposition of the latent variable case, X_i associated with name i in region $C(i)$ is written as:

$$X_i = \beta_{i,C(i)} Z_{C(i)} + \sqrt{1-\rho_i^2} \times \varepsilon_i$$

Where $Z_{C(i)}$ denote the country specific factor. In a credit risk context $Z_{C(i)}$ would be an equally weighted average of returns associated with region $C(i)$. Let us denote by $C(i)$ the set of names belonging to the same region and sector as name i. Then, proceeding as above, we can still write $E\big[\varepsilon_j | Z_{C(i)}\big] = 0, \forall j \in C(i)$, which leads to determining the betas associated

[21] The constant correlation model is also the benchmark model in credit risk modelling (i.e. the Basel 2 framework and the Vasicek (2002) model) and in the pricing of CDO tranches (Laurent & Gregory (2005)). It is also considered by Ledoit and Wolf (2004) as the simplest, still useful shrinkage target, see also Elton and Gruber (1973).

with the relevant factors for bucket $C(i)$ for all names within that bucket in the same way as for the unrestricted case.

Let us remark that the residual risks ε_i are uncorrelated only with the factors associated with a specific bucket $C(i)$. To ease the exposition, let us assume that the Gaussian vector $Z_{C(i)}, Z_{C(j)}$ is non-degenerated, i.e., we deal with different factors. We can proceed with orthogonality restrictions between ε_i and $Z_{C(i)}, Z_{C(j)}$ to uniquely derive some betas such that $X_i = \bar{\beta}_{i,C(i)} Z_{C(i)} + \bar{\beta}_{i,C(j)} Z_{C(j)} + \sqrt{1 - \bar{\rho}_i^2} \times \bar{\varepsilon}_i$, and $E\left[\bar{\varepsilon}_i | Z_{C(i)}, Z_{C(j)}\right] = 0$.

Then, $E\left[\varepsilon_i | Z_{C(i)}, Z_{C(j)}\right] = E\left[X_i - \beta_{i,C(i)} Z_{C(i)} | Z_{C(i)}, Z_{C(j)}\right] = \left(\bar{\beta}_{i,C(i)} - \beta_{i,C(i)}\right) Z_{C(i)} + \bar{\beta}_{i,C(j)} Z_{C(j)}$. For the conditional expectation to be equal to zero, we must have $\bar{\beta}_{i,C(j)} = 0$ (and $\bar{\beta}_{i,C(i)} = \beta_{i,C(i)}$). Conversely, let us assume that $\bar{\beta}_{i,C(j)} = 0$. Then, $E\left[\varepsilon_i | Z_{C(i)}, Z_{C(j)}\right] = \left(\bar{\beta}_{i,C(i)} - \beta_{i,C(i)}\right) Z_{C(i)}$. Since $E\left[E\left[\varepsilon_i | Z_{C(i)}, Z_{C(j)}\right] | Z_{C(i)}\right] = E\left[\varepsilon_i | Z_{C(i)}\right] = \left(\bar{\beta}_{i,C(i)} - \beta_{i,C(i)}\right) Z_{C(i)} = 0$, we conclude that $\bar{\beta}_{i,C(i)} = \beta_{i,C(i)}$.

Therefore, when restricting some betas to zero, the residual terms cannot be independent from all the factors, unless the unrestricted model is fully consistent with the restricted one. Consequently, labelling a restricted model as a "factor model" is messy: in most cases, the residuals will be correlated with some of the factors.

2.3 Calibration

The analysis conducted in the previous setting is probabilistic and/or relates to "population" rather than "sample". Let now turn to data. The calibration dataset will involve a rectangular matrix (or design matrix) of either equity or credit default swap returns, as required by the Basel Committee (the design matrix) R. In the sequel, we only rely on equities, encompassing more names and with deeper historical data.

- n number of names
- T number of dates
- $R = (R_{ti})$, $T \times n$ matrix of demeaned and standardized historical returns where $i \in \{1, \dots, n\}$ is the number of equities to be considered and $t \in \{1, \dots, T\}$ the number of dates.
- R^*, the transpose of R.
- $\hat{C} = \frac{1}{T} \times R^* R$, $n \times n$ the empirical (sample) correlation matrix of the returns.

For ease of exposition, we assume that there are no missing data in R. For simplicity, we assume that the returns at different dates have been sampled independently from the same multivariate Gaussian distribution. Since we will not further discuss sampling issues, we will further denote the empirical correlation matrix by C to ease notations.

We remind (cf. Appendix for details) that the singular value decomposition (SVD) theorem states that the rectangular matrix of returns can be written as: $R = U\Sigma V^*$

Where:

- U is a $T \times T$ unitary matrix: $UU^* = U^*U = \mathrm{Id}_{T\times T}$,
- Σ is a $T \times n$ rectangular diagonal matrix. The diagonal terms of Σ are the singular values, the number of them is $\min(n, T)$.
- V is a $n \times n$ unitary matrix: $VV^* = V^*V = Id_{n\times n}$.

The SVD decomposition can be written in extended form as: $R = \sum_{k=1}^{\min(n,T)} \sigma_k u_k v_k^*$, where the σ_k are the singular values sorted by decreasing order, u_k and v_k are the $k-$th column of U and V, respectively. If $\sigma_1 > \sigma_2 > \cdots > \sigma_{\min(n,T)}$, the previous decomposition is unique.

The empirical correlation matrix C is provided by: $C = \frac{1}{T} \times R^* R = \frac{1}{T} \times V\Sigma^*\Sigma V^*$

The diagonal terms of the $(n \times n)$ diagonal matrix $\frac{1}{T} \times \Sigma^* \Sigma = \Lambda = (\lambda_k)$ are the eigenvalues of C $(\lambda_1 \geq \cdots \geq \lambda_n)$ and, up to a multiplicative $\frac{1}{T}$ factor, the squared singular values. Due to the assumption of no missing data, the eigenvalues of the empirical correlation matrix C are non-negative (i.e. the empirical correlation matrix is positive semidefinite).

The n columns of V (right-singular vectors of R) are the eigenvectors of C.

Remark 1: The rank of C (or equivalently of R) is at most $\min(T, n)$. If we were to consider one-year non-overlapping returns (based on the DRC one-year liquidity horizon) and using 10 years of data, we get $T = 10$. Depending on data frequency, T may range from 10 (10Y of non-overlapping yearly returns) to 2500 (10Y of daily returns). The number of names n can range from say, 2500 up to 15000. As an example, take $n = 15000$ and $T = 250$. We then deal with large n, (rather) small $\frac{T}{n}$ (the asymptotics of random matrix theory (n and T both large, with a controlled ratio $\frac{T}{n}$) as opposed to the standard asymptotics (fixed n, $T \to \infty$), the number of zero eigenvalues of $\hat{\mathrm{C}}$ is at least $\min(n - T, 0) = 14750$.

We are now to consider various calibration criteria to determine correlation matrices, consistent with the above factor specifications. As for latent factor models, two families of discrepancy functions between the empirical correlation matrix and the target correlation matrix have been considered so far. The first one involves Frobenius norms, while the other relates to maximum likelihood estimation. Both approaches involve the minimization of some discrepancy function.

2.3.1 Nearest correlation matrix with $K -$ factor structure.

There several algorithms aiming at providing the nearest (with respect to Frobenius norm) correlation matrix with $K -$ factor structure.

This approach is also known as "least squares estimation" in factor analysis[22], see Mulaik (2009, 8.3.1), Bartholomew et al (2011, 3.8). This involves minimizing (over β, Ψ) $\|C - \beta\beta^* - \Psi\|_F^2$, where Ψ is a diagonal matrix with non-negative entries. Mulaik (2009, 15.3.9) or Bartholomew et al (2011, 3.8) provide the gradient of that criteria.

The approach aiming at minimizing $\|C - C(\beta)\|_F^2$, where $\| \ \|_F$ stands for the Frobenius norm and $C(\beta) = \beta\beta^* + Id_{n\times n} - \text{diag}(\beta\beta^*)$ with $\text{diag}(\beta\beta^*) \leq Id_{n\times n}$ is known as the minres problem (Harman and Jones (1966)). Let us remark that $\|C - \beta\beta^* - \Psi\|_F^2 \geq \|C - C(\beta)\|_F^2$, with equality when $\Psi = Id_{n\times n} - \text{diag}(\beta\beta^*)$. The right-hand term of the inequality is a majorization function (see Borg and Groenen (2005)) and as outlined by Jöreskog (2003), see also Harman (1976, Subsection 9.2), Comrey and Lee (2013, Subsection 4.2), the two minimization problems are equivalent.

Besides, since the set of $C(\beta)$ matrices is closed and convex, the projection (here based on the Frobenius norm) is well defined (existence and uniqueness). The iterated principal factor is a well-known approach to the least squares problem, see Rencher and Christensen (2012), 13.3.2 and 13.3.3. It is also known as "principal axis method", see Gentle (2009, chapter 10). This is an alternating projection algorithm, iterating between optimal rank reduction and projection onto the set of matrices with diagonal elements equal to one[23]. Starting from an initial estimate Ψ^0, β^0 is computed from the low-rank decomposition of $C - \Psi$, i.e. β^0 is the $n \times K$ matrix with columns associated with the first K eigenvectors of $C - \Psi$. Ψ^1 is then set as $\text{diag}(Id_{n\times n} - \beta^0\beta^{0*})$, so that $\beta^0\beta^{0*} + \text{diag}(Id_{n\times n} - \beta^0\beta^{0*})$ is a symmetric matrix with diagonal elements equal to one ($\beta^0\beta^{0*} + \text{diag}(Id_{n\times n} - \beta^0\beta^{0*})$ will be a correlation matrix if $\text{diag}(\beta^0\beta^{0*}) \leq Id_{n\times n}$) and so on.

[22] Not to be confused with the OLS estimation methods, which is related to the residual returns and not to the correlation matrix.

[23] Morini and Webber (2006) also consider an alternating projection algorithm. While the principal axis method involves the low rank approximation of $C - \Psi$, these authors the low rank approximation of C. In terms of the minres problem, Morini and Webber (2006) consider the low rank approximation of $\beta\beta^* + Id_{(n\times n)} - \text{diag}(\beta\beta^*)$, contrasting with $C - Id_{(n\times n)} + \text{diag}(\beta\beta^*)$. Thus, they find that their method does not lead to an optimum. Their approach can be thought of as an iterated PCA, with additive normalization, see below.

Zhang and Wu (2003) connected the low rank correlation problem to least squares estimation of factor models through a Lagrange multiplier approach with an application to Libor Market Models. Andersen, Sidenius and Basu (2003) adapted the idea of iterated low rank approximations to the minres problem in a credit risk context. Let us assume that $C = \beta\beta^* + Id_{(n\times n)} - \text{diag}(\beta\beta^*)$. Then $\beta\beta^*$ is the rank K approximation of $C - Id_{(n\times n)} + \text{diag}(\beta\beta^*)$. To solve this fixed-point problem, one can start from a guessed β, say v (the first K eigenvectors of C), compute the low rank approximation of $C - Id_{(nxn)} + \text{diag}(vv^*)$, derive updated β and so-on. This method might not converge, for instance when the variance associated with the factors (the communalities) becomes larger than the variance of the returns. This corresponds to so-called Heywood cases. In most practical cases, this constraint is not binding and the principal factor method usually only requires a few iterations of PCA to provide the nearest (with respect to Frobenius norm) correlation matrix.

The gradient of $f(\beta) = \|C - C(\beta)\|_F^2$ can be written as $\nabla f(\beta) = 4 \times (\beta\beta^* \times \beta - C\beta + \beta - \text{diag}(\beta\beta^*)\beta) = -4\big(C - C(\beta)\big)\beta$. Borsdorf et al (2010) investigate several approaches, including Andersen, Sidenius and Basu (2003), Newton algorithm and the Spectral Projection Gradient (SPG) approach of Birgin, Martínez & Raydan (2000). As in Laurent et al (2016), they find good performance of the iterated principal factor, when applied to financial data, compared to other methods. As further discussed, our paper confirms previous findings.

When switching from unrestricted to restricted factor model, the iterated principal factor method can no longer be used. However, Newton and SPG algorithms are still readily applicable. The textbooks of Mulaik (2009) or Bartholomew et al (2011) show that the first order derivatives can be computed in matrix form with respect to individual factor loadings and specific variances. In the same vein, we can consider the coordinates of the $\nabla f(\beta) = -4\big(C - C(\beta)\big)\beta$ associated with the unconstrained betas.

Let us remark that the criteria to be minimized $\beta \to \|C - C(\beta)\|_F^2$ is not convex. Therefore, the SPG algorithm can converge to a local minimum. The initial value is thus of importance. We rely on Marsaglia and Olkin (1984) for randomly generating correlation matrices, then computing inception betas as the first principal components, to assess existence of local minima.

2.3.2 Maximum likelihood estimation

Under the assumption of joint normality, the derivation of the maximum likelihood estimators is quite standard. The log-likelihood involves a discrepancy function between the empirical correlation matrix C and $C(\beta)$: $-\text{trace}(C(\beta)^{-1}C) + \log|C(\beta)^{-1}C|$.

When it comes to iterative algorithms (iterated principal factors, spectral projection gradient, EM algorithm) with the aim of computing optimal factor loadings, β, an initial guess needs to be done. This might be the first K −principal components of the correlation matrix or of the correlation matrix minus the diagonal matrix of well-chosen specific variances (often used in the iterated principal factor method). Since both least-squares $\beta \to \|C - C(\beta)\|_F^2$ and maximum likelihood $\beta \to \text{trace}[C^{-1}(\beta)C] - \ln|C^{-1}(\beta)C|$ criteria are not convex, SPG algorithm and EM algorithm may lead to local optima. As for least squares approaches, we rely on Marsaglia and Olkin (1984) to compute random initial betas, to assess the existence of local minima.

Maximum likelihood estimation for isotropic model

This approach back to Young (1941) and Whittle (1952) who derive the factor loadings assuming specific variances are known. Bartlett (1950, 1951) provides a significance test for the number of principal components to retain. Lawley (1953), Anderson and Rubin (1956)

connect spectral decomposition of the empirical covariance/correlation matrix and first order conditions when maximising likelihood. Tipping and Bishop (1999) show that these do correspond to the maximum of the likelihood function. Besides truncation, the isotropic model also involves random matrix and shrinkage ideas: one potential issue with non-constant specific variances is that (the variance of) some idiosyncratic noises can (insignificantly) be close to zero. One way to better conditioning the correlation matrix is to consider homoscedastic (or isotropic) idiosyncratic noises (in factor analysis terminology, constant specific variances).

Factor representation

The probabilistic factor representation will be the starting point and we will end-up with a calibration to an empirical correlation matrix.

Let us consider the following specification of a n – dimensional random vector of standardized returns X: $X = vZ + \varepsilon$, where Z is a K – dimensional Gaussian vector ($K < n$), v is a $(n \times K)$ matrix of factor loadings, ε some idiosyncratic noise vector. ε and Z are independent and have zero mean. The covariance matrix of ε is diagonal and denoted by $\sigma^2(\varepsilon)$. The $K \times K$ covariance matrix of Z is diagonal: $\widetilde{\Lambda}_K = \left(\tilde{\lambda}_k\right)$ with non-negative diagonal terms $\tilde{\lambda}_k, k = 1, \dots, K$. The $(n \times K)$ matrix v can be such that $v^* v = \mathrm{Id}_{K \times K}$.

Remark 1: The above representation of X is not unique: Take any invertible $K \times K$ matrix Q, $X = vQ^{-1}(QZ) + \varepsilon$ provides another equivalent representation.

Remark 2: Thanks to the spectral decomposition of the covariance matrix associated with the random vector vZ (just think of a PCA), we can indeed choose Z to have orthogonal components and $v^* v = \mathrm{Id}_{K \times K}$.

Spectral decomposition

We assume that $\sigma^2(\varepsilon)$ is a scalar (or spherical) matrix, i.e., the identity matrix up to a multiplicative constant: $\sigma^2(\varepsilon) = \sigma^2 \times \mathrm{Id}_{n\times n}$. Therefore, the $n \times n$ covariance matrix of Z is provided by:

$$\Omega = v\widetilde{\Lambda}_K v^* + \sigma^2 \times \mathrm{Id}_{n\times n}$$

The K columns of v are eigenvectors of Ω and the corresponding eigenvalues $\tilde{\lambda}_k + \sigma^2$. Any n – dimensional vector orthogonal to the space generated by the columns of v (i.e., belonging to the kernel of $v\widetilde{\Lambda}_K v^*$) is an eigenvector with corresponding eigenvalue equal to σ^2. Actually, $v\widetilde{\Lambda}_K v^* + \sigma^2 \times \mathrm{Id}_{n\times n}$ is the spectral decomposition of the covariance matrix Ω.

Remark 3: If $\sigma^2 > 0$, the covariance matrix of X is of full rank. As all eigenvalues are bounded below by σ^2, we avoid some of the issues dealt with random matrix theory, i.e., very small (fake) eigenvalues for large n. On the other hand, if $\sigma^2 = 0$, $\mathrm{rank}(\Omega) \leq K$, which leads to low rank covariance matrices.

Calibration to the empirical correlation matrix C

We first build an approximate covariance matrix Ω and then normalize it to a correlation matrix. Computation of approximate covariance matrix Ω:

- Compute the first K eigenvectors of C, v (corresponding to the first K columns of V); Ω and C share the same K first eigenvectors.
- Determining σ^2: $\sigma^2 = \frac{1}{n-K}\sum_{K+1}^{n} \lambda_k$. We replace the $n - K$ smaller eigenvalues of C by their average.

- Set $\tilde{\lambda}_k = \lambda_k - \sigma^2$, $k = 1, \dots, K$, where λ_k is the k^{th} largest eigenvalue of C. Therefore, the corresponding first K eigenvalues of Ω and $\hat{\mathrm{C}}$ are equal.

Remark 4: Ω has the same first $K-$ eigen vectors as C and the same eigenvectors as $\tilde{C}_2$. The eigenvalues associated with these eigenvectors are the same as those of $\hat{\mathrm{C}}$, i.e. λ_k, $k = 1, \dots, K$.

Remark 5: σ^2 is such that the sum of the eigenvalues of Ω and $\hat{\mathrm{C}}$ are equal. Since the sum of the eigenvalues of a square matrix equals its trace and the trace of $\hat{\mathrm{C}}$ equals n, the trace of Ω is also equal to n. Unless all diagonal elements of Ω are equal to one, some elements will be larger than one. The additive normalizing approach, i.e., considering $\Omega + \mathrm{diag}(\mathrm{Id}_{n\times n} - \Omega)$ is problematic. Since some diagonal elements of Ω will be larger than one, some elements of $\mathrm{diag}(\mathrm{Id}_{n\times n} - \Omega)$ will not be positive. Therefore, we cannot readily provide a probabilistic factor representation.

Thus, to end the building, we need transform Ω to a correlation matrix through some multiplicative scaling, i.e., through the eigenvalue method of Rousseeuw & Molenberghs (1993). Let us denote by ω_{ii} the diagonal elements of Ω. The matrix:

$$\tilde{C}_3 = \mathrm{diag}\left((\omega_{ii})^{-1/2}\right)\Omega\,\mathrm{diag}\left((\omega_{ii})^{-1/2}\right)$$

is the correlation matrix associated with Ω.

Simulation

As for simulation, the random vector:

$$X = \mathrm{diag}\left((\omega_{ii})^{-1/2}\right) \times (vZ + \varepsilon)$$

has a correlation matrix equal to $\tilde{C}_3$.

This provides a third way to compute an approximate correlation matrix with underlying factor structure.

Principal Component Method

The "principal component method" to factor analysis back to Pearson (1901) and Hotelling (1933), see Gentle (2009), Rencher and Christensen (2012)).

Approximate correlation matrix

We first consider the optimal rank K approximation of C, $v\Lambda_K v^*$[24]. We then add the diagonal matrix $\mathrm{diag}(\mathrm{Id}_{n\times n} - v\Lambda_K v^*)$ to $v\Lambda_K v^*$, thus setting diagonal terms to one. The resulting approximate correlation matrix can be written as:

$$\tilde{C}_2 = v\Lambda_K v^* + \mathrm{diag}(\mathrm{Id}_{n\times n} - v\Lambda_K v^*)$$

Let us remark that the diagonal terms of $\mathrm{diag}(\mathrm{Id}_{n\times n} - v\Lambda_K v^*)$ are non-negative. To check this, just expand the spectral decomposition of C (i.e., $\hat{\mathrm{C}}$ can be associated with a $n-$ factor model). The diagonal terms of $v\Lambda_K v^*$ correspond to the variance associated with the K first factors, while the diagonal terms of $\mathrm{Id}_{n\times n} - v\Lambda_K v^*$ correspond to the variance associated with the remaining $n - K$ factors.

[24] K will be subsequently the number of factors. Determining the number of factors based on data and statistical analysis is quite standard but uneasy. In a financial context, this is prescribed by law. The FRTB requires that the number of factors accounts for two types of systematic risk factors and FAQ3 suggests a factor model with $M + N$ factors (M regions + N industries). If $M + N = 10$, say, the factor structure requirement is not binding and the idiosyncratic risk factor loadings (γ_i in FRTB FAQ wording) are equal to zero.

We denote by $\sigma_i^2 = 1 - \widetilde{\omega}_{ii}$ the generic term of

$\text{diag}(Id_{n\times n} - v\Lambda_K v^*)$. They are usually denoted as "specific variances" while $\widetilde{\omega}_{ii}$ the variance associated with the factors is the "communality". The off-diagonal terms of $v\Lambda_K v^*$ are the approximate pairwise correlations.

Probabilistic factor representation

Let (Z, ε) be a $(K + n)$ – dimensional random vector with independent zero mean and unit variance components. The K – dimensional random vector Z will correspond to latent factors and the n – dimensional random vector ε to idiosyncratic noises. Let us consider the n – dimensional random vector:

$$X = v\Lambda_K^{1/2} Z + \text{diag}(\sigma_i)\varepsilon.$$

Then the covariance/correlation matrix of Z is $\tilde{C}_2$.

Remark 1: This building consists in replacing the matrix associated with the $n - K$ eigenvectors corresponding to the smallest eigenvalues by a diagonal matrix, with the same diagonal elements. That is, we keep the variance of the residuals unchanged, but we set the correlation parameters of the residuals to zero. In factor analysis, the σ_i^2 correspond to the specific variances and the diagonal elements of $v\Lambda_K v^*$ to the communalities of the factors.

2.3.3 Observable endogenous factors: use of OLS

Let us denote by $\Sigma_\varepsilon \coloneqq E[\varepsilon\varepsilon^*]$, the covariance matrix of the residuals. The above model is used by practitioners to compute pairwise correlation coefficients by setting the non-diagonal terms of Σ_ε to zero, i.e. only considering the correlations derived from the factors. These are equal to the non-diagonal terms of $\text{var}[\beta Z] = E[\beta Z Z^* \beta^*] = CQ(Q^* CQ)^{-1} Q^* C$.

Setting the diagonal terms to one, we get the approximate (through additive transformation) correlation matrix:

$$\tilde{C} = \text{diag}(Id_{n\times n} - CQ(Q^*CQ)^{-1}Q^*C) + CQ(Q^*CQ)^{-1}Q^*C$$

Let us consider the case of observable (endogenous) factors. For simplicity, we will keep on assuming i.i.d. Gaussian returns. To our knowledge, OLS are the usual approach in that framework and the covariance matrix of residuals is assumed to be invertible. When the same observable factors are used in all regressions, the ML and GLS estimators coincide with OLS. This will correspond to the unrestricted model. We are also to consider cases where individual returns will be related to bucket specific portfolios mimicking the restricted approach when using latent factors. This multivariate setting looks like the seemingly unrelated regression (SUR) framework of Zellner. But as mentioned above, when the portfolio factor returns are linear combination of individual returns with time-invariant coefficients (e.g., average returns within a bucket), the covariance matrix of residuals cannot be of full rank, even if $T > n$. Thus, nor ML, neither feasible GLS can be implemented, while such an inconvenience does not occur when using OLS. Besides, since explanatory variables are endogenous, the covariance of residuals is determined by the covariance/correlation matrix of returns and cannot be set arbitrarily. We will thus focus later OLS, keeping in mind that the focus is the building of correlation matrices, rather than linear prediction.

2.4 Comparative study

This section aims at comparing the performances of the several factor-model settings presented above in approximating a given empirical correlation matrix. It includes both unrestricted and restricted algorithms, named “configurations” in the sequel.

This uses a subset of 186 names from the CDX and iTraxx indices (cf. Appendix), for which data could be retrieved over a period of roughly twelve years (12 January 2007 to 8 February 2019), including two periods of stress, the great financial crisis and the eurozone crisis.

The study is based on non-overlapping weekly stock returns. The cross-sectional database is composed of T=631 weekly-spaced dates. Given this dataset, one may compute the empirical Pearson correlation estimates for all pairs of names and build the (N, N) empirical correlation matrix. No spectral adjustment is necessary since the matrix is natively semi-positive defined, with the smallest eigenvalue being equal to 0.04.

In a credit risk management context, the risk horizon often used is 1 year (for regulatory capital computation for both the banking and the trading book for instance). However, using 1-year horizon is quite constraining regarding the dataset size (regulatory texts on capital assessment for trading book for instance, require using an empirical dataset of at least 10-year depth, but it is often difficult to find much longer time-series for a large set of issuers) if one wants avoiding overlapping and its biases. A natural way to deal with such issue is to use a smaller horizon for calibrating equity correlations. Weekly or daily data are then preferred by practitioners, even if the latter may suffer from microstructure noises and synchronicity issues. Indeed, mixing European (EU) and United-States (US) data may incorporate synchronization issues in calibrating correlation since the equities might be quoted on different marketplaces. This can lead to asynchronous data, biasing the correlation[25] estimation. Basically, daily data may be subject to this phenomenon. Using weekly data may reduce drastically this problem. In consequence, weekly equity returns have been preferred to daily ones for the subsequent numerical experiments.

Note also that it is quite common to use equity log returns to calibrate credit risk model since the equity market contains more names than the CDS one. From a theoretical point of view, this is in line with the Merton Model (1974), where asset of the issuer i is seen as the sum of

[25] Let us consider two stocks quoted on two different markets. For illustrative purpose, suppose they are driven by the following stochastic differential equation ($i = 1,2$): $dS_t^i = S_t^i \sigma^i dW_t^i$ where $\left(W_t^i\right)_{t\geq 0}$ are two standard Wiener processes such that $d\langle W_\cdot^1, W_\cdot^2\rangle_t = \rho dt$. Let $t(i)$ be the closing time of Equity i, h a given number of days, and define the log-return between $t(i)$ and $t(i)+h$ as $R_h(i) := \ln\left(\frac{S_{t(i)+h}^i}{S_{t(i)}^i}\right)$. If Equity 1 and Equity 2 are quoted on the same market, then $t(1) = t(2)$ leading to $\mathrm{Corr}\left(R_h(1), R_h(2)\right) = \rho$. Contrarily, if $t(1) \neq t(2)$, then $\mathrm{Corr}\left(R_h(1), R_h(2)\right) = \rho \frac{[\min(t(1)+h, t(2)+h) - \max(t(1), t(2))]}{h} \neq \rho$.

its equity and its total liabilities: i.e., $A_t^i = S_t^i + L^i(t)$ where we consider $t \mapsto L^i(t)$ as a deterministic function of time. In this context correlation among the asset is coming from correlation among stocks.

2.3.1 Constant correlation benchmark

As a benchmark, we first consider the case of a constant correlation, which is associated with the one factor constant beta model. The paper entitled "*Covariance Structure Regularization via Frobenius-Norm Discrepancy*" (Cui et al (2016)) provides some insights. Section 3.2 deals with the best approximation, with respect to Frobenius norm, of a given covariance matrix by a covariance matrix with constant diagonal terms (variances) σ^2 and constant correlation parameter ρ (so called CS – Constant Symmetry matrices). When the input covariance matrix is a correlation matrix, the best approximation is also a correlation matrix, i.e. $\sigma^2 = 1$ and the optimal correlation parameter $\bar{\rho}$ is simply the arithmetic average of pairwise correlations[26]. Thus, in our case $\bar{\rho} = 38.1\%$. This result can also be directly derived: The solution of $\min_{\rho} \sum_{ij} \left(\rho_{ij} - \rho\right)^2$ is simply $\bar{\rho} = \frac{1}{|\{(i,j), i>j\}|} \sum_{i>j} \rho_{ij}$, where $|\{(i,j), i > j\}|$ is the number of non-diagonal elements in the correlation matrix, $\frac{n(n-1)}{2}$, where $n = 186$ in our case study[27].

[26] From Theorem 3.2 and equation 3.13, $t = \sum_{i \neq j} \rho_{ij}$, $c = \frac{t}{n(n-1)} = \bar{\rho}$.

[27] For the constant correlation matrix to be PSD, we must have $-\frac{1}{n-1} \leq \bar{\rho} \leq 1$. Borsdorf, Higham and Raydan (2010), Theorem 2.2 state that the optimal constant correlation is the projection of $\bar{\rho}$ on the interval $[-1/n - 1\,, 1]$. In our case study, we have an inner optimum, i.e., $-1/185 < 0.381 < 1$. Cui et al (2016) project a covariance matrix, which always result to an inner optimum, which may not be the case in Borsdorf et al (2016).

The distance between the empirical and the best fit constant correlation matrix is based on the Frobenius norm, i.e., is the square root of the mean of squared differences between the empirical correlation matrix and the optimal constant correlation matrix. The outcome is $\left(\frac{1}{|\{(i,j),i>j\}|}\sum_{i>j}\left(\rho_{ij}-\bar{\rho}\right)^2\right)^{\frac{1}{2}} = 12.30\%$.

2.3.2 Block correlation matrices

The technicalities of the approach are detailed in the companion note "estimation of block correlation matrices". In our case study, we have four buckets (intersecting two sectors, financial vs non-financial, and two regions, EU vs NA). This leads to four intra-bucket correlation parameters and six inter-bucket correlations. Up to some extra constraints, not binding here, this results in a straightforward extension of the constant correlation case.

It turns out that the block correlation matrix that minimizes the Frobenius norm is such that intra and inter-bucket correlations are set to their average: The squared Frobenius norm is equal to the sum of the squared Frobenius norm over the 10 (4 + 6) blocks. Besides, we have 10 parameters (one correlation per block) to calibrate. The global minimization can be decomposed into ten independent unidimensional minimisation problems. As for the four inner (intra-bucket) correlation matrices, we know from the above that the average intra-bucket correlations provide the best fit. As for the six matrices associated with inter-bucket correlations, similarly, the best fit is achieved by choosing the corresponding average correlation. This leads to the following group level correlation matrix 0:

$$0 = \begin{pmatrix} 60.3\% & 44.3\% & 43.3\% & 36.1\% \\ 44.3\% & 52.3\% & 36.2\% & 40.7\% \\ 43.3\% & 36.2\% & 42.3\% & 34.3\% \\ 36.1\% & 40.7\% & 34.3\% & 37.6\% \end{pmatrix}$$

The diagonal elements of the group level correlation matrix correspond to the average of the intra-bucket correlations, while the non-diagonal terms to the inter-bucket correlations.

The average of intra-bucket correlations is 48.1%, while the average of inter-bucket correlation equals 39.1%. When the bucketing has economic significance, we do expect that intra-bucket correlations are larger than inter-bucket correlations.

Besides, we have the following nice result which states if 0 is positive semidefinite (PSD), then the global (186×186) correlation matrix is also PSD. This is the case here since all eigenvalues are positive[28].

Demey et al (2004), Huang & Yang (2010) provide a $4-$ factor representation for Gaussian distributions with the above block structure. The group level correlation matrix 0 is the matrix of the four global factors (see companion paper "*About the estimation of block correlation matrices* "). This parametrization is parsimonious, it involves only $4 \times 4 = 16$ beta parameters, to be compared with the 186 betas involved in a one factor model.

Nevertheless, 4 global factors are not compliant with the FRTB requirement of two types of systematic risk factors. We can reduce dimensionality by truncating the eigenvalue matrix, as in the first step of Rebonato and Jäckel (2000), see also Huang & Yang (2010). Thanks to Higham (1988), this truncation leads to the nearest covariance matrix to the group level matrix 0, with respect to the Frobenius norm. Besides, the block homogeneous structure of the global correlation matrix is kept unchanged.

2.3.3 One factor calibrated with the centroid method

As in the constant correlation model, the aim is to provide a straightforward implementation of a one factor model, to be used as a benchmark against more involved approaches. The latent random vector $X = (X_i)$ stems from a one factor model: $X_i = \beta_i Z + \sigma_i \varepsilon_i$, with $Z, \varepsilon_1, \dots, \varepsilon_n$ being independent and having unit variance and $\sigma_i = \sqrt{1 - \beta_i^2}$.

[28] The eigenvalues are equal to 0.017, 0.084, 0.155, 1.67.

The correlation between X_i and X_j, $i \neq j$ is $c_{ij} = \beta_i \times \beta_j$. Besides, $c_{ii} = \beta_i \times \beta_i + \sigma_i^2$. Thus, $\sum_{j=1}^{n} c_{ij} = \beta_i \times \left(\sum_{j=1}^{n} \beta_j\right) + \sigma_i^2$.

The calibration of the β_i relies on neglecting the σ_i^2 terms (i.e. the variance of the idiosyncratic risks): $\sum_{j=1}^{n} c_{ij} \approx \beta_i \times \left(\sum_{j=1}^{n} \beta_j\right)$. Therefore, $\sum_{i,j=1}^{n} c_{ij} \approx \left(\sum_{j=1}^{n} \beta_j\right)^2$. This leads to:

$$\beta_i \approx \frac{\sum_{j=1}^{n} c_{ij}}{\left(\sum_{i,j=1}^{n} c_{ij}\right)^{1/2}}$$

This calibrated pairwise correlations for $i \neq j$ are such that:

$$\tilde{c}_{ij} = \beta_i \times \beta_j = \frac{\left(\sum_{j=1}^{n} c_{ij}\right)\left(\sum_{i=1}^{n} c_{ij}\right)}{\sum_{i,j=1}^{n} c_{ij}}$$

The numerator of the fraction involves multiplying the averages of pairwise correlations involving i and involving j. The denominator is the average of all the terms of the correlation matrix (including diagonal elements). This is a straightforward approach to the calibration of a one factor model from a given empirical correlation matrix.

Forthcoming analysis focus on several aspects of the modelling. First subsection presents the model fit, measured via the so-called Root Mean Square Error (RMSE), considering all pairwise correlations in all factor-model settings. Subsequent subsections then focused on the 3-N beta case, comparing only models that assign 3 non-null betas to all issuers, linear regression being taken as the benchmark methodology for comparison. Analysis is performed first on the ability to fit inter and intra class empirical correlations, then on the extreme correlation level and the sign of the betas, providing a homogeneity measure of the considered classes.

2.3.4 RMSE comparison

To illustrate the performance of the different configurations to fit the empirical correlation matrix, one may be first interested in comparing the final objective function obtained at the end of the calibration process. To make them comparable, they are all expressed in terms of the Root Mean Squared Error (RMSE), formally defined as:

$$\mathrm{RMSE} \coloneqq \sqrt{\frac{1}{\mathrm{N} \times (\mathrm{N}-1)} \sum_{\mathrm{i,j}=1}^{\mathrm{N}} \left(\mathrm{C}_{\mathrm{i,j}}^{\mathrm{Empirical}} - \mathrm{C}_{\mathrm{i,j}}^{\mathrm{Model}} \right)^2}$$

Figure 2.2 provides the RMSE of the different calibration configurations when considering 1 to 4 factors. For convenience purpose, Spectral Projected Gradient (SPG) and Andersen-Basu-Sidenius (ABS) results have been concatenated altogether in the "Least_Squares_Correlation" configuration since they lead to almost the same results in our empirical experiment. Note however, that "Least_Squares_Correlation_Restricted" only refers to the restricted version of the SPG algorithm, since a similar restricted version seems not attainable for the ABS algorithm.

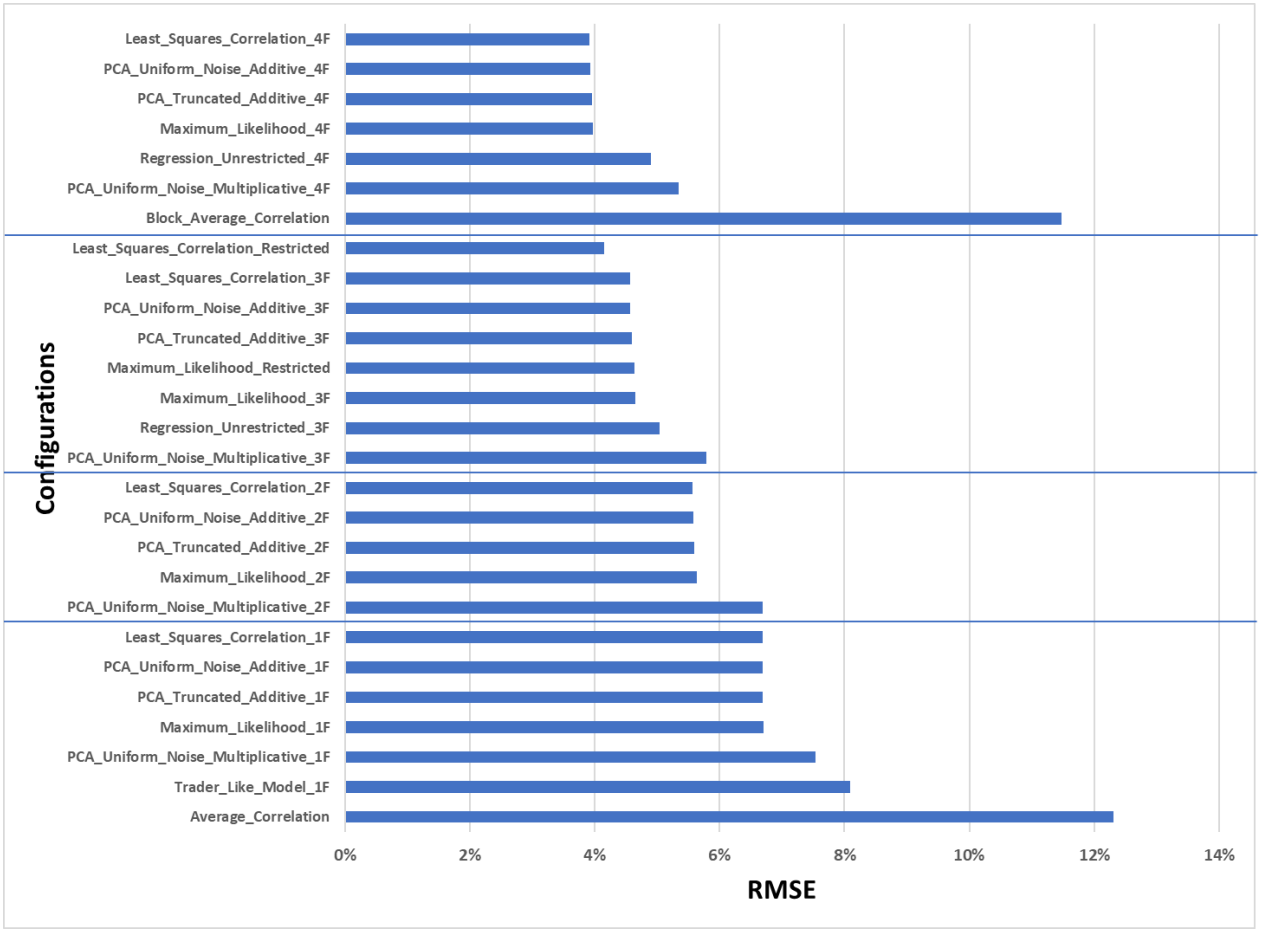


Figure 2.2 RMSE (model implied correlations vs. empirical correlations) for different calibration configurations

As expected, the RMSE is a decreasing function of the number of factors. This naïve assertion should be balanced, however, by the marginal gain to add factors. Indeed, adding one factor to a one-factor model allows a significant gain in terms of RMSE (≈ 1.1% absolute gain for the "Least_Squares_Correlation" configuration for instance), while adding one factor to a three-factor model implies only a smaller gain (≈ 0.6%). This suggests that the regulatory constraint to use a three-factor model for the capital assessment of the default risk in trading book for instance, is a good balance between accuracy and fit, or in regulatory terms: a good balance between risk sensitivity and complexity.

Going through the configurations, one observes for a given number of factors, the same type of model ordering. Basically, naïve methods ("Trader_Like_Model" and "Average_Correlation" for the one-factor setting, and "Block_Average_Correlation" for the four-factor model) imply only rough fit of the considered empirical matrix. Considering their definition, this result is not surprising. Then, PCA, Regression, ML and unrestricted LS methods perform quite equivalently for a given factor configuration. One exception occurs, however. It concerns the "PCA_Uniform_Noise_Multiplicative", for which the multiplicative transformation has a strong negative impact on the RMSE.

2.3.5 3N-beta case: conditional RMSE per class

Focusing on the 3N-beta settings, i.e., 3-factor model for unrestricted algorithms and the two restricted models, it is noteworthy that one observes a clear gain to use restricted model. Indeed, restricted regression provides better results than unrestricted ones. More importantly, restricted least square calibration improves significantly the RMSE, to a level closed to the one obtained in the unrestricted case for a four-factor configuration. By referring to the same number of non-null beta, the restricted calibration allows therefore a significant gain. Coupling with its ability to identify and explain the source of the dependency, it represents a good challenger of the unrestricted configurations, more usual in a credit risk modelling context.

To strengthen this last assertion, one may be interested in analysing the fit through the chosen classification (or segmentation). Remember that, in the restricted configuration, each issuer must be mapped onto a class (or segment). To solve this classification problem, statistical approaches, supervised or not, may be invoked. When considering a credit risk model with constraint on risk measure production for large portfolios, it is quite common instead to rely on an economical approach, based on issuer information like industrial sector or geographical region, to perform the classification. While not setting on optimal statistical procedure, this approach is readable and easily manageable since it allows a straightforward fallback procedure when considering issuers not present in the calibration dataset.

The economical classification used in this numerical analysis is based on the couple industry-region. More especially, 4 couples are considered here: Financial-Europe (FIN-EU), Financial-United-States (FIN-US), Non-Financial-Europe (NFIN-EU), Non-Financial-United-States (NFIN-US).

Besides, given this classification, the intra-class fit as well as the inter-class fit should be different in function of the restriction. These two concepts may be defined is the same way as the overall RMSE, in particular:

$$\mathrm{RMSE}_{\mathrm{Intra}} := \sqrt{\frac{1}{n_{\mathrm{intra}}}\sum_{i,j=1}^{N}\left(C_{i,j}^{\mathrm{Empirical}} - C_{i,j}^{\mathrm{Model}}\right)^2 \mathbb{I}_{\mathrm{Segment}_i=\mathrm{Segment}_j}}$$

$$\mathrm{RMSE}_{\mathrm{Inter}} := \sqrt{\frac{1}{n_{\mathrm{Inter}}}\sum_{i,j=1}^{N}\left(C_{i,j}^{\mathrm{Empirical}} - C_{i,j}^{\mathrm{Model}}\right)^2 \mathbb{I}_{\mathrm{Segment}_i\neq\mathrm{Segment}_j}}$$

Indeed, it is obvious that $\mathrm{RMSE} = \sqrt{(\mathrm{RMSE}_{\mathrm{Intra}})^2.\frac{n_{\mathrm{Intra}}}{N\text{-}(N-1)} + (\mathrm{RMSE}_{\mathrm{Inter}})^2.\frac{n_{\mathrm{Inter}}}{N\text{-}(N-1)}}$. Table 1 provides figures for this analysis.

Configuration	RMSE	RMSE Intra	RMSE Inter	Bias
Least_Squares_Correlation_Restricted	**4.15%**	**4.64%**	**3.87%**	**0.00%**
Least_Squares_Correlation_3F	4.57%	5.74%	3.82%	0.00%
PCA_Uniform_Noise_Additive_3F'	4.57%	5.75%	3.82%	0.01%
PCA_Truncated_Additive_3F'	4.60%	5.64%	3.95%	-0.27%
Maximum_Likelihood_Restricted	**4.64%**	**4.46%**	**4.73%**	**0.43%**
Maximum_Likelihood_3F	4.65%	5.96%	3.80%	0.21%
Regression_Unrestricted_3F'	5.03%	6.52%	4.06%	-0.29%
PCA_Uniform_Noise_Multiplicative_3F'	5.79%	6.93%	5.10%	0.42%

Table 2.19 RMSE for different calibration configurations according to intra and inter-class and bias

Considering the intra-class RMSE, restricted settings perform significantly better than unrestricted ones with an RMSE average of 4.55% versus 6.09%. This result is in line with expectation and confirms the relevance of the economical classification. Contrarily, for inter-class RMSE, while the "Least_Squares_Correlation_Restricted" performed the same way as the unrestricted algorithms (except the "PCA_Uniform_Noise_Multiplicative" once again), the "Maximum_Likelihood_Restricted" algorithm provides poorer fit. This is not surprising given its moderate overall RMSE and its very good intra-class RMSE.

The last column of the table provides the difference of off-diagonal element average between empirical correlation and model-implied correlation matrix. Results show that on average, for several configurations, the empirical correlation mean is not perfectly replicated, with a high (positive or negative) bias for some settings, significantly positive for the "Maximum_Likelihood_Restricted" setting for instance. Note however, that once again, the "Least_Squares_Correlation_Restricted" model performed very well on this feature.

2.3.6 3N-beta case: homogeneity score

To confirm the operational interest to use restricted factor model, one may be interested in the sign of the factor loadings. Intuitively, if the issuers affected to a given class are homogeneous enough, then the sign of the factor loading, i.e. the sign of the sensitivity to the systematic factors, should be the same.

For a given factor j, one can compute the frequency of positive sign, i.e., $\text{Freq(j)} \coloneqq \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}_{\{\text{sign}(\beta_{i,j})=1\}}$. Per complementarity, one also obtains the frequency of negative signs as $1 - \text{Freq(j)}$. Taking the maximum of both frequencies, i.e., $\max\big(\text{Freq(j)}; 1 - \text{Freq(j)}\big) \in [50\%, 100\%]$, provides us with a homogeneity measure. On the one hand, if this maximum is 100%, the homogeneity is perfect since all betas have the same sign. On the other hand, if this maximum is 50%, then it means that the there are two sub-classes of same cardinal in the considered class.

Overall, one can finally compute an aggregated metric: the average of the maximum frequency, linearly rescaled so that it has values in $[0\%, 100\%]$. Formally, this is defined as follows:

$$\text{AverageFreq} \coloneqq \sum_{j=1}^{J} 2 \times \left(\max\left(\text{Freq(j)}; 1 - \text{Freq(j)}\right) - 0.5\right) \in [0\%, 100\%]$$

Table 2 provides this homogeneity score for the different calibration configurations.

Configuration	Homogeneity-score
Least_Squares_Correlation_Restricted	**75.48%**
Least_Squares_Correlation_3F	35.13%
PCA_Uniform_Noise_Additive_3F'	35.48%
PCA_Truncated_Additive_3F'	35.48%
Maximum_Likelihood_Restricted	**74.98%**
Maximum_Likelihood_3F	43.73%
Regression_Unrestricted_3F'	30.82%
PCA_Uniform_Noise_Multiplicative_3F'	35.48%

Table 2.2 Homogeneity (i.e., same-sign betas) score

As expected, restricted algorithms provide better homogeneity in terms of beta signs than unrestricted algorithms. This result is in line with expectations and confirms the relevance of the economical classification, as previously stated.

2.3.7 3N-beta case: conditional RMSE on high pairwise correlations

It is well known that factor models, notably in their unrestricted setting, tend to underestimate high correlations (Laurent et al., 2016). These latter are of importance yet for both modelling

and classification purposes. In consequence, to study the fit to these high correlations, one is interested in this subsection by the following mapping (for each configuration):

$$[0,0.95] \ni \alpha \mapsto \mathrm{RMSE}(\alpha) := \sqrt{\frac{\sum_{i,j=1}^{N} \left(C_{i,j}^{\mathrm{Empirical}} - C_{i,j}^{\mathrm{Model}}\right)^2 . \mathbb{I}_{\left\{C_{i,j}^{\mathrm{Empirical}} \geq \mathrm{Quantile}_{C^{\mathrm{Emp}}}(\alpha)\right\}}}{\sum_{i,j=1}^{N} \mathbb{I}_{\left\{C_{i,j}^{\mathrm{Empirical}} \geq \mathrm{Quantile}_{C^{\mathrm{Emp}}}(\alpha)\right\}}}}$$

Which represents the RMSE given empirical pairwise correlations higher or equal to α. Thus, with such mapping, one may analysis the behavior of all 3N-beta model configurations when α increases (Figure 2.3).

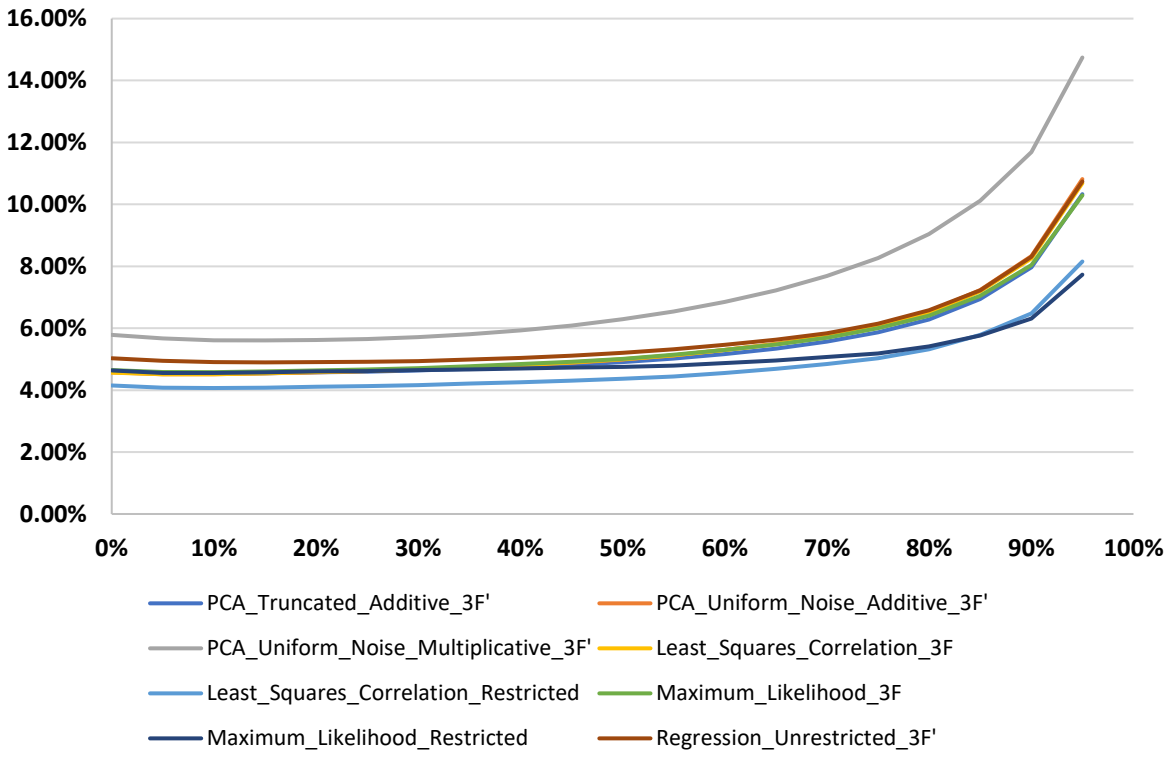


Figure 2.3 Conditional RMSE

Main comments when reading Figure 2 are the followings. First, PCA with multiplicative transformation underperforms all configurations, once again. Then, when considering low to moderate pairwise correlations, i.e., from 0% to 60%, all other methods perform equivalently even if the restricted approaches, notably the "Least_Square_Correlation_Restricted" approach already performed slightly better. A contrary, when dealing with high correlation levels, i.e. above 60%, then the restricted algorithms outperform considerably the unrestricted ones. This is mainly due to the additional information embedded in the user-provided classification. Indeed, issuers in the same class (geographical region and industrial sector) tend to exhibit higher correlations. Calibrating betas by incorporating such information clearly improve the fit. This feature argues for the use of restricted approaches since high correlations fitting is a significant issue for risk-managers as they largely impact risk measures.

2.3.8 Summary statistics

By design, the iterative principal factor method (Andersen et al, 2003) and the SPG algorithm provide the minimum distance to the empirical covariance matrix, but Principal Component and the Maximum Likelihood approach implemented with the EM algorithm perform almost as well[29].

These approaches lead to a significant improvement over the constant correlation model, used as a naïve benchmark. Block correlation approaches lag slightly behind and provide a poor fit to the empirical correlation matrix[30]. The naïve CDO pricing one does better than block correlation approaches, but it involves more betas.

[29] In https://www.r-bloggers.com/iterated-principal-factor-method-of-factor-analysis-with-r/ , one also concludes that Principal Components and Iterative Principal factors provide similar results. Hair et al (1998), p. 103 make the same claim.

[30] Considering a one factor block correlation matrix leads to small improvement to the constant correlation matrix. It was stated that constant correlation matrix had some optimality property. These statements are not inconsistent since the one factor block correlation matrix involves four beta parameters, instead of one in the case of the constant correlation model.

When considering calibration to the empirical correlation matrix, Isotropic principal component methods lags clearly behind the unconstrained Principal Component Method, due to the normalisation to a correlation matrix, through a multiplicative, rather than an additive approach.

Number factors	Block Correlation	Constant Correlation	Centroid Method	Isotropic PCA	PCA	SPG	EM
1	12.09%	12.30%	8.10%	7.54%	6.69%	6.69%	6.71%
2	11.68%			6.69%	5.59%	5.57%	5.64%
3	11.48%			5.79%	4.60%	4.57%	4.65%
4	11.47%			5.34%	3.96%	3.92%	3.98%
5				5.02%	3.46%	3.41%	3.45%

Table 2.3 Distance to empirical correlation matrix (RMSE)

The following table, based on the distances with respect to Frobenius norm, provides improvements in the adjustment to the empirical correlation matrix when adding a global factor.

Marginal improvement of adding global factors	
from 1 to 2	**17%**
from 2 to 3	**18%**
from 3 to 4	**14%**
from 4 to 5	**13%**

Table 2.4 Relative reduction in distance to the empirical correlation matrix

While the reduction in absolute error tends to be lower as the number of global factors increases, from a relative point of view, we can view a uniform impact, suggesting that two global factors (and even more) might not be enough to adequately capture dependencies.

2.4 Conclusion

We described various approaches to estimate probabilistic models with K factors and idiosyncratic noises, which can be used to simulate the latent random variables driving defaults with applications to the Basel 2.5 and 3 frameworks for credit exposures in the trading book (Incremental Risk Charge – IRC and Default Risk Charge – DRC).

From our results, one may highlight the following items:

- For a given number of factors, optimization-based approaches provide almost the same level performance in terms of RMSE. Restricted configuration being marginally better, while non optimized approaches clearly underperform.
- When dealing with the naïve PCA approach, the final transformation to recover correlation matrix must be chosen carefully. In our numerical experiment, multiplicative approach, which impact all pairwise correlations, significantly underperforms the additive approach, impacting only the diagonal elements.
- Restricted versions of optimization-based approaches behave better when dealing with intra-classes. This is mainly due to the effectiveness of the classification. Even with a rough approach, based only on geographical-sector classification, fitting is significantly improved.
- It is well known that factor-model tend to smooth the correlation matrix, leading to large error for high correlations. Numerical experiments indicate that restricted calibration performed better in that aspect too. By restricting the information onto a class of issuers that should (a priori) behaves the same way, the restricted version of SPG algorithm allows a smaller error on these high empirical correlations than other methods.
- Two last items are strengthened by the beta homogeneity for the restricted configurations, which allows a better understanding of the underlying latent-factor effect. Even more, in this configuration, latent-factor-loadings have a meaningful interpretation.

The betas with respect to the K factors provided by the principal component method can be used as a starting point of an iterative optimisation algorithm: iterated principal factor or Spectral Projection Gradient (SPG) for the nearest correlation matrix approach and Expectation Maximization (EM) for maximum likelihood estimation.

2.5 Appendices

Appendix 2.1: Principal components and optimal choice of regressors in nearest correlation matrix problem

The eigenvectors of C, v_k, $k = 1, \dots n$, form a $n \times n$ orthogonal matrix $V = (v_1 \dots, v_n)$ and the spectral decomposition of C can be written as $C = V\Lambda V^*$, where Λ is the diagonal matrix of eigenvalues $\lambda_1, \dots, \lambda_n$ sorted by decreasing order. V^*X is the $n-$dimensional random vector associated with the principal components of X. These principal components are uncorrelated, and their variances are equal to $\lambda_1, \dots, \lambda_n$. We denote by $v = (v_1 \dots, v_K)$ the $n \times K$ matrix derived from the first K columns of V and by $Z = v^*X$, the random vector associated with the K first principal components[31]. From $VV^* = \mathrm{Id}_{n\times n}$, $X = V(V^*X)$. Splitting columns in V and using orthogonality of principal components, we readily have $X = vZ + \varepsilon$, with $E[\varepsilon|Z] = 0$[32]. Given that ε and Z are uncorrelated, we have: $E[\|X\|_2^2] = E[\|vZ\|_2^2] + E[\|\varepsilon\|_2^2]$, where $\| \ \|_2$ stands for Euclidian norm. $E[\|X\|_2^2] = \sum_{i=1}^n E[X_i^2] = n$. Since $\sum_{i=1}^n E[X_i^2] = \mathrm{trace}(C)$ and by the properties of the trace operator, $\mathrm{trace}(C) = \mathrm{trace}(\Lambda) = \lambda_1 + \cdots + \lambda_n$, which stands for the "total (univariate) variance" of X.

[31] As mentioned above, in the case of the constant correlation matrix, $v_1 = \frac{1}{\sqrt{n}}e$. The first principal component can be associated to the return of an equally weighted portfolio, a typical choice in credit risk analysis, given that credit indices (CDX, iTraxx) are equally weighted.

[32] Let us denote by $Z_k, i = 1, \dots, n$ the principal components and by v_{ik} the generic term of V. $X_i = \sum_{i=1}^K v_{ik}Z_k + \sum_{i=K+1}^n v_{ik}Z_k$, $i = 1, \dots, n$. Let us denote by $\varepsilon_i = \sum_{i=K+1}^n v_{ik}Z_k$. Since the principal components are independent $E[\varepsilon_i|Z_1, \dots, Z_K] = 0$. Thus $X_i = \sum_{i=1}^K v_{ik}Z_k + \varepsilon_i$ is a linear regression model.

Let us now consider an arbitrary set of regressors $Z = Q^*X$. Let us denote by W^*X the $K-$dimensional random vector associated with the standardized principal components of Z. Therefore, the covariance matrix of W^*X is the identity matrix, which leads to $W^*CW = Id_{K\times K}$. We will further denote by $w_1 \dots, w_K$, the K columns of W. For $l \neq k$, $w_k^* C w_l = 0$ and $w_k^* C w_k = 1$. Since the principal components of Z span the linear subspace generated by the components of Z, $E[X|Z] = E[X|W^*X]$. We may look for W, that will maximise the $E[\|E[X|W^*X]\|_2^2]$. Since the components of W^*X are uncorrelated and of unit variance, decomposition of variance is straightforward: $E[\|E[X|W^*X]\|_2^2] = \sum_{i=1}^{n}(\sum_{k=1}^{K}(E[w_k^* X X_i])^2)$, corresponding to the sum of the squared regression betas. $E[XX_i] = c_i$ corresponds to the $i-$th column of C. Thus, $(E[w_k^* X X_i])^2 = w_k^* c_i c_i^* w_k$. Noticing that $\sum_{i=1}^{n} c_i c_i^* = C^2$, $E[\|E[X|W^*X]\|_2^2] = \sum_{k=1}^{K} w_k^* C^2 w_k$. Let us now define $h_k = C^{1/2} w_k$ and H the $n \times K$ matrix whose $k-$column is h_k. We are then looking for maximizing $\sum_{k=1}^{K} h_k^* C h_k = \text{trace}(H^* C H)$, under the constraints $h_k^* h_l = 0$ for $l \neq k$ and $h_k^* h_k = 1$, or equivalently $H^*H = Id_{K\times K}$. The maximum is achieved when $H = v$, see Jolliffe (2010, Sect. 2.1) and equals $\lambda_1 + \cdots + \lambda_K$. Thus, $w_k = C^{-1/2} v_k = \lambda_k^{-1/2} v_k$. The corresponding set of regressors is associated with the first K principal components. $[\|E[X|Q^*X]\|_2^2] = \text{trace}\big(\text{var}\big[E[X|Q^*X]\big]\big) = \text{trace}(CQ(Q^*CQ)^{-1}Q^*C)$ is maximised when $Q = v$ (see Rao (1964), p. 339. Then, the sum of variances of the residual terms (the total variance) in the linear regression is minimized. This shows some optimality in choosing the first principal components as optimal regressors, i.e., to build an optimal K – dimensional subspace on which to project the space spanned by the X_i.

We also have $\text{var}[vZ] = E[vZZ^*v^*] = Cv(v^*Cv)^{-1}v^*C$ and $Cv = v\Lambda_K$ where Λ_K is a diagonal $K \times K$ matrix, with diagonal terms equal to $\lambda_1, \dots, \lambda_K$. This leads to $Cv(v^*Cv)^{-1}v^*C = v\Lambda_K v^*$. Eckart-Young theorem states that $v\Lambda_K v^*$ is the best rank K approximation of C, with respect to Frobenius norm. Since Q^*C is a $K \times n$ matrix, $\text{rank}(Q^*C) \leq \min(n, K) = K$. Since $\text{var}[\beta Z] = CQ(Q^*CQ)^{-1}Q^*C$, $\text{rank}(\text{var}[\beta Z]) \leq K$. Thus $\|C - CQ(Q^*CQ)^{-1}Q^*C\|_F \geq \|C - Cv(v^*Cv)^{-1}v^*C\|_F = \sqrt{\lambda_{K+1}^2 + \cdots + \lambda_n^2}$ (see Rao (1964)).

We cannot however state that $\left\|C - \text{diag}\left(Id_{(nxn)} - CQ(Q^*CQ)^{-1}Q^*C\right) - CQ(Q^*CQ)^{-1}Q^*C\right\|_F$ is minimized when choosing $v = Q$. This criterium differs from $\|C - CQ(Q^*CQ)^{-1}Q^*C\|_F$ in that only off-diagonal elements of C and $CQ(Q^*CQ)^{-1}Q^*C$ are involved.

Besides, since $CQ(Q^*CQ)^{-1}Q^*C$ is a symmetric rank K matrix, from its spectral decomposition, it can be factorized as $\beta\beta^*$, where β is a $n \times K$ matrix. This leads to consider the minimisation of $\left\|C - (\beta\beta^* + Id_{(n\times n)} - \text{diag}(\beta\beta^*))\right\|_F$ which corresponds to the least squares problem (minres of Harman and Jones (1966))[33].

One may look for optimal regressors, i.e., some matrix Q minimizing the distance to the true correlation matrix C. We already know that: $\left\|C - \text{diag}\left(Id_{(n\times n)} - CQ(Q^*CQ)^{-1}Q^*C\right) - CQ(Q^*CQ)^{-1}Q^*C\right\|_F \geq \min_{\beta\in\mathbb{R}^{n\times K}}\left\|C - (\beta\beta^* + Id_{(n\times n)} - \text{diag}(\beta\beta^*))\right\|_F$. However, the computation of Q such that $CQ(Q^*CQ)^{-1}Q^*C = \beta\beta^*$ is not always feasible. As an example, let us go back to the case of a constant correlation matrix. By symmetry of $\|C - \text{diag}(Id_{n\times n} - CQ(Q^*CQ)^{-1}Q^*C) - CQ(Q^*CQ)^{-1}Q^*C\|_F$, a minimum of the criterium is associated with $Q = e$[34]. Since the constant correlation matrix (with $\rho \geq 0$) can be written as $C = (\beta\beta^* + Id_{n\times n} - \text{diag}(\beta\beta^*)$, with $\beta = \sqrt{\rho}e$, $\min_{\beta\in\mathbb{R}^{n\times K}}\|C - (\beta\beta^* + Id_{n\times n} - \text{diag}(\beta\beta^*))\|_F = 0$. The computation of $CQ(Q^*CQ)^{-1}Q^*C$ is then straightforward: Pairwise correlations are equal to $\rho + \frac{1-\rho}{n}$. The minimum of $\|C - \text{diag}(Id_{n\times n} - CQ(Q^*CQ)^{-1}Q^*C) - CQ(Q^*CQ)^{-1}Q^*C\|_F$ is then $(1-\rho)\sqrt{\frac{n-1}{n}}$. In such cases, factor analysis provides superior results to regression models when it comes to fitting a correlation matrix.

[33] The optimal β can be computed through the iterated principal factor method or the SPG algorithm of Borsdorf et al (2010).

[34] Multiplying Q by a non-zero scalar does not change the criterium, thus $Q = e$ consists in using the return of equally weighted portfolio as a regressor.

Besides, thanks to triangle inequality, $\left\|C - (\beta\beta^* + Id_{(n\times n)} - \text{diag}(\beta\beta^*))\right\|_F \leq \|C - \beta\beta^*\|_F + \left\|Id_{(n\times n)} - \text{diag}(\beta\beta^*)\right\|_F \leq \sqrt{\lambda_{K+1}^2 + \cdots + \lambda_n^2} + \sqrt{\sigma_1^2 + \cdots + \sigma_n^2}$, where the σ_i^2 are the specific variances.

Appendix 2.2: Low rank approximation of the correlation matrix

Low rank approximation of a covariance matrix can be achieved equivalently through the SVD decomposition of the rectangular matrix of historical returns or the spectral decomposition of the empirical correlation matrix, by setting singular value, eigen values to zero below rank K. Since the outcome is not a correlation matrix, a normalisation step is involved. Such approaches have been used in Finance, when the purpose was to cancel out idiosyncratic risks, achieve perfect dependency in arbitrage free pricing settings or when dimension reduction is required for numerical purpose. Though truncation is also involved as a first step in PCA approaches to factor models, the lack of idiosyncratic risk is not meaningful in our context.

SVD viewpoint: Eckart-Young theorem

Low rank approximations can be built equivalently either from the matrix of demeaned and standardised historical returns, from the spectral decomposition of the empirical correlation matrix (PCA based techniques).

Let $K \leq \min(T, n)$ be the targeted number of factors and set:

$$\tilde{R} = U\tilde{\Sigma}_K V^*$$

where $\tilde{\Sigma}_K$ is the same matrix as Σ except that it contains only the K largest singular values (the other ones being replaced by zero).

Eckart-Young theorem states that $\tilde{R}$ is the rank K matrix closest to $R = U\Sigma V^*$ with respect to the Frobenius norm[35]. The columns of $\tilde{R}$ can be viewed as time series of pseudo-returns associated with the n assets; this corresponds to "filtering out" the "noises". Equivalently, the time series of pseudo-returns correspond to the dynamics of the returns associated with the K first principal components. Processing these pseudo-returns is known as the Karhunen-Loève transform in signal analysis or as the Hotelling transform after independent work form these authors (see Gentle (2009)).

$R - \tilde{R}$ is the rectangular matrix of the residuals. Thus, by choosing the K first principal components, the sum of the squared residuals (across time and names), which corresponds to $\left\|R - \tilde{R}\right\|_F^2$ is minimized. Equivalently, the variance of the returns explained by K dependent variables is maximised when these are chosen as the first principal components.

This relates PCA to regression analysis and least squares estimation. For instance, the sum of the squared residuals in a one factor model, where the factor is an equally weighted portfolio (say) will be larger than when using the portfolio with weights associated to factor loadings of the first principal component. In this case $K = 1$. This readily extends to a multifactor setting.

As a side remark, Golub, Hoffman & Stewart (1987) provide additional results in an extended framework where the minimisation is done under the constraint that some of the rows of the matrix R are kept unchanged. Therefore, one can build time series of pseudo-returns for all names, under the constraint that these pseudo-returns are equal to original returns for a pre-specified subset of names.

[35] Mirsky (1960) extended Eckart and Young result to arbitrarily unitarily invariant norms, i.e., matrix norms such that $\|X\| = \|UX\| = \|XU\|$ for any $(n \times n)$ matrix X and any unitary $(n \times n)$ matrix U. Higham (2002) deals with weighted Frobenius norms. While not further considered in this paper, this approach is promising: It allows to deal with the issue of highly correlated names, for which confidence intervals are narrow, by increasing weights as pairwise correlation come close to one.

The covariance (diagonal terms are not equal to one) matrix associated with these pseudo returns is $\widetilde{\Omega} = \frac{1}{T} \times \tilde{R}^* \tilde{R} = \frac{1}{T} \times V \tilde{\Sigma}_K^* \tilde{\Sigma}_K V^*$. This is to be compared with: $\hat{\mathbb{C}} = \frac{1}{T} \times R^* R = \frac{1}{T} \times V \Sigma^* \Sigma V^*$. $\tilde{\Sigma}_K^* \tilde{\Sigma}_K$ and $\Sigma^* \Sigma$ are $(n \times n)$ diagonal matrices. The diagonal terms are the squares of the singular values, but in $\tilde{\Sigma}_K^* \tilde{\Sigma}_K$ the smallest singular values (below rank K) have been set to zero.

PCA viewpoint

As discussed below, we may equivalently consider the low rank approximation of the empirical covariance/correlation matrix $\hat{\mathbb{C}}$. We denote by λ_k, $k = 1, \dots, K$, the k^{th} largest eigenvalue of $\hat{\mathbb{C}}$ and by v the $(n \times K)$ matrix of the K corresponding eigenvectors. v corresponds to the first K columns of V. For simplicity, we assume in the following that $\lambda_k > 0$ for $k = 1, \dots, K$.

From Eckart-Young theorem (applied here to the covariance/correlation matrix, not to the rectangular matrix of historical returns), the rank K approximation of $\hat{\mathbb{C}}$ is provided by $v \Lambda_K v^*$, where Λ_K is the $(K \times K)$ diagonal matrix with diagonal terms equal to λ_k, $k = 1, .., K$.

As stated above, since $\frac{1}{T} \times V \tilde{\Sigma}_K^* \tilde{\Sigma}_K V^* = v \Lambda_K v^* = \widetilde{\Omega}$, SVD and PCA provide the same rank K approximation of $\hat{\mathbb{C}}$.

Multiplicative normalization to a correlation matrix

However, a final step needs to be considered since $\widetilde{\Omega}$ is a covariance matrix, but not a correlation matrix: the diagonal elements of $\widetilde{\Omega}$ are not equal to one.

Let us denote by $\tilde{\omega}_{ii}$ the diagonal elements of $\widetilde{\Omega}$. The matrix

$$\tilde{C}_1 = \text{diag}\big((\widetilde{\omega}_{ii})^{-1/2}\big)\widetilde{\Omega}\,\text{diag}\big((\widetilde{\omega}_{ii})^{-1/2}\big) = \text{diag}\big((\widetilde{\omega}_{ii})^{-1/2}\big)v\Lambda_K v^*\,\text{diag}\big((\widetilde{\omega}_{ii})^{-1/2}\big)$$

is the correlation matrix associated with $\widetilde{\Omega}$. $\tilde{C}_1$ can be viewed as a rank K approximation of the empirical correlation matrix $\hat{\mathbb{C}}$. This approach is described in Rousseeuw & Molenberghs (1993), section 3, as the "eigenvalue method".

Probabilistic representation

Let Z be a K – dimensional random vector with independent, zero mean and unit variance components (i.e., latent factors) and consider the n – dimensional random vector:

$$X = \text{diag}\big((\widetilde{\omega}_{ii})^{-1/2}\big)v\Lambda_K^{1/2}Z$$

The covariance/correlation matrix of X is $\tilde{C}_1$.

A K – dimensional only random vector is involved in the simulation, which is appealing from a numerical point of view. The low rank approximation might also be suitable for pricing purposes or in a counterparty credit risk framework. On the other hand, from a risk management or an asset allocation perspective, it is not comfortable to deal with many eigenvalues equal to zero since each of them is associated with seemingly risk-free portfolios.

Appendix 2.3: Data description, list of selected names from iTraxx and CDX NA IG

ANHEUSER-BUSCH INBEV SA/NV	Europe	NONFIN	DANSKE BANK A/S	Europe	FIN
BARRICK GOLD CORP	United States	NONFIN	DEUTSCHE BANK AG-REGISTERED	Europe	FIN
ACCOR SA	Europe	NONFIN	DEERE & CO	United States	NONFIN
CREDIT AGRICOLE SA	Europe	FIN	VINCI SA	Europe	NONFIN
KONINKLIJKE AHOLD DELHAIZE N	Europe	NONFIN	DIAGEO PLC	Europe	NONFIN
AMERICAN ELECTRIC POWER	United States	NONFIN	QUEST DIAGNOSTICS INC	United States	NONFIN
AERCAP HOLDINGS NV	Europe	NONFIN	DR HORTON INC	United States	NONFIN
AEGON LTD	Europe	FIN	WALT DISNEY CO/THE	United States	NONFIN
AMERICAN INTERNATIONAL GROUP	United States	FIN	DARDEN RESTAURANTS INC	United States	NONFIN
AIRBUS SE	Europe	NONFIN	DEVON ENERGY CORP	United States	NONFIN
AKZO NOBEL N.V.	Europe	NONFIN	OVINTIV CANADA ULC	United States	NONFIN
ALLSTATE CORP	United States	FIN	EDF	Europe	NONFIN
AMGEN INC	United States	NONFIN	ELECTROLUX AB-B	Europe	NONFIN
APA CORP	United States	NONFIN	EASTMAN CHEMICAL CO	United States	NONFIN
ANADARKO PETROLEUM CORP	United States	NONFIN	BOUYGUES SA	Europe	NONFIN
ARROW ELECTRONICS INC	United States	NONFIN	ENBRIDGE INC	United States	NONFIN
AVIVA PLC	Europe	FIN	ENEL SPA	Europe	NONFIN
AMERICAN EXPRESS CO	United States	FIN	ENGIE	Europe	NONFIN
AUTOZONE INC	United States	NONFIN	ENI SPA	Europe	NONFIN
BOEING CO/THE	United States	NONFIN	E.ON SE	Europe	NONFIN
BARCLAYS PLC	Europe	FIN	EQUITY RESIDENTIAL	United States	FIN
BASF SE	Europe	NONFIN	EXELON CORP	United States	NONFIN
BRITISH AMERICAN TOBACCO PLC	Europe	NONFIN	EXPEDIA GROUP INC	United States	NONFIN
BAXTER INTERNATIONAL INC	United States	NONFIN	FORD MOTOR CO	United States	NONFIN
BAYER AG-REG	Europe	NONFIN	FIRSTENERGY CORP	United States	NONFIN
BANCO BILBAO VIZCAYA ARGENTA	Europe	FIN	FORTUM OYJ	Europe	NONFIN
BEST BUY CO INC	United States	NONFIN	FOX CORP - CLASS A	United States	NONFIN
BRISTOL-MYERS SQUIBB CO	United States	NONFIN	TOTAL SE	Europe	NONFIN
DANONE	Europe	NONFIN	VALEO	Europe	NONFIN
BNP PARIBAS	Europe	FIN	ASSICURAZIONI GENERALI	Europe	FIN
BERKSHIRE HATHAWAY INC-CL A	United States	FIN	GENERAL ELECTRIC CO	United States	NONFIN
CARREFOUR SA	Europe	NONFIN	GENERAL MILLS INC	United States	NONFIN
CARDINAL HEALTH INC	United States	NONFIN	SOCIETE GENERALE SA	Europe	FIN
CATERPILLAR INC	United States	NONFIN	HALLIBURTON CO	United States	NONFIN
CHUBB LTD	United States	FIN	HOME DEPOT INC	United States	NONFIN
COMMERZBANK AG	Europe	FIN	HEIDELBERG MATERIALS AG	Europe	NONFIN
CBS CORP-CLASS B NON VOTING	United States	NONFIN	HEINEKEN NV	Europe	NONFIN
CARNIVAL CORP	United States	NONFIN	HENKEL AG & CO KGAA VOR-PREF	Europe	NONFIN
COMCAST CORP-CLASS A	United States	NONFIN	HESS CORP	United States	NONFIN
CENTRICA PLC	Europe	NONFIN	HARTFORD FINANCIAL SVCS GRP	United States	FIN
CANADIAN NATURAL RESOURCES	United States	NONFIN	HANNOVER RUECK SE	Europe	FIN
CAPITAL ONE FINANCIAL CORP	United States	FIN	HONEYWELL INTERNATIONAL INC	United States	NONFIN
CONTINENTAL AG	Europe	NONFIN	HP INC	United States	NONFIN
CONOCOPHILLIPS	United States	NONFIN	H&R BLOCK INC	United States	NONFIN
CAMPBELL SOUP CO	United States	NONFIN	HSBC HOLDINGS PLC	Europe	FIN
AXA SA	Europe	FIN	IBERDROLA SA	Europe	NONFIN
CREDIT SUISSE GROUP AG-REG	Europe	FIN	INTL BUSINESS MACHINES CORP	United States	NONFIN
CSX CORP	United States	NONFIN	ING GROEP NV	Europe	FIN
CVS HEALTH CORP	United States	NONFIN	INGERSOLL-RAND INC	United States	NONFIN
DOMINION ENERGY INC	United States	NONFIN	INTESA SANPAOLO	Europe	FIN
MERCEDES-BENZ GROUP AG	Europe	NONFIN	ITV PLC	Europe	NONFIN

3 Stress-testing: a meta-factor approach

We propose an innovative approach based on “meta-factors” to perform stress tests on portfolios, using either historical or hypothetical economic scenarios. Stress testing in finance describes the capacity of a particular portfolio to withstand adverse conditions. Stress testing can be a useful complement to usual risk measures, such as volatility, VaR, and expected shortfall, since the estimated loss is linked to a specific event, which can be more meaningful than a summary statistic of a loss distribution. For convenience reasons, traditional stress tests are mostly based on historical scenarios, i.e., reproducing past events on current portfolio holdings. However, it lacks the desired forward-looking aspect of stress testing.

Our approach relies on few parameters, remains highly flexible, easy to implement, and helps to better understand portfolio sensitivities to systematic factors. An application on actively managed equity portfolios is conducted.

3.1 Introduction

We propose a simple and intuitive but innovative approach to stress test market value of portfolios relying on either historical or hypothetical economic scenarios while staying parsimonious to limit the risk of complexity and overfitting.

"Stress-testing" in finance has been adopted as a generic term describing various techniques to gauge the potential vulnerability of a portfolio to exceptional but plausible events. Stress-tests are part of the overall risk management policy, as specified by the UCITS[36] and AIFM[37] directives. Portfolio risk models typically calculate measures such as volatility, Value at Risk (VaR) or expected shortfall – summary statistics of the forecast return distribution. While these statistics help evaluate potential losses and identify the positions that contribute most to portfolio risk, they do not reveal how the losses might occur.

The key advantage of stress-tests is that they link a loss to a specific event, which can be more meaningful than a summary statistic of the loss distribution, especially in the context of delegated portfolio management with principal-agent relationship (Stracca, 2006). "Narrative stories" about risk can help portfolio managers to better explain risk to their clients. Stress-tests help identify and manage situations that can result in extreme losses. While banks increasingly use stress-tests, and banking regulators increasingly require them, this paper focuses on the use of market stress-tests in portfolio risk management, rather than bank balance sheet management or liquidity, credit, and counterparty stress-tests.

By enhancing our understanding of portfolio losses, stress-tests can be valuable at all stages of the investment process, including portfolio construction, limit setting, and hedging. Stress tests can be used at the different stages of a fund's life from creation to liquidation, via the

[36] Article 51 of Directive 2009/65/EC of the European Parliament and of the Council of 13 July 2009 (the "UCITS Directive"), and articles 38 and 40 of Commission Directive 2010/43/EU of 1 July 2010.
[37] Articles 15 and 16 of Directive 2011/61/EU of the European Parliament and of the Council of 8 June 2011 (the "AIFM Directive"), and articles 45 and 48 of Commission Delegated Regulation (EU) 231/2013 supplementing the AIFM Directive.

investment decisions made to manage it during its lifetime. Specifically, AMF Instruction DOC-2011-15 on the calculation of global exposure states that the "results of these tests should be [...] taken into consideration when making any investment decisions".

Traditionally, scenarios were designed using two methods - historical and hypothetical. Historical scenarios are based on events in the past, for example the subprime crisis of 2007-2009 or the Covid crash of 2020. Historical scenarios are easier to formulate and to understand intuitively. They are fully articulated, and they involve little judgment in implementation. They can be useful when some aspect of a historical scenario is expected to reoccur, and the scenario is of an appropriate magnitude.

However, such scenarios are backward looking and may lose relevance through time. This approach can obviously lead to questions about the portfolio's ability to deal with future crises. Clearly, there is nothing to suggest that future crises will be like previous ones. Moreover, it is difficult to precisely replicate past crises because data (for example in terms of the historic correlation of risk factors) are often incomplete, particularly going back more than a decade, and may produce very different scenarios.

Hypothetical scenarios consider plausible future developments. They allow a flexible formulation of an event and can use a mixture of elements—a shock from a previous historical event can be combined with other developments that never occurred. The advantage of such scenarios is that they can be tailored to be relevant to the risk profile of the portfolio. However, the building of a well-articulated hypothetical scenario can be a labour-intensive process, especially if the underlying model considers many factors, and it is important to understand the implicit assumptions made in scenario construction.

While model complexity helps to fit more precisely a stress event, it comes with the drawback of calibrating many inputs which ultimately might lead to less robust results. As complexity increases, interpretability, and efficiency decrease (Blaschke et al., 2001). A too much complex model can be well fitted for one for specific stress test scenarios but failed to be generalized.

Finally, stress tests can provide counterintuitive results (ESMA, 2019), that is being far from what could be reasonably expected knowing the portfolio exposures and their related stylized facts (e.g., stress-testing could show a value fund react negatively to increasing interest rates as this factor is historical associated to market turbulences while in fact value stocks benefit from higher rates).

Our objective is to propose a new method for stress testing market value of portfolios that permits both historical and hypothetical ("forward looking") scenarios while avoiding complexity that could lead to lack of robustness of results. Our strategy relies on a simple methodology which remains easy to understand and to update (avoiding a kind of black box) with the right balance between parsimony and customization. The proposed method considers economic factors, named meta-factors, that correlates with usual portfolio-based factors such as Fama-French factors. In our framework, historical or hypothetical scenarios of these meta factors are impacting portfolios through traditional Fama-French kind of factor models. For example, a 20% crash of oil prices will impact potentially value, growth, small capitalization stocks and as such the expected return of a portfolio. Our intuitive and simple method is based on a very limited sets of hypotheses or parameters, it limits the use of computational and data resources, and provides much more intuitive and coherent results relative to usual industrial risk tools.

Section 3.2 deals with the related literature and the existing models. Section 3.3 explains the methodological framework and Section 3.4 presents data and discusses specification of the model. Section 3.5 offers an empirical illustration on some equity funds before sharing concluding remarks.

3.2 Related literature

3.2.1 From the zoo of factors …

The perpetual quest for systematic factors has led to the discovery of an impressive number of potential ones, and recent papers try to navigate within the zoo of factors, questioning the

relevance of the supposed drivers of risk (Harvey et al., 2016; McLean and Pontiff, 2016; Hou et al., 2020, etc.). McLean and Pontiff (2016) used an out-of-sample approach to study the post-publication bias of 97 discovered risk anomalies and concluded that investors learn about mispricing from academic publications, based on the empirical facts that portfolio returns are 26% lower out-of-sample and 58% lower post-publication, with the out-of-sample decline being an upper bound estimate of data mining effects. Harvey et al. (2016) found that most factors in the zoo are likely false, by adjusting critical values for statistical inference, given the underlying intensive data mining. This is like the conclusion of Hou et al. (2020) who consider a large set of anomalies and illustrate the p-hacking (i.e., manipulation of data or estimation methods in a way that generates desired p-values) concern through anomaly replication. See also Green et al. (2017) and Feng et al. (2020). The decline in factor performance following their publications can also be explained by crowding. This refers to the phenomenon where all investors play the same factor, resulting in a decrease in its performance. Arnott et al. (2019) specifically suggest that after 2003, it is likely that the returns of popular factors decreased due to crowding. Other studies also tend to demonstrate the existence of crowding (Kang et al., 2021; Baltas, 2019). However, crowding is difficult to identify due to the lack of an effective measure, as well as the fact that the reasons for underperformance of a factor after publication can be explained by the aforementioned factors such as data mining and p-hacking.

In this zoo of factors, the identification of relevant factors poses yet another challenge. In addition to the multitude of factors or anomaly, the question of the relevance and selection of "observable" variables is complex. Bernanke et al. (2005) discusses the difficulty of representing certain factors with observable variables, using the example of economic activity.

Which variable best represents this concept? Industrial production, GDP... these are merely proxies and do not fundamentally capture economic activity. Beyond that, the observable "fundamental" variables in the case of stock returns may have the disadvantage of being delivered at lower frequencies, rarely changing, and exhibiting correlation among them.

Thus, literature has focused on using more statistically oriented models in which the factors are directly inferred from the data and has then explore the relationships between observable and latent factors. Indeed, the latter, although often more suitable for explaining returns, are difficult to interpret (Pukthuanthong et al., 2019; Parker and Sul, 2016). This literature is primarily based on the notion that the true factor is a latent one. To identify the "true factors", Pukthuanthong et al. (2019) propose an interesting solution. The best observable factors are those canonically correlated with the eigenvectors of the variance/covariance matrix of returns. The combination of observable variables and latent factors can be the solution, although the economic significance behind the latent factor remains challenging to explain.

Jushan Bai and Serena Ng propose in their paper "Determining the number of factors in approximate models" (2002) a criterion for determining the optimal number of factors for large-scale models in terms of cross-section and time, which is often the case when trying to explain returns. The problem is that these criteria, although playing a major role in literature, have two drawbacks. Firstly, they are suitable for static models and assume stable betas. As a result, they completely ignore that factor sensitivities vary over time. Baltagi, Kao, and Wang (2017) noticed that, assuming stable betas, the criterion overestimates the number of factors, since an unstable beta factor model can be represented by the same stable beta factor model with an additional factor. Thus, they show that to account for structural instabilities, the number of model factors needs to be recalculated each time there is a breakpoint. However, this also complicates the model and makes its estimation more challenging. Secondly, this criterion is effective when the period is very large (the authors assume it tends to infinity in their papers). However, over such time periods, the optimal number of factors can also vary. For example, such criteria demonstrate little robustness during periods of crisis. The discontinuity in data linearity can lead to the emergence of a new factor that is not truly significant. Since then, other techniques have been proposed to estimate the optimal number of factors for dynamic models (Bai and Ng, 2007) or static models (Onatski, 2010). However, these methods often lack robustness when compared to each other.

3.2.2 … to parsimonious economic meta-factors

Fama and French (1996) suggest that we could interpret their factors as proxies for state variables, whose innovations describe the investment opportunity set. Huberman and Kandel (1987), for example, interpret the HML factor as a proxy for profitability. Weak firms with long periods of low profits tend to have high book-to-market and positive loadings on HML. Investors demand higher expected returns to buy these stocks. The opposite happens to strong firms, with long periods of high profits. There was already evidence of this in Chan and Chen (1991). Chen and Zhang (1998) presented a similar result a few years later. Subsequently, HML's ability to explain expected returns might be confused because the current value of the stock is correlated simultaneously with operational profitability and investment (Fama & French, 2015).

Many other interpretations have been made in the last decade, and most of them provide support for the risk-based explanation. Jagannathan and Wang (1996) relate the FF factor to the return on human capital and the business cycles using a conditional version of CAPM, which resembles the multi-factor model of Ross (1976). The authors show that when the conditional version of CAPM holds, a two-factor model obtains unconditionally. They also use the spread of corporate bonds as a complement to market returns in predicting future economic conditions. Lettau and Ludvigson (2001) and Vassalou (2003) show that the presence of measures of macroeconomic risk reduces the explanatory power of SMB and HML, which suggests that the FF factors are partly embodied in the macroeconomic variables. Petkova (2006) applies shocks to the aggregate dividend yield and term spread, default spread, and one-month T-bill rate and shows that they are proxies for HML and SMB factors. She shows that a model with innovations to these variables performs better than the original three-factor model in explaining the cross-section of portfolio returns. Gulen, Xing, and Zhang (2011) study the flexibility of value and growth companies in adapting to bad economic conditions. Also, Perez-Quiros and Timmermann (2000) indicate that small firms have high average returns because they are more susceptible to changes in credit market conditions. Since the availability of credit is tied to underlying economic factors, the higher

returns to small firms can be viewed as a compensation for the high sensitivity to a credit-related risk factor. The relationship between FF factors and economic state variables is further supported by Liew and Vassalou (2000) who show that the FF factors can predict future GDP growth in some countries. In a later study, Vassalou and Xing (2004) conclude that size and BM effects are related to default risk and can be viewed as default effects. In the same line, Hahn, and Lee (2006) find that changes in term and default yield spreads capture most of the systematic risks proxied by size and BM effects. Simpson and Ramchander (2008) show the ability of the FF three-factor model to capture information related to a number of macroeconomic variables, such as personal consumption, the consumer price index (CPI), retail sales, factory orders. All these results support the hypothesis that SMB and HML act as economic state variables of the ICAPM.

3.2.3 Survey of industry risk models

We compare three well-known financial services providers, Bloomberg, Barra and Riskmetrics. Both, Bloomberg, and Barra, develop stress-testing tools using multi-factor models while Riskmetrics prefers a package of risk measurements used to model the risk in a portfolio induced by a change in the value of risk factors.

A simple approach of defining a portfolio risk is to compute its sample covariance matrix from historical asset returns. But, following this approach suffers the so-called curse of dimensionality as it requires to estimate $(N^2 - N)/2$ parameters for each portfolio, where N is the number of assets within a portfolio. Factor models offer a solution as we would only need to estimate N x K sensitivities where K is the number of factors per portfolio. Factor models also allow to disentangle systematic risk (factor returns) from residual risk in a simple way.

Hence, the financial industry relies heavily on three types of factor models: explicit factor models where factor returns are known, latent factor models where factors are unobserved and statistically determined (often through a principal component analysis), and implicit factor models where exposures (sensitivities) are known but factor returns are estimated using cross-section regression. Bloomberg and Barra build their methodology on this last form of factor

models. While implicit factor models offer robust estimates of factor returns, they are data-intensive and lack parsimony.

Bloomberg latest equity risk model called MAC3 GRM (Global Risk Model) was developed in 2020[38]. While the model serves as a general risk management tool, it can also be used to forecast risks or stress-testing. It is a multi-factor model which aims to estimate a covariance matrix. An asset return in the MAC3 GRM model is given by : $R = \beta f + \epsilon$, where R is a vector of T asset returns (observed) , β is the factor loadings matrix (betas : predetermined), f is a vector of K factor returns (unobserved) and ϵ is the residual vector representing specific risk, thus the total asset covariance matrix can be estimated, given the factor loadings matrix, the factor covariance matrix and the specific risk matrix.

There are 3 types of factors retained in the MAC3 GRM equity model: style factors, country factors and industry factors. The aim of introducing country and industry factors is to increase the explanatory power of the model and remove the spurious correlation between factor returns and residual returns.

The estimation universe of Bloomberg model corresponds to the full Bloomberg database for which are applied some filters. The style factor loadings matrix is estimated using time series regression for each stock of the estimation universe. Style factors include Market, Volatility, Momentum, Size, Mid-cap, Earnings yield, Valuation, Dividend, Long term reversal, Liquidity, Growth, Variability and Leverage. Style factor loadings are computed in the form of z-scores, such that the weighted mean (mean over stocks and not over time) of these z-scores is 0.

Once the factor loadings matrix is estimated, it is plugged into the MAC3 GRM model. The factor returns f_k are estimated using cross-sectional regression. In applying this regression, the MAC3 GRM model uses a particular weighting scheme: stocks with more residual variance are underweighted in the regression whereas stocks with less noise are overweighted.

[38] MAC3 Global Equity Risk Model, Methodology Notes, Menchero et al., 2020.

The weights of the observations in the cross-sectional regression are thus the inverse of the residual variance for each stock, the residual variance being defined as the volatility of residual returns estimated from time-series regression. Note that the process explained above is used to estimate factor returns for only one date. To get a whole series of factor returns, we should repeat the process, reason why this type of model is data intensive.

Barra also computes factor exposures the same way, only that the factors considered are different. Barra does not explicitly mention the factors but describes the methodology to select them. Barra regresses the asset returns against each factor and tests their statistical significance. Those statistically significant are put together in a regression model and then the remaining are added one by one if they add to the explanatory power.

Barra and Bloomberg methodologies slightly diverge in deriving the factor covariance matrix. Barra estimates the factor covariance matrix in 2 steps. Given the factor return series estimated, the first step is to compute the sample covariance matrix of these factors (using an exponential weighting scheme). Secondly, the factor covariance matrix is scaled such that the portfolio volatility estimated through a GARCH equals the portfolio volatility estimated from the factor model. Bloomberg blends two approaches when computing the factor covariance matrix to account for serial correlations. The first approach consists of simply aggregating returns to the desired period in question and estimating the covariances using an exponential weighting scheme. The second approach is based on the Newey-West covariance matrix.

A stress scenario can then be applied either to replay a historical event by applying factor returns from an explicit historical period that is specified or specify some changes in market variables, then propagate these changes to other factors based on the correlations among the historical returns of the changing variables.

As already mentioned, Riskmetrics is not a multi-factor model, but a package of risk measurements used to model the risk in a portfolio induced by a change in the value of risk factors.

Riskmetrics defines four risk factors — equity prices, exchange rates, commodity prices, and interest rates — and makes two important assumptions: risk factor returns are normally distributed, and risk factor volatilities are best estimated by exponentially weighting past returns. Riskmetrics follows two approaches: use of historical data and assumption on the probability distribution of the risk factor. This allows Riskmetrics to define the scenarios used to model the portfolio returns. Monte Carlo methods and parametric methods are used to generate risk factor scenarios.

Riskmetrics relies on three types of scenarios: historical scenarios, user-defined simple scenarios and used-defined predictive scenarios. The difference between simple and predictive scenarios is that simple scenarios do not consider the impact (correlations) of stressed risk factors (the factors which value has been changed and which are called core factors) on the non-stressed risk factors (peripheral factors). Predictive scenarios are more realistic as they consider the correlation between variables.

Barra also proposes other methods [39] such as reverse stress-testing to define which scenarios can lead to a certain loss level. The paper explains the idea of correlated reverse stress tests, using a multi-factor model: factors in this model are considered as shocks (scenarios) and the aim is to determine the expected loss on each factor (shock) given an X total loss in the portfolio, and a factor correlation matrix. Once the loss on each factor (shock) are determined, a probabilities is assigned to each shock. Barra recommends considering only the change in the levels of variables and to consider the second moment namely the volatility and correlations as correlations are likely to rise in periods of high volatility. They also make the distinction between sensitive and scenario analysis. Sensitivity analysis corresponds to the impact of the change of one factor, whilst scenario analysis considers the impact of change of multiple factors simultaneously and considers movements in levels of variables as well as movements in their relations (through volatilities and correlations). Scenario analysis is more complete and closer to reality.

[39] Stress Testing in the Investment Process, Ruban and Melas, 2010.

3.3 Methodology

3.3.1 Definitions

Meta-factor refers to a higher-level or overarching factor that exerts a significant influence on multiple individual factors within the financial system. It explains common risk exposures of other factors and can be viewed as a factor of factors. By nature, a meta-factor is defined as a macroeconomic or financial variable that impacts in a certain degree all markets, sectors, and financial instruments.

More precisely, we expect meta-factors to be intuitive, to have a strong economic value (highly recognized financial or macroeconomic variable) and to explain the variance of systematic factors affecting asset returns. In this paper, they are higher-level factors that capture distinct characteristics or attributes of the underlying systematic factors.

A systematic factor or a risk factor, is a fundamental driver of risk and return that affects a broad set of financial assets or portfolios. It represents a pervasive and persistent influence on asset prices or investment performance that extends beyond the idiosyncratic risk.

3.3.2 Meta-factors selection

In a first step, we use a clustering approach to reduce the list of available meta-factors from a large base of macroeconomic and financial variables. The set of selected meta-factors must be parsimonious, sound, straightforward and unambiguous. The set must cover the main areas shaping the financial markets. We categorize the variables into five groups: macroeconomics, equity, credit, commodities, and currencies. For more convenience, we limit the macroeconomics variables to the US region and assume that their European equivalents exhibit the same characteristics.

In a second step, we select meta-factors by establishing linkages with systematic factors, drawing from the existing literature (Bass et al., 2017; Clark, 2022; Swade et al., 2022). Finally, it must be easy and coherent to apply a stress scenario to the set of selected meta-factors.

3.3.3 Systematic factors selection and modelling

We consider the usual extended Fama French factors as systematic factors: Market, Value/Growth, Size, Momentum, Quality and Volatility which are largely used in the financial industry and academic literature.

We propose a two-step methodology to stress-test assets through innovations in meta-factors. In a first step, meta-factors are regressed against systematic factors to estimate the factor loadings. In a second step, systematic factors are regressed against the assets we want to stress.

Let $\hat{f}_{t,t+\tau} = \{f_{s,t,t+\tau}, s = 1, \dots, S\}$ be the set of innovations in the S systematic factors estimated between time t and $t + \tau$ where τ is the length of the stress in days and t the length estimation window in days, and m is the frequency of the calibration variables (e.g. m = 30 days for a monthly frequency). The estimation of the innovation in the systematic factor s between t and $t + \tau$ is given by:

$\hat{f}_{s,t,t+\tau} = \mathrm{E}(f_{s,t,t+\tau} | F_{t,t+\tau})$ where $F_{t,t+\tau} = \{F_{k,t,t+\tau}, k = 1, \dots, K\}$ is the set of innovations (shocks) in K meta-factors between t and $t + \tau$.

We assume the following representation of the innovation of each systematic factor s between t and $t + \tau$:

$$\hat{f}_{s,t,t+\tau} = \mu_{s,t} \times \frac{\tau}{m} + \sum_{k=1}^{K} \beta_{k,t}^{s} F_{k,t,t+\tau} + \epsilon_{s,t} \quad (3.1)$$

Then, let $\hat{r}_{i,t,t+\tau} = \mathrm{E}\left(r_{i,t,t+\tau} \middle| F_{t,t+\tau}, \hat{f}_{t,t+\tau}\right)$ be the estimation of the return of the stressed asset i between t and $t+\tau$.

We assume the following representation of the return of each stressed asset i between t and $t+\tau$:

$$\hat{r}_{i,t,t+\tau} = \alpha_{i,t} \times \frac{\tau}{m} + \sum_{s=1}^{S} \beta_{s,t}^{i} \hat{f}_{s,t,t+\tau} + \epsilon_{i,t} \quad (3.2)$$

From (1) and (2), we can write:

$$\hat{r}_{i,t,t+\tau} = \alpha_{i,t} \times \frac{\tau}{m} + \sum_{s=1}^{S} \beta_{s,t}^{i} \left\{ \mu_{s,t} \times \frac{\tau}{m} + \sum_{k=1}^{K} \beta_{k,t}^{s} F_{k,t,t+\tau} + \epsilon_{s,t} \right\} + \epsilon_{i,t} \quad (3.3)$$

Hence, between t and $t+\tau$ the return of the stressed asset i is given by:

$$\hat{r}_{i,t,t+\tau} = \Phi_{i,t} + \Lambda_{i,t,t+\tau} + \Psi_{i,t} \quad (3.4)$$

Where,

$$\Phi_{i,t} = (\alpha_{i,t} + \sum_{s=1}^{S} \beta_{s,t}^{i} \mu_{s,t}) \times \frac{\tau}{m}$$

is the constant component of the estimated asset return.

$$\Lambda_{i,t,t+\tau} = \sum_{s=1}^{S} \sum_{k=1}^{K} \beta_{s,t}^{i} \beta_{k,t}^{s} F_{k,t,t+\tau}$$

is the cross-factor sensitivity component or simply the stress applied to the estimated asset return.

$$\Psi_{i,t} = \epsilon_{i,t} + \sum_{s=1}^{S} \beta_{s,t}^{i} \epsilon_{s,t}$$

is the residual component of the estimated asset return.

One can generalize the equation (3.4) to a set of N assets using the equation in the matrix form:

$$\hat{R}_{t,t+\tau} = \Phi_t + \Lambda_{t,t+\tau} + \Psi_t \quad (3.5)$$

Where Φ_t, $\Lambda_{t,t+\tau}$ and Ψ_t are vectors of size N.

We note $\mathrm{B}_t^{\mathrm{S}} = \begin{pmatrix} \beta_{1,t}^{1} & \cdots & \beta_{K,t}^{1} \\ \vdots & \ddots & \vdots \\ \beta_{1,t}^{S} & \cdots & \beta_{K,t}^{S} \end{pmatrix}$ the $(S \times K)$ matrix of sensitivities between systematic factors and meta-factors common to each asset and

$\mathrm{B}_t^{\mathrm{N}} = \begin{pmatrix} \beta_{1,t}^{1} & \cdots & \beta_{S,t}^{1} \\ \vdots & \ddots & \vdots \\ \beta_{1,t}^{N} & \cdots & \beta_{S,t}^{N} \end{pmatrix}$ the $(N \times S)$ matrix of sensitivities between assets and systematic factors.

Hence, we need to estimate $(N \times S) + (S \times K)$ parameters to apply a stress scenario to a whole set of assets between t and $t + \tau$ (excluding the constants). With,

$$\Phi_t = A_t + \mathrm{B}_t^{\mathrm{N}} \mu_t$$

$$\Lambda_{t,t+\tau} = \mathrm{B}_t^{\mathrm{N}} \mathrm{B}_t^{\mathrm{S}} F_{t,t+\tau}$$

$$\Psi_t = \mathrm{E}_t^{\mathrm{N}} + \mathrm{B}_t^{\mathrm{N}} \mathrm{E}_t^{\mathrm{S}}$$

one can rewrite (3.5) as the equivalent of (3.3):

$$\hat{R}_{t,t+\tau} = \mathrm{A}_t^{\mathrm{N}} + \mathrm{B}_t^{\mathrm{N}} (\mu_t + \mathrm{B}_t^{\mathrm{S}} F_{t,t+\tau} + \mathrm{E}_t^{\mathrm{S}}) + \mathrm{E}_t^{\mathrm{N}} \quad (3.6)$$

With $\mathrm{A}_t^{\mathrm{N}}$ the N-vector of $\alpha_{i,t} \times \frac{\tau}{m}$, μ_t the S-vector of $\mu_{s,t} \times \frac{\tau}{m}$, $F_{t,t+\tau}$ the K-vector of $F_{k,t,t+\tau}$, $\mathrm{E}_t^{\mathrm{S}}$ the S-vector of $\epsilon_{s,t}$ and $\mathrm{E}_t^{\mathrm{N}}$ the N-vector of $\epsilon_{i,t}$.

3.3.4 Stress scenario applications

After having estimated the parameters of the model at time t, one can use this procedure to apply a stress by specifying the innovation in each of the meta-factor according to the macroeconomic scenario.

One can also shock only one meta-factor to isolate its impact on the asset returns or use the meta-factors covariance matrix to diffuse the shock from one meta-factor to others. This methodology allows to diffuse the shock in several risk regimes or to shock a financial variable that would not be a meta-factor and diffuse the shock to them using covariances.

Despite its simplicity, the model is highly flexible and permits the application of a combination of shocks that has never been observed in the past in a forward looking way. But the procedure also permits to apply a historical scenario by specifying the shock observed on the meta-factors at that time. One might argue that asset could be regressed directly against meta-factors, but this step would come with a significant drawback. Indeed, it would be cumbersome or very approximate to estimate a potential change of sensitivities between the asset and the meta-factors while it is easier to make assumptions about the change of the asset sensitivities with the systematic factors (e.g., a fund manager switches her style management

from growth to value). The model offers a large degree of flexibility for the estimation of the $(N \times S) + (S \times K)$ parameters (ex-post, ex-ante, penalization, etc.).

3.4 Data

3.4.1 Meta-factors

The dataset contains 58 variables from February 2003 to June 2023 at a monthly frequency. We use a K-means clustering algorithm to partition the dataset into five distinct groups. Then, we refine the dataset by excluding variables that are too far from their corresponding cluster. 20 variables are excluded.

A dendrogram further helps to identify clusters. By visually analysing the dendrogram, we can identify clusters of closely related variables that tend to exhibit similar behaviour or patterns. Consequently, based on this dendrogram we remove some clusters with low information and keep only relevant variables for each cluster.

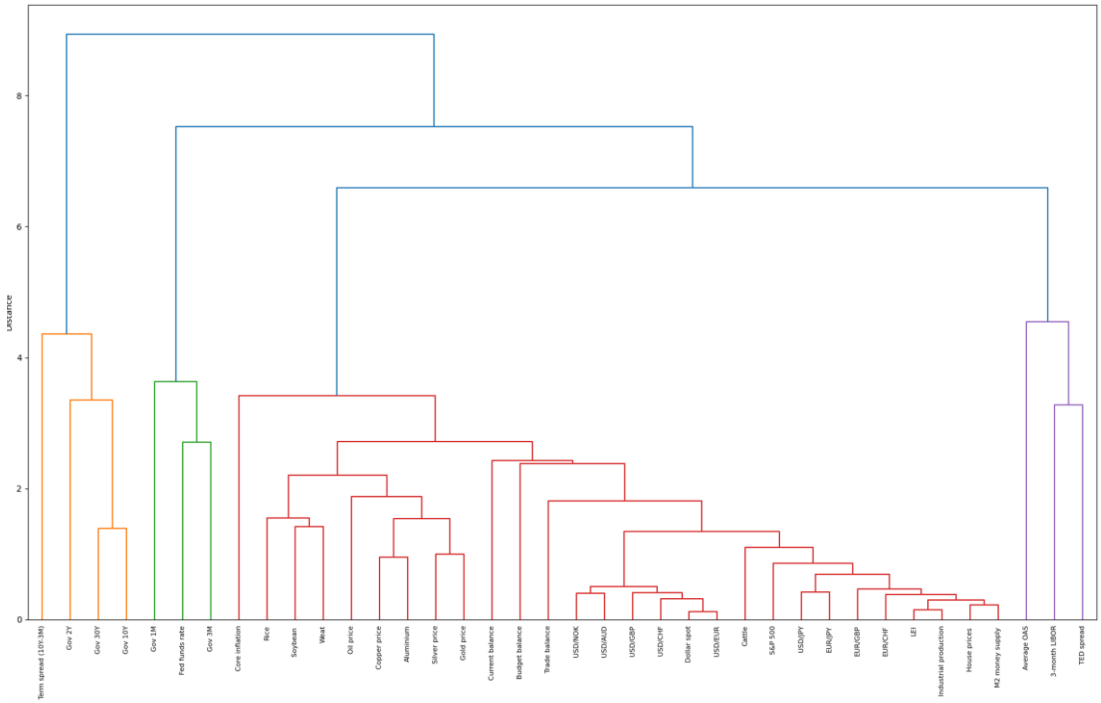


Figure 3.1 Dendrogram of meta-factor candidates

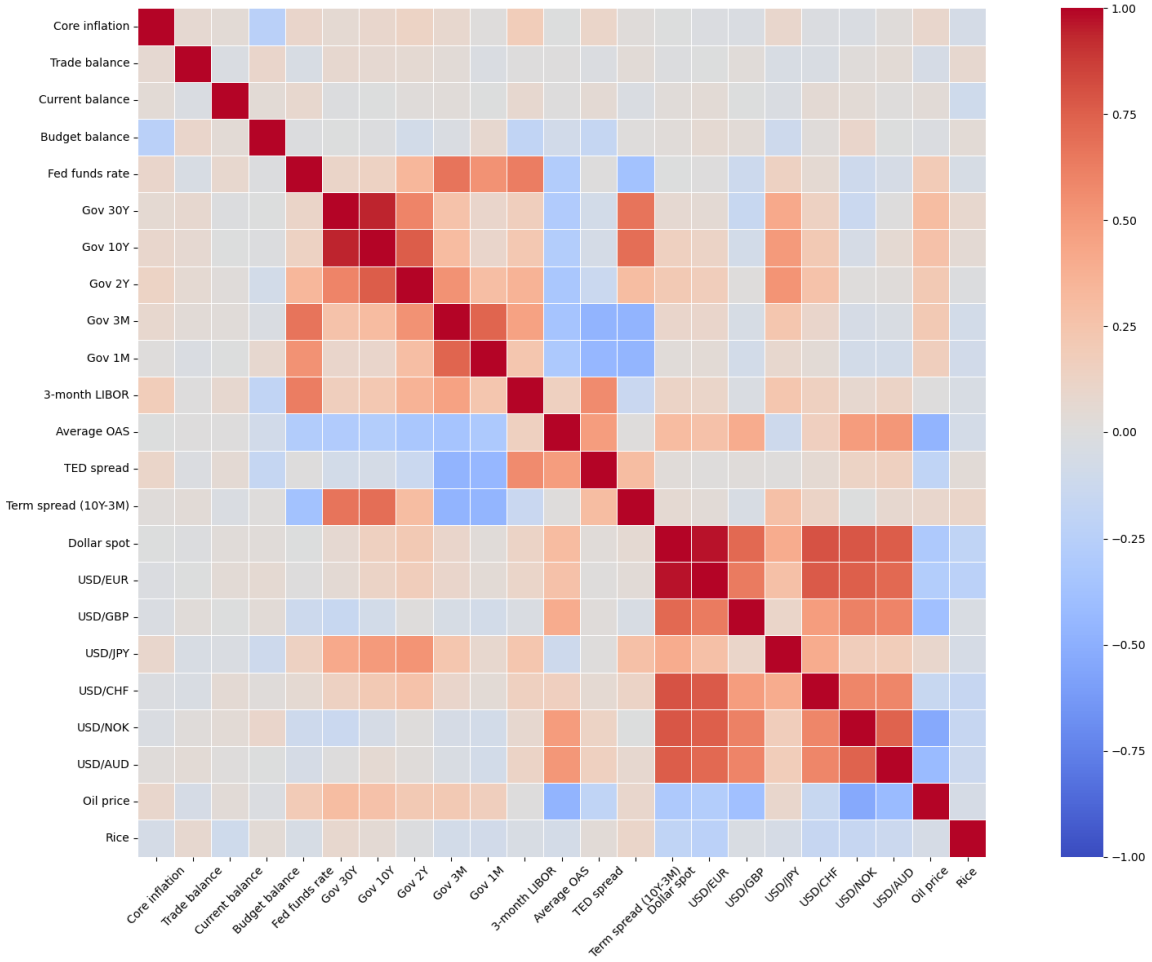


Figure 3.2 Historical correlation matrix between meta-factor candidates

Based on these results, considering the existing literature and the objective of the paper, we select five variables as meta-factors: the average option-adjusted spread, the term spread (10-year minus 3-month), the US dollar value, the oil price, and the two-year sovereign yield. These variables show the desired characteristics, and they seem to be complementary from each other. We can easily shape a stress scenario from this set of variables.

Option-adjusted spread

The option-adjusted spread (OAS) measures the impact of embedded options, such as call or put options, on the value of a fixed-income instrument. These options provide the issuer or the bondholder with the right to exercise certain actions, such as early redemption or conversion,

which can affect the bond's cash flows and overall value. The OAS adjusts the yield spread to account for the value of these options, providing a more accurate measure of the instrument's relative value. As the credit risk is higher, the price of the option increases hence the OAS.

The OAS considers market expectations regarding future interest rates, volatility, and other relevant factors. It captures the consensus view of market participants on the additional yield investors demand to compensate for the uncertainty associated with embedded options. By incorporating market expectations, the OAS provides a more accurate reflection of the economic value of the security. It captures the compensation investors require for taking on the potential cash flow volatility resulting from changes in interest rates or other factors.

The OAS is similar to other measures of implied volatilities such as the VIX Index. Both the OAS and the VIX are designed to capture market expectations and sentiment. But the OAS exhibits more stability and reflects credit risk. We often observe increasing equity implied volatility without an increase of credit risk.

We use the Bloomberg Aggregate Corporate Average OAS (euro and US) as a proxy for the overall credit risk.

Term-spread

We consider the term spread between the 10-year and 3-month Treasury yields. This variable is an important economic indicator and is closely watched by investors, economists, and policymakers. It provides insights into the yield curve slope.

A positive term spread typically occurs during periods of economic expansion when investors expect higher future inflation and higher yields on longer-term investments. A wider term spread is often associated with a healthy economic outlook and can be seen as a positive signal for economic growth.

Conversely, a negative term spread, also known as an inverted yield curve is often interpreted as a signal of potential economic weakness or an impending economic downturn. An inverted yield curve has historically been associated with recessions and is closely monitored as a potential indicator of future economic conditions.

US Dollar value

The US dollar value as a meta-factor seems an obvious choice for several reasons. The US dollar is a key consideration in assessing macroeconomic trends and expected asset returns:

- A global reserve currency: the US dollar is widely recognized as the dominant global reserve currency.
- International trade and investment: the US dollar plays a crucial role in international trade and investment activities. Most commodities, such as oil, gold, and major agricultural products, are priced and traded in U.S. dollars.
- Safe-haven status: during times of economic uncertainty the flight to safety strengthens the dollar and can impact currency exchange rates, asset prices, and global capital flows.
- Financial market dominance: US financial markets are the largest and most liquid in the world. The depth and breadth of these markets make the US dollar a preferred currency for conducting global financial transactions.

We use the DXY Index to proxy the value of the dollar. The DXY Index is a measure of the value of the US dollar relative to a basket of other major currencies. It provides a weighted average of the dollar exchange rates against six currencies: euro, Japanese yen, British pound sterling, Canadian dollar, Swedish krona, and Swiss franc. The DXY is often used as a benchmark to gauge the overall strength or weakness of the US dollar in the foreign exchange market.

Oil price

Oil price can be considered as a good meta-factor. Oil price has a significant impact on various aspects of the economy and financial assets:

- It is a key input in various industries. Fluctuations in oil prices can reflect changes in global demand and supply dynamics. Changes in oil prices affect industries that are closely tied to energy costs, but they also have less direct effects on all sectors.
- It affects consumer spending. Oil prices directly impact the cost of transportation, production, and distribution, influencing the prices of goods and services. Higher oil prices can contribute to inflationary pressures, affecting consumer spending power and business costs. Changes in inflation expectations can impact bond yields and interest rates, which can have implications for equity valuations and investor sentiment.

We use the price of the Brent as a measure of oil price.

Two-year sovereign yield

Finally, interest rate risk is represented by the two-year sovereign yield as a meta-factor.

The 2-year yield is closely watched as an indicator of market expectations regarding short-term interest rates and monetary policy. Changes in the 2-year yield can reflect shifts in market sentiment about the overall health of the economy, inflation expectations, and the outlook for future interest rate movements. Increasing borrowing costs for businesses affect their profitability and investment decisions. To some extent, higher interest rates can also make fixed-income investments more attractive compared to equities, leading some investors to reallocate their portfolios away from stocks, potentially putting downward pressure on equity prices.

The 2-year yield might also be used as a risk-free rate. An increasing risk-free rate leads to a higher discount rate applied to future cash flows of equities. This results in a decrease in equity valuations, as higher discount rates reduce the present value of expected future earnings.

3.4.2 Systematic factors

In contrast to asset returns, factor returns cannot be directly observed. We rely on regional MSCI indices (US and Europe) as proxies of systematic factors. Based on empirical evidence, these indices proved to be good candidates to proxy the systematic factors.

Market factor

As mentioned on the early CAPM-related literature, the Market factor can be viewed as the first and most important equity factor. Alone, the Market factor already explains a large amount of an equity fund variance. It represents the overall direction and performance of the overall market or a specific market index, such as the S&P 500 or a global equity index. It captures the influence of macroeconomic factors, investor sentiment, and broad market trends. The Market factor is highly correlated with corporate earnings. We use the MSCI USA and the MSCI Europe Index as proxies of the Market factor for the US and Europe region respectively. We will use also regional indices for other factors.

Value/Growth factor

The Value factor reflects the tendency for assets with lower valuations (e.g., low price-to-earnings ratio or price-to-book ratio) to generate higher returns compared to assets with higher valuations while the Growth factor represents the performance of companies or assets that exhibit strong earnings growth, sales growth, or other indicators of future growth potential.

The abnormal return associated with the Value factor has been consistently observed by researchers in various major securities markets, as documented by Hawawini and Keim (2000).

However, it is worth noting that critics of the value premium raise concerns about the potential influence of data mining on empirical evidence. They emphasize that empirical studies often rely on specific samples, as highlighted by Black (1993).

We build a long-short portfolio MSCI Value Index – MSCI Growth Index to proxy this factor. MSCI uses the book to price, earnings to price, book value, sales, earnings, cash earnings, net profit, dividends, and cash flow of individual stocks to construct these indices.

Size factor

The Size factor captures the impact of company size on investment returns. It suggests that smaller companies (by market capitalization) tend to outperform larger ones.

The small-cap premium persists even after accounting for other factors. Various theories have been proposed to explain this phenomenon, and the debate surrounding it continues. In the efficient market view, Fama and French (1992, 1993) originally suggested that small-cap stocks carry higher systematic risk, leading to a higher return premium. Subsequent research has proposed that size may proxy for other unobservable risk factors associated with smaller firms, such as liquidity (Amihud, 2002), information uncertainty (Zhang, 2006), financial distress (Chan and Chen, 1991), and default risk (Vassalou and Xing, 2004). Chan, Chen, and Hsieh (1985) argue that the spread between low and high-quality corporate bonds, reflecting the macro environment and default risk, is the most significant factor in explaining the size effect.

We build a long-short portfolio MSCI Small Index – MSCI Large Index to proxy this factor.

Momentum factor

The Momentum factor captures the persistence of price trends and the tendency for assets with positive price momentum to continue performing well in the short to medium term.

The momentum premium lacks a widely accepted efficient markets-based theory to explain its existence. Instead, behavioural theories are often invoked to account for this phenomenon. These theories suggest that investor overreaction (Barberis, Sheleifer, and Vishny 1998; Daniel, Hirshleifer, and Subrahmanyam 1998) or underreaction to news (Hong, Lim, and Stein 2000) play a role in driving the momentum effect. These behavioural biases stem from factors such as overconfidence, self-attribution, conservatism bias, aversion to realizing losses, representative heuristic, and limited analyst coverage. Additionally, theories such as herding behaviour driven by reputation concerns (Dasgupta, Prat, and Verardo 2011) have also been put forward. Other frameworks focus on institutional investing dynamics as a source of both momentum and value effects (Vayanos and Woolley 2011).

We build a long-short portfolio MSCI Momentum Index – MSCI Index (Market) to proxy this factor. We sell the Market factor to avoid any redundancy. MSCI uses relative returns (3-mth, 6-mth, 12-mth) and historical alpha of individual stocks to construct this index.

Quality factor

The Quality factor captures excess returns to stocks that are characterized by low debt, stable earnings growth, and other "quality" metrics.

The literature exploring the reasons behind the effectiveness of the Quality factor remains limited, mainly due to the lack of a standardized definition for quality. In accordance with the Fama-French model, it is believed that all systematic risks ultimately stem from economic risks. This viewpoint establishes a connection between stock returns and quality measures such as financial leverage and earnings growth. Campbell, Polk, and Vuolteenaho (2010) put

forth a "fundamental" explanation, suggesting that the primary source of systematic risks for both growth and value stocks can be traced back to cash flow fundamentals rather than market sentiments. From a corporate finance perspective, the influence of firm quality on stock prices becomes apparent. Well-managed companies exhibit careful capital management practices, mitigating the risk of excessive leverage or over-capitalization. Consistent earnings growth reduces the reliance on external financing, which contributes to supporting stock prices. This positive feedback loop strengthens the company's competitiveness in the eyes of both customers and investors.

For the same reason as for the Momentum factor, we build a long-short portfolio MSCI Quality Index – MSCI Index (Market) to proxy this factor. MSCI uses the return on equity, earnings stability, dividend growth stability, strength of balance sheet, financial leverage, accounting policies, strength of management, accruals, and cash flows of individual stocks to build this index.

Volatility factor

The volatility factor captures excess returns to stocks with lower-than-average volatility, beta, and/or idiosyncratic risk.

The low volatility anomaly directly challenges the principles of the Efficient Market Theory and the assumptions of the Capital Asset Pricing Model. To explain this phenomenon, behavioural factors are commonly put forth. One prevalent explanation is the presence of the "lottery effect," whereby investors demonstrate a tendency to engage in bets that offer a small expected loss but a potentially substantial gain, even though the probability of a loss is significantly higher than that of a win. This behaviour resembles the act of purchasing a lottery ticket, where individuals pay a small amount in the hopes of winning a large sum, despite the low likelihood of success. It is argued that this behavioural bias leads investors to overvalue high volatility stocks and undervalue low volatility stocks, reflecting an "irrational" preference for volatile stocks.

Again, we build a long-short portfolio MSCI Minimum Volatility Index – MSCI Index (Market) to proxy this factor. MSCI uses the standard deviation (1-yr, 2-yrs, 3-yrs), downside standard deviation, standard deviation of idiosyncratic returns and beta of individual stocks to build this index.

3.5 Application

This section illustrates our methodology to a real dataset of assets. We apply the methodology to a set of equity long only mutual funds managed by a European asset manager. The set is diversified enough to observe how the funds with different style exposures would react to stress scenarios. These mutual funds show more a less aggressive biases and are solely exposed to European or US stocks. We apply different stress scenarios, first we show the results for hypothetical scenarios applied to the dataset with either specified shocks on all meta-factors or shock on one meta-factor. Then, we show results for the reproduction of historical scenarios. When possible, we compare results to Bloomberg MAC 3.

3.5.1 Parameters calibration

The dataset contains 6 active mutual funds invested in US stocks and 8 active mutual funds invested in European stocks. Mutual funds return sensitivities with systematic factors returns are computed at a monthly frequency from inception. For confidentiality reasons, mutual funds are kept unnamed. Table 3.1 displays sensitivities with systematic factors and highlights the high level of diversity across the dataset. For instance, one would expect a value fund to react differently than a growth fund to an increase of interest rates or more aggressive funds to react strongly to a stress event compared to more defensive ones. Sensitivities are calibrated using a multi-linear regression as described in the methodology section over the whole life of the fund (Full) but without the constant to ease the interpretation and be free of specification of the length of the shock (τ). We also calibrate parameters under different risk regime windows using the VIX level as a threshold. We expect stressed assets to react differently in low-risk regime (Low: VIX < 30) and high-risk regime (High: VIX > 30).

Table 3.1 Equity long only mutual funds sensitivities to systematic factors (Full)

	Market	Value-Growth	Small-Large	Quality	Momentum	Min. Vol.
US tilt Value	0.95***	0.18**	0.24*	0.17	-0.12	-0.09
US Blend	0.89***	0.07	-0.12*	0.22**	-0.02	0.01
US Aggressive Growth	0.89***	-0.41***	0.04	0.14	0.04	-0.28**
US Aggresive Value	1.1***	0.87***	0.62**	-0.49***	-0.25**	-0.76***
US Value	1.01***	0.44***	0.05	-0.08	0.15	-0.11
US Defensive Value	1.08***	0.31**	-0.07	-0.52***	-0.16	0.38**
EU Aggresive Value	0.94***	0.41***	0.27**	-0.66***	-0.11	-0.49**
EU Value	1.02***	0.16**	0.16	-0.78***	-0.22*	-0.34**
EU Growth	0.94***	-0.48***	0.19	-0.33**	-0.08	-0.14
EU tilt Value	0.91***	0.05	0.19	-0.48**	-0.23**	-0.11
EU Small Cap	0.91***	-0.21**	0.58***	-0.16	-0.18*	-0.01
EU Blend	0.99***	-0.14*	0	-0.04	-0.15*	0.08
EU Defensive Growth	0.99***	-0.46**	-0.02	-0.54**	0.12*	0.13
EU Defensive Blend	0.7***	-0.26*	0.1	0.84***	-0.12*	-0.41**

Notes: *, **, *** indicate rejection of the null hypothesis of sensitivities equal to 0 at 10%, 5% and 1% confidence levels. R^2 are close to 1 for each mutual fund.

Following the same methodology above, we calibrate meta-factors sensitivities with systematic factors. Results are displayed in Table 2. OAS is the main meta-factor in terms of predictive power as shown by its high (negative or positive) sensitivity with the Market factor. The level of the slope appears to be also significant for almost all systematic factors. Results are more divergent for other meta-factors. The level of R^2 are relatively high except for the US Value-Growth. This might be explained by the strong performance contribution of technology (growth) stocks in US markets.

Table 3.2 Systematic factors sensitivities to meta-factors (Full)

		ΔOAS	ΔTerm spread	%USD	%Oil	Δ2Y rate	R^2
US	**Market**	-7.96***	0.98	-0.54**	0.02	2.39*	50%
	Value-Growth	1.07	0.35	-0.07	0.00	0.39	1%
	Small-Large	-3.27***	0.44	0.00	0.03**	1.31	20%
	Quality	1.12**	-0.28	0.10**	0.00	-0.73	21%
	Momentum	1.68**	-2.19***	-0.04	0.01	2.03**	10%
	Min. Vol.	2.73***	-2.03***	0.14**	-0.01	-0.53	41%
EU	**Market**	-11.61**	2.17	0.20**	0.04*	3.17**	50%
	Value-Growth	-3.47**	2.82***	-0.04	0.01	0.86*	18%
	Small-Large	-3.54***	-0.97	-0.07	0.02	-0.15	19%
	Quality	3.09***	-2.25***	0.07	0.00	-0.88*	33%
	Momentum	3.05***	-2.19***	0.04	0.00	0.9*	16%
	Min. Vol.	3.54***	-2.22***	0.02	-0.02**	-1.96***	49%

Notes: *, **, *** indicate rejection of the null hypothesis of sensitivities equal to 0 at 10%, 5% and 1% confidence levels.

3.5.2 Hypothetical scenarios

First, we use our methodology to model a market crash. A typical market crash scenario is usually accompanied by an increase of credit risk (OAS), a steepening of the yield curve with long term interest rates increasing faster than short term ones (term spread), a reinforcement of the US dollar which is regarded as a safe-haven asset (USD), a supply shock on oil (Oil) and an increase of financing costs (2Y rate). This would quantitatively translate to shocks on meta-factors displayed in Table 3.3.

	ΔOAS	ΔTerm spread	%USD	%Oil	Δ2Y rate
US	3.00%	1.00%	5%	70%	1.00%
EU	3.00%	1.00%	5%	70%	1.00%

Table 3.3 Shocks on meta-factors to model a typical market crash

Table 4 displays the results of the stress scenario. All mutual funds show a significant decrease, but results diverge with style exposures. Funds with higher sensitivity to Growth or Value show stronger negative expected performance as they are more sensitive to an increasing interest rate or a reversing economic cycle (Table 3.4). More defensive funds (with higher exposures to alternative factors such as Quality, Momentum and Min. Vol.) outperform aggressive funds in this stress scenario. These results are largely expected and show that our modelling is coherent with what is observed when a market crash materialized.

	Expected return
US tilt Value	-21%
US Blend	-17%
US Aggresive Growth	-21%
US Aggresive Value	-30%
US Value	-20%
US Defensive Value	-21%
EU Aggresive Value	-36%
EU Value	-37%
EU Growth	-26%
US tilt Value	-31%
EU Small Cap	-30%
EU Blend	-25%
EU Defensive Growth	-24%
EU Defensive Blend	-14%
Value	-28%
Growth	-23%
Blend	-19%
Defensive	-20%
Aggresive	-29%

Table 3.4 Typical market crash scenario results based on shocked meta-factors

One can also specify a shock on one meta-factors and diffuse the shock on other meta-factors using the covariance matrix. We apply two different shocks commonly observed: 1. an increase of interest rates and 2. an increase of oil prices. Results of the diffusion differ depending on the risk regime (Table 3.5). Without adverse market conditions, an increase in interest rates or oil prices is generally associated with positive equity performance. But, as we already mentioned, we should observe divergent returns between mutual funds. For instance, we expect Value funds to strongly outperform Growth funds in case of increasing rates. Results display in Table 3.6 confirm this point. We also compare our results to Bloomberg MAC 3 and compute the spearman coefficient to show that we obtain a similar ranking with close performance for the increase in oil price scenario while performance is significantly higher under our modelling for the increase in interest rates scenario.

		ΔOAS	ΔTerm spread	%USD	%Oil	Δ2Y rate
Full	**US**	-0.49%	0.31%	1.45%	14.70%	**1.00%**
	EU	-0.71%	-0.23%	-4.00%	11.24%	**1.00%**
Low	**US**	-0.30%	0.30%	1.42%	14.20%	**1.00%**
	EU	-0.68%	-0.27%	-5.10%	13.75%	**1.00%**
High	**EU**	-1.13%	0.46%	1.80%	17.49%	**1.00%**
	US	-0.80%	-0.22%	-2.83%	9.13%	**1.00%**

Table 3.5 Increase in interest rates scenario and diffusion under different risk regimes

	Interest up 100bps				Oil up 20%			
	MAC 3	**Full**	**Low**	**High**	**MAC 3**	**Full**	**Low**	**High**
US tilt Value	4%	7%	6%	13%	6%	4%	3%	5%
US Blend	2%	5%	4%	10%	4%	3%	3%	4%
US Aggresive Growth	1%	7%	6%	13%	6%	4%	3%	5%
US Aggresive Value	9%	12%	10%	21%	9%	6%	5%	8%
US Value	5%	8%	6%	13%	6%	4%	3%	5%
US Defensive Value	4%	7%	5%	12%	5%	3%	3%	5%
Spearman coeff.	-	**83%**	**58%**	**58%**	-	**89%**	**70%**	**89%**
EU Aggresive Value	9%	17%	16%	18%	7%	5%	4%	6%
EU Value	9%	16%	15%	18%	7%	5%	4%	6%
EU Growth	4%	12%	11%	13%	5%	4%	3%	4%
US tilt Value	7%	13%	12%	14%	5%	4%	3%	5%
EU Small Cap	3%	12%	12%	14%	4%	4%	4%	5%
EU Blend	4%	11%	10%	12%	4%	3%	3%	4%
EU Defensive Growth	4%	11%	10%	12%	5%	3%	3%	4%
EU Defensive Blend	1%	7%	7%	8%	4%	3%	2%	3%
Spearman coeff.	-	**86%**	**78%**	**79%**	-	**76%**	**51%**	**70%**

Table 3.6 Hypothetical stress scenario results

3.5.3 Historical scenarios

Our methodology also allowed for the reproduction of historical scenarios. Table 3.7 shows the shocks on meta-factors observed during the Great Financial Crisis market crash following Lehman Brothers default in 2008 (from 14/09/2008 to 14/10/2008) and the shocks observed after the downgrade of the US debt rating (from 22/07/2011 to 08/08/2011). Our results are compared to Bloomberg MAC 3 and realized returns.

Results in Table 3.8 confirm that our model reproduces accurate expected performances during stress periods. In terms of ranking, our model is close to MAC3 and realized returns (see spearman coefficients of MAC3 vs realized and Model vs realized). Our model does even

better than MAC3 during the Lehman event in terms of magnitude. Results are slightly mixed for the US debt downgrade event in terms of ranking but again we do better than MAC3 in terms of magnitude.

		ΔOAS	ΔTerm spread	%USD	%Oil	Δ2Y rate
Lehman Default (2008)	**US**	2.80%	1.40%	11%	-40%	-0.90%
	EU	2.10%	2.00%	11%	-35%	-1.80%
Debt Ceiling Crisis & Downgrade (2011)	**US**	0.60%	-0.50%	3%	-7%	-0.20%
	EU	0.80%	-0.20%	3%	-2%	-1.00%

Table 3.7 Observed shocks during 2008 and 2011 market events

	Lehman Default - 2008			Debt Ceiling Crisis & Downgrade in 2011		
	MAC 3	**Model**	**Realized**	**MAC 3**	**Model**	**Realized**
US tilt Value	-18%	-30%	-29%	-17%	-7%	-7%
US Blend	-20%	-23%	-26%	-16%	-5%	-9%
US Aggresive Growth	-22%	-29%	-	-18%	-7%	-
US Aggresive Value	-24%	-44%	-35%	-21%	-12%	-12%
US Value	-20%	-30%	-29%	-19%	-7%	-6%
US Defensive Value	-18%	-30%	-25%	-17%	-7%	-7%
Spearman coeff.	**70%**	**75%**		**60%**	**70%**	
EU Aggresive Value	-18%	-34%	-32%	-18%	-19%	-12%
EU Value	-18%	-34%	-	-17%	-19%	-12%
EU Growth	-20%	-25%	-	-16%	-12%	-
US tilt Value	-16%	-28%	-23%	-16%	-14%	-12%
EU Small Cap	-20%	-28%	-34%	-16%	-12%	-11%
EU Blend	-18%	-23%	-28%	-16%	-11%	-15%
EU Defensive Growth	-19%	-23%	-28%	-16%	-11%	-12%
EU Defensive Blend	-17%	-16%	-23%	-14%	-6%	-6%
Spearman coeff.	**83%**	**83%**		**83%**	**46%**	

Table 3.8 Historical stress scenario results

3.6. Conclusion

This article contributes by proposing a two-step framework to conduct stress tests through meta-factors and systematic factors. We show that we can keep the model simple, easy to implement and parsimonious while allowing a high level of flexibility. Indeed, meta-factors help to translate an economic scenario into market shocks as they exert a significant influence on multiple individual factors within the financial system (style or systematic factors in our

framework). Despite its simplicity, our methodology offers a better fit than Bloomberg MAC3 during two historical market stress. Beyond the risk management aspects of stress-testing, our framework offers the possibility to quickly implement forward or hypothetical stress scenarios helping managers to adjust their active portfolios in consequences.

Like any stress test model, our framework has its limitations. A stress test estimates the exposure to a specified event, but not the probability of such an event occurring. In addition, numerous decisions in the specification of a stress test must be made that rely on the judgement and experience of the risk manager. Therefore, there is no guarantee that the risk manager will choose the "right" scenarios or interpret the results effectively. Stress tests also impose a high computational cost, especially in collecting the data from diverse business units and from the need to revalue complex options-based positions. A further limitation is that, at present, firms cannot integrate market and credit risks in a systematic way in their stress tests, although some interviewed firms are engaged in efforts in this direction.

The methodology can be further developed in terms of parameters calibration (use more advanced tools than the multilinear regression, explore other factors). Additionally, more applications can be conducted on different equity universes and asset classes.

3.7 Appendices

Appendix 3.1: List of meta-factors candidates

Variable	Type	Transformation
Core inflation	Macroeconomic	Difference
Inflation surprises	Macroeconomic	Difference
House prices	Macroeconomic	Percentage change
LEI	Macroeconomic	Difference
PMI Manufacturing	Macroeconomic	Difference
PMI Services	Macroeconomic	Difference
Trade balance	Macroeconomic	Percentage change
Current balance	Macroeconomic	Percentage change
Budget balance	Macroeconomic	Percentage change
M2 money supply	Macroeconomic	Percentage change
FED NY forecast GDP growth	Macroeconomic	Difference
Industrial production	Macroeconomic	Difference
Consumer confidence	Macroeconomic	Difference
Unemployment Rate	Macroeconomic	Difference
Economic surprises	Macroeconomic	Difference
S&P 500	Equity	Percentage change
Average dividend yield	Equity	Difference
Average earnings growth	Equity	Difference
Average corporate margin	Equity	Difference
Average P/E	Equity	Percentage change
Average Financial leverage	Equity	Difference
Volatility 30d	Equity	Difference
Volatility 90d	Equity	Difference
Volatility 260d	Equity	Difference
VIX	Equity	Difference

Variable	Type	Transformation
Credit impulse	Credit	Difference
Fed funds rate	Credit	Difference
Gov 30Y	Credit	Difference
Gov 10Y	Credit	Difference
Gov 2Y	Credit	Difference
Gov 3M	Credit	Difference
Gov 1M	Credit	Difference
3-month LIBOR	Credit	Difference
US IG yield	Credit	Difference
US HY yield	Credit	Difference
Average OAS	Credit	Difference
TED spread	Credit	Difference
Term spread (10Y-3M)	Credit	Difference
Dollar spot	Forex	Percentage change
USD/EUR	Forex	Percentage change
USD/GBP	Forex	Percentage change
USD/JPY	Forex	Percentage change
USD/CHF	Forex	Percentage change
USD/NOK	Forex	Percentage change
USD/AUD	Forex	Percentage change
EUR/GBP	Forex	Percentage change
EUR/CHF	Forex	Percentage change
EUR/JPY	Forex	Percentage change
Gas price	Commodities	Percentage change
Oil price	Commodities	Percentage change
Silver price	Commodities	Percentage change
Gold price	Commodities	Percentage change
Copper price	Commodities	Percentage change
Aluminium	Commodities	Percentage change
Soybean	Commodities	Percentage change
weat	Commodities	Percentage change
rice	Commodities	Percentage change
cattle	Commodities	Percentage change

General Conclusion

Through three essays, this thesis demonstrates the applications of factor models in various contexts. The articles introduce innovative methods aimed at resolving current issues affecting financial institutions (banks and asset managers) and regulators. The following questions were addressed:

(i) Could we use ESG metrics to forecast idiosyncratic risk?

(ii) What are the most efficient calibration techniques to derive correlation matrices associated with factor models?

(iii) Could we propose a simple but appropriate framework to perform stress tests on equity portfolios?

A test was proposed to evaluate how well extra-financial metrics (ESG ratings) help predict firm-specific risks. The focus was on idiosyncratic volatility estimated through factor models, revealing a negative relationship between ESG ratings and idiosyncratic risk, with higher ratings predicting lower levels of idiosyncratic volatility.

The findings carry significant implications for both investors and researchers. The backtesting approach provides a valuable framework for evaluating ESG rating providers before incorporating these ratings into investment strategies. The results underscore the importance of being cautious when faced with divergent ESG ratings. For researchers in asset pricing, it is imperative to rigorously assess the quality of ESG ratings. Additionally, the observed connection between ESG ratings and idiosyncratic volatility, particularly when ratings converge, suggests that ESG investing extends beyond investor preferences. ESG ratings can

serve as indicators of future fundamentals and risks. A prospective application for investors involves comparing ratings from different extra-financial metrics providers. Our inferential procedure can be easily adapted for this purpose, assisting investors in selecting the most suitable agency in non-nested comparisons. It also facilitates the consideration of additional competing agencies for integration with existing ratings in nested comparisons, providing a comprehensive approach to decision-making in the realm of ESG investments.

Subsequently, various approaches to estimating probabilistic models with K factors and idiosyncratic noises are described. These models can be used to simulate latent random variables driving defaults with applications to the Basel 2.5 and 3 frameworks for credit exposures in the trading book.

It is shown that restricted versions of optimization-based approaches perform better when dealing with intra-classes (correlations), primarily due to the effectiveness of classification. Even with a rough approach based only on geographical-sector classification, fitting is significantly improved. Factor models are known to smooth the correlation matrix, leading to large errors for high correlations. Empirical results indicate that restricted calibration performs better in this aspect as well. By restricting information to a class of issuers that should (a priori) behave the same way, the restricted version of the SPG algorithm allows for a smaller error on these high empirical correlations than other methods. Finally, beta homogeneity for the restricted configurations provides a better understanding of the underlying latent-factor effect. In this configuration, latent-factor loadings have a meaningful interpretation.

The starting point of the iterative optimization could be discussed. The betas with respect to the K factors provided by the principal component method can be used as a starting point for an iterative optimization algorithm: iterated principal factor or Spectral Projection Gradient (SPG) for the nearest correlation matrix approach and Expectation Maximization (EM) for maximum likelihood estimation. Further work can also be done on the performance of these algorithms before considering implementation in the banking industry.

As a last application of factor models, a two-step framework was proposed to conduct stress tests through meta-factors and systematic factors. Despite its simplicity, the methodology offers a better fit than some well-established models. Beyond the risk management aspects of stress-testing, the framework offers the possibility to quickly implement forward or hypothetical stress scenarios, helping managers adjust their active portfolios accordingly.

Like any stress test model, the framework has limitations independent of factor modeling. A stress test estimates exposure to a specified event but not the probability of such an event occurring. Additionally, numerous decisions in the specification of a stress test must be made, relying on the judgment and experience of the risk manager. A further limitation is that, at present, firms cannot integrate market and credit risks systematically in their stress tests, although some interviewed firms are engaged in efforts in this direction. The methodology can be further developed in terms of parameters calibration (using more advanced tools than multilinear regression) and exploring other factors. Additionally, more applications can be conducted on different equity universes and asset classes, as the methodology remains flexible and easy to adapt.

Bibliography

Agrawal, R., Roy, U., & Uhler, C. (2019). Covariance matrix estimation under total positivity for portfolio selection. *Journal of financial econometrics*, 20, 367-389.

Akgun, O., Pirotte, A., Urga, G., Yang, Z. (2020). Equal predictive ability tests for panel data with an application to OECD and IMF forecasts. *Unpublished Manuscript*.

Albuquerque, R., Koskinen, Y., Chendi, Z. (2019). Corporate social responsibility and firm risk: theory and Empirical evidence. *Management Science*, 65, 4451–4469.

Amel-Zadeh, A., Serafeim, G. (2018). Why and how investors use ESG information: evidence from a global survey. *Financial Analyst Journal*, 74(3), 87–103.

Andersen, L., Sidenius, J., & Basu, S. (2003). All your hedges in one basket. *Risk*, 16(11), 67-72.

Anderson, T. W., & Rubin, H. (1956). Statistical inference in factor analysis. *In Proceedings of the third Berkeley symposium on mathematical statistics and probability* (Vol. 5, pp. 111-150).

Andreou, E., Ghysels, E. et A. Kourtellos (2013). Regression Models with Mixed Sampling Frequencies. *Journal of Econometrics*, 158.

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59, 817–858.

Ardia, D., Bluteau, K., Boudt, K., Inghelbrecht, K. (2022). Climate change concerns and the performance of green vs. brown stocks. *Management Science*. INFORMS.

Arnott, R., Harvey, C. R., Kalesnik, V., & Linnainmaa, J. (2019). Alice's adventures in factorland: three blunders that plague factor investing. *The Journal of Portfolio Management*, 45(4), 18-36.

Bai, J., & Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1), 191-221.

Baltagi, B. H., Kao, C., & Wang, F. (2017). Identification and estimation of a large factor model with structural instability. *Journal of Econometrics*, 197(1), 87-100.

Baltas, N. (2019). The impact of crowding in alternative risk premia investing. *Financial Analysts Journal*, 75(3), 89-104.

Barkemeyer, R., Revelli, C., Douaud, A. (2023). Selection bias in ESG controversies as a risk for sustainable investors. *Journal of Cleaner Production*, 405, 137035. Elsevier.

Bartholomew, D. J., Knott, M., & Moustaki, I. (2011). Latent variable models and factor analysis: A unified approach (Vol. 904). John Wiley & Sons.

Bartlett, M. S. (1950). Tests of significance in factor analysis. *British Journal of statistical psychology*, 3(2), 77-85.

Bartlett, M. S. (1951). A further note on tests of significance in factor analysis. *British Journal of Statistical Psychology*, 4(1), 1-2.

Basilevsky, A. T. (2009). Statistical factor analysis and related methods: theory and applications (Vol. 418). John Wiley & Sons.

Bass, R., Gladstone, S., Ang, A. (2017). Total Portfolio Factor, Not Just Asset Allocation. *Journal of Portfolio Management*, Special Issue.

Becchetti, L., Ciciretti, R., Hasan, I. (2015). Corporate social responsibility, stakeholder risk, and idiosyncratic volatility. *Journal of Corporate Finance*, 35, 297–309. Elsevier.

Berg, F., Koelbel, J., Rigobon, R. **(**2022). Aggregate confusion: The divergence of ESG ratings. *Review of Finance*, 26(6), 1315–1344. Oxford University Press.

Berg, F., Koelbel, J., Pavlova, A., Rigobon, R. (2022). ESG confusion and stock returns: Tackling the problem of noise. *National Bureau of Economic Research*.

Bernanke, B. S., Boivin, J., & Eliasz, P. (2005). Measuring the effects of monetary policy: A Factor-Augmented Vector Autoregressive (FAVAR) approach. *The Quarterly Journal of Economics*, 120(1), 387-422.

Billio, M., Costola, M., Hristova, I., Latino, C., Pelizzon, L. (2019). Inside the ESG Ratings: (Dis)agreement and performance. *Unpublished Manuscript*.

Birgin, E. G., Martínez, J. M., & Raydan, M. (2000). Nonmonotone spectral projected gradient methods on convex sets. *SIAM Journal on Optimization*, 10(4), 1196-1211.

Blaschke, W., Jones, M. T., Majnoni, G., & Martinez Peria, M. S. (2001). Stress testing of financial systems: An overview of issues, methodologies, and FSAP experiences. *IMF Working Paper* No. 88.

Borsdorf, R. (2012). Structured matrix nearness problems: Theory and algorithms doctoral dissertation, *Manchester Institute for Mathematical Sciences*, The University of Manchester.

Borsdorf, R., Higham, N. J., & Raydan, M. (2010). Computing a nearest correlation matrix with factor structure. *SIAM Journal on Matrix Analysis and Applications*, 31(5), 2603-2622.

Bouslah, K., Kryzanowski, L., M'Zali, B**.** (2013). The Impact of the dimensions of social performance on firm risk. *Journal of Banking and Finance*, 37, 1258–1273.

Boutin-Dufresne, F., Savaria, P. (2004). Corporate social responsibility and financial risk. *The Journal of Investing*, 13(1), 57–66.

Campbell, J. W., Lo, A. W., & MacKinlay, A. C. (1997). The econometrics of financial markets. *Princeton University Press*.

Chamberlain, G., & Rothschild, M. (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica*, 1281-1304.

Champagne, C., Coggins, F., Sodjahin, A. (2019). Can extra-financial ratings serve as an indicator of ESG risk? *Global Finance Journal*, 54, 100638. Elsevier.

Chan, K. C., & Chen, N. (1991). Structural and return characteristics of small and large firms. *Journal of Finance*, 46, 1467-1484.

Chatterji, A., Levine, D., Toffel, M. (2009). How well do social ratings actually measure corporate social responsibility? *Journal of Economics and Management Strategy*, 18(1), 125–169.

Chatterji, A., Durand, K., Levine, D., Touboul, S. (2016). Do ratings of firms converge? Implications for managers, investors and strategy researchers. *Strategic Management Journal*, 37(8), 1597–1614.

Chen, N., & Zhang, F. (1998). Risk and return of value stocks. *Journal of Business*, 71(4), 71-102.

Chen, Z., Petkova, R. (2012). Does idiosyncratic volatility proxy for risk exposure? *The Review of Financial Studies*, 25(9), 2745–2787. Oxford University Press.

Clark, C. (2022). The level, slope and curve factor model for stocks. *Journal of Financial Economics*, 143(1).

Connor, G., & Korajczyk, R. A. (1986). Performance measurement with the arbitrage pricing theory: A new framework for analysis. *Journal of financial economics*, 15(3), 373-394.

Connor, G., & Korajczyk, R. A. (1988). Risk and return in an equilibrium APT: Application of a new test methodology. *Journal of financial economics*, 21(2), 255-289.

Connor, G., & Korajczyk, R. A. (1993). A test for the number of factors in an approximate factor model. *Journal of Finance*, 48(4), 1263-1291.

Cui, X., Li, C., Zhao, J., Zeng, L., Zhang, D., & Pan, J. (2016). Covariance structure regularization via Frobenius-norm discrepancy. *Linear Algebra and Its Applications*, 510, 124-145.

Davies, A., Lahiri, K. (1995). A new framework for analyzing survey forecasts using three-dimensional panel data. *Journal of Econometrics*, 68, 205–228.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1), 1-22.

Diebold, F. X., Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13, 253–263.

Diemont, D., Moore, K., Soppe, A. (2016). The downside of being responsible: Corporate social responsibility and tail risk. *Journal of Business Ethics*, 137(2), 213–229. Springer.

Dimson, E., Marsh, P., Staunton, M. (2020). Divergent ESG ratings. *The Journal of Portfolio Management*, 47(1), 75–87.

Dyck, A., Lins, V., Roth, L., Wagner, H. (2019). Do institutional investors drive corporate social responsibility? *Journal of Financial Economics*, 131(3), 693–714.

Eckart, C., & Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3), 211-218.

Elton, E. J., & Gruber, M. J. (1973). Estimating the dependence structure of share prices--implications for portfolio selection. *Journal of Finance*, 28(5), 1203-1232.

Elton, E. J., Gruber, M. J., Brown, S. J., & Goetzmann, W. N. (2014). Modern portfolio theory and investment analysis. John Wiley & Sons.

ESMA. (2019). Stress simulation for investment funds. ESMA, Paris.

Fama, E. F. (1991). Efficient Capital Markets: II. *Journal of* Finance, 46, 1575–1617.

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

Fama, E. F., & Kenneth, R. (1993). French, 1993, Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, 33(1), 3-56.

Fama, E. F., & French, K. R. (1996). Multifactor explanations of asset pricing anomalies. *Journal of Finance*, 51, 55-184.

Fama, E. F., French, K. R. (2007). Disagreement, tastes, and asset prices. *Journal of Financial Economics*, 83(3), 667–689.

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1).

Fan, J., Fan, Y., & Lv, J. (2008). High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics*, 147(1), 186-197.

Fan, J., Liao, Y., & Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(4), 603-680.

Feng, G., Giglio, S., & Xiu, D. (2020). Taming the factor zoo: A test of new factors. *The Journal of Finance*, 75, 1327-1370.

Giacomini, R., White, H. (2006). Tests of conditional predictive ability. *Econometrica*, 74(6), 1545–1578.

Gibbons, M. R., Ross, S. A., & Shanken, J. (1989). A test of the efficiency of a given portfolio. *Econometrica*, 1121-1152.

Gibson, B., Krueger, P., Schmidt, P. (2021). ESG rating disagreement and stock returns. *Financial Analysts Journal*, 77(4), 104–127. Taylor & Francis.

Glasserman, P., & Suchintabandid, S. (2007). Correlation expansions for CDO pricing. *Journal of Banking & Finance*, 31(5), 1375-1398.

Gollier, C., Pouget, S. (2022). Investment strategies and corporate behaviour with socially responsible investors: A theory of active ownership. *Economica*, 89(356), 997–1023. Wiley Online Library.

Golub, G. H., Hoffman, A., & Stewart, G. W. (1987). A generalization of the Eckart-Young-Mirsky matrix approximation theorem. *Linear algebra and its applications*, 88, 317-327.

Gregory, J., & Laurent, J. P. (2004). In the core of correlation. *Risk*, 17, 87-91.

Green, J., Hand, J., & Zhang, X. (2017). The characteristics that provide independent information about average U.S. monthly stock returns. *Review of Financial Studies*, 4389-4436.

Grubišić, I., & Pietersz, R. (2007). Efficient rank reduction of correlation matrices. *Linear algebra and its applications*, 422(2-3), 629-653.

Gulen, H., Xing, Y., & Zhang, L. (2011). Value versus growth: Time-varying expected stock returns. *Financial Management*, 40(2), 381-407.

Hahn, J., & Lee, H. (2006). Yield spreads as alternative risk factors for size and book-to-Market. *Journal of Financial and Quantitative Analysis*, 41, 245–69.

Harman, H. H., & Jones, W. H. (1966). Factor analysis by minimizing residuals (minres). *Psychometrika*, 31(3), 351-368.

Hartzmark, S., Sussman, A. (2019). Do investors value sustainability? A natural experiment examining ranking and fund flows. *Journal of Finance*, 74(6), 2789–2837.

Harvey, C. R., Liu, Y., & Zhu, H. (2016). ...and the cross-section of expected returns. *Review of Financial Studies*, 29, 5–68.

Higham, N. J. (2002). Computing the nearest correlation matrix—a problem from finance. *IMA Journal of Numerical Analysis*, 22(3), 329-343.

Hoffmann, J. P. (2016). Regression models for categorical, count, and related variables: An applied approach. *Univ of California Press*.

Hotelling, H. (1957). The relations of the newer multivariate statistical methods to factor analysis. *British Journal of Statistical Psychology*, 10(2), 69-79.

Hou, K., Xue, C., & Zhang, L. (2020). Replicating Anomalies. *Review of Financial Studies*, 33(5), 2019-2133.

Huang, J., & Yang, L. (2010). Correlation matrix with block structure and efficient sampling methods. *Journal of Computational Finance*, 14(1), 81.

Huberman, G., & Kandel, S. (1987). Mean-Variance Spanning. *Journal of Finance*, 42, 873-888.

Jagannathan, R., & Wang, Z. (1996). The Conditional CAPM and the cross-section of expected returns. *Journal of Finance*, 51, 3–53.

Jegadeesh, N., Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance*, 48(1), 65–91.

Jeffers, J. N. R. (1967). Two case studies in the application of principal component analysis. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 16(3), 225-236.

Jo, H., Na, H. (2012). Does CSR reduce firm risk? Evidence from controversial industry sectors. *Journal of Business Ethics*, 110, 441–456. Springer.

Jolliffe, I. T. (2010). Principal component analysis. Springer.

Jöreskog, K. G. (1969a). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34(2), 183-202.

Jöreskog, K. G. (1969b). Factoring the multitest-multioccasion correlation matrix. *ETS Research Bulletin Series*, 1969(2), i-32.

Kakushadze, Z. (2015). Russian-doll risk models. *Journal of Asset Management*, 16(3), 170-185

Kang, W., Rouwenhorst, K. G., & Tang, K. (2021). Crowding and factor returns. Available at SSRN.

Kercheval, A. N. (2008). On Rebonato and Jäckel's parametrization method for finding nearest correlation matrices. *International Journal of Pure and Applied Mathematics*, 45(3), 383.

King, B. F. (1966). Market and industry factors in stock price behavior. *Journal of Business*, 39(1), 139-190.

Laurent, J. P., & Gregory, J. (2005). Basket default swaps, CDOs and factor copulas. *Journal of risk*, 7(4), 103-122.

Laurent, J. P., Sestier, M., & Thomas, S. (2016). Trading book and credit risk: How fundamental is the Basel review? *Journal of Banking & Finance*, 73, 211-223.

Ledoit, O., & Wolf, M. (2004). Honey, I shrunk the sample covariance matrix. *Journal of Portfolio Management*, 30(4), 110-119.

Ledoit, O., & Wolf, M. (2020). The power of (non-) linear shrinking: a review and guide to covariance matrix estimation. University of Zurich, Department of Economics, *Working Paper*, (323).

Lee, D., Faff, R. (2009). Corporate sustainability performance and idiosyncratic risk: A global perspective. *Financial Review*, 44(2), 213–237. Wiley Online Library.

Lettau, M., & Ludvigson, S. C. (2001). Consumption, Aggregate Wealth, and Expected Stock Returns. *Journal of Finance*, 56(3), 815-849.

Liew, J., & Vassalou, M. (2000). Can Book-to-Market, Size, and Momentum be Risk Factors that Predict Economic Growth? *Journal of Financial Economics*, 57, 221–245.

MacKinlay, A. C. (1987). On multivariate tests of the CAPM. *Journal of Financial Economics*, 18(2), 341-371.

Marsaglia, G., & Olkin, I. (1984). Generating correlation matrices. *SIAM Journal on Scientific and Statistical Computing*, 5(2), 470-475.

McLean, R. D., & Pontiff, J. (2016). Does academic research destroy stock return predictability? *Journal of Finance*, 71, 5–32.

Menchero, et al. (2020). MAC3 Global equity risk model, *Methodology Notes*.

Miranda-Agrippino, S., & Rey, H. (2015). World asset markets and the global financial cycle.

Mirsky, L. (1960). Symmetric gauge functions and unitarily invariant norms. *The quarterly journal of mathematics*, 11(1), 50-59.

Mishra, S., Modi, S. (2013). *Positive and negative corporate social responsibility, financial leverage, and idiosyncratic risk. Journal of Business Ethics*, 117, 431–448. Springer.

Morini, M., & Webber, N. (2006). An EZI method to reduce the rank of a correlation matrix in financial modelling. *Applied Mathematical Finance*, 13(4), 309-331.

Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *The Review of Economics and Statistics*, 92(4), 1004-1016.

Parker, J., & Sul, D. (2016). Identification of unknown common factors: Leaders and followers. *Journal of Business & Economic Statistics*, 34(2), 227-239.

Pástor, L., Stambaugh, R., Taylor, L. (2021). Sustainable investing in equilibrium. *Journal of Financial Economics*, 142(2), 550–571. Elsevier.

Pástor, L., Stambaugh, R., Taylor, L. (2022). Dissecting green returns. *Journal of Financial Economics*, 146(2), 403–424. Elsevier.

Pedersen, L., Fitzgibbons, S., Pomorski, L. (2021). Responsible investing: The ESG-efficient frontier. *Journal of Financial Economics*, 142(2), 572–597. Elsevier.

Perez-Quiros, G., & Timmermann, A. (2000). Firm size and cyclic variation in stock returns. *Journal of Finance*, 55, 1229-1262.

Petkova, R. (2006). Do the Fama–French factors proxy for innovations in predictive variables? *The Journal of Finance*, 61(2), 581-612.

Pietersz, R., & Groenen, P. J. (2004). Rank reduction of correlation matrices by majorization. *Quantitative Finance*, 4(6), 649-662.

Pukthuanthong, K., Roll, R., & Subrahmanyam, A. (2019). A Protocol for factor identification. *The Review of Financial Studies*, 32(4), 1573-1607.

Pykhtin, M. (2004). Portfolio credit risk multi-factor adjustment. *Risk*, 17(3), 85-90.

Qi, H., & Sun, D. (2006). A quadratically convergent Newton method for computing the nearest correlation matrix. *SIAM journal on matrix analysis and applications*, 28(2), 360-385.

Qi, H., & Sun, D. (2010). Correlation stress testing for value-at-risk: an unconstrained convex optimization approach. *Computational Optimization and Applications*, 45(2), 427-462.

Rao, C. R. (1964). The use and interpretation of principal component analysis in applied research. *Sankhyā: The Indian Journal of Statistics*, Series A, 329-358.

Rapisarda, F., Brigo, D., & Mercurio, F. (2007). Parameterizing correlations: a geometric interpretation. *IMA Journal of Management Mathematics*, 18(1), 55-73.

Rebonato, R., & Jäckel, P. (2000). The most general methodology to create a valid correlation matrix for risk management and option pricing purposes. *Journal of Risk*, (2) 17-27.

Reiersøl, O. (1950). On the identifiability of parameters in Thurstone's multiple factor analysis. *Psychometrika*, 15(2), 121-149.

Rencher, A. C., & Christensen, W. F. (2012). Multivariate regression. *Methods of multivariate analysis*, 339-384.

Rendlen, B., Weber, B. (2019). Guidance Note: Integrating ESG factors into financial models for infrastructure investments. *WWF Working Paper*.

Riedl, A., Smeets, P. (2017). Why do investors hold socially responsible mutual funds. *Journal of Finance*, 72(6), 2505–2550.

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13, 341-360.

Rousseeuw, P. J., & Molenberghs, G. (1993). Transformation of non-positive semidefinite correlation matrices. *Communications in Statistics-Theory and Methods*, 22(4), 965-984.

Ruban, & Melas. (2010). Stress Testing in the investment process. *MSCI Barra Research Paper*. No 2010-28.

Rubin, D. B., & Thayer, D. T. (1982). EM algorithms for ML factor analysis. *Psychometrika*, 47(1), 69-76.

Semenova, N., Hassel, L. (2015). On the validity of environmental performance metrics. *Journal of Business Ethics*, 132(2), 249–258.

Serafeim, G., Yoon, A. (2021). Stock price reactions to ESG news: The role of ESG ratings and disagreement. *Review of accounting studies*, 1–31. Springer.

Serafeim, G., Yoon, A. (2022). Which corporate ESG news does the market react to? *Financial Analysts Journal*, 78(1), 59–78. Taylor & Francis.

Shanken, J. (1992). On the estimation of beta-pricing models. *The review of financial studies*, 5(1), 1-33.

Simpson, M., & Ramchander, S. (2008). An inquiry into the economic fundamentals of the Fama and French equity factors. *Journal of Empirical Finance*, 15, 801–815.

Sodjahnin, A., Champagne, C., Coggins, F., Gillet, R. (2017). Leading or lagging indicators of risk? The informational content of extra-financial performance scores. *Journal of Asset Management*, 18(5), 347–370.

Stracca, L. (2006). Delegated portfolio management: A survey of the theoretical literature. *Journal of Economic Surveys*, 20, 823-848.

Swade, L., Shackleton, Nolte, Hixon, Jay. (2022). Macro investing with style. *Journal of Portfolio Management*, QES Special Issue.

Timmermann, A., Zhu, Y. (2019). Comparing forecasting performance with panel data. *Unpublished Manuscript*.

Vassalou, M. (2003). News related to future GDP growth as a risk factor in equity returns. *Journal of Financial Economics*, 68(1), 47–73.

Vassalou, M., & Xing, Y. (2004). Default risk in equity returns. *Journal of Finance*, 59, 831–868.

West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica*, 64, 1067–1084.

White, H. (2001). Asymptotic theory for econometricians. *Revised Edition*. Academic Press.

Wu, L. (2002). Fast at-the-money calibration of the LIBOR market model through Lagrange multipliers. *Journal of Computational Finance*, 6(2), 39-77.

Yang, R. (2022). What do we learn from ratings about corporate social responsibility? New evidence of uninformative ratings. *Journal of Financial Intermediation*, 52, 100994. Elsevier.

Young, G. (1940). Maximum likelihood estimation and factor analysis. *Psychometrika*, 6(1), 49-53.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American statistical Association*, 57(298), 348-368.

Zerbib, O. D. (2022). *A Sustainable Capital Asset Pricing Model (S-CAPM): Evidence from environmental integration and sin stock exclusion*. *Review of Finance*, 26(6), 1345–1388. Oxford University Press.

Zhang, Z., & Wu, L. (2003). Optimal low-rank approximation to a correlation matrix. *Linear algebra and its applications*, 364, 161-187.

Résumé en français

Les modèles à facteurs sont un concept fondamental en modélisation statistique. Ils fournissent un cadre pour comprendre les relations complexes entre les variables. Ces modèles sont particulièrement utiles en finance où les variables étudiées sont interconnectées.

Une des applications clés des modèles à facteurs en finance est leur utilisation intensive dans l'évaluation des actifs. Le Modèle d'Évaluation des Actifs Financiers (MEDAF) de William Sharpe (1964) et John Litner (1965) est considéré comme une forme simple de modèle à facteurs, où les rendements d'un actif sont expliqués par un facteur de marché (les rendements d'un portefeuille représentatif d'actions, appelé le portefeuille de marché). Le modèle marque l'émergence des modèles à facteurs en finance et reste largement utilisé de nos jours. La force du MEDAF, et des modèles à facteurs en général, réside dans leur simplicité, leur robustesse et leur capacité à prédire des relations (entre le risque et les rendements attendus dans le cas du MEDAF). Des extensions de ce modèle incluent les modèles multifactoriels, tels que le modèle à trois facteurs de Fama-French (Fama, 1993), qui introduit des facteurs supplémentaires tels que la taille (capitalisations boursières) et la valeur (valorisation boursière) pour mieux saisir la complexité des rendements boursiers.

En économétrie, les modèles à facteurs sont utilisés pour discerner les tendances économiques sous-jacentes. Par exemple, un modèle à facteurs macroéconomiques peut inclure des variables telles que l'inflation, les taux d'intérêt et la croissance du PIB pour expliquer les fluctuations observées dans divers indicateurs économiques. Au-delà de la finance et de l'économie, les modèles à facteurs sont également largement utilisés dans d'autres domaines exploitant des ensembles de données volumineux et complexes. En psychologie, par exemple, un modèle à facteurs pourrait être appliqué pour comprendre les facteurs sous-jacents influençant les scores sur divers tests psychologiques. D'ailleurs, nous nous appuierons sur la littérature psychométrique dans le deuxième chapitre de cette thèse.

Nous distinguons trois principaux types de modèles à facteurs en finance (Connor, 1995) : macroéconomiques, statistiques et fondamentaux. Les modèles à facteurs macroéconomiques utilisent des séries temporelles économiques observables comme mesures de facteurs influençant une variable financière, en supposant que cette variable réponde de manière linéaire aux chocs macroéconomiques ou financiers. Les modèles à facteurs statistiques utilisent plusieurs techniques statistiques pour estimer les paramètres des modèles à facteurs. L'analyse en composantes principales (ACP) (Jolliffe, 2002 ; Stock et al., 2002) est une méthode courante utilisée pour identifier les facteurs sous-jacents en capturant les directions de la variance maximale dans les données. La méthode d'estimation du maximum de vraisemblance (MLE) et les méthodes bayésiennes sont également utilisées (Watson and Engle, 1983). Enfin, les modèles à facteurs fondamentaux reposent sur les attributs des actifs financiers pour expliquer leur dynamique (par exemple, le rendement des dividendes, le ratio cours-valeur comptable et la classification sectorielle peuvent expliquer une proportion substantielle de rendements communs entre plusieurs actions ou obligations).

Au-delà de l'évaluation des actifs, les modèles à facteurs ont de nombreuses applications en finance. Dans la gestion de portefeuille, ils aident à optimiser les allocations d'actifs en identifiant les sources de diversification (Meucci, 2009). En gestion des risques, ces modèles sont cruciaux pour quantifier et gérer diverses sources de risque. Par exemple, ils aident les institutions à comprendre l'impact de différents facteurs sur la *Value at Risk* (VaR) et le risque de défaut (Jorion, 2007). Dans la modélisation du risque de crédit et l'analyse des titres à revenu fixe, les modèles à facteurs peuvent être utilisés pour évaluer la solvabilité des emprunteurs. Ils aident à identifier les facteurs communs qui influent sur le risque de crédit des prêts (ou des obligations) et peuvent être utilisés pour modéliser le risque lié aux taux d'intérêt (Altman, 1968 ; Litterman and Scheinkman, 1991). Les modèles à facteurs trouvent de nombreuses autres applications, par exemple, dans l'évaluation des options (Black and Scholes, 1973) ou dans l'attribution de performance (Brinson, 1986). Ces applications démontrent la polyvalence des modèles à facteurs dans leur capacité à répondre aux problématiques actuelles de la recherche en finance.

Les modèles à facteurs présentent de nombreux avantages qui justifient leur utilité. La réduction de dimension, dans laquelle les modèles à facteurs excellent, est peut-être l'avantage le plus important. En effet, les modèles à facteurs aident à réduire la dimensionnalité des données en capturant les similitudes (Fan and Liao, 2014). La réduction de la dimensionnalité permet également d'améliorer l'interprétabilité des structures latentes qui génèrent les dynamiques observées ou de confirmer l'influence des facteurs observables. Cette caractéristique rend les modèles à facteurs toujours très utiles aujourd'hui malgré l'émergence de modèles basés sur l'apprentissage automatique. En effet, les facteurs identifiés ont souvent un sens économique ou théorique bien identifié, ce qui facilite la compréhension et l'explication de la modélisation. Au contraire, certains modèles d'apprentissage automatique, en particulier les plus complexes tels que les réseaux neuronaux profonds, sont souvent considérés comme des "boîtes noires", ce qui rend difficile leur interprétation. Les modèles à facteurs ont également tendance à être plus efficaces sur le plan du traitement des données, même lorsqu'ils sont appliqués à des données de haute dimensionnalité (Bai and Li, 2012). Ils peuvent capturer et résumer efficacement l'essentiel de l'information à l'aide d'un ensemble plus restreint de facteurs, tandis que les modèles d'apprentissage automatique peuvent nécessiter d'énormes quantités de données, un coût non négligeable pour les institutions financières.

Les modèles à facteurs présentent cependant plusieurs limitations. La plupart de ces modèles suppose des relations linéaires entre les variables observées et les facteurs. Cette hypothèse pourrait ne pas être valable dans tous les cas, notamment lorsqu'il s'agit de relations complexes et non linéaires dans les données. La faible robustesse de certains modèles à facteurs constitue également une limitation. La performance de ces modèles peut être sensible au choix des facteurs et aux spécifications du modèle. Choisir un nombre inapproprié de facteurs ou utiliser une méthode d'estimation inadéquate peut conduire à des résultats biaisés et à des interprétations erronées. Enfin, deux limitations importantes doivent être prises en compte lors de la manipulation de modèles à facteurs : l'hypothèse selon laquelle les variables suivent une distribution gaussienne et l'hypothèse d'homoscédasticité. Certains modèles à facteurs supposent que les facteurs sous-jacents et les composants idiosyncratiques suivent une distribution gaussienne. Cette hypothèse peut ne pas être valable dans tous les cas,

notamment lorsqu'il s'agit de données financières, ou des événements extrêmes où des valeurs aberrantes peuvent avoir un impact significatif sur les résultats (Murray, 2013). Les modèles à facteurs supposent également souvent l'homoscédasticité, i.e., la variance des composants idiosyncratiques est constante pour toutes les observations. Dans des scénarios réels, cette hypothèse peut ne pas être valable, surtout pendant des périodes de grande instabilité, entrainant une incapacité à capturer des ruptures structurelles. Dans cette thèse, nous utilisons les modèles à facteurs pour estimer : (i) la volatilité idiosyncratique, (ii) les corrélations et (iii) les scénarios de stress de marché.

Comprendre et quantifier la volatilité idiosyncratique est crucial dans divers en finance et en économie. La volatilité idiosyncratique fait référence à la partie de la volatilité totale d'un actif qui est spécifique à cet actif et qui n'est pas expliquée par des facteurs communs. Les modèles à facteurs offrent un cadre robuste pour estimer et décomposer cette volatilité idiosyncratique, fournissant une estimation du risque spécifique associé aux actifs étudiés. Dans le premier chapitre de cette thèse, nous cherchons à mesurer le contenu informatif d'une variable pour expliquer la volatilité idiosyncratique des actions. Nous avons ainsi besoin d'un cadre pour quantifier la volatilité idiosyncratique. Deux modèles à facteurs largement utilisés à ces fins sont le CAPM et le modèle à trois facteurs de Fama-French mentionnés précédemment. Selon l'hypothèse jointe (Fama, 1991), si les marches sont efficients alors le seul risqué pour lequel les investisseurs doivent être compenser est le risque systématique. Ainsi, la présence de volatilité idiosyncratique peut indiquer l'existence d'un facteur qui n'est pas capturé par le modèle.

Pour estimer la volatilité idiosyncratique à l'aide de modèles à facteurs, nous suivons généralement un processus en deux étapes. Tout d'abord, nous estimons les sensibilités aux facteurs à l'aide de données historiques des rendements de l'actif et des facteurs pertinents. Ensuite, les sensibilités estimées sont utilisées pour calculer la composante du risque systématique, qui est ensuite soustraite de la volatilité totale observée pour obtenir la volatilité idiosyncratique. Plusieurs études ont exploré l'efficacité des modèles à facteurs dans l'estimation de la volatilité idiosyncratique. Les recherches menées par Ang, Hodrick, Xing et Zhang (2006) offrent un certain éclairage sur l'application du modèle à trois facteurs de Fama-

French pour estimer la volatilité idiosyncratique. Ils constatent que l'incorporation des facteurs taille et valeur améliore significativement la puissance explicative du modèle, renforçant la précision des estimations de la volatilité idiosyncratique. De plus, le travail de Bali et Cakici (2008) explore le rôle des facteurs macroéconomiques dans l'explication de la volatilité idiosyncratique. Ils étendent les modèles à facteurs traditionnels en incluant des variables macroéconomiques, démontrant que ces facteurs contribuent de manière significative à l'estimation de la volatilité idiosyncratique. Les modèles à facteurs offrent ainsi un cadre polyvalent et efficace pour estimer la volatilité idiosyncratique, offrant aux chercheurs et aux praticiens la capacité de démêler le risque spécifique aux mouvements systématiques du marché.

Les modèles à facteurs peuvent être spécifiquement utilisés pour révéler les structures sous-jacentes dans les ensembles de données multivariées. La forme générale d'un modèle à facteurs pour l'estimation des matrices de corrélation peut être exprimée comme suit :

$$\Omega = \beta\beta' + \mathcal{E}$$

où Ω est la matrice de corrélation observée, β représente la matrice des facteurs communs, et $\mathcal{E}$ est une matrice diagonale capturant les composants idiosyncratiques.

L'utilisation des modèles à facteurs est recommandée par le Comité de Bâle sur le Contrôle Bancaire (BCBS) pour les institutions financières. En effet, les banques doivent traiter une grande quantité de données très complexes. Les modèles à facteurs offrent une approche parcimonieuse, particulièrement pertinente pour la gestion du risque de contreparties. Cependant, les exigences de Bâle 3 impliquent l'utilisation d'une matrice de corrélation associée à une représentation factorielle probabiliste, avec au moins "deux types de facteurs de risque systématiques". Il est précisé que "les corrélations doivent être basées sur les spreads de crédit ou sur les prix des actions cotées en bourse".

Bien que cela puisse faire référence aux modèles à facteurs latents utilisés par les banquiers centraux et les chercheurs, les régressions linéaires sur des facteurs observables, généralement les rendements de portefeuille, sont souvent utilisées pour construire des matrices de corrélation conformes aux exigences réglementaires.

Les banques utilisent alors les sensibilités aux facteurs (des facteurs latents) générant la matrice de corrélation la plus proche de la matrice de corrélation historique pour modéliser le risque de défaut des contreparties en portefeuille. Sur la base de cette modélisation, les banques capitalisent une charge de capital. Ces exigences ont des implications pratiques sur les modèles à deux facteurs et les contraintes de calibration des corrélations (Laurent et al., 2016).

Les stress tests sont un outil crucial pour évaluer la robustesse des systèmes financiers et des institutions dans des conditions adverses. Les modèles à facteurs jouent un rôle clé dans les stress tests en fournissant un cadre structuré pour évaluer comment divers facteurs peuvent impacter la performance et la stabilité des actifs financiers. Ces modèles aident à simuler des scénarios extrêmes et identifient les vulnérabilités potentielles.

Dans un stress test, un modèle à facteurs intègre généralement un ensemble de variables macroéconomiques et financières susceptibles d'être des sources de stress. Le modèle vise à comprendre comment les changements dans ces facteurs peuvent affecter un portefeuille et ses sous-jacents. Ces facteurs peuvent inclure des variables telles que les taux d'intérêt, les taux de croissance économique, les taux de change et d'autres indicateurs pertinents.

Identifier les principaux facteurs pertinents pour les actifs en question est d'une importance primordiale. Ces facteurs doivent capturer les principales sources de risque. Le modèle à facteurs est ensuite utilisé pour évaluer l'impact des scénarios de stress sur le portefeuille. Cette évaluation peut comprendre l’étude des variations de prix des actifs, la liquidité du marché, le risque de crédit, etc.

Les régulateurs imposent aux banques de se soumettre à des stress tests pour garantir leur capacité à résister aux conditions économiques adverses (comme mentionné ci-dessus). Les enseignements tirés des stress tests au travers des modèles à facteurs aident les institutions

financières à développer des stratégies efficaces de mitigation des risques (Jokivuolle, 2015). Les stress tests peuvent également s'avérer être un outil de gestion qui permet aux investisseurs d'ajuster leurs positions en portefeuille en prévision des chocs potentiels.

Cette thèse, organisée en trois chapitres, exploite les modèles à facteurs pour répondre à des problématiques actuelles de l'industrie financière. Nous abordons divers domaines de la recherche financière : ESG, gestion des risques, gestion de portefeuille et réglementation bancaire. Les trois articles de recherche présentés par la suite introduisent des méthodes innovantes visant à répondre aux questions suivantes :

(i) Pouvons-nous utiliser des métriques ESG pour prévoir le risque idiosyncratique ?

(ii) Quelles sont les techniques de calibration les plus efficaces pour dériver les matrices de corrélation associées à des modèles à facteurs ?

(iii) Pouvons-nous proposer une méthodologie simple mais efficace pour réaliser des stress tests de portefeuilles d'actions ?

Dans le chapitre 1, nous utilisons les modèles à facteurs pour leur capacité à capter les inefficiences de marché, notamment au travers de l'estimation de la volatilité idiosyncratique. Le contenu informationnel des notations ESG dans la prédiction de cette volatilité est testé. Plus précisément, nous formulons une procédure de backtesting pour évaluer l'efficacité des notations ESG dans la prédiction du risque idiosyncratique d'une entreprise. Cela implique d'étendre le test de capacité prédictive conditionnelle proposé par Giacomini et White (2006) à un contexte de données de panel. Notre approche est appliquée dans le but de prédire la volatilité idiosyncratique des rendements boursiers, en comparant deux systèmes de notation ESG - Sustainalytics et Asset4 - sur trois régions d'investissement (Europe, Amérique du Nord et région Asie-Pacifique). Alors que ce premier chapitre fournit une étude empirique, le chapitre 2 se situe à l'intersection entre développements théoriques et empiriques. Il présente un aperçu des différentes techniques de spécification et calibration pour dériver des matrices de corrélation grâce aux modèles à facteurs. Nous nous concentrons à la fois sur des modèles non restreints et restreints, où des bêtas spécifiques (sensibilités) sont fixés à zéro, conduisant à des facteurs régionaux ou sectoriels. Nous abordons les facteurs latents par la méthode du

maximum de vraisemblance en utilisant l'algorithme Expectation-Maximisation (EM) et l'approche des moindres carrés avec l'algorithme Spectral Projection Gradient (SPG). À l'aide d'un ensemble de données de rendements d'actions provenant des principaux indices de crédit américains et européens, nous effectuons un exercice de calibration en faisant varier le nombre et la nature des facteurs. Nous évaluons la performance de l'ajustement, la parcimonie des paramètres, la facilité d'interprétation et la capacité à gérer les corrélations élevées. Enfin, le chapitre 3 fournit une méthodologie robuste pour les praticiens du secteur financier. Il présente une approche novatrice s'appuyant sur des "méta-facteurs" pour réaliser des stress tests de portefeuilles à travers des scénarios économiques historiques ou hypothétiques. Les stress tests en finance évaluent la résilience d'un portefeuille face à des conditions adverses, constituant un complément précieux aux mesures de risque conventionnelles telles que la volatilité, la VaR et l'*expected shortfall*. Contrairement aux mesures de risque statistiques, les tests de résistance fournissent des estimations liées à des événements spécifiques, apportant ainsi un supplément d'information significatif. Alors que les stress tests traditionnels s'appuient souvent sur des scénarios historiques en reproduisant des événements passés sur des positions financières actuelles, ils manquent souvent d'une approche proactive. Notre approche, reposant sur quelques paramètres, conserve une grande flexibilité, est facile à mettre en œuvre et améliore la compréhension des sensibilités d'un portefeuille d'actifs aux facteurs systématiques. Nous appliquons cette méthodologie à des gestions actives d'actions.

Le chapitre 1 " Are ESG ratings informative to forecast idiosyncratic risk?" révèle une relation négative entre les notations ESG et le risque idiosyncratique, des notations plus élevées prédisant des niveaux plus bas de volatilité idiosyncratique. De plus, l'efficacité de la prédiction est généralement plus élevée pour la dimension environnementale des notations. L'application du test réservé aux entreprises pour lesquelles il existe un haut degré de consensus entre les agences de notation ESG révèle un pouvoir informationnel plus élevé pour les trois univers. Au-delà de l'évaluation des systèmes de notation ESG, ce dernier résultat suggère que les informations recueillies auprès de plusieurs fournisseurs de notations ESG devraient être vérifiées avant que ces critères ne soient intégrés aux processus d'investissement. Les résultats ont des implications importantes pour les investisseurs et les chercheurs. Pour les investisseurs, notre procédure de backtest offre un cadre pratique pour prendre en compte la qualité des notations ESG avant d'intégrer celles-ci dans le processus

d'investissement. Nos résultats suggèrent de la prudence quant au contenu informationnel des notations ESG lorsqu'elles divergent. De plus, le lien entre les notations ESG et la volatilité idiosyncratique lorsque les notations convergent suggèrent que l'investissement ESG n'est pas seulement une question de préférences des investisseurs, mais que les notations ESG peuvent également fournir des informations sur les fondamentaux et les risques futurs. Le chapitre 2 " Correlation matrices with factor structure for credit risk exposures " montre que, outre la performance, la parcimonie des paramètres, la facilité d'interprétation et la capacité à générer des corrélations élevées, les modèles à facteurs latents restreints surpassent ses concurrents dans l'estimation des matrices de corrélation. Nous constatons que, lorsqu'il s'agit de l'approche naïve PCA, la transformation finale pour retrouver la matrice de corrélation doit être choisie avec soin et que les versions restreintes des approches basées sur l'optimisation se comportent mieux sur les corrélations intra-classes. Il est bien connu que les modèles à facteurs ont tendance à lisser la matrice de corrélation, entraînant une erreur significative pour les corrélations élevées. Les expériences numériques indiquent que la calibration contrainte est plus performante à cet égard également. En restreignant l'information à une catégorie d'émetteurs qui devrait (a priori) se comporter de la même manière, la version contrainte de l'algorithme SPG génère une erreur d'estimation plus faible sur les corrélations empiriques élevées. Le chapitre 3 " Stress-testing: A meta-factor approach " démontre que l'utilisation de méta-facteurs aident à traduire un scénario économique en chocs de marché grâce à leur influence significative sur les facteurs observables des actifs financiers (style ou facteurs systématiques dans notre étude). Malgré sa simplicité, notre méthodologie offre de meilleurs résultats que certains modèles largement utilisés par les praticiens. Au-delà de la gestion des risques, notre méthodologie offre la possibilité de mettre en œuvre rapidement des scénarios de stress hypothétiques aidant ainsi les gestionnaires à ajuster leurs positions actives en conséquence.